SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(119th Edition)

NOVEMBER 2022

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 Nineteenth** 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 **November, 2022**. 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 Comparative Study of the Feasibility of Alternative Fuel Vehicles for Sustainable Transportation in India: A Hybrid Approach of DEMATEL and TOPSIS, ***8.Aman Raj***, ***8.Ankita Dan***, ***8.Vrinda*** and ***3.Pravin Kumar***, Mechanical, DTU

2. A Complete Internet of Things based Home Security System, ***8.Haren Nagdewani*** and ***3.Pawan Singh Mehra***, CSE, DTU

3. A Decentralized Coordination Algorithm for Patrolling and Target Tracking in Internet of Robotic Things (IoRT) using Dynamic Waypoints and Self-triggered Communication, Janardan Kumar Verma and ***3.Virender Ranga***, IT, DTU

4. A feasibility study and cost-benefit analysis of an off-grid hybrid system for a remote area electrification, ***6.Upma Singh*** and ***3.Mohammad Rizwan***, Electrical, DTU

5. A Review of Deep Learning Techniques for Protein Function Prediction, ***6.Divyanshu Aggarwal*** and ***3.Yasha Hasija***, Biotechnology, DTU

6. A SPREADER RANKING ALGORITHM FOR EXTREMELY LOW-BUDGET INFLUENCE MAXIMIZATION IN SOCIAL NETWORKS USING COMMUNITY BRIDGE NODES, ***8.Aaryan Gupta***, ***8.Inder Khatri***, ***8.Arjun Choudhry***, ***8.Pranav Chandhok***, ***3.Dinesh Kumar Vishwakarma*** and ***8.Mukesh Prasad***, IT, DTU

7. Adsorptive Removal of Congo Red Using Organically Modified Zinc–Copper–Nickel Ternary Metal Hydroxide: Kinetics, Isotherms and Adsorption Studies, ***6.Jigyasa Pathak*** and ***3.Poonam Singh***, Applied Chemistry, DTU

8. ADVANCED ENERGY EFFICIENT PEGASIS BASED ROUTING PROTOCOL FOR IoT APPLICATIONS, Priyanka chugh, Meenu Gupta, ***3.Dr. S. Indu***, Gopal Chaudhary, Manju Khari and Vimal Shanmuganathan, ECE, DTU

9. AN EMOTION-AWARE MULTI-TASK APPROACH TO FAKE NEWS AND RUMOUR DETECTION USING TRANSFER LEARNING, ***8.Arjun Choudhry***, ***8.Inder Khatri***, ***3.Minni Jain*** and ***3.Dinesh Kumar Vishwakarma***, CSE and IT, DTU

10. An Iterative Analytic Approximation for a Class of Non-Linear Singularly Perturbed Parabolic Partial Differential Equations, ***6.Kartikay Khari*** and ***3.Vivek Kumar***, Mathematics, DTU

11. An Overview of Machine Learning, ***8.Wazid Ansari***, DTU

12. Analog Performance Analysis of a Novel 5nm Stacked Oxide Top Bottom Gated Junctionless FinFET, ***8.L Gangwani*** and ***8.S Hajela***, Applied Physics, DTU

13. Analysing the Impact of Covid-19 on Automobile Sales in India Using ARIMA Modelling, ***8.Suyash SRIVASTAVA***, ***8.Vasu SOOD*** and ***3.N. YUVRAJ***, Production and Industrial and Mechanical, DTU

14. Application to reduce distraction caused by phone calls while driving, ***8.Avinash Kumar***, ***8.Mohak Gupta***, ***8.Omkar Singh*** and ***3.Anamika Chauhan***, IT, DTU

15. Biosynthesis, characterization, and evaluation of antibacterial and photocatalytic dye degradation activities of silver nanoparticles biosynthesized by Chlorella sorokiniana, ***6.Lakhan Kumar*** and ***3.Navneeta Bharadvaja***, Biotechnology, DTU

16. Broadband plasmonic switches based on nanodisc-dimers with progressively increasing diameters on a plasmonic film with a VO2 spacer, ***6.Kirti Dalal*** and ***3.Yashna Sharma***, Electronics, DTU

17. Compact circularly polarized 2 and 4 port multiple input multiple output antennas with bandstop filter isolation technique, Hashinur Islam, ***7.1.Saumya Das***, Tanweer Ali, Tanushree Bose, Sourav Dhar, Bijay Rai and Pradeep Kumar, Electronics, DTU

18. Computational Analysis of Natural Compounds as Cyclin-Dependent Kinase-5 Inhibitors for Alzheimer's and Parkinson's Disease, ***6.Dia Advani*** and ***3.Pravir Kumar***, Biotechnology, DTU

19. Control algorithm for coordinated operation of wind-solar microgrid standalone generation system, ***6.Sombir Kundu***, ***3.Madhusudan Singh***, and Ashutosh K. Giri, Electrical, DTU

20. Correction to: Integrating the Particle Swarm Optimization (PSO) with machine learning methods for improving the accuracy of the landslide susceptibility model, Sunil Saha, Anik Saha, Bishnu Roy, ***3.Raju Sarkar***, ***7.Dhruv Bhardwaj*** and Barnali Kundu, Civil, DTU

21. COVID-19 Detection from Chest X-rays Using Trained Output Based Transfer Learning Approach, ***3.Sanjay Kumar*** and ***8.Abhishek Mallik***, CSE, DTU

22. Crowdsourcing a wellspring of value co-creation: an integration of social capital and organizational learning mechanisms, ***6.Divya Mishra*** and ***3.Nidhi Maheshwari***, USME, DTU

23. Deep auto encoder based on a transient search capsule network for student performance prediction, ***6.Rahul*** and ***3.Rahul Katarya***, CSE, DTU

24. Development of A Fully Faired Recumbent Bike using A Three-Piece Mold, ***8.Hrishabh Jha***, ***8.Avhishek Jha***, ***8.Kunal Nath*** and ***3.Rajesh Kumar***, Mechanical, DTU

25. DFIG-Based Wind Plant Coupled Series Compensated Transmission Line: Modeling and SSR Stability Analysis, ***6.Neevatika Verma*** and ***3.Narendra Kumar***, Electrical, DTU

26. Distributed Network of Adaptive and Self-Reconfigurable Active Vision Systems, ***6.Shashank*** and ***3.Indu Sreedevi***, Electronics, DTU

27. Dual Input Common LLC-Resonant Tank based off-Grid Charger, ***3.Vishal Verma*** and ***7.Prajeet Shukla***, Electrical, DTU

28. Effect of Plasma on the Nucleation and Growth of Vertical Carbon Nanotube Array on 2-Dimensional Graphene, Vedant Choukseya, Meetali Choudhary, Agastya Sethi, Neha Jain and ***3.Suresh C. Sharma***, Applied Physics, DTU

29. Effect of Polyvinyl Alcohol on the Growth, Structure,Morphology, and Electrical Conductivity of Polypyrrole Nanoparticles Synthesized via Microemulsion Polymerization, ***8.1.Anurag Krishna***, ***8.Amit Kumar*** and Rajiv Kumar Singh, Applied Chemistry, DTU

30. Effective Capacity Analysis over Cascaded Rayleigh Fading Channels with Arbitrary Correlation, Sandeep Kumar, Puspraj Singh Chauhan, ***8.Manpreet Kaur*** and Tarun Lata, CSE, DTU

31. Effects of Er3+ dopant on polymorphic phase boundaries and electrical properties of Sn4+ modified Ba1− yErySn0.06Ti0.94O3 multifunctional ceramics, Mohd Azaj Ansari, A. R. James, K. Zimik, ***6.Renuka Bokolia***, R. S. Kundu and K. Sreenivas, Applied Physics, DTU

32. End-to-End Convolutional Neural Network Feature Extraction for Remote Sensed Images Classification, ***6.Abebaw Alem*** and ***3.Shailender Kumar***, CSE, DTU

33. Enhancing wind power forecasting from meteorological parameters using machine learning models, ***6.Upma Singh*** and ***3.M. Rizwan***, Electrical, DTU

34. Explicate Toxicity By eXplainable Artificial Intelligence, ***7.Nakul Tanwar***, ***6.Jaishree Meena*** and ***3.Yasha Hasija***, Biotechnology, DTU

35. Fabrication of a Molecularly Imprinted Nano-Interface-Based Electrochemical Biosensor for the Detection of CagA Virulence Factors of H. pylori, Kirti Saxena, Bayu Tri Murti, Po-Kang Yang, ***3.Bansi Dhar Malhotra***, Nidhi Chauhan and Utkarsh Jain, Biotechnology, DTU

36. Fake News Detection Using Passive-Aggressive Classifier, ***8.Saloni Gupta*** and ***3.Priyanka Meel***, IT, DTU

37. Framework for detection of probable clues to predict misleading information proliferated during COVID-19 outbreak, ***6.Deepika Varshney*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

38. Geometric Properties of Functions Containing Derivatives of Bessel Functions, Sercan Kazmoglu and ***6.Kamaljeet Gangania***, Applied Mathematics, DTU

39. Green Emission from Calcium Zirconium Silicate Doped with Trivalent Terbium Phosphor, Vijay Singh, Ch. B. Annapurna Devi, B. Rupa Venkateswara Rao and ***3.A. S. Rao***, Applied Physics, DTU

40. Image Caption Generator in Hindi Using Attention, ***8.Abhishek SETHI***, ***8.Aditya JAIN*** and ***3.Chhavi DHIMAN***, Electronics, DTU

41. Implementation of Secure Communication System Using Chaotic Masking, ***6.S Suchit*** and ***3.Kriti Suneja***, Electronics, DTU

42. Improving the Scalability of GPU Synchronization Primitives, ***8.1.Preyesh Dalmia***, Rohan Mahapatray, Jeremy Intanz, Dan Negrut and Matthew D. Sinclair, Electronics, DTU

43. INFLUENCE MAXIMIZATION IN SOCIAL NETWORKS USING DISCRETIZED HARRIS' HAWKS OPTIMIZATION ALGORITHM AND NEIGHBOR SCOUT STRATEGY, ***8.Inder Khatri***, ***8.Arjun Choudhry***, ***8.Aryaman Rao***, ***8.Aryan Tyagi***, ***3.Dinesh Kumar Vishwakarma*** and Mukesh Prasad, IT, Mechanical, DTU

44. Influence of calcination and sintering temperature on the microstructure, dielectric, ferroelectric and piezoelectric properties of the lead-free KNN ceramics, ***6.Komal Verma***, Sahil Goel and ***3.Richa Sharma***, Applied Physics, DTU

45. Investigating the Economic and Environmental Sustainability of Logistic Operations in India Using ARDL Procedure, ***6.Akanksha SINGH***, ***6.Shweta KUMARI***, Sanjay kumar PANDEY, Bharat SINGH and ***3.Nand KUMAR***, Humanities, DTU

46. Isolation Enhancement using CSRR Slot in the Ground for Compact Two-Element Textile MIMO Antenna, Navneet Sharma, Anubhav Kumar, ***3.Asok De*** and R. K. Jain, Electronics, DTU

47. M&A performance in emerging markets: do they behave in unison or otherwise?, ***6.Sakshi Kukreja***, ***3.Girish C. Maheshwari*** and ***3.Archana Singh***, DSM, DTU

48. MCD: A MODIFIED COMMUNITY DIVERSITY APPROACH FOR DETECTING INFLUENTIAL NODES IN SOCIAL NETWORKS, ***8.Aaryan Gupta***, ***8.Inder Khatri***, ***8.Arjun Choudhry*** and ***3.Sanjay Kumar***, Applied Mathematics, IT and CSE, DTU

49. Measuring Reliability of Large Language Models through Semantic Consistency, ***8.Harsh Raj***, Domenic Rosati and Subhabrata Majumdar, Electronics, DTU

50. Mechanical and Tribological Properties of AA6061/SiC/Aloe Vera Powder Hybrid Al Composites Fabricated by Stir Casting, ***3.N. Yuvaraj***, Yashwant Koli, G. Vedabouriswaran and ***7.1.Plash Issar***, Mechanical, DTU

51. Mechanical properties of high strength concrete incorporating Chopped Basalt fibers: Experimental and Analytical Study, ***6.Mohit Gupta***, ***3.Dr. Ritu Raj*** and ***3.Dr. Anil Kumar Sahu***, Civil, DTU

52. Metasurface Inspired Wideband High Isolation THz MIMO Antenna for Nano Communication including 6G Applications and Liquid Sensors, ***6.1.Gaurav Saxena***, Sanjay Chintakindi, Mohsin Ahmed Kasim, Praveen Kumar Maduri, Y K Awasthi, Sanjay Kumar, Sahil Kansal, Rishabh Jain, Manish Kumar Sharma and Charul Dewan, Electronics, DTU

53. Numerical Analysis of Aerodynamic Drag for the Hyperloop System, ***8.Anmol PRABHAKAR***, ***8.Arya SHEKHAR***, ***8.Bhuvan JAIN***, ***3.Mohammad ZUNAID*** and ***3.Naushad Ahmad ANSARI***, Mechanical, DTU

54. Performance Analysis of Non-Ideal DC-DC Positive Output Elementary Superlift Luo Converter, ***6.Ritambhra Katoch*** and ***3.Dheeraj Joshi***, Electrical, DTU

55. PLC and SCADA based Real Time Monitoring and Train Control System for the Metro Railways Infrastructure, ***6.Ishu Tomar***, ***3.Indu Sreedevi*** and ***3.Neeta Pandey***, Electronics, DTU

56. Power Quality Improvement with Seven Level Packed-E Cell for Renewable Applications, Sanjay Upreti, Shivam Kumar Yadav, Bhim Singh and ***3.Narendra Kumar***, Electrical, DTU

57. Probabilistic multiplicative unbalanced linguistic term set and its application in matrix games, ***6.Tanya Malhotra*** and ***3.Anjana Gupta***, Applied Mathematics, DTU

58. Product Packaging by E-commerce Platforms: Impact of COVID-19 and Proposal for Circular Model to Reduce the Demand of Virgin Packaging, ***8.Tanya Arora***, ***8.Sarvani Reddy Chirla***, ***8.Nimisha Singla*** and ***3.Lovleen Gupta***, Civil and Environmental Engineering, DTU

59. Pulse splitting cum width control for transient handling of LLC Resonant Converter, ***7.Prajeet Shukla*** and ***3.Vishal Verma***, Electrical, DTU

60. Reconfigurable Manufacturing System: A Review, ***6.Vipin Kumar***, ***6.Sreeraj Ramesan***, ***6.Vinay Kumar*** and ***3.Dr. A. K Madan***, Mechanical, DTU

61. Reversible Data Hiding Using an Improved Pixel Value Ordering and Complementary Strategy, ***3.Rajeev Kumar***, Neeraj Kumar and Ki-Hyun Jung, CSE, DTU

62. Sales and operations planning for delivery date setting in engineer-to-order manufacturing: a research synthesis and framework, ***8.1.Swapnil Bhalla***, Erlend Alfnes, Hans-Henrik Hvolby and Olumide Oluyisola, Mechanial, DTU

63. Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model with Convolution Network, Le Hoang Son, ***3.Akshi Kumar***, ***8.Saurabh Raj Sangwan***, ***8.Anshika Arora***, Anand Nayyar and Mohamed Abdel-Basse, CSE, DTU

64. Sensor network driven novel hybrid model based on feature selection and SVR to predict indoor temperature for energy consumption optimisation in smart buildings, Sachin Kumar, Zairu Nisha, ***3.Jagvinder Singh*** and Anuj Kumar Sharma, USME, DTU

65. Simulation of Compression Molding of Thermoplastics Using CAE Tools: Assessment of Operational Conditions for Mold Design Development, ***8.Annu Pawar*** and ***3.Roli Purwar***, Applied Chemistry, DTU

66. Simulation Tools and Testbeds for Internet of Things(IoT): "Comparative Insight", ***7.Abhishek Singh***, ***7.Himanshu Nandanwar*** and ***3.Anamika Chauhan***, IT, DTU

67. Solar Panels Crack Detection using Overhead Images, ***8.Saksham Checker***, ***8.Saksham Gupta*** and ***3.J. P. Kesari***, Applied Physics and Mechanical, DTU

68. Solid Waste Management in India: A State-of-the-Art Review, ***3.Anunay A. Gour*** and ***3.S.K. Singh***, Environmental, DTU

69. Structural and optical studies on Dy3+ doped Gd2Ti2O7 pyrochlore as white light emission, ***6.Umang Berwal***, ***3.Vinod Singh*** and ***3.Rinku Sharma***, Applied Physics, DTU

70. Supervision Software using Artificial Intelligence, ***8.Shantanu Shishodia***, ***8.Tanmay Garg***, ***8.Shubhanshu Chaudhary*** and ***3.Dr. Kapil Sharma***, IT, DTU

71. Sustainable Xylose-Based Non-Isocyanate Polyurethane Foams with Remarkable Fire-Retardant Properties, ***6.Pooja Singh*** and ***3.Raminder Kaur***, Applied Chemistry, DTU

72. SwitchNet: A modular neural network for adaptive relation extraction, Hongyin Zhu, Prayag Tiwari, Yazhou Zhang, ***7.1.Deepak Gupta***, Meshal Alharbi, Tri Gia Nguyen and Shahram Dehdashti, CSE, DTU

73. Tailoring the output performance of PVDF based piezo-tribo hybridized nanogenerator via B, N co-doped reduced graphene oxide, ***6.Shilpa Rana***, ***6.Vishal Singh*** and ***3.Bharti Singh***, Applied Physics, DTU

74. Textbooks and Their Free Availability on WWW: A Case Study of University of Delhi, ***3.LALITA*** and NARENDER KUMAR, Central Library, DTU

75. Toeplitz Determinants for a Class of Holomorphic Mappings in Higher Dimensions, ***7.Surya Giri*** and ***3.S. Sivaprasad Kumar***, Applied Mathematics, DTU

76. Underwater image enhancement using multiscale decomposition and gamma correction, ***6.Amarendra Kumar Mishra***, ***3.Mahipal Singh Choudhry*** and ***3.Manjeet Kumar***, Electronics, DTU

77. University of Delhi Library : Celebrating the Century, Narender Kumar and ***3.Dr. Lalita***, Central Library, DTU

78. Validation of an Information Diffusion based Health Social Media on Yoga and Ayurveda, ***3.RK Yadav***, ***3.Girish Kumar***, ***3.Naveen Kumar***, ***8.Pratyaksh Varshney***, ***8.Natwar Jha*** and ***8.Amit***, CSE and Mechanical, DTU

79. Voltage-tunable surface-enhanced Raman scattering substrates based on electroactive polymeric membranes containing plasmonic nanoparticles, ***3.YASHNA SHARMA***, ENES CAKMAK, HSIANG-KUO YUAN, ROBERTO GARCIA, DALE BATCHELOR, TUAN VO-DINH, TUSHAR GHOSH AND ANUJ DHAWAN, Electronics, 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*

Transportation in Developing Economies (2023) 9:2
https://doi.org/10.1007/s40890-022-00171-6

ORIGINAL ARTICLE

A Comparative Study of the Feasibility of Alternative Fuel Vehicles for Sustainable Transportation in India: A Hybrid Approach of DEMATEL and TOPSIS

Aman Raj[1] · Ankita Dan[1] · Vrinda[1] · Pravin Kumar[1]

Received: 26 March 2022 / Accepted: 15 September 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract
Due to the increasing contribution of vehicles to air pollution, rapid depletion of petroleum sources, considerable fluctuations in gasoline prices, and high dependency on politically volatile oil-producing nations, alternative fuel vehicles (AFVs) are gaining popularity amongst users as well as manufacturers. This study explores all the major factors that impact the feasibility of AFVs for cleaner and sustainable transportation. This study is based on the opinions of both the users and industry experts regarding screening the factors influencing the feasibility of AFVs in their adoption. An integrated framework of Decision-Making Trial and Evaluation Laboratory (DEMATEL) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) integrated tools has been used to determine the importance of the factors influencing the adoption of AFVs and compare the AFVs against these factors. It has been observed that people are giving more preference to electric vehicles (EVs) compared to solar, hydrogen, and biofuel-based vehicles to replace fossil-fuel vehicles. However, the adoption of EVs has major constraints, such as long charging time, insufficient charging infrastructure, and the high initial cost of vehicles. This study may help the governments to formulate effective policy and take initiatives for cleaner and sustainable transportation in India.

Keywords Government initiatives · Sustainability · DEMATEL · TOPSIS · Alternative fuel vehicles

Introduction

The world's three most challenging issues are control of greenhouse gas (GHG) emissions, use of renewable and unconventional sources of energy, and prevention of environmental degradation [1]. The transportation sector is found to be one of the major contributors to environmental pollution due to its dependency on gasoline as fuel [2]. With the development of environmental consciousness among the users, researchers and scientists have been forced to focus their research on alternative and clean technologies for transportation [3, 4]. Various AFVs have been identified as a promising way to decrease transportation-related pollution [5]. Growing demand for EVs has been observed due to the reliance on oil-producing nations which are politically unstable, rapid exhaustion of fuel supply, and severe fluctuation in fuel prices [6]. Major contributors to the energy crunch in emerging economies are overconsumption, insufficient infrastructure for power production and distribution, the disparity between energy supply and demand, inadequate utilization of renewable energy sources, inadequate energy storage, etc. [7]. AFVs may solve the problems related to environmental sustainability and efficient energy utilization [8]. The ever-increasing utility and consumer demand for passenger cars have resulted in substantial consumption of petroleum fuels, with the corresponding output of GHG. Governments have attempted to reduce these pollutants by moving to AFVs such as electric vehicles (EVs), hydrogen vehicles (HVs), biofuel vehicles (BFVs) [9].

✉ Pravin Kumar
pravin.dce@gmail.com

Aman Raj
ar7722j@gmail.com

Ankita Dan
ankitadan@hotmail.com

Vrinda
svrin94@gmail.com

[1] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

Published online: 31 October 2022

Among the various studies conducted for AFVs, EVs have emerged as one of the feasible options for sustainable transportation. The government of India has planned to develop hundreds of smart cities. Sustainable transportation and power supply are the important features of smart cities. EVs are advantageous over gasoline vehicles in areas such as maintenance, zero-emission, improved efficiency of powertrain, and, finally, better urban air quality [10]. The transport sector in India accounts for 20% of total commercial energy in the form of liquid and gaseous fuel [11]. In the reduction of global emissions, the application of electricity as power/fuel for transportation provides a contrasting contribution in comparison with fossil fuels. It has been observed that the potential savings in GHG emissions are approximately 90%, 25%, and 50–80% using EVs, hybrid electric vehicles (HEVs), and plug-in hybrid electric vehicles (PHEVs), respectively [12, 13]. Hydrogen is considered one of the alternatives for the replacement of gasoline, enhancing energy security. Some of the environmental challenges caused by the extensive use of fossil energy, like climate changes and pollutant emissions, can be resolved with hydrogen [14]. With a calorific value of 3 times higher than that of gasoline, hydrogen is the most eco-friendly fuel [15]. However, numerous studies are required to remove the technological constraints before the industrial usage of hydrogen as fuel [16, 17].

Biofuels and solar cells have received great attention from researchers' sources as renewable sources of energy [18]. Solar energy is a clean and renewable source of energy but, currently contributes just a small portion of total energy consumption and confronts many financial and environmental issues related to solar panels and photovoltaic cells [19]. India being an agriculture-based country can successfully develop biofuel technologies based on agricultural wastes for a significant reduction in the usage of conventional energy sources and, to some extent, help to address pollution issues [20]. Biofuels may be utilized to provide transportation fuel, electricity, and heat, which helps to reduce carbon emissions and promote rural development [21].

Many studies have been conducted to compare the feasibility of various AFVs. Consumer surveys, life cycle assessment, cost analysis, and well-to-wheel (WTW) analysis are considered as most prominent methods in comparative studies of AFVs. Li et al. [22] focused on a consumer survey to determine the willingness to pay (WTP) for various vehicle types. The study suggests that refueling time and accessibility, driving range, and emission reduction are key elements that influence the purchaser's decision. Sagaria et al. [23] examined the performance, energy efficiency, and cost-effectiveness of hydrogen vehicles and EVs. Acar and Dincer [24] analyzed gasoline, electric, biofuel, and hydrogen vehicles on each aspect of sustainability, vehicle range, fuel cost, and thermodynamic properties to establish hydrogen's potential as a sustainable fuel for transportation. Li et al., [25] also compared gasoline, electric, and hydrogen vehicles using a life-cycle cost model that incorporates tangible, intangible, and exterior expenses to examine the cost-effectiveness of traffic and environmental regulations. The majority of these research articles focus on environmental, economic, and infrastructure factors; however, these factors are not sufficient enough to assess the feasibility of each vehicle type for long-term sustainability. Extensive research is required to identify and incorporate all possible factors for the feasibility study of AFVs. Considering the above facts, this study has been conducted with the following objectives:

(a) To explore the factors influencing the feasibility of the use of different AFVs.
(b) To find the importance of these factors based on the users' perspectives and industry experts.
(c) To compare the AFVs based on the screened factors for performance evaluation.

The rest parts of the paper are arranged as: Literature review represents the literature review; DEMATEL- and TOPSIS-Integrated Model represents the research methodology; Sect. Research methodology shows the findings and discussion; and Sect. Conclusions concludes the research with limitations and scope for future research.

Literature Review

Alternative fuel-based vehicle technologies, such as EVs, hydrogen vehicles, biofuels, and solar vehicles, have been causing a substantial shift in the transportation sector and a reduction in fossil fuel dependencies [6]. EVs and hydrogen vehicles (HVs) are preferred over gasoline-based vehicles for lowering fossil fuel imports and consumer fuel expenses. Biofuel has an advantage over EVs and HVs in terms of consumer economic benefits and the percentage of alternative fuels [26].

Factors Influencing the Feasibility of AFVs

The opinions of the researchers vary regarding the suitability of different types of AFVs due to variations in preferences. These preferences may be the cost of the vehicle, fuel price, availability of fuel, availability of charging infrastructure and concerned technologies, etc. Onat et al. [27] have ranked different vehicles using the TOPSIS approach considering social, environmental, and economic factors for sustainability life cycle assessment. They observed that hybrid electric vehicles (HEVs) are the best option when all factors are taken into account. The ranking of the vehicles changes significantly when the weight assigned to environmental,

economic, and social factors changes. Acar and Dincer [24] observed that when comparing the fuel expenses of gasoline with electric, biofuel, and hydrogen vehicles, gasoline vehicles perform poorly, whereas EVs have the lowest fuel/electricity costs. In terms of WTW energy and emissions, comparing gasoline and hydrogen vehicles reveals that hydrogen vehicles consume less WTW energy and release less amount of greenhouse gases [28]. Table 1 shows the factors considered by the different researchers for the evaluation and comparison of the AFVs.

The major factors influencing the performance of AFVs and the feasibility of the adoption of AFVs were explored through the literature review and discussed with the users and industry experts. Based on the experts' opinions, some major factors were shortlisted for the analysis as discussed in the following paragraphs.

Economic Sustainability (F1)

A gradual transition from conventional to AFVs necessitates a comprehensive assessment of the economic implications of establishing a sustainable transportation system. The cost differentials between different vehicle types (gasoline, HEVs, PHEVs, and EVs) are insufficient. Tax credits play an important role in addressing the problem of inadequacies in cost differentials. With tax incentives, EVs and HEVs can be inexpensive, and their lifetime total costs are equivalent to those of gasoline vehicles, except for PHEVs with extended battery range, which is approximately 11–12% more expensive than gasoline vehicles [32, 31]. Hydrogen vehicles on the other hand are roughly three times more expensive than gasoline vehicles in terms of engine cost and four times more costly when the entire supply chain is considered [33].

End-of-life Options (F2)

The end-of-life options of a vehicle must be evaluated to determine the best solutions for maximizing economic, social, and environmental sustainability [34, 35]. AFVs require a variety of critical materials, which must be recycled to address issues of circular economy, cost savings, environmental sustainability, supply chain, and energy challenges [36, 37]. The Indian government plans to implement the Vehicle Scrappage Policy (VSP), which will generate over 9 million tons of steel scrap, 1 million tons of aluminum scrap, and 3-million tons of other materials such as plastics, rubber, and glass, providing a significant boost to India's circular economy and lowering GHG emissions [38].

The rise of EVs poses a significant challenge for waste lithium-ion battery (LIBs) management. Globally, it is anticipated that by 2030, the repurposed LIBs will produce 275 GWh/year [39]. Despite having a strong electronic waste management system, photovoltaic (PV) panel trash is still classed as general waste in India, and with India intending to install 100GW of solar panels by 2022, the waste generated would be significant in the coming years [40].

Government Incentives (F3)

The initial cost of EVs accounts for more than 70% of the total ownership cost. Hydrogen vehicles have significant upfront and operating expenditures [23]. AFVs become more competitive as a result of financial incentives (Jones et al. 2020). Government financial incentives on the initial cost of AFVs are preferred by customers [41, 42]. The average increase in EVs' sales is observed as 2.6–3.0% for every $1000 increase in rebates or tax incentives [43, 44]. In the case of solar technology, a similar outcome has been observed. Apart from the upfront cost discounts, consumers appreciate other benefits like tax refunds, energy bill rebates, free parking, and preferential road priority [41, 42, 45].

Environmental Sustainability (F4)

Despite the substantial incentives for EVs, the environmental benefits of transitioning from conventional gasoline vehicles to EVs in China are meager. During the pump-to-wheels (PTW) stage, BEVs and hydrogen vehicles have zero tailpipe emissions and hence outperform conventional gasoline vehicles in terms of environmental efficacy. In addition to this, the social cost of carbon in hydrogen vehicles is found to be the lowest compared to gasoline vehicles [24]. However, for most emission types, BEVs and hydrogen vehicles perform poorly throughout the well-to-pump and vehicle cycle stages, resulting in unobvious environmental benefits across the life cycle [25].

Biofuels are considered one of the possibilities for reducing emissions in Europe. Following a period of stagnation from 2012 to 2016, European biofuel usage increased in 2017 and 2018 and even accounted for 4.5% of the energy used in road transportation. Biohydrogen and upgraded biomethane are the most potential biofuels for minimizing GHG emissions [46]. Bioethanol has lower non-renewable energy requirements (NER) and GHG emissions than gasoline. Cellulosic bioethanol offers the best performance in terms of reducing the use of fossil fuels while utilizing biodiesel as a replacement for diesel can lead to a net reduction in fossil energy usage and carbon footprint [47].

PV (Photovoltaic) systems are environmentally safe once they have completed the production and installation phases. However, from the manufacturing phase to disposal, the possible environmental repercussions of such systems must not be neglected. PV production entails the use of hazardous

Table 1 Factors influencing the comparison of different types of vehicles for the feasibility study

Authors	Factors	Types of vehicles studied	Description of the study
Li et al. [25]	Initial cost of the vehicle, driving restriction rate, Environmental sustainability, Government subsidies, and fuel cost, intangible cost, external cost, life cycle cost and assessment, purchase tax, vehicle use tax, environmental tax	Hydrogen, electric, and gasoline vehicle	Conventional, electric, and hydrogen vehicles using a Life cycle cost model which included tangible, intangible, and external costs are compared
Sagaria et al. [23]	Initial cost of the vehicle, maintenance cost, vehicle Range, vehicle purchase cost and taxes, energy consumption, average distance traveled per year. technical specification of vehicles	Hydrogen and electric vehicles	The performance, energy efficiency, and cost-effectiveness of hydrogen and electric vehicles are examined
Verma et al. [13]	Initial cost of the vehicle, disposal cost, Environmental cost, operation cost, failure cost, support cost, maintenance cost, and fuel cost	Electric and gasoline vehicles	The Life Cycle Costs of Electric Vehicles and Gasoline Vehicles are compared
Forrest et al. [29]	Fuel efficiency and ranges, availability of charging/refueling infrastructure, environmental Sustainability, charging rates, initial cost of the vehicle, vehicle range, energy demand	Hydrogen and Electric vehicles	The viability of electric and hydrogen vehicles for satisfying the transportation needs of medium and heavy-duty vehicles is evaluated
Acar and Dincer [24]	Environmental impact, Social cost of carbon emission, energy efficiency, Range, Fuel/electricity consumption and cost	Hydrogen, Biofuels, and Electric vehicles	A comparative study between hydrogen and other alternative fuel technologies is conducted
Li et al. [22]	Purchase price of the vehicle, Environmental sustainability, Driving range, Refueling time, emission reduction, refueling accessibility	Hydrogen, gasoline, and electric vehicles	The willingness to pay for hydrogen vehicles comparing electric, hydrogen, and gasoline vehicles is measured based on several factors
Onat et al. [27]	Economic sustainability, Environmental Sustainability, Social Sustainability, Government taxes, employment, injuries	Electric and gasoline vehicles	The conventional and alternative vehicle technologies are compared based on sustainability indicators
Shafiei et al. [26]	Fuel cost, Charging/Refueling infrastructure, Environmental Sustainability, Initial cost of vehicle, reduction of total fuel demand, mitigation cost, and profitability of fuel supply, emissions reduction, share of alternative fuels	Hydrogen, biofuels and electric vehicles	The transitional approaches to future sustainable road transportation employing electric, hydrogen, and biofuels are compared
Onat et al. [30]	End-of-life option, Economic sustainability, Environmental sustainability, and Social Sustainability	Electric and gasoline vehicles	A comprehensive macro-level sustainability evaluation framework for alternative passenger vehicles is proposed
Tseng et al. [31]	Economic sustainability, Environmental sustainability	Electric and gasoline vehicles	A life cycle cost analysis is conducted to study the economic and environmental consequences of electric and hybrid electric vehicles

chemicals, including acids and solvents which can pose health and environmental threats [48, 49].

Maintenance Cost (F5)

The cheaper maintenance cost of AFVs is one of the major factors driving their adoption. EVs have a cheaper maintenance cost per km as compared to gasoline vehicles and require minimal maintenance as well [50]. EVs have advantages over gasoline vehicles in terms of zero tailpipe emissions, minimal maintenance, improved powertrain efficiency, and, finally, reduced urban air pollution [10]. Concerning the adoption of hydrogen vehicles, lower reliability, higher cost of maintenance, and a scarcity of modern hydrogen refueling stations are some of the major barriers. Effective health monitoring of hydrogen refueling stations and vehicles can lower maintenance costs while increasing reliability [51]. The deployment of solar energy systems is supported by the costs of installation, maintenance, and operation, as well as monetary reforms. One of the significant factors influencing the adoption of solar technology is the cost of maintenance [52].

Charging/Refueling Time (F6)

Users are accustomed to refueling gasoline vehicles, which is quite a rapid process; this is especially problematic in the case of EVs because they need to wait substantially longer to recharge their vehicles. They are not ready to wait for a longer charging time for the batteries. People desire to have the charging time under 5 min [53]. The majority of the users favor faster EV charging technology and facilities. While it takes around 4 min to refuel a gasoline vehicle, EVs take around 4–8 h to charge at a lower power level and at least 30 min to charge at a fast-charging station. The desire to purchase an EV can be increased considerably by reducing the charging time [54]. The time of refueling hydrogen is equal to that of gasoline vehicles. Hydrogen vehicles can overcome the shortcomings of EVs like long charging/refueling time [22].

Fuel/Electricity Cost (F7)

One of the driving factors toward the adoption of AFVs is low fuel cost [55]. The performance of gasoline vehicles is poor, while EVs have the lowest cost in terms of electricity prices. The idea of acquiring the cheapest available electricity from an already existing and well-developed grid link is one factor for EV's good performance [24]. In the case of biofuel vehicle adoption, biofuel pricing is considered one of the important factors influencing customer willingness to pay [56].

Although EVs are the best option for minimizing fuel demands, hydrogen vehicles are seen as a viable alternative for lowering fossil fuel imports and consumer fuel expenses [26].

Vehicle Range (F8)

AFVs are becoming increasingly popular around the world as a result of their low fuel consumption and lower local pollution. However, in comparison with conventional automobiles, the range of these vehicles is still limited [57]. EVs meet the range needs of a significant portion of the driving population [58]. The majority of the customers demand long-distance travel, i.e., they prefer vehicle range as one of the driving factors when purchasing any vehicle [59]. Customers are ready to pay a 20% premium for increased vehicle range [22]. Customers considered hydrogen vehicles as a viable choice and preferred them over EVs because of their range [60, 59]. Vehicle range is affected by other factors like environmental and vehicle operating conditions, giving gasoline vehicles a significant advantage over the EV in this regard [57]. The limited range of EVs is cited as one of the key bottlenecks to adoption. Limited vehicle range of solar vehicles is considered a major factor in their low feasibility as AFVs [61].

Supply Chain Complexity (F9)

With the increasing demand for AFVs, it is imperative to assess the uncertainties and complexities in their supply chain. The growth in the use of EVs has been found to have a significant increase in the demand for batteries. Policymakers face enormous hurdles in controlling operations for sustainable EV battery supply chains [6]. Considering the hydrogen supply chain, the physicochemical properties of hydrogen pose significant difficulties for mass implementation and usage in the current environment. The hydrogen supply chain faces problems such as rising costs and a lack of infrastructure in the global energy shift [62]. The biofuel supply networks face problems in their operation due to the seasonal production and annual unpredictability of biomass availability, uncertainty in production and transportation, and unpredictability in demand and price [63, 64].

The forward and reverse supply chains for solar panels have received very little attention. A closed-loop supply chain for solar panels is very imperative since it is both cost-effective and environmentally friendly [65]. The growing reliance on only a few countries like China for the import of critical raw materials required for solar panels may lead to supply bottlenecks and price instability [66].

Possibility of Usage in Public Transport (F10)

Governments' actions should be oriented to facilitate public transportation following the concepts of sustainable urban development and "smart city" [67]. The suggested biodiesel blend of 20–50% in public transportation has been found to have a significant impact on urban city air quality [68]. Hydrogen buses can reduce CO_2 emissions by 93% when compared to diesel buses [69]. Hydrogen buses have the same range and operating hours per day as diesel buses, as well as the same capabilities [70]. Battery-powered buses are a viable option. Hydrogen-powered buses can excel in the medium to long term [69]. The main aspects of the electric bus fleet are the optimal bus routes, charging technique, and charging infrastructure availability in the deployment of electric buses [71, 72].

Social Sustainability (F11)

The social dimension of sustainability is understudied in the studies of AFVs [73, 30]. Only focusing on the economic or environmental dimensions of sustainability may not be the optimal decision considering the social issues [27, 30]. While examining the long-term viability of EV charging stations, experts placed the greatest emphasis on social and economic factors [74]. As many firms in the automobile industry strive to include all three parts of sustainability in their operations, the focus should shift to integrating the social aspects of vehicles with the environmental and economic aspects [75].

Safety (F12)

The users' confidence in the equipment and its safety measures is more important than the other parameters [76]. The rising number of EV accidents has shifted the focus to its safety. The breakdown of the battery system and the low thermal stability of elements used in LIB are the key causes of EV failure [1]. Similarly, the physicochemical properties and combustion nature of hydrogen gas, which include wide flammability limits (4–75%), low minimum ignition energy, and fast laminar burning velocity, make it more harmful than the hydrocarbons widely used as automotive fuels [77–79]. The whole safety of the refueling station must also be investigated for the safe usage of hydrogen as a fuel [76, 78].

Initial Cost of Vehicle (F13)

The high initial cost of EVs is identified as a major hurdle to the adoption of EVs. People with poor purchasing power are unable to purchase high-priced HEVs and E4Ws (Electric four-wheelers). E4Ws and HEVs will not gain appeal in poor countries unless economies of scale bring them down to a reasonable price [80]. Hydrogen vehicles have significant upfront and operating expenditures, making them the least popular among customers. Due to the high upfront price of hydrogen vehicles, high fuel costs, and a lack of refueling facilities, hydrogen vehicles emerged as the most expensive option among light-duty vehicles [23]. The relatively modest initial cost of the vehicle offsets the rising biofuel price. The initial cost of a hydrogen vehicle is the highest, which is a substantial barrier to adoption [26].

Charging/Refueling Infrastructure (F14)

If the hydrogen vehicles are to gain sufficient market penetration beyond the early adopters, refueling availability for hydrogen vehicles (and any other AFVs) must be perceived as adequate by the general users, which necessitates the consideration of their preferences for hydrogen refueling infrastructure design. The most effective option for the quick deployment of hydrogen vehicles is to create a hydrogen refueling infrastructure [81]. Battery performance is linked to driving range limitation, recharging time, acceleration, and cost savings, whereas the price of an EV is heavily influenced by its battery cost. Consumers' biggest complaints about EVs, are long recharging times, limited driving range, and expensive pricing [82].

Several studies have identified a significant number of impediments to the adoption of EVs and hydrogen vehicles, both in India and globally. There has been relatively little research comparing all of the different vehicle alternatives. In most of the research, vehicle comparisons have been made only based on economic, environmental, social, and infrastructure factors. However, factors like supply chain complexity, the possibility of usage in public transport, end-of-life options, and safety have not been given much preference. In the present study, we have considered a broad view for the selection of criteria to evaluate the feasibility of alternate fuel vehicles.

DEMATEL- and TOPSIS-Integrated Model

It is used to find the feasibility ranking of the AFVs. DEMATEL is used to provide the weights to the different factors affecting the feasibility of the adoption of AFVs and TOPSIS is used to rank the AFVs. The DEMATEL- and TOPSIS-integrated framework has many applications in the area of decision-making. It is prescribed that the sample size for the ideal usage of DEMATEL should be within the range of 8–45 [83]. According to Asadi et al. [84], one of the main advantages of DEMATEL compared to other techniques is its potential to provide probable outcomes using minimal data. Additionally, Sekhar et al. [85] have used sample size of 52 for TOPSIS without affecting the results. Some of the major works are summarized in Table 2.

Table 2 Applications of the integrated model of DEMATEL and TOPSIS

References	Application areas
Li et al. [86]	Comprehensive risk assessment of hydrogen production unit
Ortiz-Barrios et al. [87]	Identification of high- and low-performing pork suppliers
Mohammed et al. [88]	Ranking of suppliers based on supply chain resilience
Kilic and Yalcin [89]	Ranking of municipal performance based on environmental sustainability and other aspects
Chang et al. [90]	Ranking the risk of failure identified through Failure Mode and Effects Analysis (FMEA)
Aragão et al. [91]	A selection of initiatives showcasing smart city business models

Research Methodology

This study is based on the opinion of vehicle owners/users, experts from auto manufacturing firms, and service providers on charging stations and service stations of vehicles. A questionnaire is used to take the opinions of the different respondents. These opinions are used to screen the major factors influencing the feasibility of the adoption of AFVs. To find the importance rating/weightage of the factors, DEMATEL is used and these weights are used in TOPSIS to find the rank of AFVs. The framework used for the research methodology is shown in Fig. 1.

Experts' Opinions Through a Questionnaire Survey

The complete analysis of the factors affecting the feasibility of the adoption of alternative fuel vehicles is based on the users' perspectives. The opinions were taken from the concerned people of different parts of India through a questionnaire survey. Initially, 60 users including the people concerned with charging and fueling stations and the automobile manufacturing sector were contacted but only 43 respondents responded with their opinions regarding the pairwise comparison of the factors and rating of the different types of vehicles with respect to these factors. Out of 43 respondents,

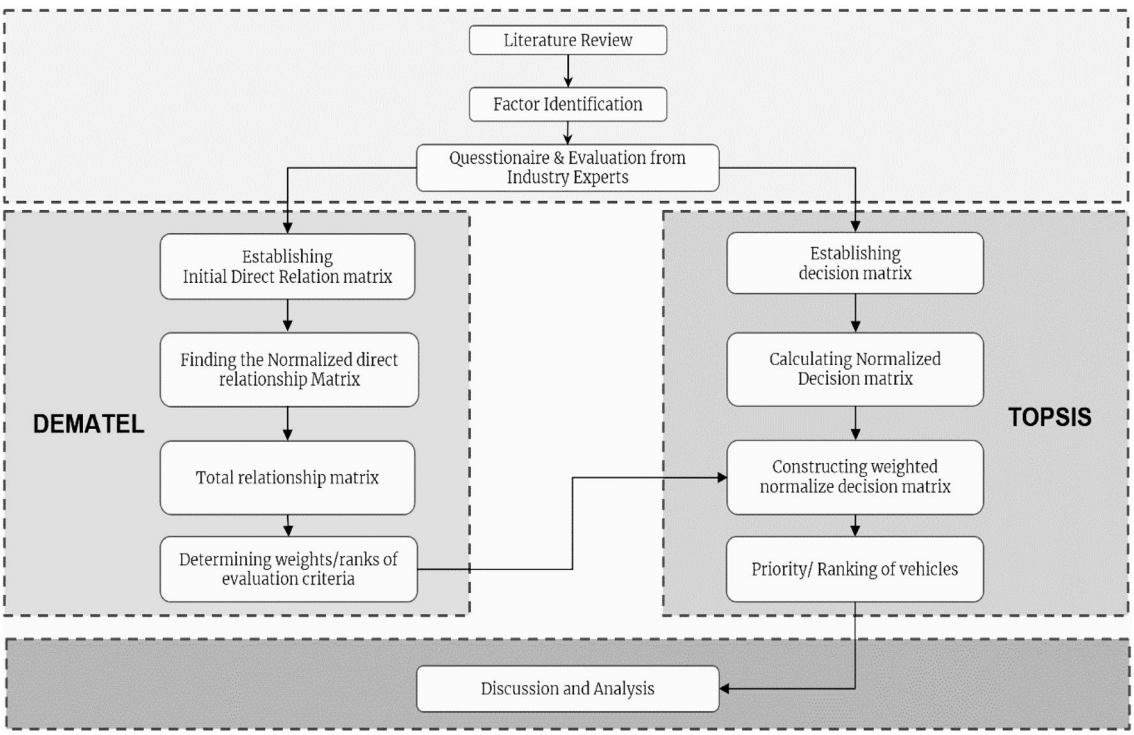


Fig. 1 Framework for the proposed research methodology

7 respondents were concerned with the automobile manufacturing companies, 13 respondents were concerned with the charging/fueling stations, and 23 respondents were owning their passenger vehicles. They were provided all the factors explored through the literature review as shown in Column 2, Table 1, and asked to rate the importance of these factors on a 5-point rating scale. Based on the ratings, only 14-factors having an average rating of more than 3 on a 5-point rating scale are incorporated in the analysis.

In the second part of the questionnaire, they were asked to make a pairwise comparison of all the 14 factors to show the influence of one factor over another for DEMATEL. The most common responses were incorporated into the analysis to ignore the fractional average value of the response. In the third part of the questionnaire, the respondents were asked to rate the AFVs against all the 14 factors for rating the AFVs using TOPSIS. Again, the most common responses were considered for the rating of AFVs. DEMATEL provides the importance rating of the feasibility factors, which is used as inputs in TOPSIS for the rating of AFVs. The adoption of AFVs in India is in a very initial phase; therefore, a limited number of respondents are involved in this survey as per their involvement in AFV operations.

DEMATEL (Decision Making Trial and Evaluation Laboratory)

DEMATEL was first proposed by Gabus and Fontela in the early 1970s [92]. It uses matrices or digraphs to illustrate intricate, structural, and causal relationships and can transform a relationship between causes and effects of criteria into a unique structural model [93]. The DEMATEL approach can be employed to quantify the interdependence of several interacting risks, which includes both direct and indirect influences from numerous causes [86]. The following mathematical steps are followed in the DEMATEL approach:

Step 1 Identification of the set of factors influencing the feasibility of AFVs through the literature survey.

Step 2 Construction of initial direct relationship matrix A using Eq. (1).

$$A = \left[a_{ij}\right]_{n\times n}, \tag{1}$$

where a_{ij} shows the direct relationship of factor i with factor j.

Step 3 Normalization of direct relationship matrix using Eq. (2).

$$D = k \times A, \tag{2}$$

where $k = \dfrac{1}{\max\left(\sum_{j=1}^{n} a_{ij}\right)}$.

Step 4 Establishing total relationship matrix: with the help of normalized matrix *'D'*, the total relationship matrix '*T*' is calculated as shown in Eq. (3).

$$T = D + D^2 + \cdots + D^m = D(I - D)^{-1}, \tag{3}$$

where I is the Identity matrix.

Step 5 Calculation of casual parameters *'R'* and *'C'*: these casual parameters are calculated using Eqs. (4) and (5)

$$T = \left[t_{ij}\right], i, j = 1, 2, \ldots, \ldots, n,$$

$$R = \left[r_i\right]_{n\times 1} = \left[\sum_{j=1}^{n} t_{ij}\right]_{n\times 1}, \tag{4}$$

$$C = \left[c_j\right]_{n\times 1} = \left[\sum_{i=1}^{n} t_{ij}\right]_{n\times 1}. \tag{5}$$

Here r_i is the sum of the ith row in the matrix T, which represents the sum of the impact of a factor i on the other factors. Similarly, c_j is the sum of the jth column in the matrix T, which represents the sum of the impact of a factor j on the other factors.

Step 6 Calculation of factors' weight.

To calculate the factors' weight, Eq. (6) is used [94].

$$w_i = \sqrt{(R + C)^2 + (R - C)^2}. \tag{6}$$

Using Eq. (8), the w_i weights are normalized,

$$W_i = \frac{w_i}{\sum_{i=1}^{n} w_i}. \tag{7}$$

TOPSIS (Technique for Order Performance by Similarity to Ideal Solution)

TOPSIS, one of the most well-known classical Multi-Criteria Decision Attributes (MCDA) approaches, was presented by Hwang and Yoon [95]. The method's main goal is to discover the optimal alternative that is the lowest distance from the positive ideal solution while also being the farthest away from the negative solution [96]. Its key advantage over other MCDA methods is the lower number of parameters required for implementation, which reduces subjectivity [97]. The following steps are used for the application of TOPSIS.

Step 1 The decision matrix A is built using expert opinion and real measurable data as shown in Eq. (8). The columns reflect the factors for evaluating the feasibility

of each vehicle, whereas the rows represent the decision-making unit (DMU), i.e., AFVs.

$$X = \left[x_{ij}\right]_{m\times n}, \tag{8}$$

where $i = 1,2,\ldots, m$ represents the decision-making units and $j = 1,2,\ldots,n$ represents the factors against which the DMUs are rated.

Step 2 The decision matrix X is normalized using Eq. (9):

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}, \forall i = 1 \ldots m, \forall j = 1 \ldots n, \tag{9}$$

where x_{ij} is the rating of the ith DMU for the jth criterion.

Step 3 The weighted normalized decision matrix V is used using Eq. (10)

$$V = \left[v_{ij}\right]_{m\times n}, \tag{10}$$

where $v_{ij} = w_j \cdot r_{ij}, \forall i = 1 \ldots m$ and $\forall j = 1 \ldots n$ and $\sum_{j=1}^{n} w_j = 1$.

Step 4 The positive ideal solution V^+ and negative ideal solution V^- were determined using Eqs. (11) and (12), respectively.

$$V^+ = \left\{v_1^+, \ldots, v_n^+\right\} = \left\{\left[\max v_{ij} | j \in I'\right]\right\}, \left\{\left[\min v_{ij} | j \in I''\right]\right\}, \forall i, \tag{11}$$

$$V^- = \left\{v_1^-, \ldots, v_n^-\right\} = \left\{\left[\min v_{ij} | j \in I'\right]\right\}, \left\{\left[\max v_{ij} | j \in I''\right]\right\}, \forall i, \tag{12}$$

where I' and I'' are the benefit and cost factors, respectively. The cost factors are also represented as benefit criteria in this study because each DMU is rated on a five-point scale, with 1 being the worst and 5 being the best, and the cost criteria follow the same rating scale. As a result, the ideal positive V^+ and ideal negative solutions V^- are customized as Eqs. (13) and (14):

$$V^+ = \left\{v_1^+, \ldots, v_n^+\right\} = \left\{\left[\max v_{ij} | j \in I'\right]\right\}, \tag{13}$$

$$V^- = \left\{v_1^-, \ldots, v_n^-\right\} = \left\{\left[\min v_{ij} | j \in I'\right]\right\}, \tag{14}$$

Step 5 The relative distances from positive and negative ideal solutions are calculated using Eqs. (15) and (16).

$$S_i^+ = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^+)^2}, \forall i = 1 \ldots m, \tag{15}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^-)^2}, \forall i = 1 \ldots m. \tag{16}$$

Step 6 The final step combines the two distances and uses Eq. (17) to calculate the closeness coefficient C_i

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-}; 0 \le C_i \le 1. \tag{17}$$

Results and Discussion

Based on the opinion of the experts', 14 factors were screened for the analysis of AFVs. These factors are Economic Sustainability (F1), End-of-life options (F2), Government Incentives (F3), Environmental sustainability (F4), Maintenance cost (F5), Charging/Refueling time (F6), Fuel /Electricity cost (F7), Vehicle Range (F8), Supply

Table 3 Initial direct relation matrix D

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
F1	0	5	4	2	3	4	4	2	3	2	3	2	3	2
F2	2	0	4	4	4	3	5	5	5	4	4	4	4	5
F3	3	3	0	4	4	2	4	3	4	4	5	4	4	5
F4	4	3	3	0	5	5	5	4	5	4	4	5	4	5
F5	2	2	2	2	0	2	3	2	3	2	3	3	2	3
F6	5	3	3	4	4	0	4	5	5	5	5	4	4	5
F7	2	2	2	2	3	2	0	2	3	2	2	2	2	3
F8	4	4	2	3	4	4	4	0	4	4	4	3	4	4
F9	3	3	2	3	3	2	3	2	0	2	3	3	3	2
F10	3	3	3	4	3	3	3	3	3	0	4	2	3	2
F11	2	3	3	2	2	2	3	3	2	3	0	3	3	2
F12	4	4	3	4	4	3	4	3	4	4	4	0	5	5
F13	5	3	3	3	5	4	4	4	5	5	5	3	0	4
F14	5	4	3	4	4	4	5	3	4	4	5	4	5	0

Table 4 Causal parameters R and C

	R	C	$R+C$	$R-C$	Cause/effect	Weights
F1	3.298	3.674	6.973	− 0.376	Effect	0.066
F2	4.399	3.529	7.928	0.870	Cause	0.076
F3	4.062	3.130	7.192	0.932	Cause	0.069
F4	4.620	3.429	8.049	1.191	Cause	0.077
F5	2.622	3.990	6.612	− 1.368	Effect	0.064
F6	4.611	3.360	7.970	1.251	Cause	0.077
F7	2.468	4.235	6.703	− 1.767	Effect	0.066
F8	3.989	3.428	7.418	0.561	Cause	0.071
F9	2.879	4.140	7.020	− 1.261	Effect	0.068
F10	3.288	3.737	7.026	− 0.449	Effect	0.067
F11	2.827	4.221	7.048	− 1.394	Effect	0.068
F12	4.258	3.515	7.773	0.743	Cause	0.074
F13	4.316	3.825	8.141	0.490	Cause	0.078
F14	4.454	3.876	8.330	0.578	Cause	0.079

Table 5 Decision matrix X

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
Weights →	0.066	0.076	0.069	0.077	0.064	0.077	0.066	0.071	0.068	0.067	0.068	0.074	0.078	0.079
EVs	3	5	5	4	5	3	4	2	3	5	3	2	1	3
Hydrogen vehicles	2	3	4	5	3	5	4	4	2	4	5	2	4	1
Solar vehicles	2	3	4	5	5	1	5	1	3	1	5	5	2	4
Bio-fuel-based vehicles	1	2	3	3	1	5	2	5	2	2	3	4	5	1

chain complexity (F9), Possibility of usage in public transport (F10), Social Sustainability (F11), Safety (F12), Initial Cost of Vehicle (F13), and Charging/Refueling infrastructure (F14). All these factors were assigned the weights through DEMATEL. Based on the common opinion of the respondents, a direct relationship matrix is produced as shown in Table 3 using Eq. (1).

Using Eq. (2), the direct relationship matrix as shown in Table 3 is normalized. The normalized matrix is shown in Table 7 in Appendix. Using Eqs. (3–5), a total-relationship matrix is produced which is shown in Table 8 in Appendix. Finally, using Eqs. (4–7), weights of the factors and causal parameters are determined as shown in Table 4. The weights of the factors are used as input in the TOPSIS used for finding the ranks of the AFVs. Here,

The weights of the factors determined through DEMATEL are used for calculation in TOPSIS to find the priority/rank of AFVs. In this study, electric vehicles (EVs), hydrogen vehicles (HVs), solar vehicles (SVs), and bio-fuel vehicles (BFVs) are considered AFVs. Using Eq. (8) of TOPSIS, a decision matrix 'X' is prepared for the rating of the AFVs with respect to the factors as shown in Table 5. These ratings are based on the most common opinion of the respondents on the 5-point rating scale.

Table 6 Rank of DMU based on closeness coefficient C_i

	S_i^+	S_i^-	C_i	Rank
EVs	0.071628	0.083655	0.538723871	1
Hydrogen vehicles	0.069109	0.078674	0.532360887	2
Solar vehicles	0.082741	0.078674	0.487402313	3
Bio-fuel based vehicles	0.089242	0.077099	0.463498208	4

Using Eq. (10), the decision matrix is normalized which is shown in Table 9 in Appendix. The weighted normalized matrix with positive and negative ideal solutions is shown in Table 10 in Appendix. Using Eqs. (15–16), the distance between positive and negative ideal solutions is produced for all the AFVs. Finally, closeness indices are calculated for all the AFVs using Eq. (17). The distance from the positive and negative ideal solutions and closeness to the ideal solution are summarized in Table 6.

All the factors can be divided into three tiers impacting the feasibility of AFVs. The *first-tier factors* include Charging/Refueling infrastructure (F14), Initial Cost of Vehicle (F13), Environmental sustainability (F4), and Charging/Refueling time (F6). In the *second-tier factor*, there are a variety of aspects to consider. The five factors in the second tier include End-of-life options (F2), Safety (F12), Vehicle

Range (F8), Government Incentives (F3), and Social Sustainability (F11). The *third-tier factors* include Supply chain complexity (F9), Possibility of usage in public transport (F10), Economic Sustainability (F1), Fuel /Electricity cost (F7), and Maintenance cost (F5). The factors in the first tier are the directly impacting factors when deciding the suitable alternative for gasoline vehicles. These second-tier factors have a strong interdependence on the first-tier as well as third-tier factors influencing the feasibility of AFVs. The third-tier factors have a direct influence on the second-tier factors and indirectly on the top-tier factors, via second-tier factors.

It has been observed that the most feasible alternative to gasoline vehicles is EVs; it is cleaner than other alternative vehicles. EVs also provide less running cost in terms of electricity bills for charging the batteries but major constraints are the supply of electricity, charging infrastructure, availability of fast charging technologies, and the high initial cost of the vehicles, etc. Knez et al. [98] observed EVs as zero-emission vehicles. But they also observed the high initial cost as a major challenge in the adoption of EVs. Zhao et al. [99] suggested that mini-electric cars would always be the most credit cost-effective. Moreover, 350 km will be the optimal driving range under the credit regulation in China. If the constraints related to infrastructure are eliminated, the EVs will be the first choice of the people.

Hydrogen vehicles are the second-best alternative to gasoline as they provide a longer range and have significantly lesser refueling time, in comparison with the recharging time of EVs. The feasibility of hydrogen vehicles is hampered by poor refueling infrastructure, vehicle's high initial cost, safety, and fuel costs. The remaining two vehicle types, solar and biofuel, are not significantly viable alternatives due to their poor effectiveness in the top and mid-tier factors. The economic and environmental performance of the biofuel vehicle, as well as the availability of cultivable land, are key factors impacting its viability. The primary reason that biofuel vehicles are the least viable option is that increasing biofuel production would result in the conversion of non-cropland lands, such as forests, wetlands, and peatlands, to agricultural lands. A shift in land use might greatly diminish the potential advantages of GHG savings from increasing biofuel production.

Implications of the Study for Industry and Policymakers

Gasoline vehicle is one of the significant sources of environmental pollution; it is still popular among the uses only due to the lower initial cost and availability of sufficient infrastructure. To make EVs and HVs more feasible, the auto manufacturing companies and the government will have to develop the infrastructure related to charging and hydrogen fueling stations, improving the safety provision, sufficient power production and production of hydrogen fuel, provision for tax relaxation, incentives, and subsidies on the alternative fuel vehicles, etc. This research shows that the major problem of the adoption of alternative vehicles is the high initial cost, lack of infrastructure, and lack of advanced charging technologies which can be resolved in collaboration with the government and industry.

Implications for Academia

This study brings the attention of the researcher to the viability of hydrogen vehicles with electric vehicles. EVs cannot fully replace gasoline vehicles due to a shortage of power supply. Hydrogen vehicles and EVs jointly can be the alternative for gasoline vehicles. The innovative research may support the production of hydrogen fuel with less environmental effect and the use of non-conventional sources of energy for the charging of EVs. It is a major question for the researchers on how to make these alternative vehicles more feasible in urban and rural areas including metro cities. Biofuel vehicle is not feasible in India due to the unavailability of sufficient crop land and high production costs.

Conclusions

This study is focused on identifying and incorporating all the possible factors for feasibility analysis and ranking of all the vehicle types to find the most viable alternative to conventional gasoline vehicles. To examine the viability of different types of vehicles in the Indian context, an integrated model of DEMATEL-TOPSIS is used. Extensive research was conducted to find the major possible factors that affect the viability of the vehicles. DEMATEL was used to rank the different factors and to find the relative weightage of the factors. As per our DEMATEL Analysis, the four top-most profound factors include charging/refueling infrastructure, initial cost of vehicle, environmental sustainability, and charging/refueling time. They have the greatest impact on all other factors when determining the appropriate alternative for gasoline vehicles.

Furthermore, TOPSIS was used incorporating the relative weightage of the factors as well as making use of the expert knowledge to rank the different types of vehicles. EVs emerge as the most practicable alternative to conventional gasoline vehicles, followed by hydrogen vehicles, solar vehicles, and biofuel vehicles. Due to the better environmental sustainability, end-of-life options, government incentives, safety, etc., the EVs outrank hydrogen, solar, and biofuel vehicles. Due to their longer range and less refueling time, hydrogen vehicles are the second-best alternative to

conventional gasoline vehicles as per our study. While comparing EVs and hydrogen vehicles, the latter is seen as a less feasible option comparatively, because of issues such as scarcity of charging infrastructure, safety issues, end-of-life options, and comparatively higher fuel costs. The effectiveness of solar and biofuel vehicles in terms of top and mid-tier factors makes them a less viable option [39].

Limitations and Future Scope of the Research

This research is based on a few common mode of energy for transportation. For a wider acceptance of the results, the findings of the study might be further validated for some additional modes of energy for transportation in developing economies considering the broad perspectives of sustainability. Some advanced modelling of feasibility analysis may be used to incorporate the bigger size of respondents' opinions.

Appendix

See Tables 7, 8, 9, 10.

Table 7 Normalized decision matrix of DEMATEL

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
F1	0.000	0.089	0.071	0.036	0.054	0.071	0.071	0.036	0.054	0.036	0.054	0.036	0.054	0.036
F2	0.036	0.000	0.071	0.071	0.071	0.054	0.089	0.089	0.089	0.071	0.071	0.071	0.071	0.089
F3	0.054	0.054	0.000	0.071	0.071	0.036	0.071	0.054	0.071	0.071	0.089	0.071	0.071	0.089
F4	0.071	0.054	0.054	0.000	0.089	0.089	0.089	0.071	0.089	0.071	0.071	0.089	0.071	0.089
F5	0.036	0.036	0.036	0.036	0.000	0.036	0.054	0.036	0.054	0.036	0.054	0.054	0.036	0.054
F6	0.089	0.054	0.054	0.071	0.071	0.000	0.071	0.089	0.089	0.089	0.089	0.071	0.071	0.089
F7	0.036	0.036	0.036	0.036	0.054	0.036	0.000	0.036	0.054	0.036	0.036	0.036	0.036	0.054
F8	0.071	0.071	0.036	0.054	0.071	0.071	0.071	0.000	0.071	0.071	0.071	0.054	0.071	0.071
F9	0.054	0.054	0.036	0.054	0.054	0.036	0.054	0.036	0.000	0.036	0.054	0.054	0.054	0.036
F10	0.054	0.054	0.054	0.071	0.054	0.054	0.054	0.054	0.054	0.000	0.071	0.036	0.054	0.036
F11	0.036	0.054	0.054	0.036	0.036	0.036	0.054	0.054	0.036	0.054	0.000	0.054	0.054	0.036
F12	0.071	0.071	0.054	0.071	0.071	0.054	0.071	0.054	0.071	0.071	0.071	0.000	0.089	0.089
F13	0.089	0.054	0.054	0.054	0.089	0.071	0.071	0.071	0.089	0.089	0.089	0.054	0.000	0.071
F14	0.089	0.071	0.054	0.071	0.071	0.071	0.089	0.054	0.071	0.071	0.089	0.071	0.089	0.000

Table 8 Total relation matrix *T* of DEMATEL

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
F1	0.185	0.261	0.226	0.209	0.253	0.236	0.282	0.209	0.261	0.224	0.264	0.213	0.244	0.232
F2	0.283	0.238	0.277	0.299	0.336	0.277	0.368	0.314	0.362	0.319	0.352	0.304	0.325	0.344
F3	0.280	0.271	0.195	0.280	0.315	0.244	0.331	0.264	0.324	0.299	0.346	0.286	0.305	0.323
F4	0.328	0.301	0.272	0.243	0.365	0.321	0.382	0.309	0.376	0.331	0.366	0.331	0.338	0.357
F5	0.182	0.177	0.160	0.172	0.160	0.169	0.221	0.172	0.217	0.184	0.220	0.192	0.189	0.206
F6	0.343	0.302	0.272	0.309	0.348	0.238	0.366	0.324	0.374	0.346	0.381	0.315	0.337	0.355
F7	0.174	0.168	0.153	0.164	0.202	0.162	0.160	0.164	0.207	0.176	0.194	0.168	0.179	0.197
F8	0.292	0.283	0.227	0.260	0.311	0.273	0.326	0.210	0.320	0.295	0.325	0.265	0.300	0.303
F9	0.213	0.207	0.173	0.202	0.227	0.183	0.238	0.186	0.183	0.199	0.237	0.206	0.219	0.206
F10	0.236	0.229	0.209	0.239	0.252	0.221	0.265	0.224	0.260	0.189	0.280	0.213	0.243	0.230
F11	0.194	0.204	0.187	0.184	0.208	0.180	0.235	0.200	0.215	0.214	0.183	0.203	0.217	0.204
F12	0.307	0.298	0.256	0.291	0.328	0.271	0.344	0.275	0.337	0.311	0.343	0.230	0.333	0.335
F13	0.325	0.285	0.258	0.277	0.345	0.289	0.346	0.292	0.355	0.328	0.361	0.282	0.252	0.320
F14	0.333	0.308	0.265	0.300	0.339	0.296	0.372	0.285	0.349	0.322	0.370	0.306	0.343	0.264

Table 9 Normalized decision matrix *R* of TOPSIS

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
EVs	0.707	0.729	0.615	0.462	0.645	0.387	0.512	0.295	0.588	0.737	0.364	0.286	0.147	0.577
Hydrogen vehicles	0.471	0.438	0.492	0.577	0.387	0.645	0.512	0.590	0.392	0.590	0.606	0.286	0.590	0.192
Solar vehicles	0.471	0.438	0.492	0.577	0.645	0.129	0.640	0.147	0.588	0.147	0.606	0.714	0.295	0.770
Bio-fuel based vehicles	0.236	0.292	0.369	0.346	0.129	0.645	0.256	0.737	0.392	0.295	0.364	0.571	0.737	0.192

Table 10 Weighted normalized decision matrix *V* of TOPSIS

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
EVs	0.047	0.055	0.042	0.036	0.041	0.030	0.034	0.021	0.040	0.049	0.025	0.021	0.011	0.046
Hydrogen vehicles	0.031	0.033	0.034	0.045	0.025	0.049	0.034	0.042	0.027	0.039	0.041	0.021	0.046	0.015
Solar vehiscles	0.031	0.033	0.034	0.045	0.041	0.010	0.042	0.010	0.040	0.010	0.041	0.053	0.023	0.061
Bio-fuel based vehicles	0.016	0.022	0.025	0.027	0.008	0.049	0.017	0.052	0.027	0.020	0.025	0.042	0.057	0.015
V^+	0.047	0.055	0.042	0.045	0.041	0.049	0.042	0.052	0.040	0.049	0.041	0.053	0.057	0.061
V^-	0.016	0.022	0.025	0.027	0.008	0.010	0.017	0.010	0.027	0.010	0.025	0.021	0.011	0.015

References

1. Hu G, Huang P, Bai Z, Wang Q, Qi K (2021) Comprehensively analysis the failure evolution and safety evaluation of automotive lithium ion battery. eTransportation 10: 100140.
2. Bicer Y, Dincer I (2018) Life cycle environmental impact assessments and comparisons of alternative fuels for clean vehicles. Resour Conserv Recycl 132:141–157
3. Ahmad L, Khordehgah N, Malinauskaite J, Jouhara H (2020) Recent advances and applications of solar photovoltaics and thermal technologies. Energy 207:118254
4. Alizadeh R, Soltanisehat L, Lund PD, Zamanisabzi H (2020) Improving renewable energy policy planning and decision-making through a hybrid MCDM method. Energy Policy 137:111174
5. Liang H, Ren J, Lin R, Liu Y (2019) Alternative-fuel based vehicles for sustainable transportation: a fuzzy group decision supporting framework for sustainability prioritization. Technol Forecast Soc Chang 140:33–43
6. Kumar P, Singh RK, Paul J, Sinha O (2021) Analyzing challenges for sustainable supply chain of electric vehicle batteries using a hybrid approach of delphi and best-worst method. Resour Conserv Recycl 175:105879
7. Poudyal R, Loskot P, Nepal R, Parajuli R, Khadka SK (2019) Mitigating the current energy crisis in Nepal with renewable energy sources. Renew Sustain Energy Rev 116:109388
8. Lee W, Xiang L, Schober R, Wong VW (2013) Analysis of the behavior of electric vehicle charging stations with renewable generations. In: 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm) (pp. 145–150). IEEE.
9. Liu Z, Song J, Kubal J, Susarla N, Knehr KW, Islam E, Ahmed S (2021) Comparing total cost of ownership of battery electric vehicles and internal combustion engine vehicles. Energy Policy 158:112564
10. Hawkins TR, Singh B, Majeau-Bettez G, Strømman AH (2013) Comparative environmental life cycle assessment of conventional and electric vehicles. J Ind Ecol 17(1):53–64
11. Majumdar D, Majhi BK, Dutta A, Mandal R, Jash T (2015) Study on possible economic and environmental impacts of electric vehicle infrastructure in public road transport in Kolkata. Clean Technol Environ Policy 17(4):1093–1101
12. Tagliaferri C, Evangelisti S, Acconcia F, Domenech T, Ekins P, Barletta D, Lettieri P (2016) Life cycle assessment of future electric and hybrid vehicles: a cradle-to-grave systems engineering approach. Chem Eng Res Des 112:298–309
13. Verma S, Dwivedi G, Verma P (2021) Life cycle assessment of electric vehicles in comparison to combustion engine vehicles: a review. Mater Today Proc.
14. Cipriani G, Di Dio V, Genduso F, La Cascia D, Liga R, Miceli R, Galluzzo GR (2014) Perspective on hydrogen energy carrier and its automotive applications. Int J Hydrogen Energy 39(16):8482–8494
15. Acar C, Dincer I (2017) Experimental investigation and analysis of a hybrid photoelectrochemical hydrogen production system. Int J Hydrogen Energy 42(4):2504–2511
16. AlRafea K, Fowler M, Elkamel A, Hajimiragha A (2016) Integration of renewable energy sources into combined cycle power plants through electrolysis generated hydrogen in a new designed energy hub. Int J Hydrogen Energy 41(38):16718–16728
17. Ehret O, Bonhoff K (2015) Hydrogen as a fuel and energy storage: success factors for the German Energiewende. Int J Hydrogen Energy 40(15):5526–5533
18. Mohammed H, Al-Othman A, Nancarrow P, Tawalbeh M, Assad MEH (2019) Direct hydrocarbon fuel cells: a promising technology for improving energy efficiency. Energy 172:207–219
19. Jing R, Kuriyan K, Lin J, Shah N, Zhao Y (2020) Quantifying the contribution of individual technologies in integrated urban energy systems—a system value approach. Appl Energy 266:114859
20. Narwane VS, Yadav VS, Raut RD, Narkhede BE, Gardas BB (2021) Sustainable development challenges of the biofuel industry in India based on integrated MCDM approach. Renew Energy 164:298–309
21. Altan A, Karasu S, Bekiros S (2019) Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos Soliton Fract 126:325–336
22. Li W, Long R, Chen H, Chen F, Zheng X, He Z, Zhang L (2020) Willingness to pay for hydrogen fuel cell electric vehicles in China: a choice experiment analysis. Int J Hydrogen Energy 45(59):34346–34353

23. Sagaria S, Neto RC, Baptista P (2021) Assessing the performance of vehicles powered by battery, fuel cell and ultra-capacitor: application to light-duty vehicles and buses. Energy Convers Manage 229:113767
24. Acar C, Dincer I (2020) The potential role of hydrogen as a sustainable transportation fuel to combat global warming. Int J Hydrogen Energy 45(5):3396–3406
25. Li J, Liang M, Cheng W, Wang S (2021) Life cycle cost of conventional, battery electric, and fuel cell electric vehicles considering traffic and environmental policies in China. Int J Hydrogen Energy 46(14):9553–9566
26. Shafiei E, Davidsdottir B, Leaver J, Stefansson H, Asgeirsson EI (2015) Comparative analysis of hydrogen, biofuels and electricity transitional pathways to sustainable transport in a renewable-based energy system. Energy 83:614–627
27. Onat NC, Gumus S, Kucukvar M, Tatari O (2016) Application of the TOPSIS and intuitionistic fuzzy set approaches for ranking the life cycle sustainability performance of alternative vehicle technologies. Sustain Prod Consump 6:12–25
28. Liu X, Reddi K, Elgowainy A, Lohse-Busch H, Wang M, Rustagi N (2020) Comparison of well-to-wheels energy use and emissions of a hydrogen fuel cell electric vehicle relative to a conventional gasoline-powered internal combustion engine vehicle. Int J Hydrogen Energy 45(1):972–983
29. Forrest K, Mac Kinnon M, Tarroja B, Samuelsen S (2020) Estimating the technical feasibility of fuel cell and battery electric vehicles for the medium and heavy duty sectors in California. Appl Energy 276:115439
30. Onat NC, Kucukvar M, Tatari O (2014) Towards life cycle sustainability assessment of alternative passenger vehicles. Sustainability 6(12):9305–9342
31. Tseng HK, Wu JS, Liu X (2013) Affordability of electric vehicles for a sustainable transport system: an economic and environmental analysis. Energy Policy 61:441–447
32. Rapa M, Gobbi L, Ruggieri R (2020) Environmental and economic sustainability of electric vehicles: life cycle assessment and life cycle costing evaluation of electricity sources. Energies 13(23):6292
33. Veziroglu A, Macario R (2011) Fuel cell vehicles: state of the art with economic and environmental concerns. Int J Hydrogen Energy 36(1):25–43
34. Hua Y, Liu X, Zhou S, Huang Y, Ling H, Yang S (2021) Toward sustainable reuse of retired lithium-ion batteries from electric vehicles. Resour Conserv Recycl 168:105249
35. Slattery M, Dunn J, Kendall A (2021) Transportation of electric vehicle lithium-ion batteries at end-of-life: a literature review. Resour Conserv Recycl 174:105755
36. Albertsen L, Richter JL, Peck P, Dalhammar C, Plepys A (2021) Circular business models for electric vehicle lithium-ion batteries: an analysis of current practices of vehicle manufacturers and policies in the EU. Resour Conserv Recycl 172:105658
37. Chowdhury MS, Rahman KS, Chowdhury T, Nuthammachot N, Techato K, Akhtaruzzaman M, Amin N (2020) An overview of solar photovoltaic panels' end-of-life material recycling. Energ Strat Rev 27:100431
38. Arora M, Shinde AM, Yadav V, Maji KJ, Singh V, Myers RJ (2021) Ramifications of Indian vehicle scrapping policy across the mobility sector. Resour Conserv Recycl 174:105845
39. Zhang Q, Tang Y, Bunn D, Li H, Li Y (2021) Comparative evaluation and policy analysis for recycling retired EV batteries with different collection modes. Appl Energy 303:117614
40. Majewski P, Al-shammari W, Dudley M, Jit J, Lee SH, Myoung-Kug K, Sung-Jim K (2021) Recycling of solar PV panels-product stewardship and regulatory approaches. Energy Policy 149:112062
41. Gong S, Ardeshiri A, Rashidi TH (2020) Impact of government incentives on the market penetration of electric vehicles in Australia. Transp Res Part D Transp Environ 83:102353
42. Khan U, Yamamoto T, Sato H (2020) Consumer preferences for hydrogen fuel cell vehicles in Japan. Transp Res Part D Transp Environ 87:102542
43. Jenn A, Springel K, Gopal AR (2018) Effectiveness of electric vehicle incentives in the United States. Energy Policy 119:349–356
44. Yan S (2018) The economic and environmental impacts of tax incentives for battery electric vehicles in Europe. Energy Policy 123:53–63
45. Wang N, Tang L, Pan H (2019) A global comparison and assessment of incentive policy on electric vehicle promotion. Sustain Cities Soc 44:597–603
46. Puricelli S, Cardellini G, Casadei S, Faedo D, Van den Oever AEM, Grosso M (2021) A review on biofuels for light-duty vehicles in Europe. Renew Sustain Energy Rev 137:110398
47. Liu H, Huang Y, Yuan H, Yin X, Wu C (2018) Life cycle assessment of biofuels in China: status and challenges. Renew Sustain Energy Rev 97:301–322
48. Rabaia MKH, Abdelkareem MA, Sayed ET, Elsaid K, Chae KJ, Wilberforce T, Olabi AG (2021) Environmental impacts of solar energy systems: a review. Sci Total Environ 754:141989
49. Tawalbeh M, Al-Othman A, Kafiah F, Abdelsalam E, Almomani F, Alkasrawi M (2021) Environmental impacts of solar photovoltaic systems: a critical review of recent progress and future outlook. Sci Total Environ 759:143528
50. Lopez NS, Tria LA, Tayo LA, Cruzate RJ, Oppus C, Cabacungan P, Biona JBM (2021) Societal cost-benefit analysis of electric vehicles in the Philippines with the inclusion of impacts to balance of payments. Renew Sustain Energy Rev 150:111492
51. Kurtz J, Sprik S, Bradley TH (2019) Review of transportation hydrogen infrastructure performance and reliability. Int J Hydrogen Energy 44(23):12010–12023
52. Devabhaktuni V, Alam M, Depuru SSSR, Green RC II, Nims D, Near C (2013) Solar energy: trends and enabling technologies. Renew Sustain Energy Rev 19:555–564
53. Krishna G (2021) Understanding and identifying barriers to electric vehicle adoption through thematic analysis. Transp Res Interdiscip Perspect 10:100364
54. Higueras-Castillo E, Guillén A, Herrera LJ, Liébana-Cabanillas F (2021) Adoption of electric vehicles: which factors are really important? Int J Sustain Transp 15(10):799–813
55. Krishnan VV, Koshy BI (2021) Evaluating the factors influencing purchase intention of electric vehicles in households owning conventional vehicles. Case Studies on Transport Policy.
56. Leibensperger C, Yang P, Zhao Q, Wei S, Cai X (2021) The synergy between stakeholders for cellulosic biofuel development: perspectives, opportunities, and barriers. Renew Sustain Energy Rev 137:110613
57. Zhang Z, Wang D, Zhang C, Chen J (2018) Electric vehicle range extension strategies based on improved AC system in cold climate—a review. Int J Refrig 88:141–150
58. Franke T, Krems JF (2013) What drives range preferences in electric vehicle users? Transp Policy 30:56–62
59. Stotts R, Lopez-Jaramillo OG, Kelley S, Krafft A, Kuby M (2021) How drivers decide whether to get a fuel cell vehicle: an ethnographic decision model. Int J Hydrogen Energy 46(12):8736–8748
60. Lopez JO, Stotts R, Kelley S, Kuby M (2019) Content analysis of interviews with hydrogen fuel cell vehicle drivers in Los Angeles. Transp Res Rec 2673(9):377–388
61. Chaturvedi A, Kushwaha K, Kashyap P, Navani JP (2015) Solar powered vehicle. Int J Electrical Electron Res 3(2):270–273

62. Ratnakar RR, Gupta N, Zhang K, van Doorne C, Fesmire J, Dindoruk B, Balakotaiah V (2021) Hydrogen supply chain and challenges in large-scale LH2 storage and transportation. Int J Hydrogen Energy.
63. Awudu I, Zhang J (2012) Uncertainties and sustainability concepts in biofuel supply chain management: a review. Renew Sustain Energy Rev 16(2):1359–1368
64. Yue D, You F, Snyder SW (2014) Biomass-to-bioenergy and biofuel supply chain optimization: overview, key issues and challenges. Comput Chem Eng 66:36–56
65. Nili M, Seyedhosseini SM, Jabalameli MS, Dehghani E (2021) A multi-objective optimization model to sustainable closed-loop solar photovoltaic supply chain network design: a case study in Iran. Renew Sustain Energy Rev 150:111428
66. Rabe W, Kostka G, Stegen KS (2017) China's supply of critical raw materials: risks for Europe's solar and wind industries? Energy Policy 101:692–699
67. Wathne MW, Haarstad H (2020) The smart city as mobile policy: insights on contemporary urbanism. Geoforum 108:130–138
68. Bai A, Jobbágy P, Popp J, Farkas F, Grasselli G, Szendrei J, Balogh P (2016) Technical and environmental effects of biodiesel use in local public transport. Transp Res Part D Transp Environ 47:323–335
69. Ajanovic A, Haas R (2021) Renewable energy systems implementation in road transport: prospects and impediments. Renew Energy Environ Sustain 6:39
70. Lozanovski A, Whitehouse N, Ko N, Whitehouse S (2018) Sustainability assessment of fuel cell buses in public transport. Sustainability 10(5):1480
71. Kubański M (2020) Prospects for the use of electric vehicles in public transport on the example of the city of czechowice-dziedzice. Transp Res Proc 44:110–114
72. Lajunen A (2018) Lifecycle costs and charging requirements of electric buses with different charging methods. J Clean Prod 172:56–67
73. Kamali FP, Borges JAR, Osseweijer P, Posada JA (2018) Towards social sustainability: screening potential social and governance issues for biojet fuel supply chains in Brazil. Renew Sustain Energy Rev 92:50–61
74. Guo S, Zhao H (2015) Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective. Appl Energy 158:390–402
75. Omahne V, Knez M, Obrecht M (2021) Social aspects of electric vehicles research—trends and relations to sustainable development goals. World Electric Veh J 12(1):15
76. Hienuki S, Hirayama M, Hirayama Y, Kamada H, Kasai N, Shibutani T, Otaki T (2021) Public acceptance for the implementation of hydrogen self-refueling stations. Int J Hydrogen Energy 46(72):35739–35749
77. Frolov SM, Medvedev SN, Basevich VY, Frolov FS (2013) Self-ignition of hydrocarbon–hydrogen–air mixtures. Int J Hydrogen Energy 38(10): 4177–4184
78. Hirayama M, Shinozaki H, Kasai N, Otaki T (2018) Comparative risk study of hydrogen and gasoline dispensers for vehicles. Int J Hydrogen Energy 43(27):12584–12594
79. Le H, Liu Y, Mannan MS (2013) Lower flammability limits of hydrogen and light hydrocarbons at subatmospheric pressures. Ind Eng Chem Res 52(3):1372–1378
80. Rajper SZ, Albrecht J (2020) Prospects of electric vehicles in the developing countries: a literature review. Sustainability 12(5):1906
81. Meyer PE, Winebrake JJ (2009) Modeling technology diffusion of complementary goods: the case of hydrogen vehicles and refueling infrastructure. Technovation 29(2):77–91
82. Kim S, Lee J, Lee C (2017) Does driving range of electric vehicles influence electric vehicle adoption? Sustainability 9(10):1783
83. Prakash G, Srivastava S (2019) Developing a care coordination model using a Hybrid DEMATEL and PLS-SEM approach. IIM Kozhikode Soc Manag Rev 8(1):34–49
84. Asadi S, Nilashi M, Iranmanesh M, Ghobakhloo M, Samad S, Alghamdi A, Mohd S (2022) Drivers and barriers of electric vehicle usage in Malaysia: a DEMATEL approach. Resour Conserv Recycl 177:105965
85. Sekhar C, Patwardhan M, Vyas V (2017) Developing a curriculum for entrepreneurship education: Prioritizing the content using TOPSIS method. In: Entrepreneurship Education (pp. 85–102). Springer, Singapore.
86. Li X, Han Z, Zhang R, Zhang Y, Zhang L (2020) Risk assessment of hydrogen generation unit considering dependencies using integrated DEMATEL and TOPSIS approach. Int J Hydrogen Energy 45(53):29630–29642
87. Ortiz-Barrios M, Miranda-De la Hoz C, López-Meza P, Petrillo A, De Felice F (2020) A case of food supply chain management with AHP, DEMATEL, and TOPSIS. J Multi Crit Decis Anals 27(1–2):104–128
88. Mohammed A, Naghshineh B, Spiegler V, Carvalho H (2021) Conceptualising a supply and demand resilience methodology: a hybrid DEMATEL-TOPSIS-possibilistic multi-objective optimization approach. Comput Ind Eng 160:107589
89. Kilic HS, Yalcin AS (2021) Comparison of municipalities considering environmental sustainability via neutrosophic DEMATEL based TOPSIS. Socioecon Plann Sci 75:100827
90. Chang KH, Chang YC, Lee YT (2014) Integrating TOPSIS and DEMATEL methods to rank the risk of failure of FMEA. Int J Inf Technol Decis Mak 13(06):1229–1257
91. Aragão FV, Gomes PFDO, Chiroli DDG, Zola FC, Rocha Loures EDF, Santos EAP, Colmenero JC (2021) Projects aimed at smart cities: a hybrid MCDA evaluation approach. Technol Anal Strat Manag: 1–13.
92. Thakkar JJ (2021) Decision-making trial and evaluation laboratory (DEMATEL). In: Multi-criteria decision making. Studies in systems, decision, and control, 336. Springer, Singapore. https://doi.org/10.1007/978-981-33-4745-8_9
93. Zhao G, Ahmed RI, Ahmad N, Yan C, Usmani MS (2021) Prioritizing critical success factors for sustainable energy sector in China: a DEMATEL approach. Energ Strat Rev 35:100635
94. Dalalah D, Hayajneh M, Batieha F (2011) A fuzzy multi-criteria decision making model for supplier selection. Expert Syst Appl 38(7):8384–8391
95. Hwang CL, Yoon K (1981) Methods for multiple attribute decision making. In: Multiple attribute decision making (pp. 58–191). Springer, Berlin, Heidelberg
96. Micale R, Marannano G, Giallanza A, Miglietta PP, Agnusdei GP, La Scalia G (2019) Sustainable vehicle routing based on firefly algorithm and TOPSIS methodology. Sustain Fut 1:100001
97. Robles JO, Azzaro-Pantel C, Garcia GM, Lasserre AA (2020) Social cost-benefit assessment as a post-optimal analysis for hydrogen supply chain design and deployment: application to occitania (France). Sustain Prod Consump 24:105–120
98. Knez M, Jereb B, Jadraque Gago E, Rosak-Szyrocka J, Obrecht M (2021) Features influencing policy recommendations for the promotion of zero-emission vehicles in Slovenia, Spain, and Poland. Clean Technol Environ Policy 23(3):749–764

99. Zhao F, Chen K, Hao H, Wang S, Liu Z (2019) Technology development for electric vehicles under new energy vehicle credit regulation in China: scenarios through 2030. Clean Technol Environ Policy 21(2):275–289

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

Springer Nature or its licensor 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.

A Complete Internet of Things based Home Security System

Haren Nagdewani[1], Pawan Singh Mehra[2]

[1,2]*Department of Computer Science and Engineering*
Delhi Technological University
New Delhi, India

harennagdewani_2k18co138@dtu.ac.in
pawansinghmehra@gmail.com

2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA) | 978-1-6654-5834-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCSEA54677.2022.9936168

***Abstract*— Security has been a big concern especially in the unsafe areas of urban India. Theft, invasion and fire are some of the most common concerns among individuals. The existing methods of preventing the above-mentioned concerns are flawed and/or don't address all three problems together. Most of the existing solutions rely on surveillance using cameras. In this paper, we present a complete solution which can inform the owner about invasion, send alerts for fire hazard and authenticate a visitor. Our solution uses Raspberry Pi 3B+, MQ2 gas detection module, PIR sensor, ultrasonic sensor, Arduino uno microcontroller, 3x4 keypad and a webcam. Our system is better than existing solutions as it uses ultrasonic and PIR sensors as opposed to just PIR sensors for motion detection which can lead to false positives due to high sensitivity. Also, use of arduino to connect the smoke sensors gives us a modular approach which will enable the user to connect multiple smoke sensors to the main system. Lastly, iris recognition is used for authenticating an individual. EfficientNet-B1 with SDGR optimizer was used with 300 epochs. The model had training accuracy of 99.80% and validation accuracy of 97.66%.**

***Keywords*— IoT Based Home Security System, Iris Recognition Raspberry Pi, PIR Sensor, Ultrasonic Sensors, Motion Detection, Smoke Detection.**

I. INTRODUCTION

The idea of being able to monitor objects over the internet gave rise to IoT technology. The vision of connecting the real world with the virtual world to improve the overall quality of life with the help of sensors, wireless communication, nanotechnology, modular design and a lot of technical additions has made application of IoT practical in every field [10]. IoT based solutions are being adopted and developed in every industry - logistics, supply chains, manufacturing, surveillance, automation, security, agriculture, aggregation, data collection, etc

Home automation systems are solutions which assist a user to control traditional appliances such as lights, music, fan, air conditioner, etc with the help of smartphone or voice control. In India, home automation systems are growing at a rapid pace, specially with the likes of Google Home, Alexa, Amazon Echo and other smart home technologies. Home automation has been made accessible to a large number of people due to reduction in cost and more affordable systems. Home security systems can be considered a subset of home automation systems which deals particularly with an individual's security practices with the help of hardware and software placed on the user's premises. Unfortunately, even though the use of home automation has seen wide adoption for entertainment and convenience, we are yet to witness the same for home security using smart systems and internet of things, yet.

Currently, majority houses in India do not have any security system apart from bolts and locks. According to Census 2011, there are 11,01,39,853 houses in urban India and according to Sterlite Technologies limited, the Indian security and surveillance stands at 9000 crores out of which 85% is government and industrial security. The above stats represent that opportunity and growth possibility for smart security systems is huge in the upcoming decades.

Keeping in mind the huge opportunity available, in this paper, we propose a home security system based on IoT which is affordable and practical. This work plans on reducing the threats in the society related to home theft, invasion and fire. Being able to remotely monitor the premises when no resident is in the house is one of the main functionalities of our project. [1] It is important to create a system that can inform the owner about any theft or intrusion which has taken place when the owner is not at home. Apart from an alert system for unauthorised access to one's premises, we also propose a secure way to enter one's house with the help of iris recognition. One more thing which we kept in mind is to create an affordable system so that it can be adopted on a large scale as security is a necessity, not a luxury. [2] Lastly, preventing theft and unauthorised access is one part of home security. Our project will also consist of smoke and fire detection in order to prevent losses incurred due to fire accidents. [3] The main improvements of our system over the existing solutions are as follow:

- Improved motion detection with the help of a PIR sensor supported by an ultrasonic sensor.
- Modular approach to smoke sensors, following star topology. Therefore, it is easy to extend and add more smoke sensors to the main system.
- Adapting the usage of iris recognition from a pre-trained network to improve authentication. Additionally, providing a keypad in case the biometric system fails.

II. PROBLEM STATEMENT

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:40:31 UTC from IEEE Xplore. Restrictions apply.

The objective of the project is to create a home security system which will have following attributes:

- Authorised Access - The system should let only authorised individuals access the premises using biometric or alternative technology.
- Inform about Intrusion - If there is any forced entry or unauthorised entry, the system should raise an alert to the owner.
- Inform about Fire / Smoke - If any fire or smoke is detected inside the premises then the owner should be alerted to avoid damage.
- Quick Response - The security system should be able to quickly capture the picture of the intruder and relay it to the owner without any delay.
- Secure System - System should be secure irrespective of the source of data, information transmission and positioning of sensors and computers. It should not be easily hacked [11]
- Communication - The sensors should work as desired, not very close or far off.
- Reliability - The system should be reliable in all scenarios [18].
- Affordable - Security is a necessity, not a luxury. Therefore, the system should be affordable for large-scale adoption.

III. RELATED WORKS

S. Tanwar et al. [4] created a home security system which detects an intruder or any unusual event at home. This system utilises a PIR module along with Raspberry Pi. Whenever an intruder enters, it is detected using an infrared sensor (PIR) and an image of the intruder is captured using a webcam. All these modules are linked together using the WIFI module. The owner of the home is immediately alerted over email. The advantages of their system are: affordable, easy to set up and high accuracy. However, one major drawback of the system is that it is easy to disable the entire system.

Rizwan Majeed et al. [5] have not only created a Home security system but an entire Smart home automation system. They have identified the key challenges which arise in any smart home system like: authentication of IoT devices, data security, privacy issues, continuous monitoring and connectivity, secure identification and intelligent decision making. They've tried to tackle most of these issues in this project. The use of machine learning for making decisions in the system will lead to an intelligent system, other than that, blockchain technology is used to ensure identification and authentication of IoT devices. The system is created with the help of Raspberry Pi microcontroller, relay circuit and other modules. They've also developed a mobile application which can control the entire system for Android operating system.

Panda, Saroja & Sahu, Sushanta [6] offers a practical method to handle real-time theft in their research paper. The project used Raspberry Pi to create a real-time video surveillance system. They've also added functionality of password locking. The system is made up using webcam, Raspberry pi, Keypad, and Passive Infrared sensor. Whenever someone enters an incorrect password, the webcam captures the image and sends it over SMTP mail server (email). PIR sensor is used to detect motion, and as soon as motion is detected, the webcam is activated.

R. G. Anvekar and R. M. Banakar [7] created a home security system different from others. Unlike other research papers which utilise SMTP or SMS service to alert the homeowner, the authors of this paper have utilised secure end-user applications such as Telegram for end-to-end communication. Telegram provides it's API to build Bots compatible with Raspberry Pi. The authors have built a system which deals in presence detection, identification and authentication of strangers.

Dahmen J, Cook DJ, Wang X and Honglei W. [8] took a look at home security for smart homes rather than just automation and ease-of-use in their research paper. The paper starts with discussing various ways in which smart homes can be made more secure for their residents. Further the paper provides a comparative study between various methods available. The paper takes into consideration not only commercially available products but also multiple research papers and ideas proposed in them to provide a concrete result along with future possibilities.

P. Bhatia et al. [9] developed a face recognition system for identification for entry into the premises. Their machine learning model uses LBPH to recognize the stranger. A local database is created of known family members and the image of the stranger is matched with the database. The key components of the system is access to only authorised people, real time monitoring and control over automation. The hardware which was used to implement the system is: Raspberry Pi, webcam, speaker and stepper motor.

Kodali et al. [12] identified the shortcomings of using PIR sensors in intrusion detection systems. Every time the sensor detects the movement, the owner receives an alert. As a result, there are a lot of false alarms because the motion detected may not be a trespasser all the time. The proposed solution relies on the fact that the owner will be connected to the home access point rather than mobile data when inside or around the premises. The system uses PIR sensors and is also made aware of its surroundings. Authors have used ESP8266 wifi module to detect the MAC address of the owner's phone to know his/her presence. This mechanism is built in order to reduce the false alarm rate.

Das S. et al. [13] have created a system which can detect motion and presence of LPG / toxic gases. For motion detection, a PIR sensor is used and for LPG / gas detection, MQ2 sensors are used. These sensors are installed in various places and communicate with the main Raspberry Pi using a transmitter and receiver connected via an Arduino microcontroller. An automated lock system is also created which is operated using two servo motors. The lock can be accessed locally after entering the correct password on the keypad.

A. B. Dorothy, S. B. R. Kumar and J. J. Sharmila [14] discusses an approach for authenticating the people entering the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:40:31 UTC from IEEE Xplore. Restrictions apply.

premises. The system detects motion and captures the image of a person who is trying to enter the house. Then the image is sent to a database where it is processed and matched with the authenticated people and pets. If a match is not found then an alert message is sent to the owner. The image processing techniques used are linear filtering and fourier transforms. Template matching and Cooley-Tukey algorithm is used for identification.

Irving Vitra Paputungan et al. [16] identified the limitations of using traditional CCTV security systems and proposed an IoT based solution. The authors are using PIR and Accelerometer sensors to detect motion in a specified range. The sensors are placed as nodes at various locations and they communicate with the node station. Node station is created using ESP8266 NodeMCU. Once motion is detected, the node station is responsible for sending data to Adafruit IoT cloud and forwards an email to the owner. Moreover, the authors have also uploaded the data to a firebase server which can further be accessed using the mobile application.

Rui Yu, Xiaohua Zhang and Minyuan Zhang [17] have created a smart home security analysis system as opposed to home security systems developed by other authors discussed in this paper. The main objective is to be able to identify and defend against wireless attacks that smart homes may receive.The authors have created the system by separating it into four modules namely, traffic capture, traffic analysis, system detection and warning modules. Once data packets are captured using the first module, they are sent for analysis on the basis of content, protocol and implant attack. If deemed dangerous, the warning notification module is called. The system detection module is used to find malicious and potential threats in the system.

IV. PROPOSED WORK

The proposed work is divided into 4 sections. Each section addresses a unique problem. The system proposed in this paper is cost-effective, practical and easy to use.

A. Authorised Access

To make sure that only authorised individuals can enter the premises, we are using biometric technology. We use iris detection to identify an individual and match it with the existing dataset of authorised individuals. If iris detection fails, we have also added a keypad to enter a password manually.

B. Iris Recognition

For Iris recognition, we have used the CASIA V1.0 publicly available dataset. It consists of 756 iris images of 108 different subjects. In Raspberry Pi we are using a pre-trained model for iris recognition. The original model was trained using EfficientNet-B1 and Stochastic Gradient Descent was used as the optimizer and 300 epochs. The dataset used to train the model had clear images captured in a controlled environment. However, when testing the system in a real time environment where the factors such as lighting, camera angle, distance of subject from camera, etc can introduce additional noise in the clicked images. This may lead to reduced accuracy of the iris recognition module.

C. Intrusion Information

The entire security system is made active with the help of a PIR sensor. The HC-SR501 PIR sensor has a range upto 7m and 140 degree viewing angle. Due to the long range and wide angle of the PIR sensor, there is a high chance of false negatives. The sensor may detect some motion in the background and consider it as an intrusion. In order to avoid false positives, we have combined the PIR sensor with an ultrasonic sensor (HC-SR04) to monitor the distance. If the distance is under 1m and some motion is detected, then we activate the security system. The intruder is given 2 minutes to identify himself using Iris recognition or entering the password in the keypad. If he fails to do so, an email is sent to the owner along with a picture of the intruder .

D. Fire and Smoke detection

To inform the owner about threats of possible fire and smoke in the premises we use MQ2 gas detection module. The MQ2 sensor is connected to Arduino which communicates with Raspberry Pi [15]. The current model uses wires to transmit data between Arduino and Pi. However, it can be made wireless using transreceivers.

V. IMPLEMENTATION DETAILS

To understand the working of the proposed system, we discuss the system components and architecture in detail. The architecture diagram represents the placement of various components with respect to each other and surroundings.

A. System Components

The security system uses the following system components:

TABLE I
SYSTEM COMPONENTS

Components	**Function**
Raspberry Pi 3B+	Main computer
Arduino Uno	Connect MQ2 to Pi
MQ2 Gas Sensor	Smoke detection
USB Webcam	Image capturing
HC-SR501 PIR Sensor	Motion detection
HC-SR04 Ultrasonic Sensor	Distance detection
3x4 Keypad	Entering pin
Jumper Wires	Connections between components
Breadboard	To use jumper wires
Resistors	Regulate the voltage
LEDs	Authentication state

B. System Architecture

The architecture of the system is as follow:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:40:31 UTC from IEEE Xplore. Restrictions apply.

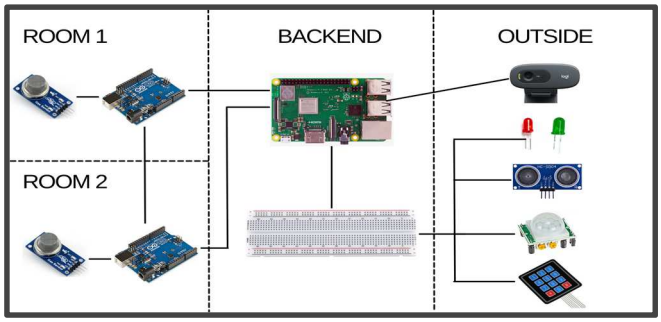


Fig. 1 System architecture

Webcam, PIR sensor, Ultrasonic sensor and keypad is placed outside as they interact with the subject directly. However, the raspberry pi and breadboard are placed inside (hidden from the subject). Raspberry Pi is responsible for handling communication between various sensors and informing the owner about intrusion over email. Multiple gas sensors can be used in various locations. The gas sensors communicate with Raspberry Pi with the help of Arduino Uno. In the current version of the system, wired connections are used between Arduino Uno and Raspberry Pi. However, to make the system more portable, wireless communication between Arduino Uno and Pi can be set up with the help of transceiver modules. Raspberry Pi uses raspbian OS. Once switched on, the entire system runs indefinitely on loop till the time it is manually switched off. As intrusion detection and fire prevention are two separate modules, we've presented two separate flowcharts for each:

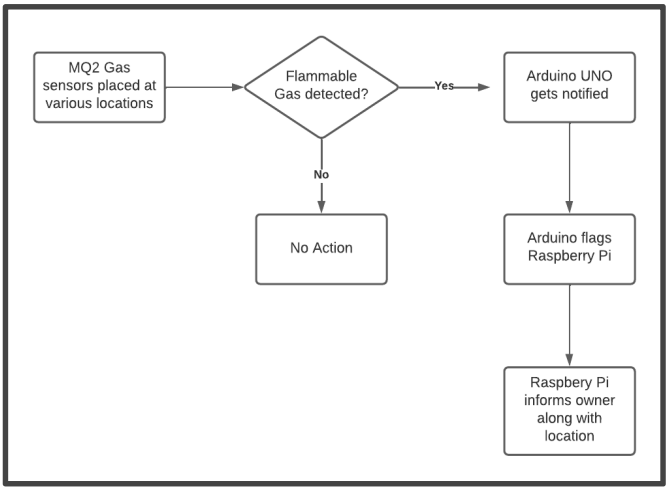


Fig. 2 Smoke detection flowchart

VI. SIMULATION AND RESULTS

The system was tested in a real time environment. All the features of the system were working however, iris recognition struggled sometimes depending on the distance and lighting conditions. For this exact reason, we provided a keypad to enter the pin manually and gain access. Theoretically, the iris recognition model had a training accuracy of 99.80% and a validation accuracy of 97.66%. However, when testing with the help of a 720p logitech webcam in a real time environment, the iris recognition system accurately recognized the subject 26 out of 30 times, an accuracy of 86.67%.

We simulated fire hazard by burning a matchstick close to the MQ2 sensor, the results can be seen in the table below. System uses smtp library and informs the owner about fire / subject over email along with date and time which is captured using strftime in-built python library.

For testing the motion detection module, we moved past the system multiple times and recorded the observation.

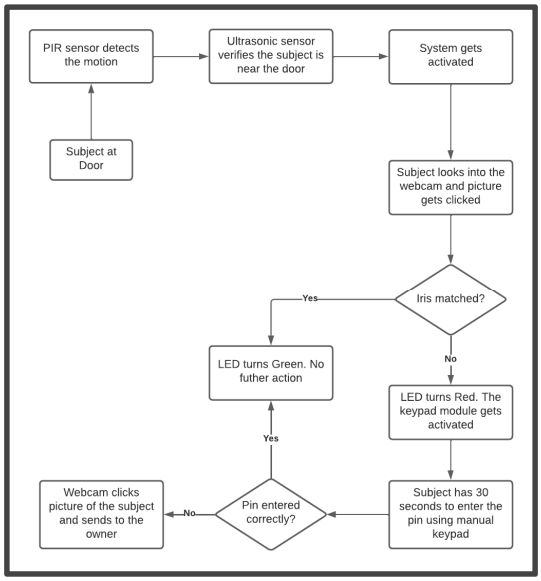


Fig. 3 Motion detection flowchart

TABLE II
RESULTS FOR IRIS RECOGNITION

Result	**Number of times (Total = 40)**
True Positive	26
True Negative	10
False Positive	0
False Negative	4

TABLE III
RESULTS FOR MOTION DETECTION

Result	**Number of times (Total = 30)**
True Positive	17
True Negative	8
False Positive	2
False Negative	3

TABLE IV
RESULTS FOR SMOKE DETECTION

Result	**Number of times (Total = 20)**
True Positive	13
True Negative	5
False Positive	0
False Negative	2

VII. CONCLUSION AND FUTURE WORK

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:40:31 UTC from IEEE Xplore. Restrictions apply.

Security is one of the biggest concerns for any property owner. We're slowly witnessing a shift from traditional bolts and locks in urban households to more and more secure systems. In this paper, we discussed an easy to implement approach for an affordable security system using IoT technology. Our system successfully detects the presence of an intruder and any potential fire. It can reliably communicate the information to the owner so that preventive measures can be taken. Additionally, we used iris recognition from a pre-trained model which uses EfficientNet B1 with a training accuracy of 99.80% and validation accuracy of 97.66%. The future possibilities for this project is immense, one can integrate the system with smart locks and utilise the authentication system for even unlocking the doors. Not only that, if multiple smart locks are used in various sections of premises, then, only selective access can be given depending on individual permissions. Lastly, the webcam can also be broadcasted over the internet [19] and the owner can monitor the premises from the live video feed.

REFERENCES

[1] M. N. N. A. Asghar, M.H., "Principle application and vision in internet of things (iot)," in Communication Technologies (GCCT), 2015 Global Conference on, may 2015.

[2] Lin H, Bergmann NW. IoT Privacy and Security Challenges for Smart Home Environments. Information. 2016; 7(3):44. https://doi.org/10.3390/info7030044

[3] S. Sarkar, S. Gayen and S. Bilgaiyan, "Android Based Home Security Systems Using Internet of Things(IoT) and Firebase," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 102-105, doi: 10.1109/ICIRCA.2018.8597197.

[4] S. Tanwar, P. Patel, K. Patel, S. Tyagi, N. Kumar and M. S. Obaidat, "An advanced Internet of Thing based Security Alert System for Smart Home," 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), 2017, pp. 25-29, doi: 10.1109/CITS.2017.8035326.

[5] Rizwan Majeed, Nurul Azma Abdullah, Imran Ashraf, Yousaf Bin Zikria, Muhammad Faheem Mushtaq, Muhammad Umer, "An Intelligent, Secure, and Smart Home Automation System", Scientific Programming, vol. 2020, Article ID 4579291, 14 pages, 2020. https://doi.org/10.1155/2020/4579291

[6] Panda, Saroja & Sahu, Sushanta. (2021). Design of IoT-Based Real-Time Video Surveillance System Using Raspberry Pi and Sensor Network. 10.1007/978-981-33-6081-5_11.

[7] R. G. Anvekar and R. M. Banakar, "IoT application development: Home security system," 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2017, pp. 68-72, doi: 10.1109/TIAR.2017.8273688.

[8] Dahmen J, Cook DJ, Wang X, Honglei W. Smart Secure Homes: A Survey of Smart Home Technologies that Sense, Assess, and Respond to Security Threats. J Reliab Intell Environ. 2017;3(2):83-98. doi:10.1007/s40860-017-0035-0

[9] P. Bhatia, S. Rajput, S. Pathak and S. Prasad, "IOT based facial recognition system for home security using LBPH algorithm," 2018 3rd International Conference on Inventive Computation Technologies (ICICT), 2018, pp. 191-193, doi: 10.1109/ICICT43934.2018.9034420.

[10] Mehra, P.S. LBECR: load balanced, efficient clustering and routing protocol for sustainable internet of things in smart cities. J Ambient Intell Human Comput (2022). https://doi.org/10.1007/s12652-022-03706-2

[11] Dhiraj Sunehra and Ayesha Bano, " An Intelligent Surveillance with Cloud Storage for Home Security", Annual IEEE India Conference (INDICON), 2014, pp. 1-5.

[12] Kodali, Ravi & Rajanarayanan, Sasweth & Koganti, Anvesh & Boppana, Lakshmi. (2019). IoT based security system. 1253-1257. 10.1109/TENCON.2019.8929420.

[13] Das S., Budia M., Ghosh R., Dhar S., Mruthyunjaya H.S. (2020) Design and Development of IoT-Based Low-Cost Robust Home Security System. In: Bera R., Pradhan P.C., Liu CM., Dhar S., Sur S.N. (eds) Advances in Communication, Devices and Networking. ICCDN 2019. Lecture Notes in Electrical Engineering, vol 662. Springer, Singapore. https://doi.org/10.1007/978-981-15-4932-8_51

[14] A. B. Dorothy, S. B. R. Kumar and J. J. Sharmila, "IoT Based Home Security through Digital Image Processing Algorithms," 2017 World Congress on Computing and Communication Technologies (WCCCT), 2017, pp. 20-23, doi: 10.1109/WCCCT.2016.15.

[15] A. Kumar Dwivedi, A. Kumar Sharma, and P. Singh Mehra, "Energy Efficient Sensor Node Deployment Scheme for Two Stage Routing Protocol of Wireless Sensor Networks assisted IoT", ECTI-EEC, vol. 18, no. 2, pp. 158–169, Aug. 2020.

[16] I. V. Paputungan, M. R. Al Fitri and U. Y. Oktiawati, "Motion and Movement Detection for DIY Home Security System," 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET), 2019, pp. 122-125, doi: 10.1109/CSUDET47057.2019.9214684.

[17] R. Yu, X. Zhang and M. Zhang, "Smart Home Security Analysis System Based on The Internet of Things," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 596-599, doi: 10.1109/ICBAIE52039.2021.9389849.

[18] Gupta, Priyanshu and Raj, Pallav and Tiwari, Shwetanshu and Kumari, Puja and Mehra, Pawan Singh, Energy Efficient Diagonal Based Clustering Protocol in Wireless Sensor Network (April 1, 2020). Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020, Available at SSRN: https://ssrn.com/abstract=3565781 or http://dx.doi.org/10.2139/ssrn.3565781

[19] R. Chanana, A. K. Singh, R. Killa, S. Agarwal and P. S. Mehra, "Blockchain Based Secure Model for Sensor Data in Wireless Sensor Network," 2020 6th International Conference on Signal Processing and Communication (ICSC), 2020, pp. 288-293, doi: 10.1109/ICSC48311.2020.9182776.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:40:31 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/364935735

A Decentralized Coordination Algorithm for Patrolling and Target Tracking in Internet of Robotic Things (IoRT) using Dynamic Waypoints and Self-triggered Communication

Article · October 2022

DOI: 10.12785/ijcds/120180

CITATIONS
0

READS
16

2 authors:

Virender Ranga
Delhi Technological University
177 PUBLICATIONS **1,360** CITATIONS
SEE PROFILE

Janardan Kumar Verma
3 PUBLICATIONS **24** CITATIONS
SEE PROFILE

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

Sensor enabled Internet of Things for smart cities View project

Hybrid Framework for FANets View project

All content following this page was uploaded by Virender Ranga on 31 October 2022.

The user has requested enhancement of the downloaded file.

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Oct-2022)

https://dx.doi.org/10.12785/ijcds/120180

A Decentralized Coordination Algorithm for Patrolling and Target Tracking in Internet of Robotic Things (IoRT) using Dynamic Waypoints and Self-triggered Communication

Janardan Kumar Verma[1] **and Virender Ranga**[2]

[1]*Department of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana, India*
[2]*Department of Information Technology, Delhi Technological University, Delhi, India*

Received 22 Jan. 2021, Revised 15 Jul. 2022, Accepted 23 Jul. 2022, Published 31 Oct. 2022

Abstract: The Internet of Robotic Things (IoRT) deals with autonomous objects such as robots and sensors together. It can be used for efficient patrolling and target tracking; however, mobile sensors (robots) must coordinate to decide their optimal actions. This paper proposes a decentralized coordination strategy that can enable multiple robots to perform both patrolling and target tracking in the deployed scenario. The basic idea is that the whole site is divided into local zones, and a single robot is deployed in each zone. All robots are logically divided into two groups, namely category $C1$ and category $C2$ robots. A dynamic waypoint generation algorithm is proposed to assist category $C1$ robots in the perimeter patrol. It produces the waypoints such that certain locations can be prioritized and intruders cannot predict the patrolling trajectory. Category $C2$ robots are responsible for area patrolling within the zone. Here, we also propose a strategy such that the places with a high probability of unusual acts can be visited frequently. We use the distributed Extended Kalman filter (EKF) to estimate and predict the position of the targets. Each robot has a self-triggered communication mechanism to share the necessary information with the neighbors, such as the estimated position of the intruder, EKF parameters, asking for help, etc. We have also developed an Internet of Thing (IoT) based web application to monitor and control the robots. In this app, robots subscribe to the server for necessary information and commands and publish their position, patrolled area, intruders' position, battery status, etc., to the server for real-time monitoring and control. The proposed solution is validated through simulations in the Robot Operating System (ROS) and Gazebo. The results show the patrolling and target tracking performance using idleness and error in the target's estimated position, respectively, as metric.

Keywords: Cooperative target tracking, Cooperative patrolling, Internet of Robotic Things Coordination, Multi-Robot System, Multi-Robot Coordination.

1. Introduction

Nowadays, the expedite advancements in Multi-Robot Systems (MRS) are motivated by robotics' progress and cooperative behavior present in various living beings such as flocks of birds, fish swarm, etc. An increasing interest in the MRS coordination [1] research domain for monitoring and surveillance applications [2] such as mapping and exploration, monitoring wildfire and environment, security, and disaster management has emerged. The multi-robot systems employ robot groups that communicate and execute various tasks need to coordinate. The mobility aspect of mobile robots enhances their inherent sensor coverage and renders supervision in areas where the deployed static nodes do not provide adequate coverage. Therefore, the combination of static sensors and MRS i.e. IoRT [3],[4] offers additional benefits, enhanced efficiency, and broader application domains. Deploying networked MRS [5], [6] and static sensor networks [7], [8] are two popular ways for tracking mobile targets. The robots detect and track the moving target by sharing the target's information with neighbor robots. These robots estimate the target's position as the tracking destination and regulate their motion towards it.

Along with target tracking, this paper focuses on engaging multi-robots in the patrolling problem also. In patrolling problems, robots monitor the environment for a longer extent of time. They move and sense the environment for any unusual activity. In some cases, they can collect data from the deployed sensors in the environment and transmit the captured data to the base station for further actions. Integrating of patrolling and tracking can provide various benefits such as: when the intruder is detected by patrolling team there is no need of additional communication and team to track, it can enhance the response time of the action taken on the intruder, and it eliminates the need for separate detection system to send alert when intruder enters the zone. Combining tracking and patrolling is widely applicable in civilian, military environments such as surveillance around a sensitive site, house, livestock farms, for security at airport, etc.

E-mail address: janardan18@gmail.com, virenderranga@dtu.ac.in

In our proposed work, initially, robots operate in a patrolling mode where multiple robots patrol a large premise, subdivided into locally protected zones that are assigned to each robot (as shown in Figure 1). Robots cannot enter the zones of other robots unless they are requested (through help message) for helping. When the intruder is detected and needed to be tracked, the corresponding robot immediately switches to the tracking mode. After completing their tracking task, they switch back to the patrolling mode. In this proposed solution, we have tried to minimize the idleness time of sensitive places and maximize the target's observation time with minimum estimation error and communication cost.

2. Related Work

Several multi-robot coordination algorithms for patrolling and target tracking have been proposed. The detailed information of such algorithms is presented in [9], [10], [11]. Chang et al. [12] developed a patrolling algorithm in a wireless sensor network (WSN). In the first step, the tours are planned in the algorithm. The next step consists of sending data mules on these precomputed tours, and in the last step, the speed of the dispatched nodes is regulated to fulfill the revisit constraints of the points of interest.

Authors in [13] proposed a patrolling algorithm that accounts for the robots joining and leaving the linear perimeter with dynamic length. The work shown in [14] focuses on the movement of robots along with their segments. The robots coordinate locally with their neighbors to reciprocate the fluctuations in the team size, speed, and perimeter length. A velocity controller for the robots following their tours is presented in [15]. The main focus is to reduce the uncertainty, which increases with different rates in an environment. In [16], the tours for other sensing locations are determined to meet the idleness constraint. Furthermore, the periodicity property of these tours is investigated. Nigam et al. [17] proposed a method to convert the problem of idleness minimization into the short horizon control law. This method finds the locations that the robot should visit. The patrolling problem using multiple robots can be categorized into two subparts: 1. Planning paths for multi-robots in the environment, 2. Coordinating and regulating the movement of the robots in these planned paths. Minimizing idleness is the optimization criteria in patrolling problems. Pasqualetti et al. [18], proposed algorithms to compute minimal idleness segments for the robots in a given cyclic graph or tree and chain. Authors in [19] have calculated a tour for the environment consisting of sensor locations, including various priorities. The control law is also developed for coordinating the movement of robots in the precalculated tour with the main focus on reducing weighted idleness.

Generally, target tracking approaches are based on approximating the target's position and coordinating the robots' or sensors to estimate the target position and move towards it. Various proposed works use Distributed Kalman Filter (DKF) with slight variations to evaluate the target's location [20], [21]. In such works, the information in the DKF expresses numerous measurements along with the covariance matrices. This information form in DKF can be signified to summarize each sensor's measurements and covariance matrix. In [20], a consensus algorithm is proposed to implement the Kalman filter in a distributed way and estimate the sums. Consensus algorithms provide various ad hoc applications for Kalman filters in a distributed manner. Consensus Kalman Filter (CKF) is proposed in [22] to optimize the consensus matrix, and Kalman gains design parameters. Speranzon et al. [23] presented a distributed adaptive algorithm to analyze the effect of packet losses and measurement noises. The adaptive weights are computed at each node locally to reduce the estimation error variance. The proposed algorithm also considers dispersing conditions for the weights to ensure the whole network's estimation stability. Some approaches need the formation control algorithms after estimating the target's position so that robots can congregate towards the approximate target's position. Hu et al. [24] implemented a vision-guided control strategy on two robot fishes for target tracking. In [25] and [26], path following and sliding-mode method are employed for formation control, respectively.

In our work, the proposed algorithm can handle tracking and patrolling simultaneously. Moreover, the waypoint generation (for patrolling purpose) can handle prioritized patrolling (keeping lower idleness of sensitive locations) and the patrolling trajectories are probabilistic thus making it difficult to be predicted by the intruder. This is largely neglected in the presented works. We have also developed self-triggered communication stagey to reduce the communication cost, which is not considered in the above works. For target position and estimation, we have used EKF, however our modified EKF is distributed and the EKF parameters are shared through communication messages (only at triggering conditions) only with the neighbor robots. Thus, minimizes the complexity and communication cost of distributed EKF.

3. Problem Description

A large site is considered where premises are divided into locally protected zones. Each robot can patrol in its local area of the site, ensuring comprehensive control over its most remote parts. In some cases, the robot can leave (for a fixed time interval) its local area to assist other robots. Each robot can operate in two modes patrolling and tracking. When the target needs to be followed, the robots can switch to tracking mode and then back to patrolling automatically.

The problem is to cooperatively patrol the entire perimeter of the site by the robots whose local area has one or more common edges with the site's perimeter. Robots which do not share any edge of their local area with the perimeter are responsible for area (not the perimeter) patrolling of the site. Whenever the presence of an unusual object or intruder (target to be tracked) is detected, all the robots should track the target cooperatively and share the target's live location. Robots update this to the security administrator for taking further actions. The main purpose of this proposed coordination approach is to minimize idleness time (significantly higher priority locations) and maximize the observation time of intruders (target) with the least communication cost and movement

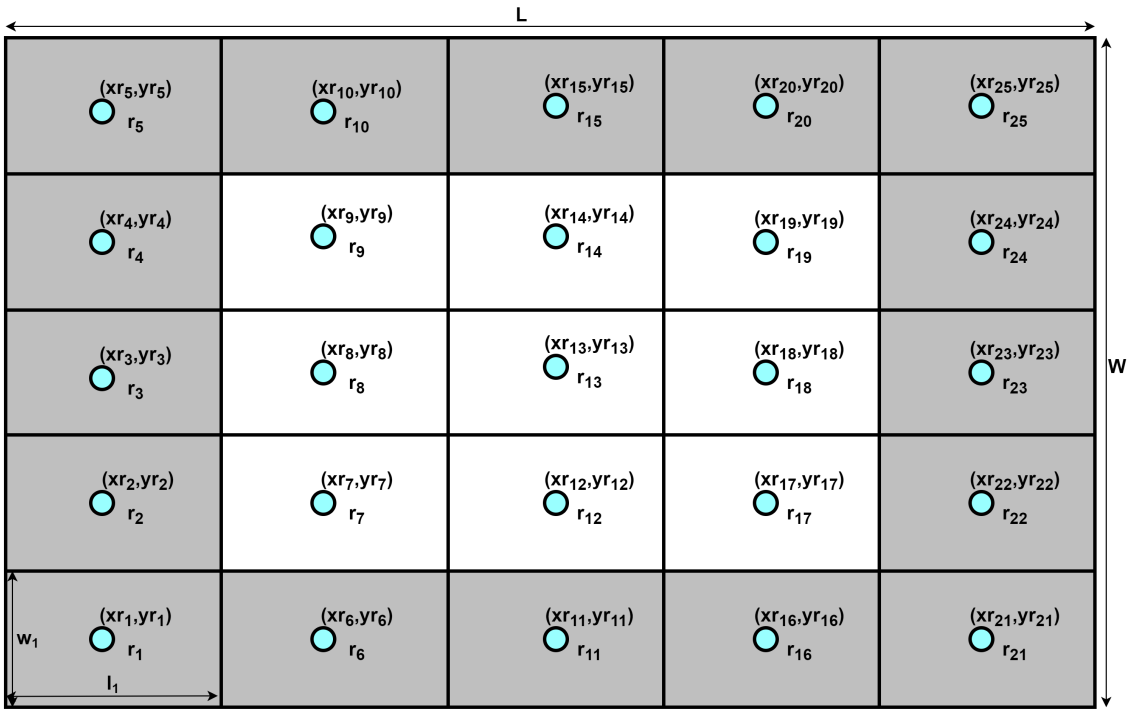


Figure 1. Problem scenario

of the robots. Coordination becomes more challenging when the target leaves the region of a robot and enters into the area of another robot.

For simplicity, the local area of each robot is represented as a rectangle or square. We divide the site of interest (S) into N number of local regions and deploy one robot into each local area of the site (S). $R = (r_1, r_2, r_3 \ldots . r_N)$ is the set of robots deployed on the site. The local area of the robot r_i can be represented by LA_i.

$$LA_i = ((xr_i, yr_i), (w_i, l_i)) \tag{1}$$

Here (xr_i, yr_i) is the center and (w_i, l_i) is the width and length of the local area. Robots are equipped with necessary sensors such as laser, camera, lidar for detecting the target and capable of communicating with monitoring station and neighbor robots. At time t the position and maximum speed of a robot can be denoted by $r_i(t) = (x_i^t, y_i^t, \theta_i^t)$ and $V_{max} r_i$ respectively. Once the target is in the sensing range of the robot, its position can be estimated using EKF. The estimated location of the target T at time t is represented by $T_t = (x_t, y_t)$.

Once the intruder has been detected inside the site's perimeter afterward, the target (intruder) can be tracked until it remains within the perimeter or neutralized. Between the time of first detection and neutralization, if any robot does not detect the target, then its predicted location can be used for tracking for a maximum time period of P_t seconds, without getting any new measurement. The predicted location of the target is denoted by $T_t^p = (x_t^p, y_t^p)$. We assume that robots can localize themselves, and the maximum speed of the target cannot be more than the maximum speed of robots $(V_{max} T < V_{max} r_i)$, i.

If two robots have one or more common edges between their local zones, such robots are defined as the neighbor of each other. For rectangular zones, any robot can have a maximum of four and a minimum of two neighbors. Due to the decentralized nature of the coordination approach, each robot maintains its neighbor list Nr_i.

$$Nr_i = [(r_j, LA_j)]\, j \tag{2}$$

Here j is the id of the neighbor robot. The neighbor list is a set containing pair of all the neighbors r_j, and their local area LA_j. Local zones assigned to each robot are fixed. Therefore, the neighbors of each robot are unchanged. Though any change in the neighbors can be communicated to the corresponding robots through the central monitoring server. The detection of target T at time t by the robot r_i can be defined as O_i^t :

$$O_i^t = \begin{cases} 1 & \text{if } T \in FOV \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

We have also developed a web app for monitoring and control of IoRT. Each robot can send its position and location of any unusual activity or target to the IoT server. The administrator can view live location and send commands to robots over the internet from any device capable of running web apps. Each robot has a self-triggered communication mechanism to inform (when the target is intended to enter the region of some robot) its neighbors about the presence of the target to be tracked.

4. Proposed Approach

This section presents our proposed coordination algorithm for patrolling and target tracking in an environment where the site is divided into local zones. We logically categories the robots into two types, based on their local

area (zone). First category robots are denoted by $C1$. The local boundary of these robots is part of the entire perimeter. The second category is $C2$. Robots of this category do not share any portion of their provincial border with the perimeter. Each robot patrols in its assigned area. $C1$ robots are primarily responsible for perimeter patrol only, though sometimes they may visit other parts of their zone also. $C2$ robots are only accountable for area patrolling within their local zone. By default, all robots are in patrol mode. The tracking mode is activated once the robot detects the object of interest or receives a tracking command from another robot.

For target tracking purposes, all the robots are equivalent. Each one can perform target tracking whenever an intruder or target of interest is detected. Robot switches to tracking mode and estimates the position of the target. Then the object can be continuously tracked and followed while it remains in its local area. Once the object leaves from the local zone of the robot, it switches back to patrolling mode. If the object is intended to enter the local area of some other robot, then prior information can be given to the corresponding robot to keep track of the target.

A. Self-triggered Communication Policy

We use three kinds of messages to coordinate the patrolling and tracking tasks. The communication is triggered whenever any of the following three events occur: 1. the target appears near to the local boundary of any robot, 2. the target enters in the local zone of a robot, 3. when an outer robot switch back to the tracking mode. These three events correspond to three types of messages, first two responsible for tracking, and the last is used for patrolling purposes. The $type1$ and $type2$ messages can only be sent by a robot who has a target in its sensing range or it has sufficiently accurate prediction (using EKF) of the target's position. This also means that it had the target in the sensing range just a little before. If the estimated position of the target is within the local zone of the robot, then $type1$ message can be sent to all the neighbors. Otherwise, $type2$ message can be sent only to the robot in whose local zone target has entered. In other words, the message can be sent to the robot, satisfying both equations (4) and (5).

When a robot receives $type1$ message, it computes the orthogonal distance of the target from all of its edges. If the target is found near a threshold value, the robot starts moving move towards the corresponding edge. Upon receiving $type2$ message, the robot can start moving towards the target to get it within the sensing range. The $type3$ message can only be sent by $C1$ robots. Whenever an outer robot goes in tracking mode, it sends $type3$ msg to its $C1$ category neighbors (always two in case of local rectangular zones). Therefore, $type3$ messages can only be received by $C1$ robots. Upon receiving this kind of message, the robot will assist the sender by patrolling some portion of the sender's perimeter. This message is used to ask assistance from neighbors whenever some outer robot ($C1$) switches to tracking mode.

$$Tx_i = \begin{cases} \text{true} & \text{if } \left(xr_i - \frac{l_i}{2} < x_t \leq xr_i + \frac{l_i}{2}\right) \\ \text{false} & \text{otherwise} \end{cases} \tag{4}$$

$$Ty_i = \begin{cases} \text{true} & \text{if } \left(yr_i - \frac{w_i}{2} < y_t \leq yr_i + \frac{w_i}{2}\right) \\ \text{false} & \textit{otherwise} \end{cases} \tag{5}$$

B. Dynamic Waypoint Generation

The purpose of waypoints is to guide the robot for patrolling. Dynamic waypoint generation makes it difficult or impossible to predict the patrolling pattern by intruders. Both categories ($C1$ and $C2$) of robots differ in how they generate waypoints, $C1$ robots need to have their waypoint near the perimeter, as shown as a dark shaded region in Figure 2(a). Sometimes, they also need to visit the inner area of their local zones (shown as a light shaded region in Figure 2(a)). However, each robot of this category majorly patrols near the perimeter (part of the entire perimeter). $C2$ robots apply two criteria for waypoint generation. First, uniformly distribute the waypoint within the local zone such that the entire local zone can be covered with minimum idealess time. Second, give extra emphasis near the edges, which are common with a $C1$ robot. Suppose $C2$ robot has neighbors of category $C1$ then near the corresponding edges. In that case, more waypoints should be generated because that side of the local zone has a high probability of intruders being spotted, as shown in Figure 2(a). If there are no such neighbors, then waypoints should be uniform throughout the zone, as shown in Figure 2(c). There are four $(\alpha, w_dis_{min}, w_dis_{max}, c_{rate})$ parameters for waypoints generation algorithm. Using these parameters, we can control how far the next waypoint be and how fast robot want to converge towards the destination. The lower value of c_{rate} means the slow convergence towards the destination, and α controls the arbitrariness in the direction or number of direction a waypoint can be generated. A significant value means more alternate paths. The random distance (r_{dis}) of the next waypoint can be minimum w_dis_{min} and maximum w_dis_{max} from the current position. Every time the waypoint is generated at a random distance from the current position (step 3).

Then we find (using equation (7)) α number of equally spaced points on the circumference of circle with center (st_x, st_y) and radius r_{dis}.

$$\theta = \left\{ n\frac{2\pi}{\alpha} \right\}, n0, 1, 2 \ldots \alpha - 1 \tag{6}$$

Where θ is set of equally spaced angles in radian corresponding to each equally spaced point on circumference and n is an integer ($0n < \alpha$).

$$PO = \{(st_x + r_{dis}\cos\delta,\ \ st_y + r_{dis}\sin\delta)\ \ \ \delta\theta\} \tag{7}$$

Where PO is the set of all equally spaced points on the circumference. Now we need to calculate distance of each point in PO from the destination and find the point which

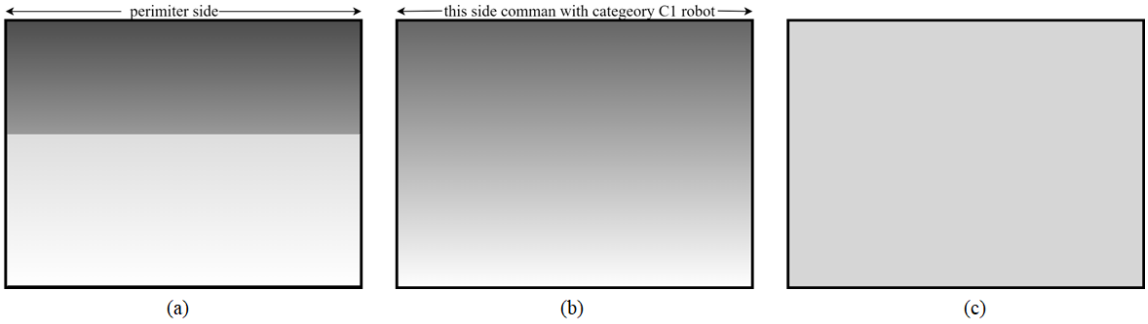


Figure 2. Varying density of waypoint generation

Algorithm- 1:

1. ***generate _waypoints***$(start,\ destination)$:
2. $define\ \alpha, w_dis_{min}, w_dis_{max}, c_{rate}$
3. $get\ random\ distance\ r_{dis}\ where\ (w_{dis_{min}} \leq r_{dis} \leq w_{dis_{max}})$
4. $PO = find\ \alpha\ points\ using\ (6) and\ (7)$
5. ***for each*** $point$ ***in*** PO
6. $find\ its\ distance\ from\ destination$
7. $calculate\ weight\ and\ probability\ of\ the\ point\ using\ (9)\ and\ (10)$
8. ***end for***
9. $PO_u = select\ a\ random\ point\ as\ next\ waypoint\ using\ (11) and\ (12)$
10. $start = PO_u$
11. ***end***

has minimum distance from it.

$$b = MIN\left(\left\{\sqrt{(PO_{n,x} - dst_x)^2 + \left(PO_{n,y} - dst_y\right)^2}\right\} \quad \forall n \in \{0, 1, 2 \ldots, \alpha - 1\}\right) \tag{8}$$

Where $PO_{n,x}$ is x coordinate of n^{th} point in the set PO, similarly $PO_{n,y}$ is n^{th} y coordinate. Here z is the index of point which has minimum distance to the destination point (dst_x, dst_y).

$$W = \left\{e^{(c_{rate} * (\cos(\theta_i - \theta_z) - 1))}\right\} \quad \forall i \in \{0, 1, 2, \ldots \alpha - 1\} \tag{9}$$

Where W is the set of weights assigned to each points of set PO. θ_z is angle of the point which has minimum distance to the destination. Now we need to calculate the probabilities of each point. After that, a probability is assigned to each point in PO based on its distance from the destination and c_{rate} (step 5-8). Points closer to the destination gets a higher probability.

$$P = \left\{\frac{W_i}{\sum_{u=0}^{\alpha} W_u}\right\} \quad \forall i \in \{0, 1, 2, \ldots \alpha - 1\} \tag{10}$$

Where P is the set of probabilities for each point under consideration, W_i is the weight of i_{th} point from set W.

$$E = [E_0,\ E_1,\ E_2 \ldots . E_\alpha] \tag{11}$$

Where $E_0 = 0$, $E_j = \sum_{i=0}^{j-1} P_i \quad \forall j \in \{0, 1, 2, \ldots \alpha\}$ Now generate a random number rd between zero and one. After that find the index u such that equation (12) is satisfied.

$$E_u rd < E_{u+1} \tag{12}$$

Where $u\alpha - 1$. Therefore, the next generated waypoint is PO_u, points having higher weight in W have high chance to be selected however others can also be selected.

Value of c_{rate} plays a vital role in this regard, large c_{rate} means the difference between the probability, assigned to points close to the endpoint, and points farther can be significant. Therefore, the chance of selecting a waypoint close to the endpoint may be higher. However, for smaller values of c_{rate} this difference can be lesser. Thus, the chance of choosing a waypoint close to the endpoint cannot be as high.

C. Coordination Algorithm

Robots need to coordinate their actions to perform an efficient patrolling and tracking. The complete coordination algorithm is presented in *Algorithm* – 2. The default mode of all the robots deployed in the vicinity is patrolling. Each robot depending on its category generates dynamic waypoints (Wl_i). Procedure shown in *Algorithm* – 1 is the same for all the robots to generate waypoints. However, the density of waypoints generated at different places inside the LA_i differs for each category of robot. if there is no intruder detected on the site, robots can continue their patrolling (step 3-9). Whenever an intruder is detected, the corresponding robot switches in the tracking mode and starts following the intruder; all other robots keep patrolling in their respective local zones. The extended Kalman Filter (EKF) is used for the estimation and prediction of the target position (equation (13) - (17)).

Prediction :

$$X^{'} = DX + \beta \tag{13}$$

$$P^{'} = DPD^T + \gamma \tag{14}$$

Update :

$$K = P^{'} H^T (HP^{'} H^T + R)^{-1} \tag{15}$$

$$X = X^{'} + K * (z - HX^{'}) \tag{16}$$

$$P = (I - KH)P^{'} \tag{17}$$

Where X is the state of the target, K is Kalman gain matrix, γ is noise covariance matrix, P is process covariance matrix, $P^{'}$ is predicted process covariance matrix, z is the measurement vector, H is measurement function, and I is the identity matrix equal to the size of state vector X.

Algorithm- 2:

Inputs

LA_i: Local zone of Robot i,

Nr_i: Neighbours of Robot i,

Wl_i: Dynamic List of waypoints to follow,

P_t: Prediction Timer (how long to predict)

F: Frequency of sending type1 messages

1. mode = Patroling;
2. **Repeat** (at each time step t):
3. **if** (mode = Patroling)
4. **if** (Wl_i list is empty):
5. Wl_i = generate waypoints using Algorithm 1
6. **end if**
7. move to next waypoint (w_x, w_y) in the Wl_i
8. $Wl_i = Wl_i - (w_x, w_y)$
9. **end if**
10. **if** (mode = Patroling and category = C1)
11. **if** (help message received)
12. update Wl_i to patrol some area of other robot
13. **end if**
14. **if** (stop_help message received)
15. update Wl_i to patrol own area only
16. **end if**
17. **end if**
18. **if** (type1 message received)
19. compute target distance from all edges
20. **if** (target is very near to any edge)
21. move towards the edge nearest to target
22. **end if**
23. **end if**
24. **if** (type2 message received)
25. move towards target
26. update EKF using Int_message.EKF_data
27. start P_t
28. **if** (mode = Patroling and category = C1)
29. Send help msg to Nr_i
30. **end if**
31. mode = Tracking
32. **end if**
33. **if** (target in sensing range)
34. detect and estimate target position
35. update EKF using measurment
36. start P_t
37. **if** (estimated position is within LA_i)
38. move towards target;
39. **if** (mode = Patroling and category = C1)
40. Send help msg to Nr_i
41. **end if**
42. mode = Tracking
43. **if** (target is near edge)
44. send type1 msg to Nr_i once in F secs
45. **end if**
46. **end if**
47. **else**
48. if mode = tracking: start patroling timer
49. R_j = estimated target position is within LA_j
50. send type2 msg to R_j
51. **end else**
52. **end if**
53. **else**
54. **if** (P_t is valid)
55. predict target position
56. **if** (predicted position is within LA_i)
57. move towards target
58. **if** (target is near edge)
59. send type1 msg to Nr_i once in F secs
60. **end if**
61. **end if**
62. **else**
63. if mode = tracking: start patroling timer
64. R_j = predicted position is within LA_j
65. send type2 msg to R_j
66. **end else**
67. **end if**
68. **end else**
69. **if** (patroling timer expired)
70. **if** (mode = Tracking and category = C1)
71. Send stop_help msg to Nr_i
72. **end if**
73. mode = Patroling
74. **end if**
75. **End Repeat**

$$D = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} \frac{a_x \Delta t^2}{2} \\ \frac{a_y \Delta t^2}{2} \\ a_x \Delta t \\ a_y \Delta t \end{pmatrix}$$

Here β is random acceleration vector and $a_x \sim N(0, \sigma_{ax}^2)$ and $a_y \sim N(0, \sigma_{ay}^2)$ are random acceleration in x and y direction.

Whenever a $C1$ category robot switches from patrolling to tracking mode, it must send a help message to its neighbors (step 31-34, 43-46) to patrol its perimeter while it's in tracking mode. While tracking the target when the target/intruder reaches very near (predefined threshold) to any edge of the local zone, the robot sends a $type1$ message to neighbors (step 47-49, 63-65). The purpose of this message is to alert the neighbors that intruded may enter their local zone. Therefore, neighbors should move towards the corresponding edge from where intruders can come (step 20-25). When a robot detects the target in another robot's local zone, it sends a $type2$ message, and receiving robot moves towards the location, observed in the message (step 51-55, 67-71). It also starts its patrolling timer if the robot is in tracking mode. This means that the intruder is out of its local zone. Therefore, this robot will return to patrolling mode after waiting a certain amount of time, i.e., when the patrolling timer expires. Whenever such patrolling timer expires for a $C1$ robot, it sends a $stop_help$ message to its neighbors (step 74-77), upon receiving this message, neighbors stop patrolling the perimeter of other robots (step 15-17).

We have used a central server, which maintains the global information of our IoRT system. It has information about robots. If the target gets lost, then the server can estimate its approximate location and direct the nearby robot to search.

D. Algorithm Analysis

In the tracking part of proposed algorithm, EKF is well known for the estimation and prediction of dynamic state of a linear and non-linear system. The acceleration of the target is modelled as noise ($a_x \sim N(0, \sigma_{ax}^2)$, $a_y \sim N(0, \sigma_{ay}^2)$). We have used timer P_t to ensure the accuracy of predicted state.

Scalability: The algorithm is well scalable due to its decentralized (all robots other than neighbors do not need to take part in any communication) nature.

Communication cost: Number of messages has been greatly reduced due to self-triggered policy. The number of type1 messages depends on F and Nr_i. Which can be reduced (by increasing F) further at cost of tracking quality.

$$No.of\ type1\ msgs\ \propto \frac{(Nr_i)}{F}$$

The number of $type2$ messages is directly proportional to the number of times target moves from the region of one robot to another. $help$ and $stop_{help}$ messages are only sent by outer robots when target passes through their region, which is less likely to happen once the target is inside the perimeter. Therefore, such messages are needed when a new target first enters in the site.

In the patrolling part of the proposed algorithm, $(\alpha, w_dis_{min}, w_dis_{max}, c_{rate})$ provide the required control over waypoint generation. Random number rd used in equation (12) ensures that patrolling path is unpredictable (any point can be selected as next waypoint even if it's not at minimum distance from end point). Distance between the waypoints, randomness in their direction can be controlled by c_{rate} and α.

The algorithm will always find the next waypoint close to the end point because we look in every direction (equation (7)).

$$\theta = \left\{ n\frac{2\pi}{\alpha} \right\}, n0, 1, 2 \ldots \alpha - 1 \tag{18}$$

After few iterations the waypoints will converge near the end point. This is ensured by weights assigned as per following:

$$Weights = \left\{ e^{(c_{rate} * (\cos(\theta_i - \theta_z) - 1))} \right\}$$

θ_z is the angle of point which has minimum distance to the end point. θ_i is angle of all other points.

5. Simulation and Results

We have implemented and simulated our proposed coordination algorithm with Gazebo, ROS, and Matlab. The robots (Pioneer-3AT [27]) are differential drive-based, and each robot is equipped with a camera to detect the target, having a sensing range $6m$ and fov $1.5rad$. Each robot has a maximum linear speed of $1.4m/s$, and angular velocity of $0.6rad/s$. Each robot has assigned a local zone of $30 * 30m^2$. For most of the experiments site area is $120 * 150m^2$, which means twenty robots (grid of $4 * 5$) need to be deployed to cover the entire area. We have also simulated nine robots with $3 * 3$ grid to verify case when $C2$ robot has four neighbors of type $C1$. Each robot has a peer to peer connection with all (maximum four) its neighbors.

Gazebo[28] is a 3D robotics simulator, it can simulate the robot dynamics, environment, and targets. Figure 3, shows Gazebo environment with twenty robots and moving targets. Targets are simulated as blue color ball; they are continuously moving on random (as shown in Figure 7) and circular path. For some time, they remain inside the perimeter then move out and after some time some of them again enter the site. The simulator is capable of generating sensor data (such as image, IMU data, location etc.) and motion, with due consideration of gravity, friction etc. ROS can interact with Gazebo thus it can receive and send commands to entities (robots and target) present in the Gazebo environment. we have developed a ROS node which is deployed on each robot, this node is responsible for communication with IoT server using MQTT protocol. Matlab can also interact directly with ROS, and through ROS it can communicate with Gazebo.

The area of the local zone can be decided based on the sensor range, required idealness in patrolling, and accuracy in tracking. If the local zone is large, then the

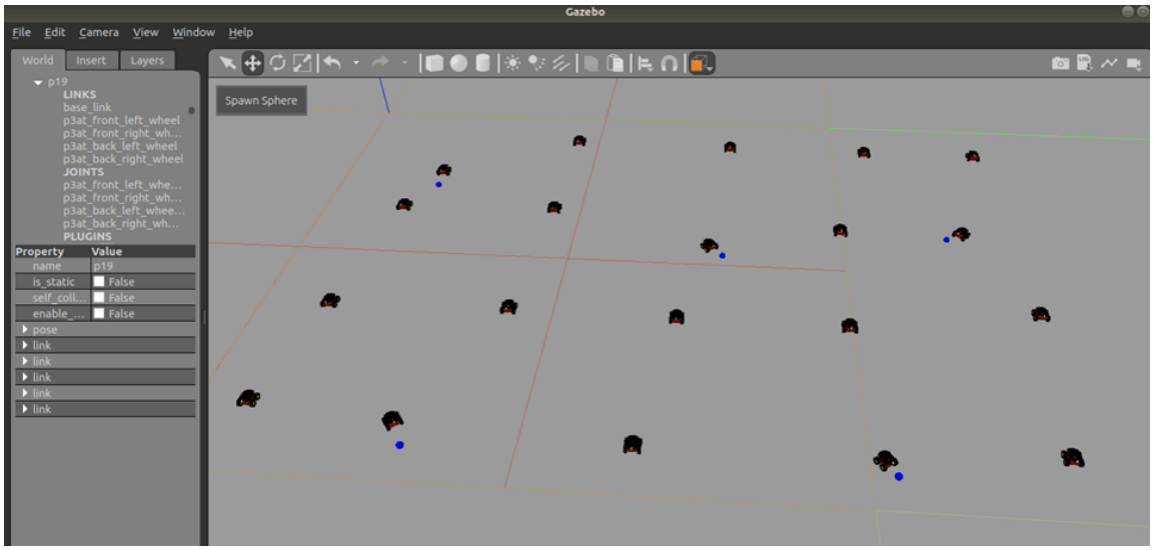


Figure 3. Gazebo Simulation Environment

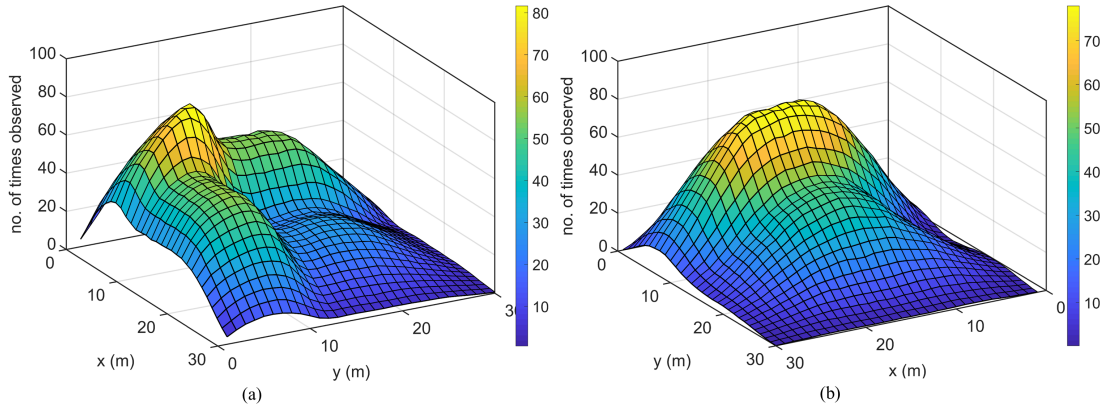


Figure 4. Observation count of each point (within the local zone) (a) of C1 with two open edges (b) of C1 with one open edges

idealness may increase. However, it does not affect the tracking quality if triggering condition (threshold distance) for sending $type1$ msg is also adjusted accordingly. Larger the zone area larger the threshold should be. However, it can increase the communication cost. In our scenario, we do not need the idealness of each point in the local zone to be equal or fair. In this case, robots require to visit some places frequently as compare with other places. For category $C1$ robots, need to visit points near the site's perimeter frequently, and category $C2$ robots need to visit those places near the local zone of $C1$ robots, i.e., frequent visits near the edges command with $C1$ robots.

Figure 4(a) shows how many times different points in the local zone are observed (during one patrol cycle) by the robot of category $C1$ and two edges of its local zone from the perimeter. Figure 4(b) shows the same for the robot of category C1 but only one edge of its local zone forms the perimeter. Each point within the local zone is considered as an area of $1 * 1m^2$. As expected in both cases robots are frequently patrolling the area near the perimeter. The peaks indicate the highest visited points. In Figure 4(a), there are points along the line around $y = 4$ *to* 12, and another line around $x = 4$ *to* 12. These points are near the perimeter therefore visited at high frequency compared to other points in the zone. However, in Figure 4(b), only one edge forms perimeter, only points along the line around $y = 4$ *to* 15 are visited the most. This satisfies our requirement, the waypoints are dynamic so robot does not move from fixed places, and places that need more attention are observed frequently.

The local zone of robots of category $C2$ does not have any perimeter. However, they can have a varying number of neighbors depends on the position of their local zone. Here we are only concerned about neighbors of category $C1$. As can be seen in Figure 1, r_7 has two such neighbours (r_2 *and* r_6), r_{12} has one such neighbour (r_{11}) etc. Therefore there can be at most four cases in the case of the local rectangular zone; having four (only exists if nine robots are deployed in a $3 * 3$ grid) or two (such as r_{19}) or one (such as r_{14}) or none (such as r_{13}) of such neighbours. Figure 5 shows the number of times different points in the local zone are observed (during one patrol cycle) by the robot of category $C2$. Figure 5(a) presents the case when there are four neighbors of category $C1$ that means that there are high chances that intruder may come from any of the four edges. Therefore, the robot visits near the four edges frequently, it can be seen in Figure 5(a). however, it leads to four peaks because corner places

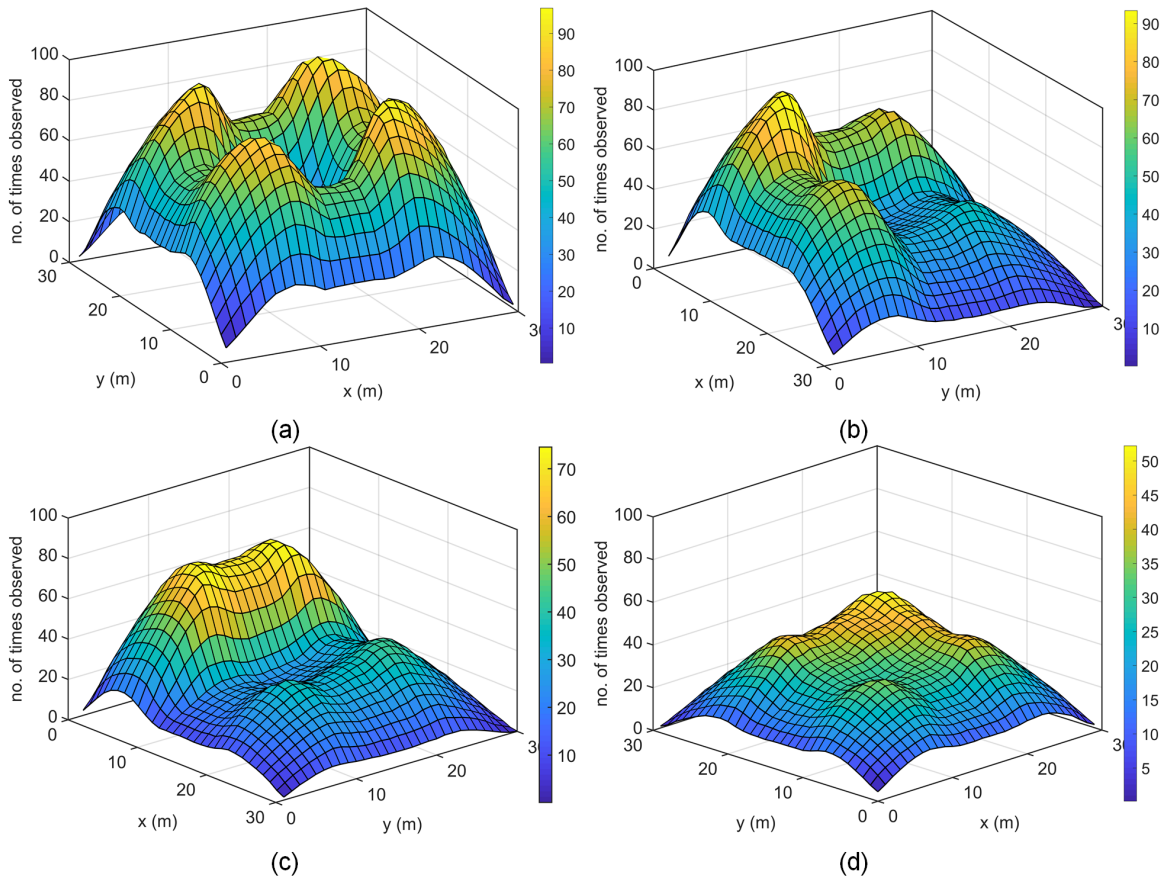


Figure 5. Observation count of each unit point (within the local zone) of C2 robots (a) with four neighbors (b) with two neighbors (c) with one neighbor (d) with no neighbor

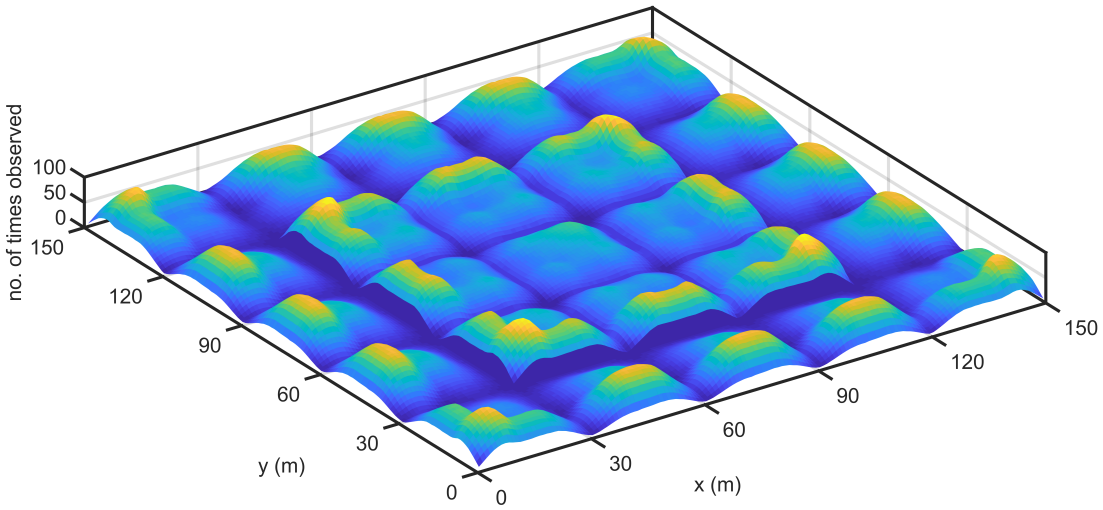


Figure 6. Observation count (each unit point) of entire site with twenty-five robots

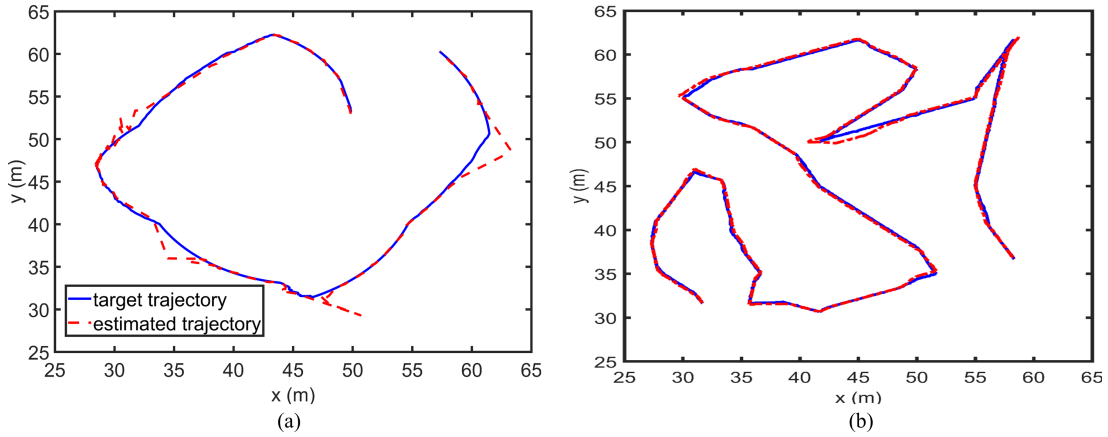


Figure 7. Tracking performance with the target moving on a random path (a) type1 message only (b) both type1 and type2 messages

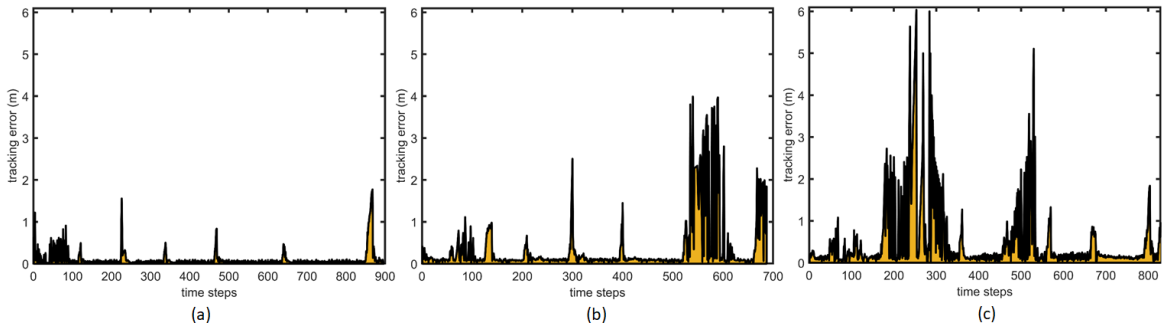


Figure 8. Tracking error with the target moving at (a) $v_T = 0.9m/s$ (b) $v_T = 1.2m/s$ (c) $v_T = 1.4m/s$

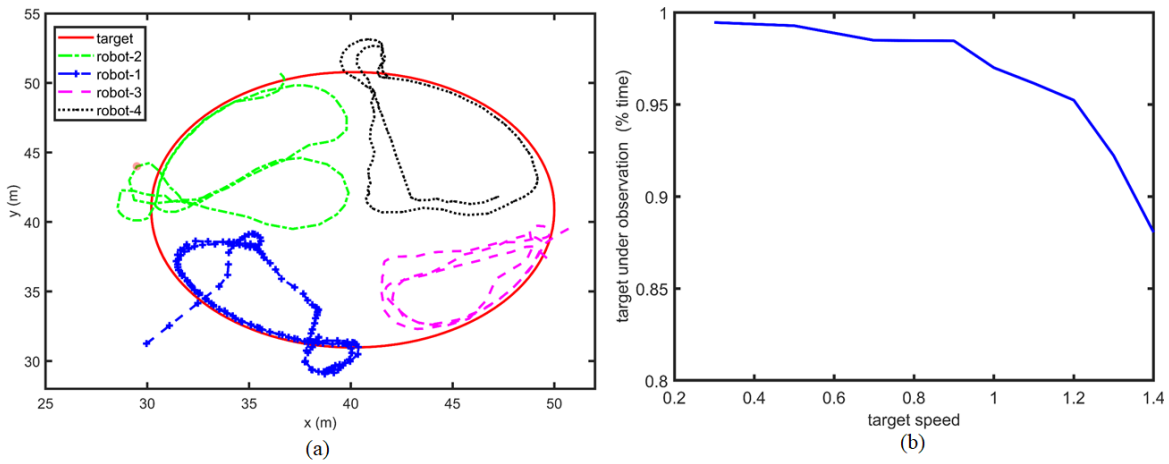


Figure 9. (a) Path followed by robots while tracking the target (b) Effect of observation time with the increasing speed of the target

act as an exchange point for two edges. Thus, visiting frequency of corner points is approximately thirty percent higher than points along the edges. Figure 5(b) shows the result for a robot that has two such neighbors. Therefore, it visits points along with the two of its borders frequently. It seems similar to Figure 4(a); however, the difference is here, the remaining (other than near the edges) area is visited more frequently as compared to Figure 4(a).

Robots are having only one of such neighbors need to worry about only one of its edge. Figure 5(c) shows the result for such a case. One side is visited frequently at other places visiting frequency is around fifty percent lesser. Figure 5(d) is the case when there is no such neighbor. Therefore, the robot is only responsible for area patrolling without special attention near the edges. Therefore, curve is almost flat, except for some places with small peaks near the corners.

Simulation has been performed with extreme settings such as target moving with complex and sharp turns, speed of target very close to the maximum speed of robots, target moving from one zone to another very frequently, to verify that the algorithm can perform coor-

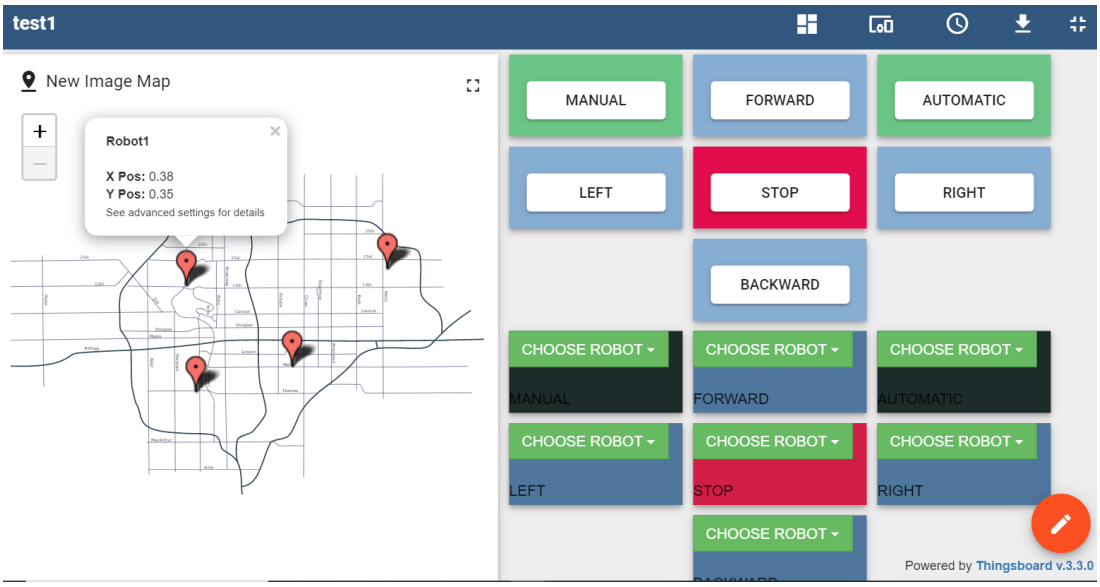


Figure 10. Snap of web-app using IoT server

dinated tracking in extreme situations.

Figure 6, shows the result of simulation over forty iterations with twenty-five robots. It can be seen in Figure 6 that all outer robots patrol in such a way that entire perimeter has a high (yellowish peaks) observation count. $C2$ robots have also formed a wall like defense to detect any intruder, which may have not be detected by $C1$ robots. Edges having less observation count are in turn have higher count by neighbor robot.

In Figure 7, tracking performance is shown; the difference between estimated and actual trajectory shows the deviation from real values. Results are shown in Figure 7(a) shows poor performance (because it only uses $type1$ message) as compared to Figure 7(b) (using both $type1$ and $type2$ message). This shows the importance of $type2$ messages. Whenever intruder/target move from the local zone of one robot to other, the chances of error in the estimated position of the target increases. This is because the robot needs to hand over the target to another robot when it reaches the end of its local zone. Figure 7(b) target takes sharp turns, and crossing the local zones is frequently still. The performance is quite good. It can be seen that the performance decreases when the target crosses the grid lines (switches the zone). Sometimes there can be a higher error in estimation due to a sudden change in target speed, direction.

Figure 8 shows an error in the estimated position of the target with respect to the time with varying target speeds. As the speed of the target is increasing from $0.9m/s$ to $1.4m/s$, the maximum error (at the crossing area) in the estimation is also increasing. At a speed of $0.9m/s$ the maximum error is around $1.9m$, at speed of $1.2m/s$ it is around $4m$, and at speed of $1.4m/s$ (this the maximum speed of robots), it's about $6m$. Along with the increment in the maximum error, the persistence time of the error is also increasing. At a lower speed, the error persists for a shorter time spawn. Accurate estimation of target position becomes more complex at a higher rate. However, the error while the target is within the local zone (not crossing it) is relatively low, as shown in Figure 8(a). The error is around $0.1-0.2m$ only. Even when the target speed is equal to the maximum speed of the robot, the error is below $0.5m$ most of the time and the robots were able to successfully track and follow the intruder.

Figure 9(a) shows the path of the target and path of the four (from whom local zone target passes) robots while tracking. It can be observed here that the robots are successfully coordinating while remaining in their respective local zones. Here, the local zone of all four robots is $40 * 40m$. Figure 9(b) shows how the target observation time decreases with an increase in speed. It explains why the error in the estimation increases, as shown in Figure 8.

Figure 10 shows a screenshot of the web app connected to the IoT server. The left part shows the current location in the environment and other related information of each robot. The right part shows the control menus of the robots. The algorithm allows robots to coordinate for both patrolling and tracking simultaneously in the application. When an intruder is present in a robot's local zone, it performs monitoring while other robots (except neighbors) keep patrolling and vice versa.

In future we would like to address tracking of multiple targets, which is a challenging problem and it becomes harder when patrolling is performed simultaneously. How to select and assign robots when there are multiple targets? Another challenge arises when the speed of the intruder is higher than the robots in that case how to communicate and pre plan the robots which can possibly detect the intruder, it may need a prediction algorithm to trigger the robots which are already deployed near to the trajectory (future) of the intruder.

6. Conclusion

In this paper, we proposed a decentralized coordination algorithm that can perform both patrolling and target tracking in the environment where the site is divided into local zones, and each robot has assigned one zone. Patrolling is based on dynamic waypoint generation; therefore, routes taken by the robot are sufficiently distinct at the same time, it provides higher visiting frequency (or lower idleness) at certain locations as per the requirements. Robots are able to assist each other by extending and shrinking the local perimeter to be patrolled. To initiate and complete this assistance process, exactly four messages (by category C1 only) are required. Any intruder, if detected by any robot, can be tracked and followed by the robot. The information regarding the position and EKF parameters are shared with the neighbors only when the target has crossed the threshold. This, in turn, alerts the neighbors. The results show that the error in estimating the target's position at a higher speed is around 0.5 m. It increases only when the target switches from one local zone to other. Robots are successfully able to track, follow, and share the required information with other neighbors. At any point, the remote user can use a web application to monitor the position and status of robots and intruders (if present), and each robot can also be controlled by it. While patrolling, the exchange position from where a robot passes several times while moving from one side of the local zone to another is visited more than the required frequency, this can be improved in the future. The detection and tracking of an intruder can also be improved by using static sensors in the environment.

Acknowledgment

This work is supported by the University Grant Commission, Government of India, and Nvidia GPU Grant.

References

[1] J. K. Verma and V. Ranga, "Multi-Robot Coordination Analysis, Taxonomy, Challenges and Future Scope," *Journal of Intelligent Robotic Systems*, vol. 102, no. 1, p. 10, may 2021.

[2] B. Rao, H. Durrant-Whyte, and J. Sheen, "A Fully Decentralized Multi-Sensor System For Tracking and Surveillance," *The International Journal of Robotics Research*, vol. 12, no. 1, pp. 20–44, feb 1993.

[3] P. Simoens, M. Dragone, and A. Saffiotti, "The Internet of Robotic Things: A review of concept, added value and applications," *International Journal of Advanced Robotic Systems*, vol. 15, no. 1, pp. 1–11, 2018.

[4] P. P. Ray, "Internet of Robotic Things: Concept, Technologies, and Challenges," *IEEE Access*, vol. 4, pp. 9489–9500, 2016.

[5] M.-F. Ge, Z.-H. Guan, B. Hu, D.-X. He, and R.-Q. Liao, "Distributed controller–estimator for target tracking of networked robotic systems under sampled interaction," *Automatica*, vol. 69, pp. 410–417, jul 2016.

[6] X.-Y. Yao, H.-F. Ding, and M.-F. Ge, "Task-space tracking control of multi-robot systems with disturbances and uncertainties rejection capability," *Nonlinear Dynamics*, vol. 92, no. 4, pp. 1649–1664, jun 2018.

[7] M. Z. A. Bhuiyan, G. Wang, and A. V. Vasilakos, "Local Area Prediction-Based Mobile Target Tracking in Wireless Sensor Networks," *IEEE Transactions on Computers*, vol. 64, no. 7, pp. 1968–1982, 2015.

[8] M. Elhoseny and A. E. Hassanien, "Mobile Object Tracking in Wide Environments Using WSNs," in *Studies in Systems, Decision and Control*, 2019, vol. 165, pp. 3–28.

[9] A. Ez-Zaidi and S. Rakrak, "A Comparative Study of Target Tracking Approaches in Wireless Sensor Networks," *Journal of Sensors*, vol. 2016, no. i, pp. 1–11, 2016.

[10] D. Portugal and R. Rocha, "A Survey on Multi-robot Patrolling Algorithms," in *IFIP Advances in Information and Communication Technology*, 2011, vol. 349 AICT, pp. 139–146.

[11] A. Khan, B. Rinner, and A. Cavallaro, "Cooperative Robots to Observe Moving Targets: Review," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 187–198, jan 2018.

[12] C.-Y. Chang, G. Chen, G.-J. Yu, T.-L. Wang, and T.-C. Wang, "TCWTP: Time-Constrained Weighted Targets Patrolling Mechanism in Wireless Mobile Sensor Networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 6, pp. 901–914, jun 2015.

[13] D. Kingston, R. W. Beard, and R. S. Holt, "Decentralized perimeter surveillance using a team of UAVs," *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1394–1404, 2008.

[14] J. J. Acevedo, B. C. Arrue, I. Maza, and A. Ollero, "Cooperative perimeter surveillance with a team of mobile robots under communication constraints," in *IEEE International Conference on Intelligent Robots and Systems*, 2013, pp. 5067–5072.

[15] S. L. Smith, M. Schwager, and D. Rus, "Persistent robotic tasks: Monitoring and sweeping in changing environments," *IEEE Transactions on Robotics*, vol. 28, no. 2, pp. 410–426, apr 2012.

[16] J. L. Fargeas, B. Hyun, P. Kabamba, and A. Girard, "Persistent visitation under revisit constraints," in *2013 International Conference on Unmanned Aircraft Systems, ICUAS 2013 - Conference Proceedings*, vol. 28, no. 2, apr 2013, pp. 952–957.

[17] N. Nigam, S. Bieniawski, I. Kroo, and J. Vian, "Control of multiple UAVs for persistent surveillance: Algorithm and flight test results," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 5, pp. 1236–1251, 2012.

[18] F. Pasqualetti, A. Franchi, and F. Bullo, "On cooperative patrolling: Optimal trajectories, complexity analysis, and approximation algorithms," *IEEE Transactions on Robotics*, vol. 28, no. 3, pp. 592–606, 2012.

[19] F. Pasqualetti, J. W. Durham, and F. Bullo, "Cooperative patrolling via weighted tours: Performance analysis and distributed algorithms," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1181–1188, 2012.

[20] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proceedings of the IEEE Conference on Decision and Control*, 2007, pp. 5492–5498.

[21] R. Rahman, M. Alanyali, and V. Saligrama, "Distributed tracking in multihop sensor networks with communication delays," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4656–4668, sep 2007.

[22] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622–633, may 2008.

[23] A. Speranzon, C. Fischione, B. Johansson, and K. H. Johansson, "Adaptive distributed estimation over wireless sensor networks with packet losses," in *Proceedings of the IEEE Conference on Decision and Control*, 2007, pp. 5472–5477.

[24] Y. Hu, W. Zhao, and L. Wang, "Vision-based target tracking and collision avoidance for two autonomous robotic fish," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 5, pp. 1401–1410, 2009.

[25] J. Ghommam and F. Mnif, "Coordinated path-following control for a group of underactuated surface vessels," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 10, pp. 3951–3963, 2009.

[26] M. Defoort, T. Floquet, A. Kökösy, and W. Perruquetti, "Sliding-mode formation control for cooperative autonomous mobile robots," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 11, pp. 3944–3953, 2008.

[27] "Pioneer-3AT, http://wiki.ros.org/Robots/AMR˙Pioneer˙Compatible (accessed Nov. 05, 2021)."

[28] "Gazebo, http://gazebosim.org/ (accessed Nov. 05, 2021)."

Janardan Kumar Verma Janardan Kumar Verma received M.Tech degree in Computer Engineering in 2014 from National Institute of Technology, Kurukshetra, India. Currently he is pursuing Ph.D. at the Department of Computer Engineering, National Institute of Technology, Kurukshetra, India. His current research interests include Mobile Sensor Networks, Multi-robot System, Autonomous vehicles, and Artificial Intelligence.

Virender Ranga Virender Ranga has received his Ph.D. degree in 2016 from Computer Engineering Department of National Institute of Technology, Kurukshetra, Haryana, India. He worked as Assistant Professor in the Computer Engineering Department, National Institute of Technology Kurukshetra, India since March 20, 2008 to October 25, 2021. Currently, Dr. Virender Ranga is working as Associate Professor in IT Department of Delhi Technological University, Delhi. He has been published more than 100 research papers in various SCI/SCIE/SCOPUS Journals and reputed International Conferences in the area of Computer Communications and Computer Security. He has been conferred by Young Faculty Award in 2016 for his excellent contributions in the field of Computer Communications. He is a member of editorial board of various reputed journals like Journal of Applied Computer Science Artificial Intelligence, International Journal of Advances in Computer Science and Information Technology (IJACSIT), Circulation in Computer Science (CCS), International Journal of Bio Based and Modern Engineering (IJBBME) and International Journal of Wireless Networks and Broadband Technologies. Currently, he is acting as Ambassador of Asia Region for Bentham Science. He is an active reviewer of many reputed journals of IEEE Transactions, Springer, Elsevier, Talyor Francis, Wiley and InderScience. His research area includes Wireless Sensor Adhoc Networks Security, IoT security, FANET security, SDN security, IoRT etc.

View publication stats

Future Energy
Open Access Journal
ISSN 2832-0328

May 2023| Volume 02 | Issue 01 | Pages 12-20

Journal homepage: https://fupubco.com/fuen

https://doi.org/10.55670/fpll.fuen.2.2.2

Future Publishing LLC

Article

A feasibility study and cost-benefit analysis of an off-grid hybrid system for a remote area electrification

Upma Singh[1,2*], Mohammad Rizwan[1]

[1]Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India
[2]Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, Delhi 110058, India

ARTICLE INFO

Article history:
Received 10 November 2022
Received in revised form
20 November 2022
Accepted 22 November 2022

Keywords:
Renewable Energy, HOMER, Biogas,
Solar Photovoltaic,
Hybrid Renewable Energy Sources (HRES)

*Corresponding author
Email address:
upma@msit.in

DOI: 10.55670/fpll.fuen.2.2.2

ABSTRACT

Off-grid power production utilizing renewable sources of power has become more significant and viable to meet the limited demands of remote locations. The primary goal of this study is to develop an economic and optimal hybrid PV/Biogas configuration for power production for rural common facilities, including one Primary school, Junior school, and Panchayat Ghar buildings of Sarai Jairam village in Uttar Pradesh, India. Data on the electric load was gathered for two schools and Panchayat Ghar. The PV/biogas hybrid configuration was designed utilizing the Hybrid Optimization Model for Electric Renewable (HOMER), and techno-economic analysis is carried out to fulfill the load requirements. The available biomass potential and the data on solar irradiance of the study area were utilized in the HOMER software to carry out the analysis. The HOMER analysis produced a solution that included total net present cost (NPC) and cost of electricity (COE), and these results were then further improved using sensitivity analysis. The sensitivity analysis employed sensitivity parameters like biomass potential, biomass pricing, solar irradiance, and variation in loads. Based on the NPC and COE, this analysis evaluates the system performance and demonstrates that it is techno-economically feasible.

1. Introduction

The unavailability or shortage of electrical networks in remote locations, the excessive cost of grid extension, and the harsh topography frequently lead to exploring other alternatives. One of the most promising approaches to meet these areas' need for electrification now involves standalone hybrid systems [1]. The grid-connected electricity for power delivery does not always benefit people. For instance, individuals who live in isolated villages, distant farms, and rural locations must rely on autonomous power supply systems that are based on the burning of fossil fuels [2]. The issue is that these systems increase greenhouse gas (GHG) emissions, which in turn lead to air pollution [3]. By utilizing renewable energy sources like solar, wind, biofuels, and others, GHG emissions may be decreased. Rural electrification is crucial to improving the economic and social conditions of rural residents and, eventually, to the development of the country [4-5]. The bulk of the population in India lives in isolated rural regions where they do not have access to grid energy, and expanding the grid is not practical. Furthermore, those remote locations with grid connections only receive a restricted supply once every six to nine hours [6]. In addition, using a diesel generator to get an electric supply is not a practical alternative due to the expense of fuel, maintenance, and operation, as well as the emission of greenhouse gases, etc [7-8]. Additionally, the Government of India is working to promote renewable energy sources by offering various programs, incentives, etc. The Indian government has also set a target of increasing the amount of electricity produced by RES to 450 GW by 2022 [9]. Due to the stochastic nature of RES, a hybrid system that combines two or more sources is a more advantageous and sustainable alternative [10]. Utilizing a variety of simulation and optimization tools and approaches, a great deal of work has been done in this regard. Flavio Odoi- Yorke et al. examined the possibility of using a hybrid solar PV/biogas/battery energy system to provide power to distant areas in Ghana. The objective is to employ locally accessible renewable energy sources to reduce

greenhouse gas emissions while achieving a Levelized Cost of Electricity (LCOE). The results show that in terms of cost and pollution savings, PV/biogas/battery systems outperform PV/diesel/battery and diesel-only systems [11]. Endeshaw Solomon Bayu et al. conducted a study to incorporate wind turbines, micro-hydro systems, solar photovoltaic (PV) systems, and battery systems to check the feasibility of hybrid systems to electrify the remote place [12]. Paul et al. examined the economic viability and feasibility of utilizing a hybrid-electricity system in rural areas. The findings show that, when compared to PV/Diesel Generator (DG)/B and isolated DG systems, the photovoltaic (PV)/battery (B) system based on renewable energy (RE) has the lowest net profit cost (NPC) and cost of energy (COE). Although the COE and NPC values of the diesel generator (DG) hybrid-electric system (HES) are lower than those of the PV/DG/B system, the DG system still emits the most substantial pollution [13]. Similarly, Nyagong Santino et al. evaluated the viability of a hybrid power system for a remote South Sudanese community without electricity access. Based on the community's energy requirements, average energy consumption profiles were created over the course of a year. The system was configured and optimized using the HOMER pro application, and based on the standalone mode of operation, six potential combinations were modeled and examined technically and economically. Due to the significant solar potential, the Battery/DG/PV system has the minimum Net Present Cost (NPC) and Cost of Energy (COE) and provides a 22.94% investment return [14].

For the Atacama Desert in Chile, Francisco et al. conducted a cost-benefit analysis of the TEG-HPV system under actual environmental and market circumstances. The economic, electrical, and thermal models of the TEG-HPV system are constructed and examined in MATLAB. With regard to system costs, energy losses, ordinal efficiencies of TEG and PV modules, and their contributions to the economical viability of TEG-HPV systems, five distinct cases are taken into consideration. Payback durations for every scenario are calculated at maximum and minimum PV temperatures for the Atacama Desert, taking into account the industrial and residential prices of electricity [15]. Laetitia et al. conducted a feasibility study with the goal of incorporating renewable energy sources into Popova Island's energy system. It takes an analytical strategy that entails using an energy systems model and the Monte Carlo method before assessing the financial results [16]. Ahmad et al. investigated the feasibility of meeting the load demand with the best system that produces the least amount of CO_2 and net present cost (NPC) emissions. The modeling results demonstrate that the NPC of the proposed grid/PV system is more sufficient than other configurations at the present grid tariff, resulting in a renewable proportion of more than over 50%, a payback period of 17 years, and a 54.3% decrease in CO_2. The outcomes further demonstrate that the integration of a 62 kW PV array with the primary grid is the optimal configuration that results in a minimal COE of 0.0688 $/kWh and sale back power of 9.16% of Al Baha University's total electricity consumption [17]. Mohammad Amin et al. developed the best renewable energy system possible to power a small community using only renewable energy sources. Like many remote Iranian communities, this one experiences regular power shortages. A hybrid stand-alone and on-grid renewable energy system using fuel cells, biogas generators, wind turbines, and photovoltaics is suggested. In addition to the fuel cells, batteries, a hydrogen tank, an electrolyzer or reformer, and other backup and storage components are employed. The major objective is to identify the best design that can fulfill the demand for power while being acceptable from an environmental and an economic standpoint. The findings demonstrate that the cheapest option is to use biogas, wind, and solar rather than adding a fuel cell to this design would raise prices by 33–37% while simultaneously increasing the scalability of the system [18]. It is clearly revealed from the literature that a hybrid energy system with a diesel engine has several benefits over one that is solely powered by a diesel engine, including a longer engine lifespan, lower O&M (Operation and Maintenance) costs, lower fuel consumption and less of an adverse impact on the environment. Such systems still depend on fossil fuel, which is unfriendly to the environment and necessitates logistical arrangements for delivering the fuel to the community, which is the problem with them. Local fuel availability is frequently poor because of expensive transit expenses and theft danger. The diesel engine may be totally replaced with a biogas engine to solve the aforementioned issues, and its fuel can be produced locally in a limited digester. Utilizing locally generated biogas from dung can resolve issues associated with diesel fuel. Furthermore, the literature survey also indicates that HRES is a more reliable and cost-effective source of power than conventional grid systems.

Additionally, the aforesaid research mostly focused on household electrification of rural areas while ignoring the need for power for schools and panchayat ghar. To the author's knowledge, there is no thorough study on the techno-economic analysis of such HRES in using HOMER on the selected location. Also, the creation of such models relied on the use of solar and wind energy, and there are only a small number of studies in which biogas and biomass energy sources were regarded as important sources for generating electricity. İn this regard, an HRE-based power producing system is suggested for supplying continuous electricity to two schools and Panchayat ghar, located in Sarai Jairam village, district Agra, Uttar Pradesh, India. As a consequence, the major goal of this study is to offer the most feasible and most optimized alternative for electrifying panchayat ghar and primary schools situated in distant areas without access to grid electricity. To create the HRES, local RES from the research region is employed. Data on electrical load, solar radiation and biogas were utilized to apply techno-economic analysis utilizing HOMER pro simulations based on sensitivity. The organizational structure of the paper is planned as follows. Section 2 discusses the method of modeling of hybrid systems. Section 3 describes the assessment of electrical load. Section 4 briefly explains the system design and its assessment. Section 5, contains results and discussion and Section 6 presents the conclusions.

2. Method of Modeling Hybrid Systems

For the design and construction of an optimum hybrid system for a rural region, a systematized modeling technique is an integral step since it ensures that the rural population has reliable, consistent and dependable access to electricity

[19]. The HOMER pro software has been utilized as a tool to identify the set of optimum systems that meet the load demand under specific system restrictions and input assumptions [20]. HOMER pro is a distributed power optimization tool designed by the National Renewable Energy Laboratory in the United States [21]. Due to the uncertainties, a sensitivity analysis is conducted to investigate the influence of input assumptions on the optimization outcomes. Figure 1 depicts the process and proposed framework for system modeling and analysis. In the current study, a modeling technique that is shown in the following sections includes a description of the selected location, an assessment of renewable energy sources' potential, electrical load estimation, and system design and optimization procedure.

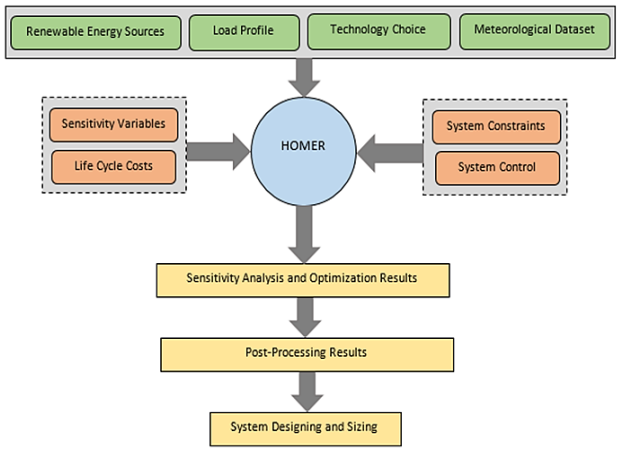


Figure 1. Process and proposed framework for system modeling and analysis

2.1 Location Description

The site for this study was Sarai Jairam village in Uttar Pradesh, India. The abundance of renewable energy and the region's strategic significance was taken into consideration while choosing the location. The location is identified by the coordinates at the longitude of 78o 10.7'E and latitude of 27o 20.2'N and time zone (GMT+05:30) from NASA meteorological data. Figure 2 shows the geographic view of the research location.

Figure 2. The geographic view of the research location

2.2 Assessment of wind and solar resources potential at the selected site

From NASA meteorological data, the solar and wind energy data for Sarai Jairam village have been extracted, and the data are shown in Table 1 with their latitude and longitude (27o 20.2'N, 78o 10.7'E), and time stamp (GMT+05:30). Figure 3 and Figure 4 provide graphical representations of the specifics of solar radiation throughout the course of the year and wind data over the same period. Table 1 provides data on the selected site's annual average daily radiation (kWh/m²/day), clearness index, and wind speed (m/s).

Table 1. Average monthly daily radiation, clearness index, and wind speed

Month	Solar Energy (SE) (kWh/m²/day)	Clearness Index	Wind Speed (WS) (m/s)
Jan	3.670	0.578	4.230
Feb	4.690	0.623	4.690
March	5.590	0.620	5.020
April	6.080	0.589	5.100
May	6.360	0.574	5.250
June	6.010	0.531	5.180
July	4.960	0.444	4.740
Aug	4.540	0.430	4.410
Sept	4.750	0.503	4.150
Oct	4.730	0.595	3.450
Nov	4.030	0.611	3.360
Dec	3.490	0.585	3.700

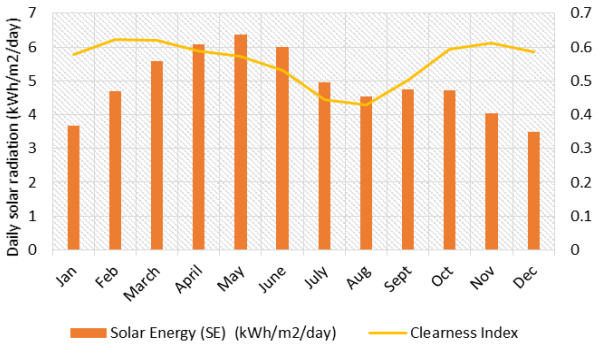


Figure 3. Monthly average solar irradiance throughout the year

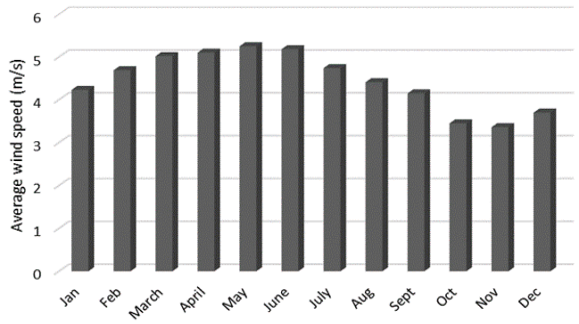


Figure 4. Monthly average wind speed throughout the year

2.3 Assessment of biomass energy resource potential at the selected location

In Sarai Jairam village animal dung may be simply used to produce biogas through the digestion and combustion processes, respectively. Anaerobic digesters are used to treat the manure produced by cattle and produce power. The total dung produced annually from the community is calculated by equation (1);

$$M_n = \sum_{n=1}^{j} N_j m_j \quad (1)$$

Where M_n is the annual total amount of manure produced, N_j is the number of the selected group of animals, n is the overall number of cattle and m_j shows manure produced per cattle [21].

Table 2 lists the number of animals and the availability of animal manure in the village of Sarai Jairam. The total amount of manure produced by all the animals is estimated to be 6.425 tons/day, which may be used to produce 166667.76 kWh of power annually. The total potential of biomass is shown in Figure 5.

Table 2. Total amount of manure produced by all animals

Description	Number of animals	Dung (kg/day/cattle)	Dung availability (kg/day)
Buffaloes	310	15	4650
Cow	175	10	1750
Goat	25	1	25

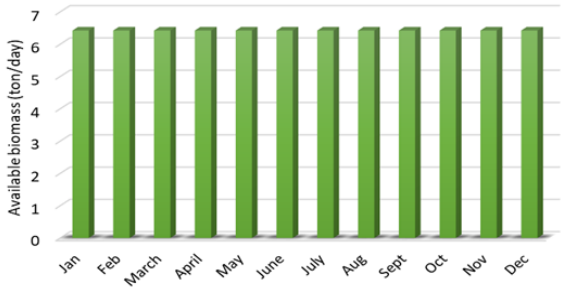


Figure 5. Technical potential of biomass (animal manure)

3. Assessment of Electrical Load

In this work, a hybrid system that employs energy from solar and biogas has been designed to meet the electrical demands of one Primary school, Junior school and Panchayat ghar. The panchayat ghar has two rooms with two fans and four tube lights. The primary school contains six rooms with twelve fans, twelve tube lights, two computers, and one submersible pump, while, five rooms, ten fans, ten tube lights, two computers, and one submersible pump make up the junior high school. Table 3 provides the quantity, power ratings, and power consumption for each load. Figure 6 depicts the electricity load of the proposed region taking into account future expansion. The data has been gathered from the school and panchayat ghar employees.

4. System Design and Assessment

The system configuration shown in Figure 7 includes a PV panel, biogas engine generator, and battery storage with a bidirectional converter. An electric load and biogas generator are connected to the AC bus. The battery storage and PV module are connected to the DC bus. Additionally, the converter is coupled with the AC and DC buses. Table 4 shows the technical specifications and capital costs of components.

Table 3. Electricity load calculation for the selected communities

S.no.		1	2	3	4
Load		Fan	Tube light	Computer	Water pump
Power rating (watts)		60	30	100	500
Panchayat ghar	No. in use	2	4	0	0
	Utilization hours	10	10	0	0
Primary school	No. in use	12	12	2	1
	Utilization hours	8	8	8	1
Junior school	No. in use	10	10	2	1
	Utilization hours	8	8	8	1
Total energy Consumed (Watt)		11760	6480	3200	1000
Total load		22.440kWh			

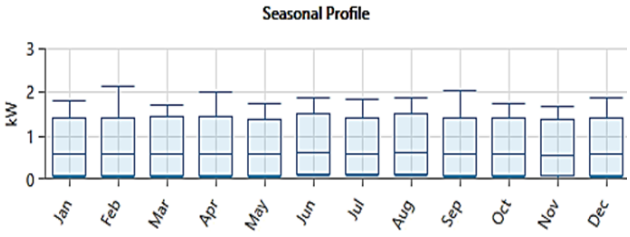


Figure 6. Monthly load profile during the whole year

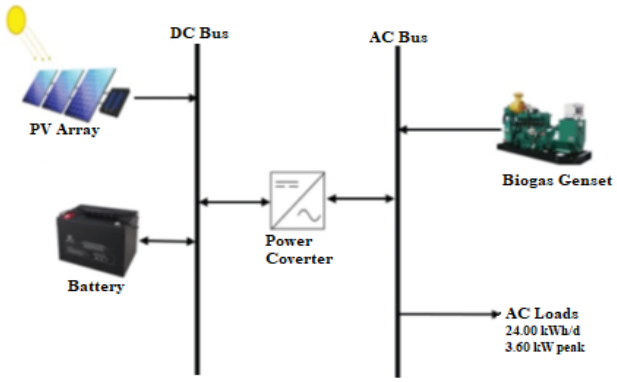


Figure 7. Configuration of off-grid hybrid PV-biogas system

5. Results and Discussion

5.1 System sensitivity analysis outcomes

For technical and economical assessment purposes, the system must be designed with certain constraints or control variables that have an impact on the operating costs and output of the HRES system. Because the developed system is sustainable and the possibilities like the changes in solar

radiations and biomass supply were taken into account for the sensitivity analysis. Since animal dung is a free source of biomass, hence, the biomass price was not considered in the system design. It enables planners and designers to select a very efficient and cost-effective method for the specified design parameters. To find out how variations in solar radiation and biomass supply might affect the system economy, both variables were varied. As the supplied biomass was 6.43 tonnes per day, it varied between 6.43-6.50 tonnes per day for the sensitivity analysis, and the solar radiation varied between 4.91-4.99 kWh/m^2/day.

Table 4. Technical specifications and cost of components

Hybrid system components	Specifications	Value
Solar PV	Capital cost	650$/kW
	O&M cost	1$/kW
	Replacement cost	0$/kW
	Sizes	0,1,1.5,2,2.5,3,3.5,4,4.5,5,6,7, 7.5,8,9,10
	Tracking system	No
	Ground reflectance	20%
	Derating factor	80%
	Life span	25 years
	Azimuth,Slope	0
Converter	Capital cost	190$/kW
	Replacement cost	190$/kW
	Sizes	0,1,1.5,2,2.5,3,3.5,4,4.5,5
	Rectifier efficiency	85%
	Capaicity relative to inverter	100%
	Inverter efficiency	95%
	Life span	20 years
	Annual rate of interest	0%
Battery	Capital cost	150$/kW
	Replacement cost	30$/kW
	Nominal capacity	1kWh
	Float life	5 years
	Round trip efficiency	80%
	Mninmum state of charge	40%
	Lifetime throughput	917kWh
Biogas generator	Capital cost	250$/kW
	O&M cost	0.08$/kW
	Sizes	0, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5
	Replacement cost	150$/kW
	Biomass cost	$3.75/tone
	Lowest load ratio	30%
	Life span	20000 hours

The sensitivity analysis outcomes of the HRE system are displayed in Table 5, which highlights how changes in the supply of biomass and solar radiation affect the NPC, COE, and operating costs. The findings reveal that as biomass is increased and solar radiation changes are taken into account, net present and operating costs also rise.

5.2 System optimization outcomes

The developed hybrid Photovoltaic/biogas system with sensitivity inputs was modeled in HOMER Pro software by ranging the potential of the biogas generator and solar radiance in order to determine the most efficient, optimized, and cost-effective system for the Primary school, Junior school, and Panchayat ghar in rural areas. To maximize the system's ability to meet the electricity demand, the capacity range of the biomass generator varied between 1kW to 5 kW, while the capacity range of the PV system varied between 4kW to 6 kW. In this study, a 1.50 kW biogas generator, a 5 kW PV array, a 3.25 kW converter, and 30 storage batteries were the optimal and most economically viable configurations evaluated for the hybrid system. This configuration is depicted in Figure 9.

The developed hybrid Photovoltaic/biogas system with sensitivity inputs was modeled in HOMER Pro software by ranging the potential of the biogas generator and solar radiance in order to determine the most efficient, optimized, and cost-effective system for the Primary school, Junior school, and Panchayat ghar in rural areas. To maximize the system's ability to meet the electricity demand, the capacity range of the biomass generator varied between 1kW to 5 kW, while the capacity range of the PV system varied between 4kW to 6 kW. In this study, an 1.50 kW biogas generator, 5 kW PV array, a 3.25 kW converter and 30 storage batteries were the optimal and most economically viable configurations evaluated for the hybrid system. This configuration is depicted in Fig. 9. As NPC, COE, and beginning capital investment are used to sort HOMER's optimum results. The total capital cost, NPC and COE for the best optimized PV-Biogas configuration are, respectively $8,743, $57,283, and $0.614. Other designs with varying equipment sizing that can meet the same energy load have low initial capital costs but high NPC, COE, and operating costs over the lifetime of the project. The various system designs with their significant economic parameters are shown in Table 6.

Table 7 also depicts the share of power production of each component throughout a year, total electricity production, renewable percentage, and excess electricity with unmet load for all different system designs. Figure 8 depicts the average monthly electricity production from a hybrid PV-Biogas system, where orange bars show the power supplied by the photovoltaic panel and green bars show the power generated by a biogas generator. The total initial cost is calculated at $8,743 as shown in Table 6. This is the first investment needed to start the project. The initial cost of Lead acid battery and photovoltaic panels are more as compared to other components as shown in Figure 9. This higher cost is a result of the substantial battery storage capacity, which is intended to offer reliable electricity dispatch when the electricity produced by the power systems is not enough to fulfill the load.

Table 5 Sensitivity outcomes for the hybrid PV/biogas system

Sensitivity		Architecture						Cost	
Solar Scaled Average (kWh/mÂ²/day)	Biomass Scaled Average (t/d)	PV (kW)	Bio (kW)	1kWh LA	Converter (kW)	NPC ($)	COE ($)	Operating cost ($/yr)	Initial capital ($)
4.91	6.43	5	1.5	30	3.25	57283	0.614	4547	8743
4.92	6.44	5	1.5	32	3.25	56901	0.610	4483	9043
4.95	6.46	5	1.5	32	3.25	55964	0.600	4396	9043
4.97	6.48	5	1.5	32	3.26	55345	0.594	4337	9044
4.99	6.5	5	1.5	32	3.25	54762	0.587	4283	9043

Table 6 Various different optimized system configurations with economic parameters.

Various Configuration	NPC ($)	COE ($)	Operating Cost ($)	Initial Cost ($)
PV/Biogas/Battery/Converter	57283	0.6145	4547	8743
PV/Biogas/Converter	169188	1.8107	15458	4178
Biogas/Battery/Converter	213798	2.3475	19884	1539
Biogas	260688	2.8009	24362	625

Table 7 Electricity generation of different feasible system designs

Parameters	PV/Biogas/Battery/ Converter	PV/Biogas/ Converter	Biogas/Battery/ Converter	Biogas
PV array (kWh/year)	7,868 (81.1%)	7,868 (56.2%)	-	-
Biogas generator (kWh/year)	1,830 (18.9%)	6,128 (43.8%)	9,013 (100%)	10,185 (100%)
Total electricity generation (kWh/year)	9,698 (100%)	13,996 (100%)	9,013 (100%)	10,185 (100%)
Renewable fraction (%)	100	100	100	0
Capacity shortage (kWh/year)	0.212 (0%)	143 (1.63%)	2.11 (0.1%)	111 (1.27%)
Excess electricity (kWh/year)	65.0 (0.67%)	5,055 (36.1%)	0	1,466 (14.4%)
Unmet electric load (kWh/year)	0.0000267 (0%)	6.77 (0.0773%)	1.10 (0%)	41.0 (0.468%)

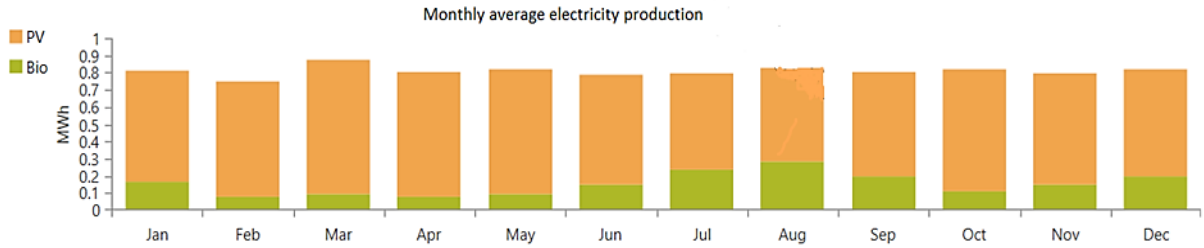


Figure 8. Monthly average electricity generation from the hybrid system

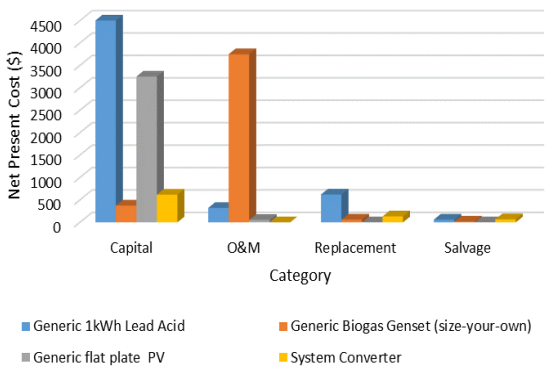


Figure 9. Cost summary of the hybrid PV/Biogas system

In addition, the photovoltaic panels, which are less expensive once installed than a biogas generator, as biomass is available for free, which is why the fuel category has not been presented. Figure 10 displays the power generated by each component, and the total electrical load served is represented in the above plot, while the lower plot shows the unmet electrical load, total renewable power output, ac primary load served, input power, and state of charge of battery over the course of July 9 to July 18 to help with a better understanding of how the system operates.

Table 8 and Figure 8 make it clear that the photovoltaic system consistently outperforms biogas in terms of electricity production. Being off the grid, the system uses solar and biogas as resources to meet the load demands because our maximum demand hours are during the daytime for community load (schools and panchayat ghar) purposes. The capacity factor of the PV modules is about 81.1%, and it operates throughout the year depending on load requirements, producing about 7,868 kWh/year, compared to 1,830 kWh/year from the biomass generator having a capacity factor of 18.9%. Due to the lack of biomass availability throughout the day, biogas power generation is reduced. Furthermore, the system is producing more power than it needs to satisfy its annual power usage of 8,733 kWh, which can be saved or used for other productive purposes.

6. Conclusion

This study offers a technical and economical assessment of different stand-alone solutions for Primary school, Junior school, and Panchayat Ghar buildings of Sarai Jairam village in Uttar Pradesh, India. HOMER pro analyzed several hybrid PV-biogas system configurations by modeling a dynamic hybrid model. An ideal solution was suggested based on the cost analysis after these hybrid designs underwent sensitivity analysis using variables such as solar radiation, biomass resource, and system sizing. In this study, the combination of a 1.50 kW biogas generator, 5 kW PV array, a 3.25 kW converter, and 30 storage batteries was found to be the most cost-effective option with a total capital cost of $8,743, Net Present Cost (NPC) of $57,283 and Cost of Energy (COE) $0.61, respectively. This hybrid renewable energy system produces roughly 9,698 kWh per year, with an additional 965 kWh per year being generated to make the study area grid-independent. Additionally, the system has an estimated payback period of 0.41 years and a favorable net current cost for a projection timeframe of 25 years. By providing rural areas with these hybrid renewable energy systems, the Indian government may significantly contribute to resolving the country's current energy crisis. Additionally, the existing legislation that supports the use of such systems only offers tax breaks or reductions, which is insufficient to allow low-income populations to make use of these systems. The government may alter its supporting policies, offer rewards for system employment, and launch a national electrification campaign. Similar studies could be undertaken for other rural places in order to electrify them, which would also help the Indian government achieve its goal of "Power to all."

Table 8. Annual generation and consumption of electricity by a hybrid system

Generation	kWh/year	%	Consumption	kWh/year	%
PV Modules	7,868	81.1	AC primary load	8,733	100
Biogas Generator	1,830	18.9	DC primary load	0	0
Total	9,698	100	Total	8,733	100

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

Datasets analyzed during the current study are available and can be given following a reasonable request from the corresponding author.

Conflict of interest

The authors declare no potential conflict of interest.

Author contributions

Mohammad Rizwan: Conceptualization and supervised the research; Upma Singh: Methodology; resources; validation; formal analysis; writing the original draft. Both authors contributed to the analysis of the findings, critically reviewed and edited the manuscript, and approved the final version for submission.

Funding

No specific grant has been given to this research by any funding agency/organizations.

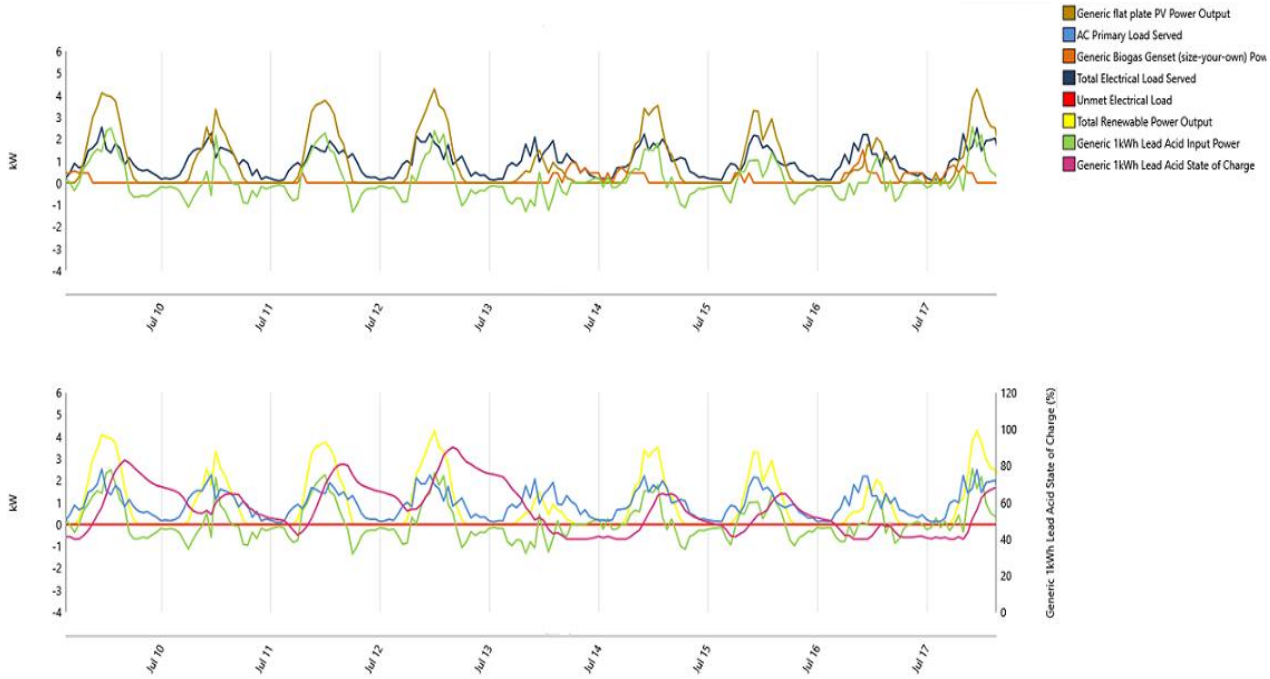


Figure 10. Total electrical load served, unmet electrical load, total renewable power output, state of charge of the battery, ac primary load served and generated power by each component during July 9 to July 18.

References

[1] K. Kusakana, J. L. Munda, and A. A. Jimoh, Feasibility study of a hybrid PV-micro hydro system for rural electrification, In AFRICON 2009, IEEE, Sep. 2009, pp. 1-5.

[2] F. Rinaldi, F. Moghaddampoor, B. Najafi, and R. Marchesi, Economic feasibility analysis and optimization of hybrid renewable energy systems for rural electrification in Peru, Clean technologies and environmental policy, 23(3), 2021, pp.731-748

[3] J. Ahmed, K. Harijan, P. H. Shaikh, and A. A. Lashari, A.A., Techno-economic feasibility analysis of an off-grid hybrid renewable energy system for rural electrification, Journal of Electrical and Electronic Engineering, 9(1), 2021, pp.7.

[4] A. K. Awopone, A.K., Feasibility analysis of off-grid hybrid energy system for rural electrification in Northern Ghana, Cogent Engineering, 8(1), 2021, pp.1981523.

[5] F. Rashid, M. E. Hoque, M. Aziz, T. N. Sakib, M. T. Islam, and R. M. Robin, Investigation of optimal hybrid energy systems using available energy sources in a rural area of Bangladesh, Energies, 14(18), 2021, p.5794.

[6] A. H. Eisapour, K. Jafarpur, and E. Farjah, Feasibility study of a smart hybrid renewable energy system to supply the electricity and heat demand of Eram Campus, Shiraz University; simulation, optimization, and sensitivity analysis, Energy Conversion and Management, 248, 2021, pp.114779.

[7] S. Iqbal, M. U. Jan, A. U. Rehman, A. Shafiq, H. U. Rehman, and M. Aurangzeb, Feasibility Study and Deployment of Solar Photovoltaic System to Enhance Energy Economics of King Abdullah Campus University of Azad Jammu and Kashmir Muzaffarabad, AJK Pakistan, IEEE Access, 10, 2022, pp.5440-5455.

[8] Q. Ma, X. Huang, F. Wang, C. Xu, R. Babaei, and H. Ahmadian, Optimal sizing and feasibility analysis of grid-isolated renewable hybrid microgrids: Effects of energy management controllers, Energy, 240, 2022, p.122503.

[9] F. Eze, J. Ogola, R. Kivindu, M. Egbo, and C. Obi, Technical and economic feasibility assessment of hybrid renewable energy system at Kenyan institutional building: A case study, Sustainable Energy Technologies and Assessments, 51, 2022, pp.101939.

[10] M. M. Seedahmed, M. A. Ramli, H. R. Bouchekara, M. S. Shahriar, A. H. Milyani, and M. Rawa, A techno-economic analysis of a hybrid energy system for the electrification of a remote cluster in western Saudi Arabia, Alexandria Engineering Journal, 61(7), 2022, pp.5183-5202.

[11] F. Odoi-Yorke, S. Abaase, M. Zebilila, and L. Atepor, L., Feasibility analysis of solar PV/biogas hybrid energy system for rural electrification in Ghana, Cogent Engineering, 9(1), 2022, pp. 2034376.

[12] E. S. Bayu, B. Khan, I. G. Hagos, O. P. Mahela, and J. M. Guerrero, Feasibility Analysis and Development of Stand-Alone Hybrid Power Generation System for Remote Areas: A Case Study of Ethiopian Rural Area, Wind, 2(1), 2022, pp.68-86.

[13] P. C. Okonkwo, E. M. Barhoumi, W. Emori, M. I. Shammas, P. C. Uzoma, A. M. Amer Mohamed, and A. M. Abdullah, Economic evaluation of hybrid electrical systems for rural electrification: A case study of a rural community in Nigeria, International Journal of Green Energy, 19(10), 2022, pp.1059-1071.

[14] N. S. D. Ladu, R. Samikannu, K. G. Gebreslassie, M. Sankoh, L.E. R Hakim, A. Badawi, and T. P..B Latio, T.P.B., Feasibility study of a standalone hybrid energy system to supply electricity to a rural community in South Sudan, Scientific African, 16, 2022, p.e01157.

[15] F. J. Montero, R. Kumar, R. Lamba, R. A. Escobar, M. Vashishtha, S. Upadhyaya, and A. M. Guzmán, Hybrid photovoltaic-thermoelectric system: Economic feasibility analysis in the Atacama Desert, Chile, Energy, 239, 2022, p.122058.

[16] L. Uwineza, H. G. Kim, and C. K. Kim, Feasibility study of integrating the renewable energy system in Popova Island using the Monte Carlo model and HOMER, Energy Strategy Reviews, 33, 2021, p.100607.

[17] A. Tazay, Techno-economic feasibility analysis of a hybrid renewable energy supply options for university buildings in Saudi Arabia, Open Engineering, 11(1), 2021, pp.39-55.

[18] M. A. V. Rad, R. Ghasempour, P. Rahdan, S. Mousavi, and M. Arastounia, Techno-economic analysis of a hybrid power system based on the cost-effective hydrogen production method for rural electrification, a case study in Iran. Energy, 190, 2020, p.116421.

[19] O. Babatunde, I. Denwigwe, O. Oyebode, D. Ighravwe, A. Ohiaeri, and D. Babatunde, Assessing the use of hybrid renewable energy system with battery storage for power generation in a University in Nigeria, Environmental Science and Pollution Research, 29(3), 2022, pp.4291-4310.

[20] L. Ji, Z. Liu, Y. Wu, and G. Huang, Techno-economic feasibility analysis of optimally sized a biomass/PV/DG hybrid system under different operation modes in the remote area, Sustainable Energy Technologies and Assessments, 52, 2022, p.102117.

[21] D. T. Hermann, N. Donatien, T. K. F. Armel, and T. René, Techno-economic and environmental feasibility study with demand-side management of photovoltaic/wind/hydroelectricity/battery/diesel: A case study in Sub-Saharan Africa, Energy Conversion and Management, 258, 2022, p.115494.

[22] A. Shiroudi, S. R. H. Taklimi, S. A. Mousavifar, and P. Taghipour, Stand-alone PV-hydrogen energy system in Taleghan-Iran using HOMER software: optimization and techno-economic analysis, Environment, development and sustainability, 15(5), 2013, pp.1389-1402.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

A Review of Deep Learning Techniques for Protein Function Prediction

Divyanshu Aggarwal
Dept of Biotechnology,
Delhi Technological University,
New Delhi, India
divyanshuggrwl@gmail.com

Yasha Hasija
Dept of Biotechnology,
Delhi Technological University,
New Delhi, India
yashahasija06@gmail.com (Corresponding author)

***Abstract*—Deep Learning and big data have shown tremendous success in bioinformatics and computational biology in recent years; artificial intelligence methods have also significantly contributed in the task of protein function classification. This review paper analyzes the recent developments in approaches for the task of predicting protein function using deep learning. We explain the importance of determining the protein function and why automating the following task is crucial. Then, after reviewing the widely used deep learning techniques for this task, we continue our review and highlight the emergence of the modern State of The Art (SOTA) deep learning models which have achieved groundbreaking results in the field of computer vision, natural language processing and multi-modal learning in the last few years. We hope that this review will provide a broad view of the current role and advances of deep learning in biological sciences, especially in predicting protein function tasks and encourage new researchers to contribute to this area.**

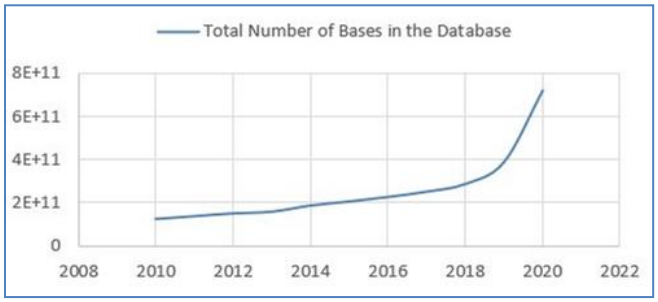
***Index Terms*—Bioinformatics, Big Data, Protein Function Classification, Deep Learning, Computational Biology**

I. INTRODUCTION

Proteins play a large role in the cellular machinery of the living organism. It is very important to know the function protein while conducting any proteomic research on that particular protein. However, more than 40% of the NCBI database's protein sequences have no assigned function as of 2013 [1]. This can be induced by the fact that it is expensive, high in processing time, and challenging to determine a protein's function using functional assays. [1] This arises the need for a computational method to determine a protein's function from the raw data obtained from the high throughput techniques, including but not limited to protein sequence, protein structure, gene expression profile and protein-protein interaction data.

Traditional approaches to classification for protein function attempt to identify the evolutionary relationship between a new protein and a query protein [1]. A high sequence similarity score can suggest a high probability of 2 proteins originating from a single evolutionary source [1]. However, it is well established that proteins with high sequence alignment or sequence similarity score may or may not show similar functions.[2] Such erroneous annotations may lead to propagation and amplification in large databases quickly.

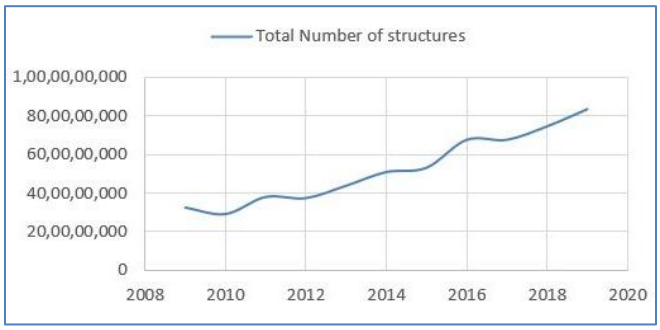
Such errors are not only due to basic transfer strategies based on homology, but also due to the manual data analysis



Graph 1. Growth of Gene Bank [3]

Process. Such problems have also been tried to tackle using different data, including but not limited to 3D structure similarity, gene expression profile similarity and genomic expression profile similarity.

In the past few years, the growth of biological databases and significant advancements in computing resources creates an opportunity for the scientific community to build novel deep learning models and architectures to tackle such problems. As the latest trend in recent developments in natural language processing research and sequential data preprocessing using deep neural networks (DNN), convolutional neural networks (CNN), recurrent neural networks (RNN), long- term and short-term memory (LSTM), and attention-based transformer models, it is important to evaluate the suitability, feasibility, sustainability and explain ability of such models for the given task. In the sense of protein function classification, this analysis aims to provide an overview of these techniques.



Graph 2. Growth of Protein Data Bank [4]

This paper outlines the latest trends in the development of automated annotation of protein function from raw data using deep learning. The flow of paper will be as follows; first, we discuss about the approached for prediction protein function from raw without using deep learning. (section 2). Next, we describe the recent deep learning and machine learning advances and how they are used in the protein

function classification perspective (Section 3). We discuss related work and concluding remarks in the section after that (Section 4 and 5, respectively).

II. CLASSICAL COMPUTATIONAL TECHNIQUES FOR PROTEIN FUNCTION PREDICTION

Understanding how the cell functions enables one to learn how the protein works. Usually, a protein function is specified as the molecular function of the protein; material transport, gene regulation, catalysis of biochemical reactions (enzymes) and others are some protein functions. The protein functions in many ways; it can interact with other proteins in order to perform its roles biologically. The mutation of the amino acid sequence may result in certain diseases from a phenotypical perspective. Factors like, but not limited to, such as, the environment of the cell, for example, temperature, subcellular protein position, may also influence the role of a protein. For standardizing functional annotation and enabling computer processing to explain various aspects of protein functions, several classification schemes have suggested specific vocabulary. Gene Ontology is the most generally recognized (GO). Three aspects of protein functions are defined by GO: molecular structure, biological mechanism, and cellular positioning. An ontology is constructed as a directed acyclic graph, here vertices reflect function of protein and their relationships are represented by the edges. For their practical predictions from this machine-readable vocabulary rather than natural language style, methods of prediction using computation should have a uniform, output, which can show errors within the annotation level. [5], [6], [7]

The most common means of identifying a protein's function by using the already annotated sequences is through BLAST. The query sequence is matched with the sequences from the database, and the similarity score is used to estimate the function of the protein with different sequences. BLAST uses heuristic algorithms to match strings. The genes with high similarity scores are said to be homologs. It is assumed that protein sequences with similar homology have a similar function, however, sequence matching is not the most accurate way of determining the protein's function; 2 highly similar protein sequences are not supposed to have same Gene Ontology in all the scenarios [8].

Gene fusion and phylogenetic profile are another 2 features which can give good accuracy in the task of prediction protein function. Gene neighborhood strategies use the idea that proteins whose translated genes are located near each other on a chromosome are deemed to be linked in functional way in different genomes. In one genome that is fused into another genome to form a single gene, the idea that group of genes are required focuses on gene fusion-based approaches. Another ground-breaking method for predicting functional relationships is to compare the phylogenetic profiles of proteins. A collection of bits that indicate a homologous presence or absence in each genome is known as protein phylogenetic profile. Similar profiling may also imply that gene which has produced the proteins have evolved together and may indicate function conservation. [9], [10]

It is also possible to use motifs and domain details. A domain is a part of the amino acid sequence of proteins that folds independently of the rest of the structure into a stable structure. A motif is a very short stretch of sequences of amino acids that theoretically encodes the structure of the protein. These are used for the metadata of protein, such as length of sequence, composition of amino acids, and physicochemical properties of a sequence. [2]

However, the structure of the protein is the most informative aspect of a protein to describe the function of a protein. Due to the Structural Genomics Initiative, we today have a database called protein data bank, which contains the structure of millions of protein sequences. The query protein is structurally matched. The protein sequences are matched in 2 ways, globally and locally. In global alignment, the entire structure is matched, which in local structural alignment, local regions like motifs and domains are matched since they are a better criterion for identifying the protein's function. The motifs and domains can capture the intramolecular interactions of a protein, while global alignment methods are able to identify the intermolecular interactions. Intermolecular interactions are generally performed on the surface, which are responsible for a protein's biochemical properties. [11]

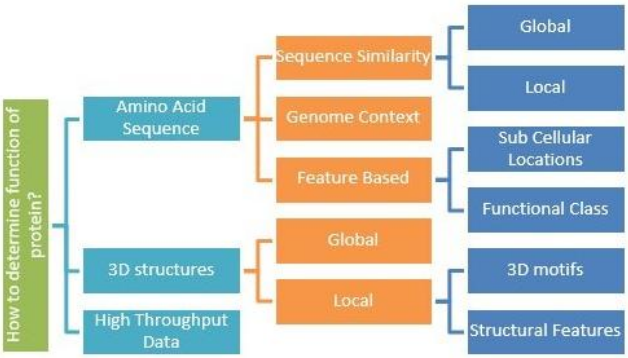


Fig. 1. .

With the emergence of high-throughput protein sequencing technologies, vast quantities of data have been made available in biological gene expression and protein-protein interaction databases. The gene expression is measured by the amount of the protein produced by a gene in specific conditions that performs a particular gene. It is probable that proteins whose genes are co-expressed have the same role. Proteins subsequently perform a particular role by communicating with other proteins; hence similar protein-protein interactions can lead to similar protein functions.

III. USING DEEP LEARNING FOR PROTEIN FUNCTION PREDICTION

Deep learning has achieved significant momentum in the past few years with the advancements in the computing power of the central processing units and graphical processing units and the rapid expansion of biological databases. Deep learning algorithms are capable of extracting hidden features and patterns from the data are also able to form non-linear classification boundaries. [12], [13] Deep learning also scales well with the growth in data volumes, and its application to computer vision and natural language processing can be adapted to the structure of 3D proteins and protein sequences, respectively. With the recent advances in neural graph networks, they can also be applied in protein-protein interaction and phylogenetic databases.

A. Data Sources

There are multiple data sources from where the data can be obtained to train the model from such tasks. Uniprot KB and PDB remains to be the open databases that can be freely accessed. [14], [15] The data from such databases can be downloaded in bulk and can be cleaned and preprocessed to train such models. However, the CAFA dataset released by The Function Special Interest Group is a dataset mainly meant for the challenge of building computational models for protein function classification. [16] However some models have also used STRING db in the past.

B. Deep Learning Paradigm

This paper presents the analysis of modern deep learning techniques in the context of protein function classification. The classification of the function of a protein can be classified as a supervised machine learning task. Since neural networks in general settings cannot use unsupervised learning tasks, we cannot use clustering techniques or other unsupervised techniques. Deep learning research has gathered momentum in the past few years. This can be attributed to the advancements in modern machines' computational capacity and availability of large publicly available databases. The sector has had a major influence, including, to name a few, banking, healthcare, social media. Recent developments in the processing of natural language and computer vision by deep learning can have an immense impact on genomics and proteomics research. [13]

Recent developments in algorithm development, such as DNNs, RNNs, CNNs, LSTMs, and attention models for image and text data, the State of the art and near-human precision have been achieved already, we can only wonder how they can do in such biological tasks where the enormous amounts of similar data are accessible. However, the challenge of data representation arises while solving such tasks using deep learn- ing. While there are pre-trained word embeddings available for natural language, it is difficult to say so for biological sequence databases. However, researchers have provided some protein embedding in recent years, and the quality of such models is not on par with natural language models, trained with similar volumes of data. The reason might be the relatively small vocabulary size of protein sequences with 20 amino acids known, whereas more than 100,000 words are taken in the vocabulary while training a natural language model. [17] The subsequent sections will explore the recent advancements in deep learning and their implication in the protein classification task.

C. Protein Function prediction using sequence only

With the advances in text classification tasks using deep learning, one can only get curious about how these techniques can be applied to biological sequence classification. Related biological problems have been further addressed by developments in word embedding methods, 1D CNN, FastText word embedding, DNNs, and RNNs by adapting methods from natural language processing tasks. We will address some of the methods mentioned in this subsection to predict a protein's function solely from raw sequence data.

1) Combination of 1D Convolutional Neural Network and Deep Neural Networks: Convolutional Neural Networks are the deep learning algorithms capable of extracting unique geographic features from the data. They are also used to reduce the data's size by amplifying and representing localized features in the final representation. The convolutional Kernel translates over an image or text and captures the localized information during translation. A 1D Convolutional Neural Network convolves through a protein sequence representation in 1 dimension. [18], [19], [20]

Fig. 2. Illustration of CNN and DNN for Function Prediction

The convolutional Kernel is translated over the word embedding of the sequence, concatenated word byword for a sequence. There are several alternatives available for natural languages like FastText or word2vec, while for protein sequences, ProtVec or prot2vec is a popular choice. [21], [22], [23]

2) Using Recurrent Neural Networks: Recurrent neural networks are neural networks that can identify the temporal characteristics of the sequence, i.e., a network graph is a graph where a directed graph forms connection between nodes along a time sequence. There are dynamic networks. They are useful in learning features from variable length features. This makes them useful for processing variable-length sequences in the input. Such models can also be used for sequence-to-sequence tasks like machine translation.

The modified version of RNN is known as Long Term Short Memory (LSTM). LSTMs are capable of storing particular temporal features in the hidden State of the cell. [24] This makes it more accurate for longer sequences and sequences with high length variability. The LSTM also tackles the problem of vanishing and exploding gradient problem better than vanilla RNNs. [12], [24], [25]

Fig. 3. Schematic Diagram of protein function classification using Recurrent Neural Networks

Bi-LSTMs, as the name suggests, are bidirectional LSTM neural network architectures. These can capture the bidirectional features of the sequence. Hence, they can have advantages in tasks with bidirectional nature like sequence classification. These models can be used by feeding embedding layers directly to the LSTM models, where they extract out the sequential features, and they can be concatenated or added further. The resulting vector is then fed to a feed-forward layer or multiple feed-forward layers, which can then be passed through a SoftMax layer for the classification task.

3) Transformers for Protein Sequence Classification: With the introduction of attention-based models in 2017 by google research, natural language processing has changed drastically. The attention-based transformers can process long sequences much better than LSTMs and can utilize the parallel GPU architecture for faster and more cost-effective training. The Transformer models can extract out location-specific features and give weights to tokens based on

context relevance. [26]

The transformer models scale up well with the increasing data and the increasing number of parameters. With GPT and BERT's introduction, large language models are becoming more and more mainstream with time. [27], [28] The models can grasp the link between sequence tokens very well and are the present State of the art for most NLP assignments. Moreover, these models can also form word embeddings through self-supervised tasks like masked language modeling and next-word prediction. [17] These representations can then be used to for finetuning for the number of other mainstream tasks. Such self-supervised learning techniques can also be used for protein sequences as demonstrated by protBERT and proBERTa. [29], [30]

D. Multi-modal Deep Learning for Protein Function Classifi-cation

In past few years, multi-modal methods of deep learning have gained huge traction. A basic multi-modal deep learning technique can be performed by combining the image features and text features. The researchers have previously used sequence data, genomic expression, 3D structures, and data on protein- protein interaction to predict the activity of the protein. With the advances in Natural Language Processing and Computer Vision, such techniques seem promising for the Particular Task. The Machines can learn in cognizance with different data representations to make predictions. The sequence data can provide information about the amino acid sequences and their positioning, while 3D structure can show structural domains and motifs, which are a significant indicator of the protein's function. The genomic expression data can capture the patterns in the genomic expression and the corresponding GO annotation, while protein-protein interaction databases are useful for finding the patterns in protein-to-protein interactions. [31], [32], [33], [34]

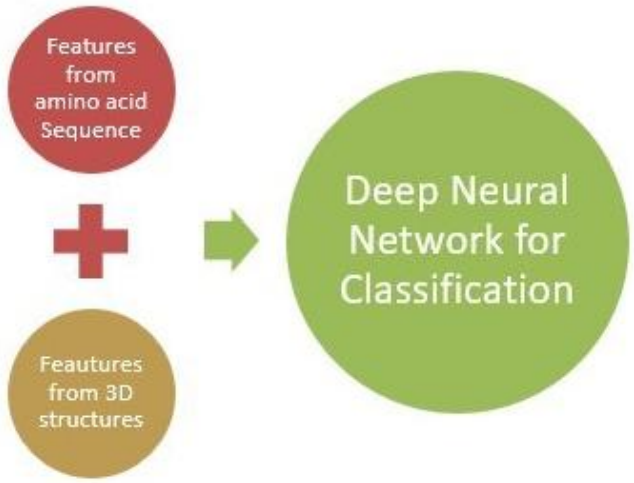


Fig. 4. Multi-modal Deep Learning Illustration.

The Multi-modal deep learning techniques, given the high compute power available these days, seems promising for such tasks. With the increase in high throughput technologies, the database for such features is only expected to grow further, providing us with more data for training these models. There is also a significant momentum gained in multi-modal deep learning research, which will further drive such models' accuracy.

E. Using Autoencoders

A few researchers in the past have also tried using denoising autoencoders for extracting features from the sequence and structure alike. [31], [35], [36], [37] These features are then fed into any other classifier of choice. This denoising autoencoder is trained to produce denoised outputs that are generated randomly with some average and standard deviation according to a normal distribution. However, the technique is not just limited to multi-modal techniques. Masked language modeling is shown in protBERT, and ProtBERTa is also denoising autoencoder task applied on raw protein sequence only. [29], [30]

Fig. 5. Autoencoder model architecture.

IV. RELATED WORK

We reviewed 9 recent models built by researchers around the world for this task. We reported AUC score of each model with respect to the CAFA challenge and human and yeast STRING networks metric. [16], [38] We have reported the methodologies and the resulting Fmax score in table 1.

According to **figure 6,** the 1D CNN combined with DNN seems to be the clear preference for function prediction task. The models are simple and perform well in the sequence classification tasks. At the same time, the transformer models seem to be least explored for this task.

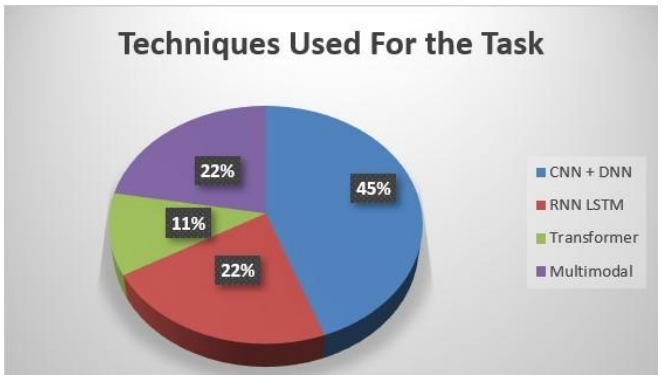


Fig. 6. Distribution of all the techniques used by the model reviewed models

From the above data we can infer that simple techniques like 1D CNN + DNN perform well on the raw sequence data. They can be improved by combining with PPI data; gene expression data are alike. Around the same time, SOTA NLP models such as transformers are not on par with other conventional techniques. However, there is not much research that is done on transformer models for such downstream tasks. It shall be interesting to see the implications of such models further in the future for such downstream tasks, especially the use of protBERT and PRoBERTa pre-trained models finetuned for protein function classification. [29], [30]

TABLE I. RECENT MODELS AND THEIR PERFORMANCE IN THE PROTEIN FUNCTION CLASSIFICATION

Model	Authors	Data Published	Methods Used	Fmax	Validation Dataset
DeepGO	Kulmanov et al.	February 2018	CNN+DNN	0.47	CAFA3
DeepNF	Gligorijevic et al.	June 2018	Multi-modal Deep Learning	0.42	STRING
DeepPred	Rifaioglu et al.	May 2019	Hierarchical DNN	0.50	CAFA3
DeepGOPlus	Kulmanov et al.	July 2019	CNN+DNN	0.54	CAFA3
UDSMProt	Strodthoff et al.	January 2020	LSTMs	0.58	CAFA3
SDN2GO	Cai et al.	April 2020	CNN+DNN	0.56	CAFA3
MultiPredGO	Giri et al.	September 2020	Multi-modal Deep Learning	0.37	CAFA3
TALE+	Cao et al.	September 2020	Transformer	0.67	CAFA3

Fmax Scores are based on the Molecular Function Prediction of the models.

V. CONCLUDING REMARKS

In the world of bioinformatics, developments in deep learn- ing have had a huge influence. The researchers can now make sophisticated deep learning Models that can derive concealed characteristics from raw biological data. This can remove the requirement of hand-crafted features for machine learning tasks. Moreover, such models can be scaled well with increasing data and compute capacity. [42] Thus, such models can be monumental in building an automated system for annotating protein functions through raw data.

It would be interesting in the future to see how some pre-trained models like prot-BERT [30] and PRoBERTa [43] performs when finetuned for the downstream tasks for protein function prediction. Such protein vector representations are said to be SOTA in the self-supervised task for masked language modelling. The recent ESM-1b protein language mode by Facebook AI Research is also another contender. [44] With recent advances in multi-modal deep learning and the production of multi-dimensional data in biological databases, the characteristics derived from various types of raw data, such as 3D structure, protein-protein interaction, are used to predict protein function, the gene expression profile can be paired with current SOTA protein sequence representations, which is likely to boost the Fmax scores on the CAFA challenge further.

REFERENCES

[1] J. Bernardes and C. Pedreira, "A Review of Protein Function Prediction Under Machine Learning Perspective," Recent Patents on Biotechnology, vol. 7, no. 2, pp. 122–141, 2013. [Online]. Available: 10.2174/18722083113079990006;https://dx.doi.org/10.2174/1872083113079990006

[2] F. Eisenhaber, 2013.

[3] W. Genbank and Statistics, 2021. [Online]. Available: https://www.ncbi. nlm.nih.gov/genbank/statistics/.Accessed28

[4] 2021. [Online]. Available: http://www.wwpdb.org/stats/download. Accessed28

[5] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene Ontology: tool for the unification of biology," Nature Genetics, vol. 25, no. 1, pp. 25–29, 2000. [Online]. Available: 10.1038/75556;https://dx.doi.org/10.1038/75556

[6] A. D.Diehl, J. A. Lee, R. H.Scheuermann, and J. A. Blake, "Ontology development for biological systems: immunology," Bioinformatics, vol. 23, no. 7, pp. 913– 915, 2007. [Online]. Available: 10.1093/bioinformatics/btm029;https://dx.doi.org/10.1093/bioinformatics/btm029

[7] J. D. Osborne, J. Flatow, M. Holko, S. M. Lin, W. A. Kibbe, L. Zhu, M. I. Danila, G. Feng, and R. L. Chisholm, "Annotating the human genome with Disease Ontology," BMC Genomics, vol. 10, no. Suppl 1, pp. S6–S6, 2009. [Online]. Available: 10.1186/1471-2164-10-s1-s6;https: //dx.doi.org/10.1186/1471-2164-10-s1-s6

[8] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," Journal of Molecular Biology, vol. 215, no. 3, pp. 403–410, 1990. [Online]. Available: 10.1016/s0022- 2836(05)80360-2;https://dx.doi.org/10.1016/s0022-2836(05)80360-2

[9] J. Huerta-Cepas, D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M. C. Walter, T. Rattei, D. R. Mende, S. Sunagawa, M. Kuhn, L. J. Jensen, C. von Mering, and P. Bork, "eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences," Nucleic Acids Research, vol. 44, no. D1, pp. D286–D293, 2016. [Online]. Available: 10.1093/nar/gkv1248;https://dx.doi.org/10.1093/nar/gkv1248

[10] B. E, "Genome-scale phylogenetic function annotation of large and diverse protein families," Genome Res, vol. 21, pp. 1969–1980.

[11] D. Cozzetto, F. Minneci, H. Currant, and D. T. Jones, "FFPred 3: feature-based function prediction for all Gene Ontology domains," *Scientific Reports*, vol. 6, no. 1, 2016. [Online]. Available: 10.1038/srep31865;https://dx.doi.org/10.1038/srep31865

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: 10.1038/nature14539;https://dx.doi.org/10.1038/nature14539

[13] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, and Z. Xie, "Deep Learning and Its Applications in Biomedicine," *Proteomics Bioinforma*, vol. 16, pp. 17–32, 2018.

[14] R. Apweiler, "The Universal Protein resource (UniProt)," *Nucleic Acids Res*, vol. 36, 2008.

[15] H. Berman, K. Henrick, and H. Nakamura, "Announcing the worldwide Protein Data Bank," *Nature Structural & Molecular Biology*, vol. 10, no. 12, pp. 980–980, 2003. [Online]. Available: 10.1038/nsb1203-980;https://dx.doi.org/10.1038/nsb1203-980

[16] A. N. Zhou, Y. Jiang, T. R. Bergquist, A. J. Lee, B. Z. Kacsoh, W. Crocker, K. A. Lewis, G. Georghiou, H. N. Nguyen, M. N. Hamid, L. Davis, T. Dogan, V. Atalay, A. S. Rifaioglu, A. Dalkiran, R. Cetin-Atalay, C. Zhang, R. L. Hurto, P. L. Freddolino, Y. Zhang, P. Bhat, F. Supek, J. M. Fernández, B. Gemovic, V. R. Perovic, R. S. Davidovic´, N. Sumonja, N. Veljkovic, E. Asgari, M. Mofrad, G. Profiti, C. Savojardo, P. L. Martelli, R. Casadio, F. Boecker, I. Kahanda, N. Thurlby, A. C. Mchardy, A. Renaux, R. Saidi, J. Gough, A. A. Freitas, M. Antczak, F. Fabris, M. N. Wass, J. Hou, J. Cheng, J. Hou, Z. Wang, A. E. Romero, A. Paccanaro, H. Yang, T. Goldberg, C. Zhao, L. Holm, P. Törönen, A. J. Medlar, E. Zosa, I. Borukhov, I. Novikov, A. Wilkins, O. Lichtarge, P. H. Chi, W. C. Tseng, M. Linial, P. W. Rose,C. Dessimoz, V. Vidulin, S. Dzeroski, I. Sillitoe, S. Das, J. G. Lees, C. T. Jones, C. Wan, D. Cozzetto, R. Fa, M. Torres, A. W. Vesztrocy, J. M. Rodriguez, M. L. Tress, M. Frasca, M. Notaro, G. Grossi, A. Petrini, M. Re, G. Valentini, M. Mesiti, D. B. Roche, J. Reeb, D. W. Ritchie, S. Aridhi, S. Z. Alborzi, M. D. Devignes, D. Koo, R. Bonneau, V. Gligorijevic´, M. Barot, H. Fang, S. Toppo, E. Lavezzo, M. Falda, M. Berselli, S. Tosatto, M. Carraro, D. Piovesan, H. U. Rehman, Q. Mao, S. Zhang, S. Vucetic, G. S. Black, J. D. Larsen, D. J. Omdahl, A. R. Sagers, L. W. Suh, E. Dayton, J. B. Mcguffin, L. J. Brackenridge, D. A. Babbitt, P. C. Yunes, J. M. Fontana, P. Zhang, F. Zhu, S. You, R. Zhang, Z. Dai, S. Yao, S. Tian, W. Cao, R. Chandler, C. Amezola, M. Johnson, D. Chang, J. M. Liao, W. H. Liu, Y. W. Pascarelli, S. Frank, Y. Hoehndorf, R. Kulmanov, M. Boudellioua, I. Politano, G, D. Carlo, S. Benso, A. Hakala, K. Ginter, F. Mehryary, F. Kaewphan, S. Björne, J. Moen, H. Tolvanen, M. Salakoski, T. Kihara, D.

Jain, A. Šmuc, T. Altenhoff, A. Ben-Hur, A. Rost, B. Brenner, S. E. Orengo, C. A. Jeffery, C. J. Bosco, G. Hogan, D. A. Martin, M. J. O'donovan, C. Mooney, S. D. Greene, C. S. Radivojac, P. Friedberg, and I, 2019. [Online]. Available: https://doi.org/10.1101/653105

[17] J. Vig, A. Madani, L. R. Varshney, C. Xiong, R. Socher, and N. F. Rajani, 2020.

[18] M. Kulmanov and R. Hoehndorf, "DeepGOPlus: Improved protein function prediction from sequence," Bioinformatics, vol. 36, pp. 422–429, 2020.

[19] M. Kulmanov, M. A. Khan, and R. Hoehndorf, "DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier," Bioinformatics, vol. 34, no. 4, pp. 660–668, 2018. [Online]. Available: 10.1093/bioinformatics/btx624;https://dx.doi.org/10.1093/bioinformatics/btx624

[20] Y. Cai, J. Wang, and L. Deng, "SDN2GO: An Integrated Deep Learning Model for Protein Function Prediction," Frontiers in Bioengineering and Biotechnology, vol. 8, pp. 1–11, 2020. [Online]. Available: 10.3389/fbioe.2020.00391;https://dx.doi.org/10.3389/fbioe.2020.00391

[21] E. Asgari and M. R. K. Mofrad, "Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics," PLOS ONE, vol. 10, no. 11, pp. e0 141 287–e0 141 287, 2015. [Online]. Available: 10.1371/journal.pone.0141287;https://dx.doi.org/10.1371/journal.pone.0141287

[22] X. Lin, "DeepGS: Deep Representation Learning of Graphs and Sequences for Drug-Target Binding Affinity Prediction," Front Artif Intell Appl, vol. 325, pp. 1301–1308, 2020.

[23] 2021. [Online]. Available: https://github.com/yotamfr/prot2vec. Accessed28

[24] X. L. Liu, 2017. [Online]. Available: https://doi.org/10.1101/103994

[25] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: 10.1162/neco.1997.9.8.1735;https://dx.doi.org/10.1162/neco.1997.9.8.1735

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in Neural Information Processing Systems. Neural information processing systems foundation, pp. 5999–6009, 2017.

[27] A. Radford, J. Wu, R. Child, D. Luan, and D. Amodei.

[28] J. Devlin, M. W. Chang, K. Lee, K. T. Google, A. I. Language, and Bert.

[29] S. C. Huang, L. Martinez-Nunez, R. T. Rupani, H. Platé, M. Niranjan, M. Chambers, R. C. Howarth, P. H. Sanchez-Elsner, T. Azodi, C. B. Tang, J. Shiu, S. H. Senior, A. W. Evans, R. Jumper, J. Kirkpatrick, J. Sifre, L. Green, T. Qin, C. Žídek, A. Nelson, A. Bridgland, A. Pene- dones, H. Petersen, S. Simonyan, K. Crossan, S. Kohli, P. Jones, D. T. Silver, D. Kavukcuoglu, K. Hassabis, D. Vig, J. Madani, A. Varsh- ney, L. R. Xiong, C. Socher, R. Rajani, N. F. Rives, A. Goyal, S. Meier, J. Guo, D. Ott, M. Zitnick, C. L. Ma, J. Fergus, R. Elnaggar, A. Heinzinger, M. Dallago, C. Rost, B. Shanehsazzadeh, A. Belanger, D. Dohan, D. Only, U. Madani, A. Mccann, B. Naik, N. Keskar, N. S. Anand, N. Eguchi, R. R. Huang, P. S. Socher, R. Lu, A. X. Zhang, H. Ghassemi, M. Moses, A. Nambiar, A. Liu, S. Heflin, M. Maslov, S. Hopkins, and M, pp. 1–11.

[30] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, Y. Wang, L. Jones, Gibbs, T. Feher, C. Angerer, M. Steinegger, D. Bhowmik, B. Rost, and Prottrans.

[31] V. Gligorijevic', M. Barot, and R. Bonneau, "deepNF: deep network fusion for protein function prediction," Bioinformatics, vol. 34, no. 22, pp. 3873–3881, 2018. [Online]. Available: 10.1093/bioinformatics/bty440;https://dx.doi.org/10.1093/bioinformatics/bty440

[32] T. Baltrušaitis, C. Ahuja, L. P. Morency, and Multimodal.

[33] S. J. Giri, P. Dutta, P. Halani, and S. Saha, "MultiPredGO: Deep Multi-Modal Protein Function Prediction by Amalgamating Protein Structure, Sequence, and Interaction," IEEE Journal of Biomedical and Health Informatics, vol. 2194, pp. 1–1, 2020. [Online]. Available: 10.1109/jbhi.2020.3022806;https://dx.doi.org/10.1109/jbhi.2020.3022806

[34] Q. Mcnamara, D. L. Vega, A. Yarkoni, and T, "Developing a comprehensive framework for multi-modal feature extraction," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1567–1574, 2017.

[35] R. Cao, C. Freitas, L. Chan, M. Sun, H. Jiang, and Z. Chen, "ProLanGO: Protein Function Prediction Using Neural Machine Translation Based on a Recurrent Neural Network," Molecules, vol. 22, no. 10, pp. 1732– 1732, 2017. [Online]. Available: 10.3390/molecules22101732;https://dx.doi.org/10.3390/molecules22101732

[36] R. Bonetta and G. Valentino, "Machine learning techniques for protein function prediction," Proteins: Structure, Function, and Bioinformatics, vol. 88, no. 3, pp. 397–413, 2020. [Online]. Available: 10.1002/prot.25832;https://dx.doi.org/10.1002/prot.25832

[37] Q. Meng, D. Catchpoole, D. Skillicorn, and P. J. Kennedy, Relational Autoencoder for Feature Extraction.

[38] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. von Mering, "STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," Nucleic Acids Research, vol. 47, no. D1, pp. D607–D613, 2019. [Online]. Available: 10.1093/nar/gky1131;https://dx.doi.org/10.1093/nar/gky1131

[39] A. S. Rifaioglu, T. Dog˘an, M. J. Martin, R. Cetin-Atalay, and V. Atalay, "DEEPred: Automated Protein Function Prediction with Multi-task Feed-forward Deep Neural Networks," Scientific Reports, vol. 9, no. 1, pp. 1–16, 2019. [Online]. Available: 10.1038/s41598- 019-43708-3;https://dx.doi.org/10.1038/s41598-019-43708-3

[40] N. Strodthoff, P. Wagner, M. Wenzel, and W. Samek, 2019. [Online]. Available: https://doi.org/10.1101/704874

[41] Y. Cao and Y. Shen, 2020. [Online]. Available: https://doi.org/10.1101/ 2020.09.27.315937

[42] J. Hestness, S. Narang, N. Ardalani, G. Diamos, H. Jun, H. Kianinejad, Patwary, Y. Yang, and Y. Zhou, 2017.

[43] A. Nambiar, S. Liu, M. Hopkins, M. Heflin, S. Maslov, and A. Ritz, 2020.

[44] A. Rives, S. Goyal, J. Meier, D. Guo, M. Ott, C. L. Zitnick, J. Ma, and R. Fergus, pp. 622 803–622 803, 2019.

arXiv:2211.09657v1 [cs.SI] 17 Nov 2022

A Spreader Ranking Algorithm for Extremely Low-Budget Influence Maximization in Social Networks using Community Bridge Nodes

Aaryan Gupta*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
aryan227227@gmail.com

Inder Khatri*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
inderkhatri999@gmail.com

Arjun Choudhry*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
choudhry.arjun@gmail.com

Pranav Chandhok
Delhi Technological University
New Delhi, India
pranav.chandhok2000@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
dinesh@dtu.ac.in

Mukesh Prasad
School of Computer Science
University of Technology Sydney
Ultimo, Australia
mukesh.prasad@uts.edu.au

November 18, 2022

Abstract

In recent years, social networking platforms have gained significant popularity among the masses like connecting with people and propagating one's thoughts and opinions. This has opened the door to user-specific advertisements and recommendations on these platforms, bringing along a significant focus on Influence Maximisation (IM) on social networks due to its wide applicability in target advertising, viral marketing, and personalized recommendations. The aim of IM is to identify certain nodes in the network which can help maximize the spread of certain information through a diffusion cascade. While several works have been proposed for IM, most were inefficient in exploiting community structures to their full extent. In this work, we propose a community structures-based approach, which employs a K-Shell algorithm in order to generate a score for the connections between seed nodes and communities for low-budget scenarios. Further, our approach employs entropy within communities to ensure the proper spread of information within the communities. We choose the Independent Cascade (IC) model to simulate information spread and evaluate it on four evaluation metrics. We validate our proposed approach on eight publicly available networks and find that it significantly outperforms the baseline approaches on these metrics, while still being relatively efficient.

Keywords Complex Networks · Influence Maximisation · Community Structures · Bridge Nodes · Entropy · Online Social Networks · Independent Cascades

1 Introduction

Complex Network Analysis can be understood as the analysis of a primary structure of a network and its dynamical characteristics (Boccaletti, Latora, Moreno, Chavez and Hwang, 2006; Newman, 2003). In these frameworks, the elements that build these networks are addressed as nodes and the connections or relations between them are addressed as edges. For example, in a reference framework, publications form the nodes, while their citations form the edges. In an internet-based social networking platform, a person's records represent a node, while the person's various connections

*Equal Contribution

with other people are represented by other nodes (like *kinship* or *follow-supporter*) form the edges. Generally, these genuine organizations observe *Power-Law Degree Circulation*, which makes their conduct eccentric and their study back-breaking (Barabási and Bonabeau, 2003).

Online Social Networks (OSNs) provide a convenient way for users to collaborate and convey their ideas to other people (Mislove, Marcon, Gummadi, Druschel and Bhattacharjee, 2007; Garton, Haythornthwaite and Wellman, 1997; Kumar, Novak and Tomkins, 2006). With the web insurgency, the unexpected upsurge in the number of users on OSNs has given the showcasing experts yet another methodology to consider, which can be more compelling and less expensive than the conventional methodologies. This methodology, also known as *Viral Marketing* (Ferguson, 2008), employs OSNs to acquire a significantly larger audience, as compared to the traditional methods. One way of performing viral marketing is Influence Maximisation, which identifies a given number of influential nodes in a network, which induces the maximum spread possible.

In recent years, several approaches have been proposed for Influence Maximisation due to the task's significant importance and applicability in the industrial domain. A lot of these approaches are local and semi-local structure-based algorithms that rely on the different local aspects of the network. Apart from these, various other works have relied on the global properties of the network, which generally require a longer time for execution. Further, some works consider *bridge nodes* to be the most influential nodes and pose Influence Maximisation (IM) as a bridge node identification problem. Some recent works also try to identify the core nodes as influential nodes.

Recently, various community structure-based approaches have been proposed, which divide the network into small sub-groups called communities, and further, find spreaders by using these structures. These community-based algorithms are found to be more efficient and simpler for real-life utilization. However, these approaches focus on extracting the core nodes, which have a good influence over their respective communities. However, in the case of extremely low-budget IM, where the number of spreaders is extremely small, it is likely that few communities get influenced using these approaches. To counter this, we instead propose the use of community bridge nodes as spreader nodes, which have connections to a large number of communities and can aid in maximising the influence spread in the case of a very small number of initial spreaders.

In this paper, we present a novel method for tackling Influence Maximisation in complex networks using a novel centrality measure that uses the concepts of community structures and K-shell Decomposition to identify the influential spreaders in a network. Our method scores a node by considering its connection to different communities, and further scores the strength of its relation to those connected communities. We define three measures, i.e., Community K-Shells, Community K-Shell Entropy, and Community K-Shell Score (CKS_Score), which are used to evaluate the connection of a node to the communities in terms of their qualitative and quantitative aspects, and finally, rank the nodes on the basis of these measures. The measures are able to identify and give a higher score to the nodes which have the maximum spreading capacity or influence over the entire network. For further experimental analysis to verify our proposed approach, we use the Independent Cascade (IC) model to simulate the influence propagation on eight different networks. Our experimental evaluations over various datasets and performance metrics validate the superiority of our proposed approach over seven previous state-of-the-art approaches across different domains and act as evidence of its efficacy and widespread applicability. The main contribution of this paper can be summarised as:

- We propose a novel approach for Influence Maximisation in low-budget scenarios by employing the concepts of Community Structure, K-Shell Decomposition, and Shannon's Entropy.
- We define three novel measures, namely Community K-Shells, Community K-Shell Entropy, and CKS_Score, that help to qualitatively and quantitatively evaluate the connections of a node to various communities.
- We verify our approach by evaluating it on eight different datasets using various performance metrics, and further compare it with the existing approaches. We also perform the Friedman and Iman-Davenport statistical tests to verify the superiority of our approach.

The remainder of the paper is presented as follows. Section 2 consists of the related works and the existing research gaps. Section 3 contains preliminary information needed to evaluate and understand our approach. Section 4 contains our proposed methodology and algorithms. Section 5 contains information about the various experimental details like the datasets used, baseline models, the performance metrics used, and the implementation environment. Section 6 consists of our experimental results, their analysis, and broad outcomes. Section 7 contains the Friedman and Iman-Davenport statistical tests to verify the superiority of our approach. Section 8 contains our concluding statement.

2 Related Works & Research Gaps

Over the last few decades, Influence Maximisation as a task has seen a significant number of contributions from researchers. A wide variety of methods have been proposed for finding the most viable and influential spreader

nodes (Li, Fan, Wang and Tan, 2018; Banerjee, Jenamani and Pratihar, 2020; Singh, Srivastva, Verma and Singh, 2021). Among these approaches, centrality measure-based methods have recently garnered significant attention from researchers due to their reduced time complexity and simpler methodology (Dey and Roy, 2019; Dey, Roy and Bhattacharya, 2021). These approaches usually consist of two steps: generating a score for each node using a defined centrality measure and selecting the top k on the basis of the calculated score.

Freeman (1978) proposed one of the earliest methods involving Degree Centrality, which directly considers the number of neighbours to generate a score for a node. Degree Centrality has linear time complexity and was found to be effective in small networks. Due to its dependence only on the immediate neighbours of the node for scoring, the efficacy of Degree Centrality is questionable while scoring the nodes with lesser but highly influential neighbour nodes.

As an alternative to degree centrality, some semi-local centrality measures like Local Centrality (LC) (Chen, Lü, Shang, Zhang and Zhou, 2012) and Local Structural Centrality (LSC) (Gao, Ma, Chen, Wang and Changming, 2014) have also been introduced, which go beyond the immediate neighbours but up to a limited range. LC ranks these on the basis of one-hop and two-hop nearest neighbours, whereas LSC further considers an additional factor, i.e. the clustering coefficient between the immediate neighbours. Some semi-local methods also try to identify the bridge nodes in a network as influential nodes. Bridge nodes are the nodes which connect different groups of nodes. Liu, Xiong, Shi, Shi and Wang (2016) defined the Degree and Importance of Lines (DIL) centrality measure, which uses the degree value and the importance of lines to highlight the bridge node. It further takes into account the number of connected triangles when determining the relevance of a link, with the bridge node taking the central role in controlling information propagation. Berahmand, Bouyer and Samadi (2019) proposed Dense and Centrality Localization (DCL), a measure which uses a node's degree, the clustering coefficient, and the relationship between its one-hop neighbours to find out the bridge nodes. Wen and Deng (2019) proposed Local Information Dimensionality (LID), an approach which identifies influencers in complex networks by considering quasi-local information, i.e., local structural properties. However, since these approaches do not take into account a node's global location and only use the information corresponding to a limited area or locality, their implications in real-life large networks become less efficient.

Some recent works have also used global centrality measures like Betweenness Centrality, Closeness Centrality, and K-shell Decomposition, among others. Freeman (1977) defined Betweenness for a node as the ratio of the shortest path which passes through the given node, and the total number of possible shortest paths in the network. Okamoto, Chen and Li (2008) proposed Closeness for a node as the inverse of the sum of the shortest path to all other nodes, hereby determining the average closeness of a node to the rest of the graph. As the calculation of Betweenness and Closeness requires the calculation of the shortest path between every node pair of the network, the time complexity becomes significantly high, which is undesirable. Eigenvector Centrality (Bonacich, 2007) and Pagerank Centrality (Brin and Page, 1998) measures are some other Global centrality measures used for IM. These measures score the nodes on the basis of their importance based on their neighbourhood nodes. According to these measures, the presence of important nodes in the surroundings makes the centre node important.

Kitsak, Gallos, Havlin, Liljeros, Muchnik, Stanley and Makse (2010) proposed K-shell decomposition, another global centrality measure, which divides the nodes into different levels on the basis of their location with respect to the core of the network. Normally, the nodes with the highest shell number are nearest to the core and are selected as the initial seeds. As K-shell decomposition assigns more than one node to a given shell number, this often makes it unreliable as the nodes still in the same shell can produce different influences.

Bae and Kim (2014) proposed an improved algorithm called the Extended Neighbourhood Coreness (ENC) measure, which ranks nodes on the basis of the sum of the K-shell of the neighbourhood nodes. This concept was further extended to get Extended Neighbourhood Coreness by taking the sum of the Neighbourhood Coreness values of a node's neighbours. Zeng and Zhang (2013) introduced Mixed Degree Decomposition (MDD), an iteration of K-shell that further distinguishes nodes in the same shell by using the removed and residual neighbours during the K-shell decomposition process.

In recent years, a lot of works have been proposed that take the help of the community structures to find out the spreader nodes that could maximise the influence spread in the network. Some of the most relevant community-based works from the last few years are Community-based Approach for Opinion Maximization (CAOM) (He, Wang, Mao, Lv, Cai, Huang and Xu, 2019), Community-based algorithm for Finding Influential Nodes (CFIN) (Khomami, Rezvanian, Meybodi and Bagheri, 2021), and Gateway Local Rank (GLR) (Salavati, Abdollahpouri and Manbari, 2018). The community-based works focus on finding out the core nodes in the significant communities, and further leverage the higher connection density of communities to increase the spread using these nodes as initial spreaders. CAOM divides the network into community structures, and the spreaders are divided among the most significant communities. The candidate nodes are hereafter generated in significant communities using a one-hop measure and further nodes are shortlisted by using a two-hop measure and *Elimination of Overlapping Influence* policy. Like CAOM, CFIN also finds the candidate nodes among the significant communities after community detection. The initial nodes are selected in

significant communities with the help of degree centrality, and the seed nodes are finally selected by applying local clustering coefficients over the initial nodes. GLR finds out the prominent nodes in the communities by simplifying the closeness centrality for the community. The best core nodes, i.e., the local critical node and the gateway node are found in each community. Then, the shortest paths of all other nodes to the best core nodes are found and the nodes are given scores based on their closeness to core nodes.

We find that most of the previous works based on the communities consider the core nodes from the significant communities to be the influential spreaders. These core nodes usually have good control over the respective communities in which they are present, and can influence a large fraction of community nodes when used as initial spreaders. In the case of real-life applications of influence maximization, where the number of spreaders is very small, it is likely that only a few selected communities would be influenced, which can be a serious drawback in the long run. To overcome such scenarios, we try to find out the community bridge nodes which are connected to a large number of communities and can simultaneously propagate information to these connected communities. There are a few other recent works like LID, DCL, and DIL, which also use bridge nodes as initial spreaders. However, these methods rely upon local information, and the selected node might not have influence beyond a given locality. Contrary to these approaches, our proposed approach relies upon global metrics like K-shell Decomposition and community structure to find out the community bridge nodes, which can act like global bridge nodes.

3 Preliminaries

This section contains preliminary information which will be utilized in our approach. We share details about the influence maximization task, the uses of community structures in social network tasks, K-shell decomposition, and the Information Propagation model used.

3.1 Influence Maximisation

Influence Maximisation (IM) can be defined as the task of distinguishing *k* potential nodes whose impact spread over the entire organization is most effective for any association of *N* users (Kempe, Kleinberg and Tardos, 2003). The aim is to boost the spread of information diffusion and increase the number of nodes affected. The initial step in IM is to pick *k* most influential spreader nodes. However, due to a significant constraint on assets and time, the number of seed nodes should be fundamentally restricted in comparison to the entire populace. Formally, IM can be considered as: $G(V, E)$, where V is the set of nodes, E is the set of edges, and k is the certain number of seed nodes. IM refers to selecting a set of k nodes such that they maximize the influence spread. The influence spread S' from a set of seed nodes S is maximum at the end of the diffusion process and is denoted as $\sigma(S)$. Mathematically, it can be presented as shown below.

$$S' = \arg\max \sigma(S) \; where \; S \subset V, |S| = k \tag{1}$$

3.2 Community Structure

A *community* is a part of a network that comprises nodes with similar behavior and characteristics. For example, if nodes are symbolized as people, then people in the same age group, or friends circle, or with the same home zip codes are treated as a community. The connections within these nodes are stronger as compared to the connections between these nodes and the rest of the network. Usually, connections within a community are dense, whereas the connections between nodes of different communities are relatively sparse. Therefore, it is assumed that a piece of information propagates efficiently and quickly when injected in a community (Newman, 2004; Banerjee, Jenamani and Pratihar, 2019; Javadpour Boroujeni and Soleimani, 2022; Gao, Yu and Zhang, 2021). Community structures are useful entities that aid not just in Influence Maximisation (Javadpour Boroujeni and Soleimani, 2022), but also in Link Prediction (Li, Song, Lu, Zeng, Shi and Liu, 2020), Viral Marketing (Ferguson, 2008), Epidemic Prevention Management (Li, Liu, Miao and Yang, 2023) and other tasks.

3.3 K-Shell Decomposition

K-Shell Decomposition (Kitsak et al., 2010) can be formally defined as the global centrality measure to divide nodes in accordance with their degree and their vicinity to the core of the network.

It comprises the following steps:

- Every node in the network is supposed to be assigned to a K-shell number.

- The nodes with degree one are recurrently removed from the network till only the higher-degree nodes are left. The removed nodes are allotted to the degree one shell.
- The process is repeated for higher degrees i.e. two, three, and so on, with the allotment of the respective shells until the highest degree, i.e., the degree of the core, is achieved.

The nodes with higher K-Shell values are closer to the core of the network and have a higher degree. These nodes are also considered more influential in comparison to the nodes with lower shell values.

3.4 Information Propagation Model

Information propagation (IP) is the process of spreading a piece of information in a complex network (Chen, Lakshmanan and Castillo, 2013; Guille, Hacid, Favre and Zighed, 2013). It is generally used to evaluate the different spreader-ranking methods. To model information propagation, epidemic models are used. Here, the nodes are treated as a population, and information is modeled as an infectious disease that spreads in the population on interaction. Various epidemic models have been used in recent years for IP processes including Susceptible-Infected-Recovered (SIR) (Hethcote, 2000) model, Independent Cascade (IC) (Goldenberg, Libai and Muller, 2001), and Linear Threshold (LT) (Granovetter, 1978). In this manuscript, we use the Independent Cascade model (IC) to evaluate our proposed approach CKS and further compare it with other state-of-the-art techniques. In the IC model, the nodes can be assigned either one of the two states, i.e., infected nodes or susceptible nodes. Information propagation starts by initially assigning a few seed nodes as infected nodes, which further interact and try to infect other neighboring nodes. Similarly, every infected node tries to infect its neighbor and the process continues till there are no new nodes infected. There is some probability associated with each interaction which describes the chances of infecting a susceptible node. Finally, the total influence of a seed node is estimated by taking the sum of all the nodes which are infected by the end of information diffusion.

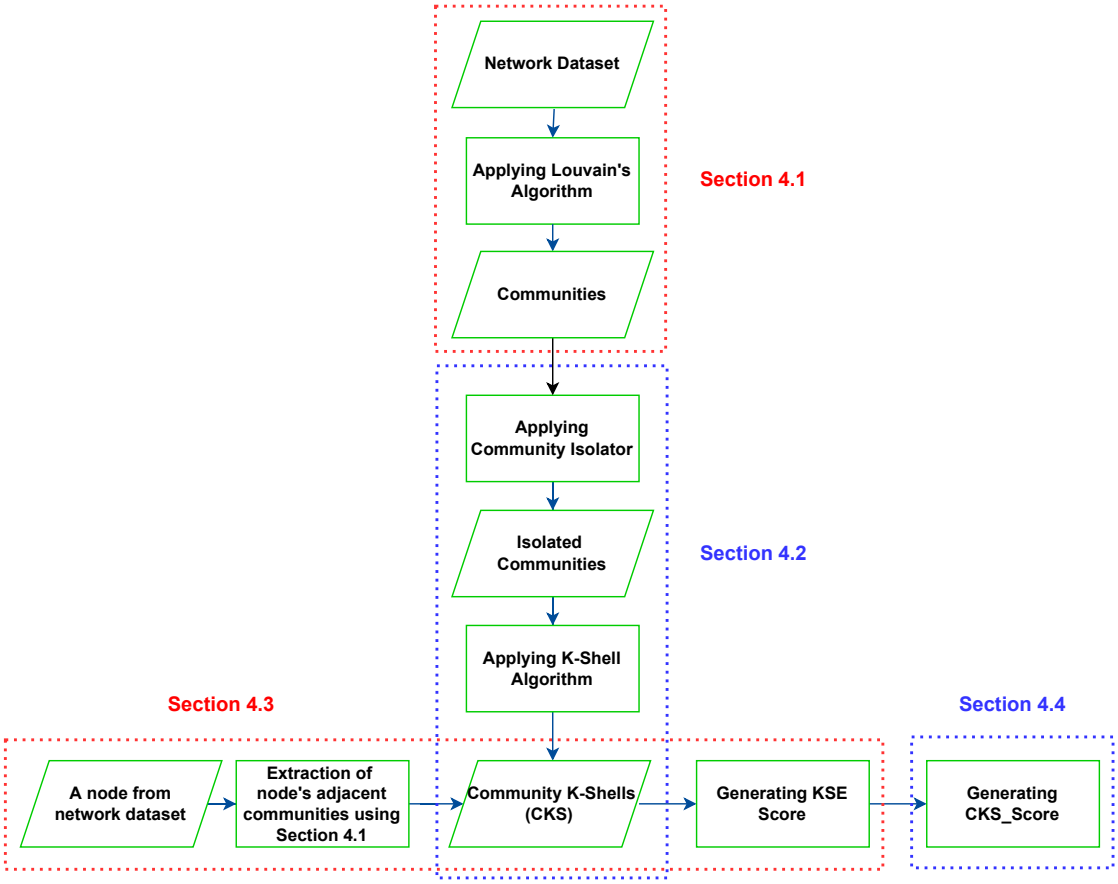


Figure 1: Flowchart representation of the steps involved for the calculation of CKS_Score.

4 Proposed Methodology

In this section, we propose Community K-Shells Score (CKS_Score), a novel algorithm for finding out the seed nodes for Influence Maximization in complex networks. This approach utilizes the concept of communities, K-shell decomposition, and Shannon's Entropy to find out the bridge nodes which are capable of simultaneously spreading information to different subgroups of the network.

Contrary to most of the previous approaches which rank nodes on the basis of their connection to other nodes, we evaluate the nodes on the basis of their connection to the different communities present in the network. We further ensure that the information is being spread in a well-structured manner throughout the community, not just a small part of it. We first compute the KSE score for the connections between a node and the corresponding communities using pre-calculated Community K-Shells. This score depicts how well a piece of information would be transmitted to the community if the given node is informed about it. Then, we compute the CKS_Score for a node by aggregating all the KSE scores between the respective node and connected communities. These aggregated scores are simultaneously weighed by the number of connections between the node and the community, and also the size of the community.

The procedure for the calculation of CKS_Score, represented pictorially in Fig. 1, is as follows:

1. Detecting the communities in the network using the Louvain algorithm (Blondel, Guillaume, Lambiotte and Lefebvre, 2008), which is a partial multi-level algorithm to detect communities in large datasets.
2. Obtaining the Community K-Shells (CKSs) by using the K-shell centrality measure over the results obtained from the Louvain algorithm.
3. Obtaining the Community K-Shell entropy using the CKSs obtained, for the connections between a node and its connected community, which results in nodes with strong distributed connections to a community.
4. Computing the CKS_Score to obtain nodes with strong distributed connections to multiple significant communities.

4.1 Identifying communities

Our proposed methodology begins by identifying communities in the network dataset, which gives a better way of tracking the flow of information in the network.

In this work, we use the Louvain algorithm (Blondel et al., 2008) to detect the communities. Louvain's algorithm is based on greedy optimization techniques, which results in efficient processing and quicker convergence. It works in a step-by-step manner of repeating 2 phases i.e. Local moving of nodes and Aggregation of the network. We use this algorithm for the following reasons:

1. It considers the size of communities, hence it ends up selecting only the impactful communities.
2. It tries to maximize the difference between the actual number of edges in a community, and the expected number of edges in the community.
3. It is based on greedy optimization techniques, which helps in faster convergence and hence makes the process very efficient.

4.2 Obtaining Community K-Shells

Finding Community K-Shells (CKS) is a crucial step in determining the influence of a node in its respective community. The algorithm for obtaining CKS consists of two major steps: isolating the communities and applying the K-Shell algorithm over the isolated community graph.

Isolating the communities helps the K-Shell algorithm handle each community separately. Since the K-Shell algorithm considers the internal structure of the network, applying it over individual communities aids in the analysis of deeper internal structures of the communities. Therefore, it is able to analyze networks not just at a macro level, but also at a micro level by tackling and handling communities individually.

- The isolation of the communities involves deleting the edges between nodes from different communities. Hence, we achieve a graph with several unconnected sub-graphs representing various communities. Any computation over these community sub-graphs will be handled by the algorithm as if each of these communities represents a separate graph.

- While applying the K-Shell algorithm to these sub-graphs, the isolated communities are passed through the K-Shell algorithm, which divides the nodes of a community into various shells as per their closeness to the core of the community. The closer a shell is to the core, the greater the influence its constituent nodes will have on the respective community. Each shell is assigned a specific community k-value, which is equal to 1 for the outermost shell, and it increases as we move toward the core of the community. Therefore, the greater the k-value of a shell, the higher the impact of its constituent nodes on the respective community.

Algorithm 1 : Community K-Shell

Input: $G = (V, E)$
Output: $Community_K - Shell$
1: $\{c_1, c_2, c_3....c_m\} \leftarrow G\ partitioned\ using\ the\ Louvain\ \ Algorithm$
2: $H \leftarrow Isolated\ communities\ by\ deleting\ inter-\ community\ connections\ in\ G$
3: $CKS \leftarrow Apply\ K - shell\ algorithm\ over\ H$
4: **return** CKS

Algorithm 1 represents the algorithmic procedure used to compute Community K-Shell (CKS) measure for graph G.

- We divide the graph into various communities using the Louvain algorithm.
- Then, the edges connecting the nodes from different communities are removed, which returns a graph consisting of isolated communities.
- Then, we execute the K-shell algorithm over the already-segregated community graph, which returns the Community K-Shell for each node v.

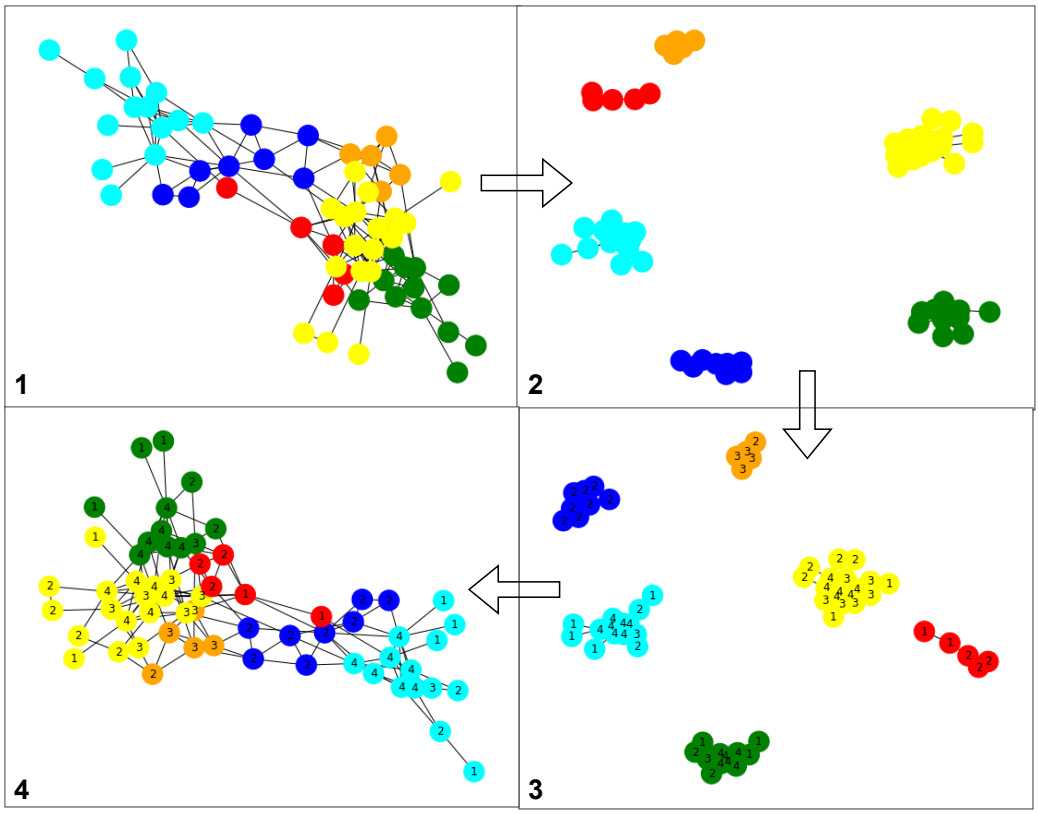


Figure 2: Illustration of the various steps involved in the identification of Community K-shells in a network graph. (1) Graph with communities (2) Isolated community graph (3) CKS in isolated graph (4) CKS in original graph

Fig. 2 portrays an illustration of the concept of CKS. The algorithm starts from a network divided into communities identified using the Louvain algorithm. Each community is represented using a different color. These communities are

further disconnected from each other using the community isolator (as discussed in subsection 4.2), resulting in isolated communities. In order to build the Community K-Shells, these isolated communities are passed through the K-Shell algorithm. Each node is labeled with a number, which represents the respective Community K-Shell (CKS) of the node for the respective community.

4.3 Calculating K-Shell Entropy (KSE)

In this step, the connections between a node and its connected community are assigned a KSE score by analyzing their distribution among various CKSs of the community. This aids in determining a node's impact and extent of influence on a given community, if the node becomes activated and acts as a spreader. Since most of the real-life communities are composed of a large number of nodes, and an activated node's impact is believed to last up to only a few hops (2 to 3 hops), just being connected to a given community does not guarantee high influence in that network. Therefore, to ensure the proper spread of information and a strong influence, a node is considered to be a seed node if and only if it has a good number of connections, and the connections are distributed over different regions of the community.

To ensure the proper distribution of connections of seed nodes, we analyze the distribution of these connections across the community shells generated in the previous step and use Shannon's Entropy over this distribution to check its connectivity to different parts of the networks. A high KSE score conveys that the information reaches the maximum number of shells, and verifies that the connections to that community are not concentrated in one part. It also reduces the possibility of overlapping spread of the same information by nodes in the same locality and ensures that the information spreads in an optimized manner, saving important spreading strength.

Moreover, we constantly ensure that the connections to core nodes are given more importance in comparison to other connections by weighing them with their k-value number, as we can't ignore the fact that the core node would be more impactful individually.

Equation 2 represents the entropy submission over the shells of the respective community weighted by the K-Value of the respective shell, where $\eta_{v,s}$ is the number of connections of node v with shell s of respective community, $shells_c$ represents the unique CKS in the community c, and K_s represents the k-value of the respective shell.

$$KSE_{v,c} = -\sum_{s=1}^{shells_c} K_s * \frac{\eta_{v,s}}{\eta_v} * log(\frac{\eta_{v,s}}{\eta_v}) \tag{2}$$

$$\eta_v = \sum_{s'=1}^{shells_c} \eta_{v,s'} \tag{3}$$

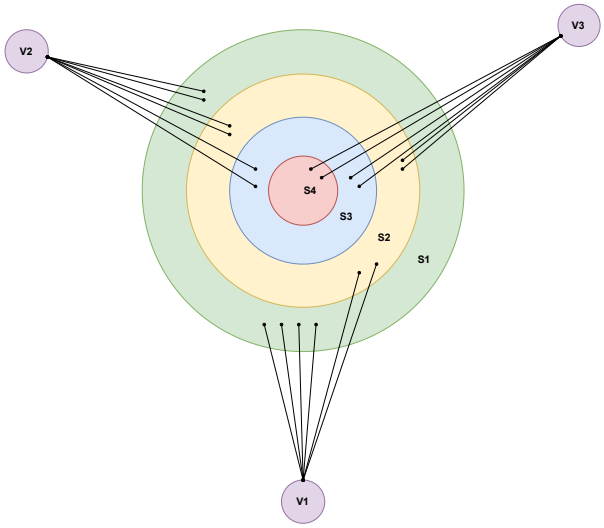


Figure 3: Toy dataset representation of a community with four community k-shells S_1, S_2, S_3 and S_4. There exist three nodes V_1, V_2, and V_3, each having six connections with the community, and each distributed differently.

A better understanding of the concept can be understood from the illustration in Fig. 3. The figure depicts a community that, when passed through the Community K-Shell algorithm, is divided into four shells, namely S_1, S_2, S_3, and S_4. We consider three scenarios where each scenario is depicted by each of the nodes V_1, V_2, and V_3, which are connected to the community by six connections distributed among the four community k-shells in different ways. The scoring for each scenario is shown below:

$$V_1 = -1 * (4/6) * log(4/6) - 2 * (2/6) * log(2/6) = 0.43547500918$$

$$V_2 = -1 * (2/6) * log(2/6) - 2 * (2/6) * log(2/6) - 3 * (2/6) * log(2/6) = 0.95424250943$$

$$V_3 = -2 * (2/6) * log(2/6) - 3 * (2/6) * log(2/6) - 4 * (2/6) * log(2/6) = 1.43136376416$$

We observe that V_2 and V_3 scored significantly higher in comparison to V_1. This can be attributed to the fact that V_2 and V_3's connections are more uniformly distributed as compared to V_1's connections, and since the concept of KSE depends on the concept of entropy, the better the distribution, the higher the score. Further, it was also observed that V_3 scored higher than V_2, due to the fact that its connections are closer to the core of the community in comparison to V_2's connections, and since the entropy score in KSE is constantly weighed by the shell's k-value, hence the closer the connections to the core, the higher the score.

4.4 Calculating CKS score

In this step, the final influence of a given node is calculated by aggregating the KSE scores of its connections with all the neighboring communities. We consider all the neighborhood communities simultaneously, which gives an overall influence of the node on the whole network. Thus, it is more likely for the algorithm to select a node that spreads information in different communities rather than a single community, hereby ensuring that the selected seed node is connected to more than one community in the network, making it a bridge node.

The scoring also considers the individual characteristics of communities, since each community might have a different impact on the whole network. Hence, the CKS_score is weighted by community size. Finally, the score for each community is multiplied by the number of connections between the given community and the given node to ensure the absoluteness of the score, since entropy is a relative term. The mathematical equation for the calculation of CKS_Score is given in Equation 4:

$$CKS_Score(v) = \sum_{c=1}^{comm} NN_c * KSE_{v,c} * \eta_v \tag{4}$$

Here, η_v is the number of connections of node v with the respective community as shown in Equation 3, $KSE_{v,c}$ represents the K-Shell Entropy for node v and community c, and NN_c represents the number of nodes in the community c.

Algorithm 2 : CKS_SCORE

Input: $G = (V, E)$, where $n = |V|$, $e = |E|$
Communities partitioned using Louvain Algorithm
CKS :Calculated using Algorithm 1
v : Node whose score needs to be computed
K_s : k-value of CKS in a community
NN_c : Number of nodes in communities
Output: CKS_Score

```
CKS_Score ← 0
for c ∈ {communities connected to v} do
    KSE_{v,c} ← −Σ_{s=1}^{shells_c} K_s * (η_{v,s}/η_v) * log(η_{v,s}/η_v)  (using Equations 2 & 3)
    CKS_Score ← NN_c * KSE_{v,c} * η_{v,c}  (using Equation 4)
end for
return CKS_Score
```

Algorithm 2 describes the computation of CKS_Score.

- It takes as input the graph G, communities $c_1, c_2, c_3...c_m$ computed using the Louvain algorithm, CKS scores computed using Algorithm 1, the dictionary of nodes in the community c, i.e., NN_c, and k-value of each CKS shell.
- It begins by the initializing score and iterating c over communities connected with v.
- Then, it loops over each unique value of CKS present in the community c.
- Further, the value of K-Shell Entropy (KSE) for a node v and a community c is calculated by applying the concept of entropy over the ratio of the number of connections between a node v and shell s, over the number of connections between node v and all the shells of community c. This is further weighted with k-value of each shell, i.e., K_s, as shown in Equation 2.
- The K-Shell Entropy of each community is further weighted with the number of connections between node v and the given community c, i.e., η_v and the number of nodes in the respective community NN_c.

5 Datasets, Baseline Models and Performance Metrics

In this section, we provide the experimental details of our in-depth analysis done for CKS and competing approaches. We elaborate on the datasets used for evaluation, the baseline approaches for comparison, the performance metrics used, and the experimental setup and parameters.

5.1 Datasets

Table 1: The different datasets used for experimentation along with the various characteristics such as Number of Nodes, Edges and Communities by Louvain Algorithm.

Dataset	Nodes (n)	Edges (m)	Communities
Wiki-Vote (Leskovec, Huttenlocher and Kleinberg, 2010a,b)	889	2914	40
Twitch (Rozemberczki, Allen and Sarkar, 2019)	7,126	35,324	20
BA (Barabási and Albert, 1999)	2000	9974	14
Soc Hamsterster (Hamsterster)	2,400	16,600	169
PGP (Boguñá, Pastor-Satorras, Diaz-Guilera and Arenas, 2004)	10,638	24,301	104
PCG (Holme and Kim, 2002)	2000	9963	21
p2p-Gnutella04 (Ripeanu, Iamnitchi and Foster, 2002)	10,876	39,994	24
Email-univ (Leskovec, Kleinberg and Faloutsos, 2007)	1100	5500	11

We evaluate our proposed approach on six real-life and two synthetic datasets of varying sizes and types. Table 1 describes the different datasets used during our evaluation.

A brief description of different datasets is given below:

- Wiki-Vote (Leskovec et al., 2010a,b): The dataset contains Wikipedia voting data right from its inception till January 2008. Nodes represent Wikipedia users and the directed edges from node *u* to node *v* represent that user *u* voted for user *v*.
- Twitch (Rozemberczki et al., 2019): This dataset is used for node classification and transfer learning based on Twitch user-user networks of gamers. Here, the nodes represent the users themselves, while their mutual friendship is represented by the edges.
- BA (Barabási and Albert, 1999): BA is a random graph generated using Barabási-Albert preferential attachment model. A network of *n* nodes is generated by adding new nodes, each having *m* edges that are preferentially coupled to existing nodes with a high degree.
- Hamsterster (Hamsterster): This network represents the friendships and family links between users on the website hamsterster.com. Here, each user is a node and their friendships are the edges.
- PGP (Boguñá et al., 2004): Pretty Good Privacy (PGP) is an encrypted communication network. PGP algorithm is used for secure information interchange. Here, the users are nodes, and the edges of this huge network are the connectivity between users.
- PCG (Holme and Kim, 2002): The Power-law Cluster (PCG) (Holme and Kim, 2002) approach was developed to generate random graphs with a power-law degree distribution and approximate average clustering. The

technique requires three parameters (n, m, p), where n denotes the number of nodes to be added, m denotes the number of random edges to be added for each new node, and p is the probability of creating a triangle after adding a random edge.

- p2p-Gnutella04 (Ripeanu et al., 2002): This dataset is one of the snapshots of the Gnutella peer-to-peer file sharing network from August 2002. Nodes represent hosts in the Gnutella network, while edges represent connections between the Gnutella hosts.
- Email-univ (Leskovec et al., 2007): This network was generated from a European research institution using their email data from October 2003 to May 2005 (18 months). Here, each node represents an email address, while directed edges represent the flow of emails.

5.2 Baseline Approaches

The baseline approaches used by us for comparison in this paper are presented below. They have been explained in detail in Section 2.

- Extended Neighborhood Coreness (ENC) (Bae and Kim, 2014): ENC extends the K-shell decomposition to a two-hop measure by taking the sum of both one and two-hop measures.
- Gateway Local Rank (GLR) (Salavati et al., 2018): GLR extends over the closeness centrality measure and simplifies it by reducing the search set to the local and gateway nodes.
- Degree and Clustering Coefficient and Location (DCL) (Berahmand et al., 2019): DCL measures the spreading ability of nodes on the basis of a node's degree, its neighbor, and the clustering coefficient.
- Local Information Dimensionality (LID) (Wen and Deng, 2019): LID measures the spreading ability of a node by considering the quasilocal structure of a node.
- Degree and Importance of Lines (DIL) (Liu et al., 2016): DIL ranks the nodes on the basis of the degree and the importance of lines.
- Betweenness Centrality (BC) (Freeman, 1977): BC ranks the nodes on the basis of their presence in the shortest path between the nodes. Despite being among the original IM approaches, it still performs comparably with newer approaches.
- Closeness Centrality (CC) (Okamoto et al., 2008): CC ranks the nodes on the basis of their closeness to the rest nodes in the graph.

5.3 Performance Metrics

- Final infected scale: The final infected scale is the ratio of infected or active nodes at the end of the spread time and the total number of nodes in the graph. The activation of nodes is caused due to activation probability working in favor of the nodes. The initial number of influential nodes plays a major role in the final number of infected nodes in the network. The fraction of total nodes that act as initial spreaders is called the spreader fraction. Usually, a high value of the spreader fraction leads to a high number of the final infected nodes.
- Average Shortest Path Length: The Average Shortest Path Length (ASPL) is the average of shortest paths between all the influential nodes when selected in pairs. It helps us determine the efficiency of data transmission in a social network. A higher value of ASPL conveys that the influential nodes are evenly spread and cover a larger part of the network, rather than being concentrated in an area and resulting in a lesser spread in the whole network.
- Final infected scale vs Activation Probability (P): Activation probability plays a big role in the total final spread. As the activation probability increases, the final infected scale should also rise. The final infected scale represents the total number of infected nodes at the end of spread time. It plays a major role in the analysis of the effect of activation probability on the total spread of information.
- Execution Time: In order to compare our proposed approach with existing techniques, the time taken to rank the nodes on the basis of their spreading ability was also taken into account. The time consumption for all these centralities was evaluated in a Jupyter notebook environment, with the code being run on a Tesla T4 GPU, with 12GB RAM available to the environment.

5.4 Experimental Setup and Parameters

We evaluated the results of our experiments 100 times for each model on each dataset for each parameter. The results plotted in Fig. 4, Fig. 5, Fig. 6, and Fig. 7 are the average results for each model for 100 simulations of each experiment.

For calculating the Final Infected State vs Spreader Fraction values for each model, initial spreader fraction values for datasets having less than 2000 nodes were chosen from the set 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, whereas for larger datasets with greater than 2000 nodes, the initial spreaders fraction were taken from the set 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04. For smaller datasets, we have taken relatively larger spreader fractions because smaller values will lead to a very small number of initial spreaders in small datasets. The infection probability, i.e., the probability for a susceptible node to be influenced was taken to be 0.1 for all methods and datasets.

For calculating the Final Infected State vs Activation Probability (P), We took the different activation probability values in the range {0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25}. The initial spreader fraction for each case was taken to be 0.03, and for each dataset, the simulation was performed 100 times.

While calculating the average spreader distance between initial spreaders, initial spreader fraction values for datasets having less than 2000 nodes were chosen from the set 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, whereas for larger datasets with greater than 2000 nodes, the initial spreaders fraction were taken from the set 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04. We set the activation probability (P) to 0.1.

6 Experimental results and analysis

In this section, we present the results of our experiments, as well as their analysis to understand why CKS outperforms competing approaches across all performance metrics while remaining reasonably efficient.

6.1 Final infected scale

As discussed in Section 5.3, the final infected scale plots the final infection at the end of the simulation, with respect to the different values for the fraction of initial spreaders. Fig. 4 shows the plots for Final Infected Scale vs Ratio for different approaches. The initial spreader fractions are plotted along the x-axis, while the corresponding final infected values are plotted along the y-axis. We observed that for most of the datasets, final infected numbers increased by increasing the spreader fraction. However, for a few datasets like Wiki-Vote and Twitch, the final infected values did not increase much on increasing the initial spreaders fraction for certain approaches other than CKS. Our proposed approach, CKS, showed significantly better performance as compared to the other novel approaches on all the datasets. This is due to the fact that the seed nodes selected using CKS are farther apart and have a wider influence spread in the network. Other approaches tend to find seed nodes in the same communities, due to which they do not show meaningful improvement in the final infected state upon increasing the spreader fraction. This is due to the overlap of information spread in the network, which occurs lesser for CKS, as its final infected state keeps increasing on increasing spreader fraction values. BC, CC, and DCL also performed reasonably well on all datasets. We also observed that ENC performed the worst among all approaches on this metric. Hence, we can conclude that CKS outperforms other competing methods for a wide range of values for the initial spreader fraction.

6.2 Final infected scale vs Activation Probability (P)

In this section, we evaluated the performance of our proposed approach for different values of infection probability. Fig. 5 shows the plots for the Final Infected Scale vs Activation Probability for the different methods on all eight datasets. The different values for the activation probabilities are taken along the x-axis, whereas the number of infected nodes is taken along the y-axis. From the given plots, we can observe that as per our expectation, the number of final infected nodes increased upon increasing the activation probability for all methods and across all datasets. CKS consistently performed as well or better than the competing approaches across all datasets, validating its superiority over other approaches. Thus, CKS can propagate information better than other approaches, and this lead only increases on increasing the activation probability.

6.3 Average Distance between Spreaders

When selecting the initial spreaders, it is desirable that the selected spreaders should have a minimal overlapping region of influence, such that they are able to influence different parts of the network. To check this, we computed the average shortest distance between the initial spreaders (L_s) for different values of spreaders fraction. Fig. 6 illustrates the L_s vs Initial spreaders plots for all eight datasets for our proposed approach as well as other competitive approaches. The x-axis shows the different initial spreader fractions, while the y-axis shows the corresponding value of L_s for each fraction. From the plots for different datasets, we observed that ENC and GLR were among the worst-performing approaches, whereas LID and DIL showed similar performance and were slightly better than ENC and GLR. We also observed that our proposed approach CKS outperformed all other approaches, with DCL and PR being the next best

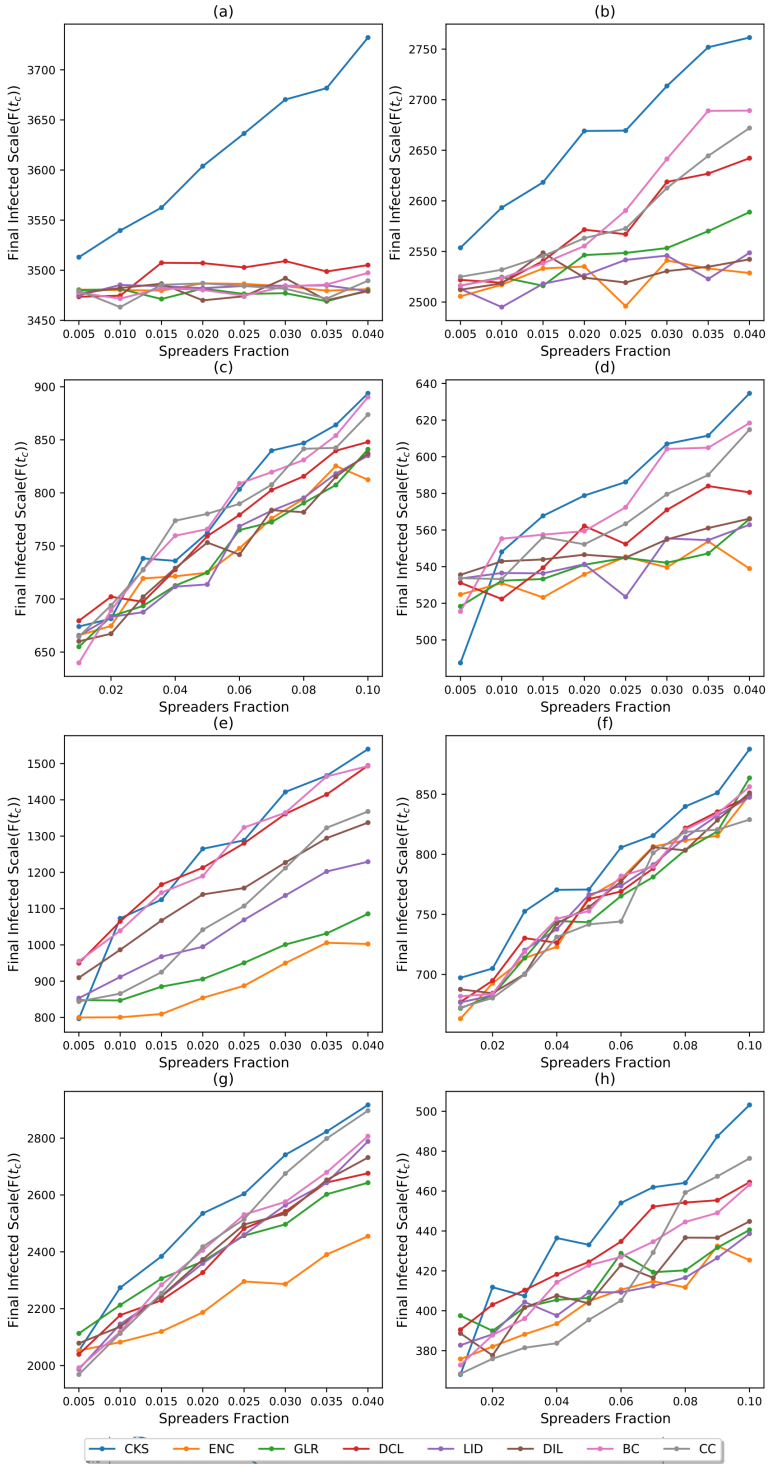


Figure 4: Final infected scale plots for various values for initial spreader fraction on the (a) Wiki-Vote dataset (b) Twitch dataset (c) BA dataset (d) Soc-Hamsterster dataset (e) PGP dataset (f) PCG dataset (g) p2p-Gnutella04 dataset (h) Email-univ dataset. The results are averaged for 100 independent simulations of the IC model with an activation probability (P) equal to 0.1.

approaches. This indicates a lesser *rich-club effect* in the case of CKS and leads to a wider spread of information in the network. We also observe that CKS as well as other approaches typically reach the maximum average spreader distance

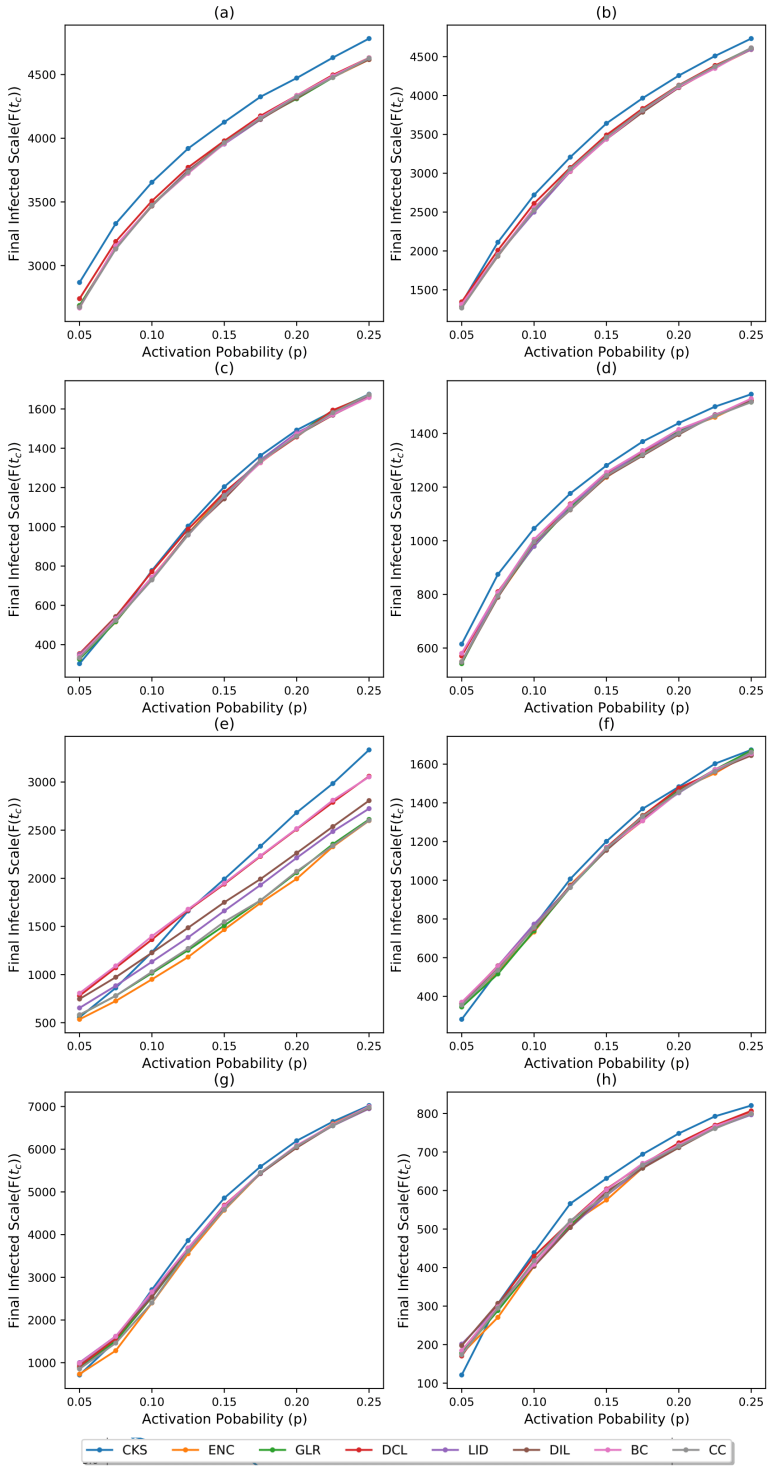


Figure 5: Final Infected Scale plots for different activation probability (P) values on (a) Wiki-Vote dataset (b) Twitch dataset (c) BA dataset (d) Soc-Hamsterster dataset (e) PGP dataset (f) PCG dataset (g) p2p-Gnutella04 dataset (h) Email-univ dataset. The results are averaged for 100 independent simulations of the IC model with an initial spreader fraction equal to 0.3.

around the spreader fraction value of 0.4. Till this point, increasing the spreader fraction only covers other communities

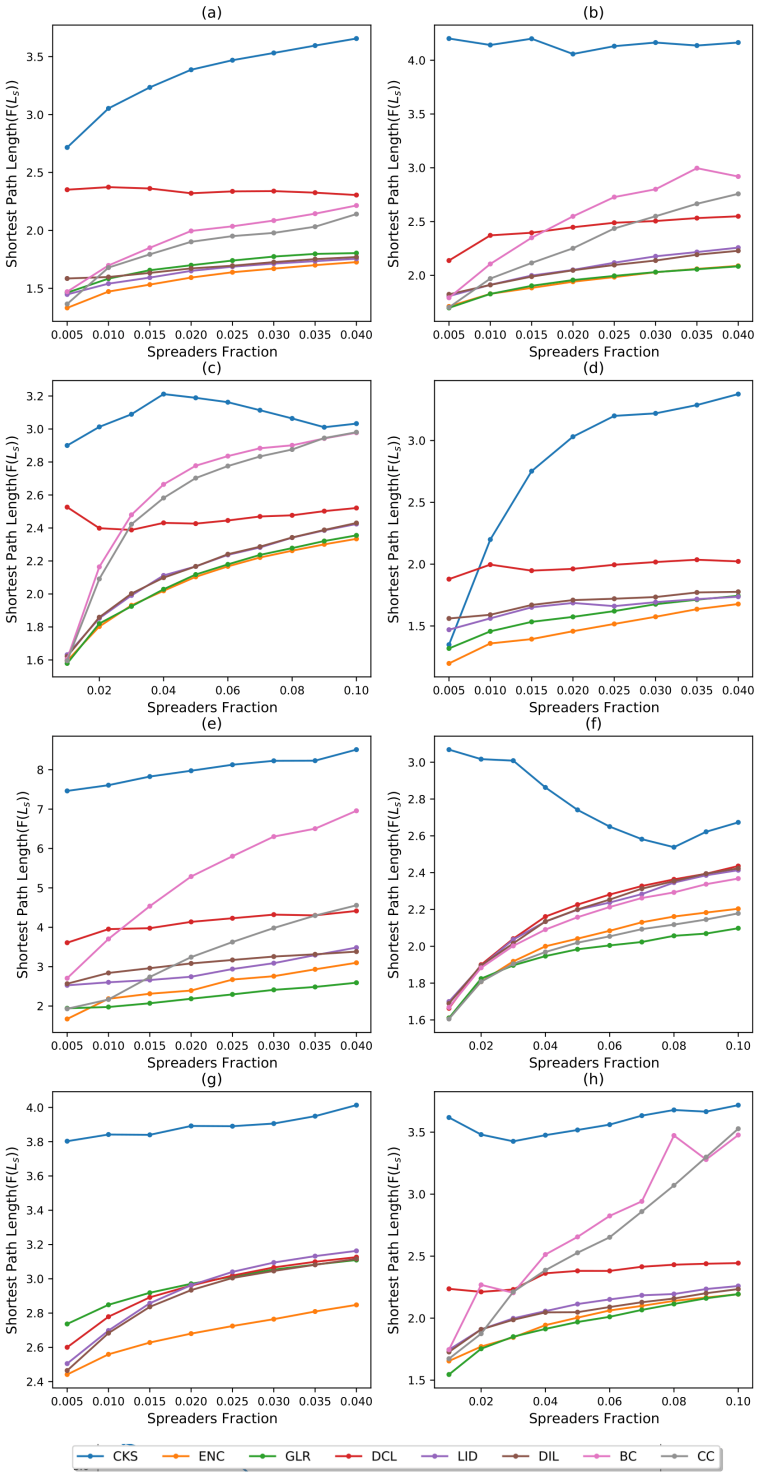


Figure 6: Average distance between spreaders with respect to different initial spreaders fraction values for (a) Wiki-Vote dataset (b) Twitch dataset (c) BA dataset (d) Soc-Hamsterster dataset (e) PGP dataset (f) PCG dataset (g) p2p-Gnutella04 dataset (h) Email-univ dataset. The results are averaged for 100 independent simulations of the IC model with an activation probability (P) equal to 0.1.

which weren't influenced at lower spreader fractions. Beyond 0.4, the density of spreaders in the network increases,

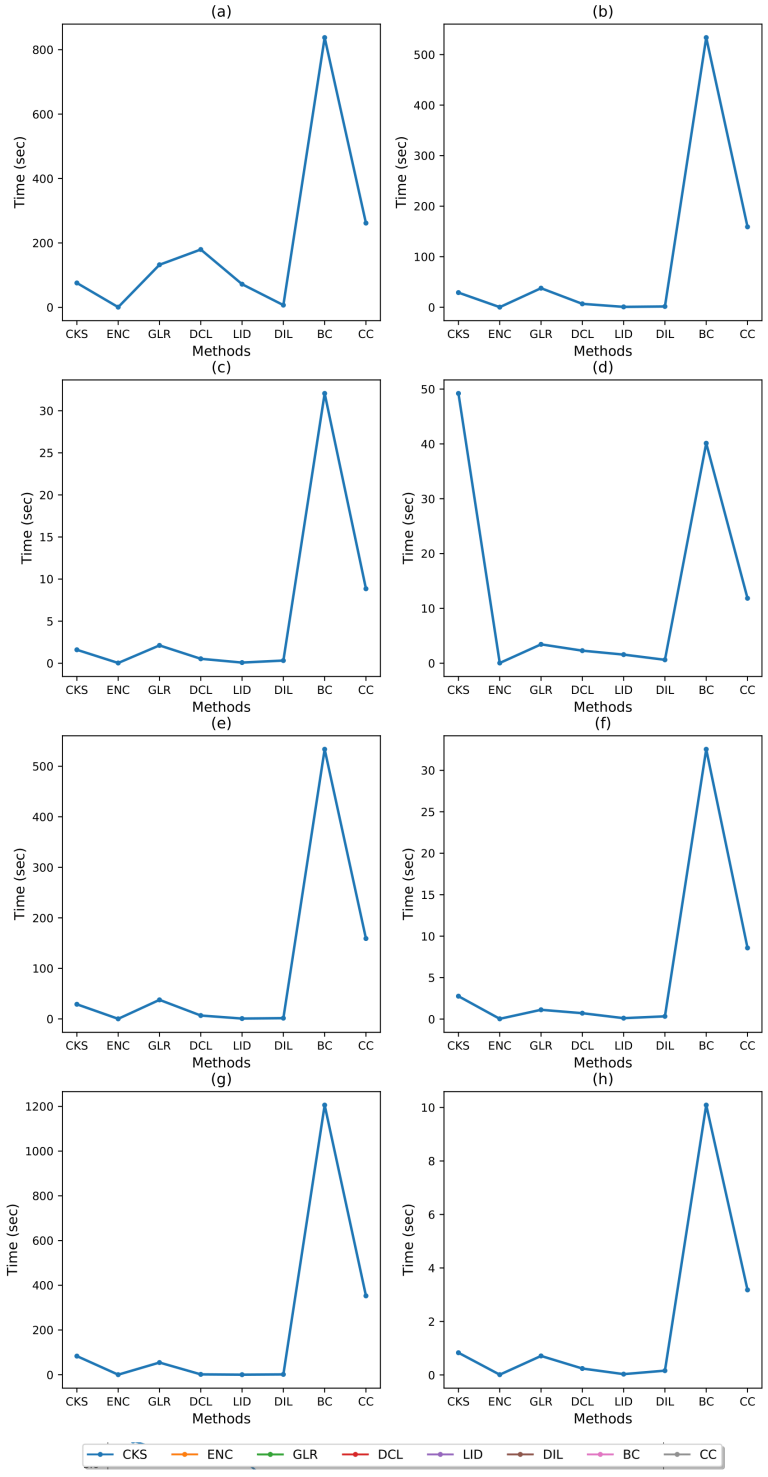


Figure 7: The execution time plots for different methods for ranking the spreaders for (a) Wiki-Vote dataset (b) Twitch dataset (c) BA dataset (d) Soc-Hamsterster dataset (e) PGP dataset (f) PCG dataset (g) p2p-Gnutella04 dataset (h) Email-univ dataset.

leading to a reduction in the average distance between spreaders. However, in almost every situation, CKS outperforms all competing approaches on all datasets, indicating the superiority of CKS.

6.4 Execution Time

In this section, we evaluated our proposed approach CKS and other competing approaches on the basis of the execution time required for ranking the nodes for all eight datasets. Their corresponding plots are shown in Fig. 7. The x-axis conveys the various algorithms, whereas the y-axis shows the execution time in seconds. We observed that BC turned out to be the most computationally expensive method, followed by CC. GLR and CKS were computationally less expensive than CC and BC and took a similar time to execute. ENC took the least amount of time for execution among all the methods, followed by DIL, LID, and DCL. From these plots, we can conclude that our proposed methodology CKS is more efficient than other global centrality measures CC and BC while performing similarly to community-based approaches like GLR. We also found CKS to be computationally more expensive than local hybrid centrality measures like DIL, DCL, and LID.

7 Statistical Testing

We further evaluate the performance of our approach CKS using the Friedman statistical test (Friedman, 1937), and Holm's P-values generated from the Iman-Davenport statistic. The former is a non-parametric statistical test that enables multiple comparisons (for two or more methods). It helps us determine if the performance of CKS is noticeably different from other approaches in sets of two or more methods. The Friedman test detects these differences based on the ranking of the methods instead of their errors. It consists of two hypotheses: the null (H_0) and the alternate hypothesis (H_1). In the former, there are no prominent differences between the algorithms (equality of medians condition), while in the latter, the algorithms have significant differences in the medians of their populations, thus negating the null hypothesis. We performed the Friedman test on the Final Infected Scale vs Spreader Fraction metric using the following procedure:

- Gather the generated results for each problem pair.
- Rank the values in ascending order from 1 (best value) to n (worst value) for each problem i for a particular algorithm j.
- For algorithm j, the average rank for each problem i is calculated using the Equation 5:

$$R_j = \frac{1}{n}\sum_{i=1}^{n}(r_i^j) \tag{5}$$

  where r^j is the rank ($1 < j < k$) and R_j is the average rank.
- Now that all algorithms are ranked according to their priority, compute the Friedman Statistic F_f using the following equation. F_f is based on a Chi-square distribution, with $k-1$ degree of freedom. F_f is computed as shown in Equation 6. In the following equation, n represents the number of rows and k represents the number of columns ($n = 8$ and $k = 6$).

$$F_f = \frac{12n}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{6}$$

Table 2 shows the rank list calculated for each approach using the Friedman test. CKS clearly has a lower average value, and thus a better rank than competing approaches.

Table 2: Average ranking of algorithm calculated using the Friedman test

Sr No.	Algorithm	Average Ranking
1	CKS	**1.828**
2	BC	3.421
3	DCL	3.500
4	CC	4.476
5	DIL	5.085
6	LID	5.515
7	GLR	5.789
8	ENC	6.382

The Friedman test produces relatively conservative results, which is undesirable. Thus, we also use Iman and Davenport (1980) statistic F_{id}, which is computed as shown in Equation 7. It follows the $F - distribution$ with the degree of

freedom as $k-1$ and $(n-1)(k-1)$.

$$F_{id} = \frac{(n-1)\chi_F^2}{n(k-1)-\chi_F^2} \tag{7}$$

The unadjusted P-value or Holm P-value, obtained from the Iman-Davenport statistic for the performance of CKS, advocates the rejection of the null hypothesis H_0. The computed P-values are below the standard significance level of $\alpha = 0.05$, indicating that there is a prominent difference between the performance of CKS and the baseline approaches. Thus, the obtained P-values help us conclude the negation of the null hypothesis H_0. However, they are not appropriate for comparison with different methods. To compare these approaches with each other, we calculated their adjusted P-values with CKS as the control algorithm.

Adjusted P-values (APVs) enable us to draw the correct correlation between these algorithms by taking into account the accumulated family error with respect to the CKS control algorithm. APVs can be directly compared with the significance level $\alpha = 0.5$. In order to calculate the adjusted P-values, we defined a few post-hoc procedures. There exist various post-hoc procedures, such as those proposed by Dunn (1961) and Holland and Copenhaver (1987). They differ from each other in their adjustment of the value of α to compensate for multiple comparisons for multiple methods. In this paper, we have used the common Holm's procedure (Holm, 1979) to evaluate the respective APVs. The values of Holm P-values and APVs are always sorted in ascending order. The equation for the same is given below, where indices i and j refers to the main hypothesis whose APVs are being computed and different hypothesis in the set respectively. P_j is the P-value for the j^{th} hypothesis. Holm APV is computed as shown in Equation 8.

$$HolmAPV_i = \min\{v, 1\}, \text{where } v = \max\{(k-j)p_j : 1 \leq j \leq i\} \tag{8}$$

Table 3: Adjusted P-values (APVs) using Holm procedure

Sr No.	Algorithm (CKS is control algorithm)	Z score	Holm p-value	Adjusted p-value
1	ENC	-10.51	3.55e-26	2.48e-25
2	GLR	-9.14	2.92e-20	1.75e-19
3	LID	-8.51	8.26e-18	4.13e-17
4	DIL	-7.52	2.66e-14	1.07e-13
5	CC	-6.11	4.79e-10	1.44e-09
6	DCL	-3.86	5.65e-05	1.18e-04
7	BC	-3.68	1.16e-04	1.16e-04

The results of the APVs using Holm's procedure are shown in Table 3. The intended APVs are less than the significance level, thus rejecting the null hypothesis. Thus, this helps us conclude the superiority of CKS over competing approaches based on two statistical tests as well.

8 Conclusion

Social networks have gained tremendous popularity among the masses due to their increasing use cases in today's world. Influence Maximisation is one of the important areas of research on social networks where optimal seed nodes are needed to be chosen for spreading maximum influence. In this paper, we proposed a novel approach for Influence Maximisation. Our approach employed the concepts of community structures, K-Shell algorithm, and Entropy to rank the influencing power of nodes in a network. Our approach can be described as a two-step procedure, where we score the connections between the community and the nodes, and the nodes with the maximum strong connections to significant communities were selected as seed nodes. To check the efficacy of our proposed approach, we simulated our approach on the Information Cascade model over eight datasets, and further compared and evaluated our approach with seven other novel approaches over Final Infected Scale vs Spreader fraction, Final Infected Scale vs Activation Probability, Average Shortest Path Length vs Initial Spreaders, and Execution Time Proficiency metrics. Our proposed approach, CKS, significantly outperformed the existing methodologies in maximizing the spread of information in the network by identifying the most suitable influential seed nodes. CKS also achieves a higher average spreader distance for the same value of spreader fraction as compared to other approaches, indicating a lower rich-club effect and more efficient spread of information in the network.

References

Bae, J., Kim, S., 2014. Identifying and ranking influential spreaders in complex networks by neighborhood coreness. Physica A Statistical Mechanics and its Applications 395, 549–559. doi:10.1016/j.physa.2013.10.047.

Banerjee, S., Jenamani, M., Pratihar, D., 2020. A survey on influence maximization in a social network. Knowledge and Information Systems 62. doi:10.1007/s10115-020-01461-4.

Banerjee, S., Jenamani, M., Pratihar, D.K., 2019. Combim: A community-based solution approach for the budgeted influence maximization problem. Expert Systems with Applications 125, 1–13. URL: https://www.sciencedirect.com/science/article/pii/S0957417419300612, doi:https://doi.org/10.1016/j.eswa.2019.01.070.

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. science 286, 509–512.

Barabási, A.L., Bonabeau, E., 2003. Scale-free networks. Scientific american 288, 60–69.

Berahmand, K., Bouyer, A., Samadi, N., 2019. A new local and multidimensional ranking measure to detect spreaders in social networks. Computing 101. doi:10.1007/s00607-018-0684-8.

Blondel, V., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics Theory and Experiment 2008. doi:10.1088/1742-5468/2008/10/P10008.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U., 2006. Complex networks: Structure and dynamics. Physics reports 424, 175–308.

Boguñá, M., Pastor-Satorras, R., Diaz-Guilera, A., Arenas, A., 2004. Models of social networks based on social distance attachment. Physical review. E, Statistical, nonlinear, and soft matter physics 70, 056122. doi:10.1103/PhysRevE.70.056122.

Bonacich, P., 2007. Some unique properties of eigenvector centrality. Social Networks 29, 555–564. doi:10.1016/j.socnet.2007.04.002.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Computer Networks 30, 107–117. URL: http://www-db.stanford.edu/~backrub/google.html.

Chen, D., Lü, L., Shang, M.S., Zhang, Y.C., Zhou, T., 2012. Identifying influential nodes in complex networks. Fuel and Energy Abstracts 391. doi:10.1016/j.physa.2011.09.017.

Chen, W., Lakshmanan, L., Castillo, C., 2013. Information and influence propagation in social networks. Synthesis Lectures on Data Management 5, 1–177. doi:10.2200/S00527ED1V01Y201308DTM037.

Dey, P., Roy, S., 2019. Influence maximization in online social network using different centrality measures as seed node of information propagation. Sādhanā 44. doi:10.1007/s12046-019-1189-7.

Dey, P., Roy, S., Bhattacharya, S., 2021. A survey on the role of centrality as seed nodes for information propagation in large scale network. ACM/IMS Transactions on Data Science 2, 1–25. doi:10.1145/3465374.

Dunn, O.J., 1961. Multiple comparisons among means. Journal of the American statistical association 56, 52–64.

Ferguson, R., 2008. Word of mouth and viral marketing: Taking the temperature of the hottest trends in marketing. Journal of Consumer Marketing 25, 179–182. doi:10.1108/07363760810870671.

Freeman, L., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35–41. doi:10.2307/3033543.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Social Networks 1, 215–239.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 32, 675–701. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1937.10503522, doi:10.1080/01621459.1937.10503522, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1937.10503522.

Gao, S., Ma, J., Chen, Z., Wang, G., Changming, X., 2014. Ranking the spreading ability of nodes in complex networks based on local structure. Physica A: Statistical Mechanics and its Applications 403, 130–147. doi:10.1016/j.physa.2014.02.032.

Gao, Y., Yu, X., Zhang, H., 2021. Overlapping community detection by constrained personalized pagerank. Expert Systems with Applications 173, 114682. URL: https://www.sciencedirect.com/science/article/pii/S0957417421001238, doi:https://doi.org/10.1016/j.eswa.2021.114682.

Garton, L., Haythornthwaite, C., Wellman, B., 1997. Studying online social networks. Journal of computer-mediated communication 3, JCMC313.

Goldenberg, J., Libai, B., Muller, E., 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters 12, 211–223. doi:10.1023/A:1011122126881.

Granovetter, M., 1978. Threshold models of collective behavior. American Journal of Sociology 83, 1420–1443. doi:10.1086/226707.

Guille, A., Hacid, H., Favre, C., Zighed, D.A., 2013. Information diffusion in online social networks: A survey. ACM Sigmod Record 42, 17–28.

Hamsterster, . Hamsterster social network. Http://www.hamsterster.com.

He, Q., Wang, X., Mao, F., Lv, J., Cai, Y., Huang, M., Xu, Q., 2019. Caom: A community-based approach to tackle opinion maximization for social networks. Information Sciences 513. doi:10.1016/j.ins.2019.10.064.

Hethcote, H.W., 2000. The mathematics of infectious diseases. SIAM Review 42, 599–653. URL: https://doi.org/10.1137/S0036144500371907, doi:10.1137/S0036144500371907, arXiv:https://doi.org/10.1137/S0036144500371907.

Holland, B.S., Copenhaver, M.D., 1987. An improved sequentially rejective bonferroni test procedure. Biometrics , 417–423.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65–70. URL: http://www.jstor.org/stable/4615733.

Holme, P., Kim, B.J., 2002. Growing scale-free networks with tunable clustering. Physical review E 65, 026107.

Iman, R.L., Davenport, J.M., 1980. Approximations of the critical region of the fbietkan statistic. Communications in Statistics - Theory and Methods 9, 571–595. URL: https://doi.org/10.1080/03610928008827904, doi:10.1080/03610928008827904, arXiv:https://doi.org/10.1080/03610928008827904.

Javadpour Boroujeni, R., Soleimani, S., 2022. The role of influential nodes and their influence domain in community detection: An approximate method for maximizing modularity. Expert Systems with Applications 202, 117452. URL: https://www.sciencedirect.com/science/article/pii/S0957417422007850, doi:https://doi.org/10.1016/j.eswa.2022.117452.

Kempe, D., Kleinberg, J., Tardos, E., 2003. Maximizing the spread of influence through a social network. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 137-146. doi:10.1145/956750.956769.

Khomami, M., Rezvanian, A., Meybodi, M., Bagheri, A., 2021. Cfin: A community-based algorithm for finding influential nodes in complex social networks. The Journal of Supercomputing 77. doi:10.1007/s11227-020-03355-2.

Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A., 2010. Identification of influential spreaders in complex networks. Nature Physics 6, 888–893. URL: https://doi.org/10.1038%2Fnphys1746, doi:10.1038/nphys1746.

Kumar, R., Novak, J., Tomkins, A., 2006. Structure and evolution of online social networks, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 611–617. doi:10.1145/1150402.1150476.

Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010a. Predicting positive and negative links in online social networks. Proceedings of the 19th International Conference on World Wide Web, WWW '10 doi:10.1145/1772690.1772756.

Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010b. Signed networks in social media, in: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 1361–1370.

Leskovec, J., Kleinberg, J., Faloutsos, C., 2007. Graph evolution: Densification and shrinking diameters. ACM transactions on Knowledge Discovery from Data (TKDD) 1, 2–es.

Li, S., Song, X., Lu, H., Zeng, L., Shi, M., Liu, F., 2020. Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm. Expert Systems with Applications 139, 112839. URL: https://www.sciencedirect.com/science/article/pii/S095741741930541X, doi:https://doi.org/10.1016/j.eswa.2019.112839.

Li, Y., Fan, J., Wang, Y., Tan, K.L., 2018. Influence maximization on social graphs: A survey. IEEE Transactions on Knowledge and Data Engineering 30, 1852–1872. doi:10.1109/TKDE.2018.2807843.

Li, Z., Liu, A., Miao, J., Yang, Y., 2023. A three-phase method for spherical fuzzy environment and application to community epidemic prevention management. Expert Systems with Applications 211, 118601. URL: https://www.sciencedirect.com/science/article/pii/S0957417422016542, doi:https://doi.org/10.1016/j.eswa.2022.118601.

Liu, J., Xiong, Q., Shi, W., Shi, X., Wang, K., 2016. Evaluating the importance of nodes in complex networks. Physica A: Statistical Mechanics and its Applications 452. doi:10.1016/j.physa.2016.02.049.

Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B., 2007. Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pp. 29–42.

Newman, M., 2004. Detecting community structure in networks. Eur Phys J 38.

Newman, M.E., 2003. The structure and function of complex networks. SIAM review 45, 167–256.

Okamoto, K., Chen, W., Li, X.Y., 2008. Ranking of closeness centrality for large-scale social networks, in: Frontiers in Algorithmics, pp. 186–195. doi:10.1007/978-3-540-69311-6_21.

Ripeanu, M., Iamnitchi, A., Foster, I., 2002. Mapping the gnutella network. IEEE Internet Computing 6, 50–57. doi:10.1109/4236.978369.

Rozemberczki, B., Allen, C., Sarkar, R., 2019. Multi-scale attributed node embedding. arXiv:1909.13021.

Salavati, C., Abdollahpouri, A., Manbari, Z., 2018. Ranking nodes in complex networks based on local structure and improving closeness centrality. Neurocomputing 336. doi:10.1016/j.neucom.2018.04.086.

Singh, S., Srivastva, S., Verma, M., Singh, J., 2021. Influence maximization frameworks, performance, challenges and directions on social network: A theoretical study. Journal of King Saud University - Computer and Information Sciences doi:10.1016/j.jksuci.2021.08.009.

Wen, T., Deng, Y., 2019. Identification of influencers in complex networks by local information dimensionality. Information Sciences 512. doi:10.1016/j.ins.2019.10.003.

Zeng, A., Zhang, C.J., 2013. Ranking spreaders by decomposing complex networks. Physics Letters A 377, 1031–1035.

Journal of Polymers and the Environment
https://doi.org/10.1007/s10924-022-02612-0

ORIGINAL PAPER

Adsorptive Removal of Congo Red Using Organically Modified Zinc–Copper–Nickel Ternary Metal Hydroxide: Kinetics, Isotherms and Adsorption Studies

Jigyasa Pathak[1] · **Poonam Singh**[1]

Accepted: 26 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

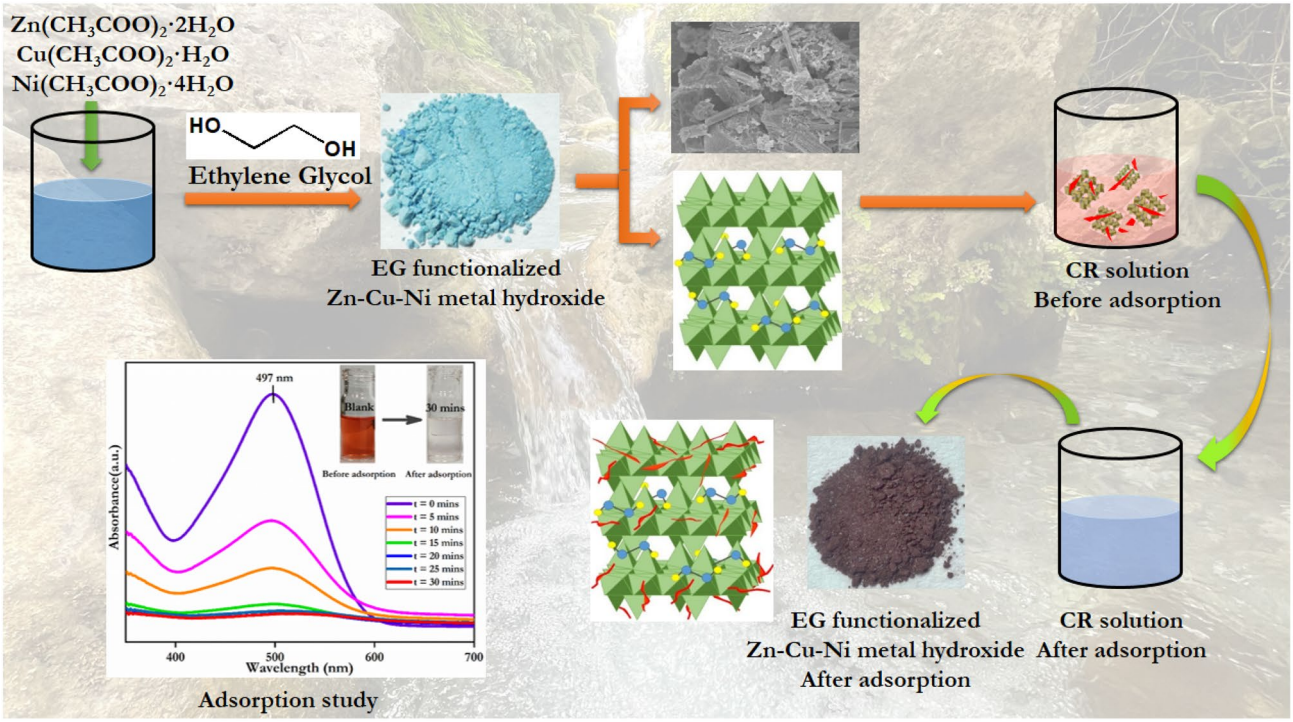
Abstract
In the current study synthesis of composite material consisting of ethylene glycol (EG) functionalized Zinc-Copper-Nickel (Zn–Cu–Ni) ternary metal hydroxide was successfully carried out and was characterized using various techniques. The formation of layered structure was confirmed using Powder X-Ray Diffraction (PXRD), Fourier Transform Infrared spectroscopy (FTIR) and Scanning Electron Microscopy (SEM) while thermal stability was assessed using Thermogravimetric Analysis (TGA). The presence of metal ions and their respective oxidation states was confirmed using Energy Dispersive X-Ray Analysis (EDX), elemental mapping, and X-Ray Photoelectron Spectroscopy (XPS). Pore size of the synthesized material, estimated using Brunauer–Emmett–Teller (BET) analysis technique, was found to be 15.08 nm g^{-1}. The synthesized material was subsequently utilized for the sequestration of toxic Congo red (CR) dye from untreated water. Investigation of controlling parameters i.e. concentration, contact time and pH was undertaken to see their influence on the adsorption performance. The maximum dye removal efficacy of the adsorbent was observed at pH 7, suggesting good buffering ability of EG functionalized Zn–Cu–Ni composite material. The experimental data for equilibrium studies was favoured by the Langmuir model, suggesting monolayer adsorption while the kinetic study was best-correlated with the model of pseudo-second order. The capacity of synthesized material for dye uptake was ascertained to be 127.71 mg g^{-1} along with the good reutilization capacity even after five consecutive cycles. These results demonstrate that EG functionalized Zn–Cu–Ni ternary metal hydroxide may be effectively used as an adsorbent for the separation of dye effluents from aqueous medium.

✉ Poonam Singh
poonam@dtu.ac.in

[1] Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India

Published online: 29 October 2022

Graphical Abstract



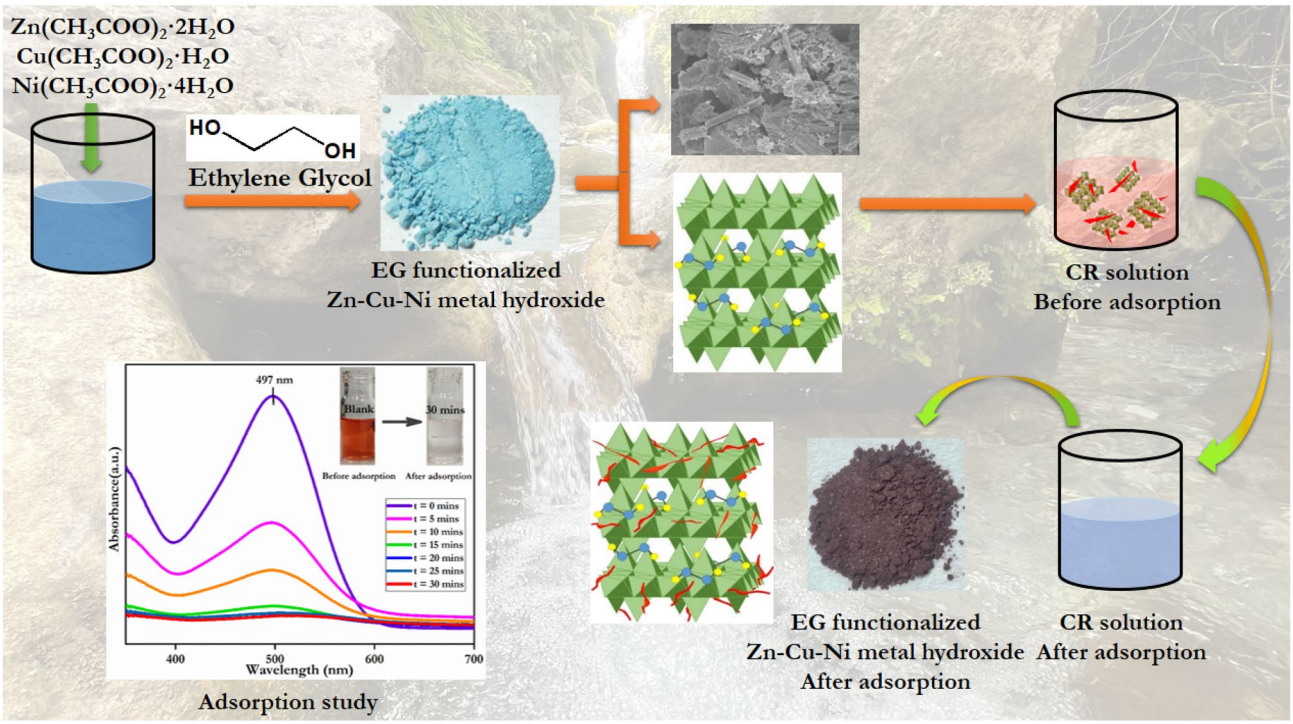
Keywords Ternary metal hydroxide · Ethylene glycol · Composite · Adsorptive performance · Congo red · Water treatment

Introduction

In last few decades, rapid globalization has led to increased water pollution levels contributing immensely to the surmounting problem of potable water availability. According to available global trends, nearly 700,000 tonnes of various dyes and pigments are utilized in the cosmetics, dyeing, food processing, paper and textile, and leather industries, out of which, nearly 5–10% of toxic pollutants are released into water bodies without purification or chemical treatment [7]. The extensive release of such highly stable, non-biodegradable, and toxic effluents puts both aquatic and human health in jeopardy as these pollutants exhibit carcinogenic, mutagenic and teratogenic effects having detrimental impact on the aesthetic appeal of sources of water [37]. Thus, in an attempt to avoid these ecological hazards, industrial wastewater must be absolved of these toxic pollutants before being discharged into water bodies. Although, various traditional water purification technologies like ion exchange, flocculation, photocatalysis, membrane filtration, biological processes, etc. involving the use of physiochemical and biological mechanistics exist but limiting factors such as low efficiency towards various pollutants and high operational costs hinder their employability at a large-scale.

However, during recent years, adsorption has emerged as a promising method for toxic waste removal from waste water since it offers several advantages over other traditional methods including high efficiency, cost-effectiveness, simple set-up and easy scalability to field scale applications, reduced amount of chemical and biological waste production along with the benefit of adsorbent regeneration. Typically, conventional materials used for adsorptive removal of toxic waste include pristine and modified activated carbon, metal oxides, biosorbents, silicon-based materials, zeolites, polymers, and natural clays, etc. [25, 27, 36]. But their profitable usage is limited due to their high costs, poor adsorption capacities and difficulty in product reuse and regeneration.

In addition, metal hydroxides, layered double hydroxides (LDH) and hydroxy double salts (HDS) have become focal points of attention for usage as versatile adsorbents by virtue of their large specific surface areas, significant structural and thermal stability, high porosity, superb anion exchange ability, availability of multiple binding sites, easy synthesis, and biocompatibility. Their layered structures possess two dimensional network comprising of positively charged layers and commutable anionic species present in between layers, offering enhanced adsorptive



Scheme 1 Schematic representation of synthesis of EG functionalized Zn–Cu–Ni ternary metal hydroxide composite and uptake of CR dye

character to these materials. The nature and composition of these materials offer a wide range of modification and tuning options due to their intralayer and interlayer compositions, resulting in the advancement of properties. But still the use of metal oxides and hydroxides in adsorption applications is limited since they are susceptible to aggregation, which tends to block active binding sites, thereby reducing adsorption efficiency [29]. Hence, there is immense need for further research to tailor eco-friendly, reusable, and effective adsorbents to remove toxic dyes such as amaranth, crystal violet, alizarin red, celestine blue, etc. from aqueous systems [1, 8].

Nevertheless, functionalization of metal hydroxides with organic moieties such as carboxylate anions, polyols, surfactants, and polymers increases the interlayer spacing and lipophilicity, and leads to improved adsorptive capacities, thereby improving their hydrophobic interactions and retention affinity for organic molecules [4, 23, 26]. These moieties may be adhered to the surface by interaction with surface hydroxyl groups or by intercalation in the interlayer region by exchange with mobile anions. In literature, incorporation of many inorganic, organic, and bio-organic molecules in metal hydroxides for adsorption applications has been extensively reported. Liu et al. prepared nickel doped layered cobalt hydroxide nanocones intercalated with sodium dodecyl sulphate (SDS) [22]. Zhu et al. fabricated a Fe–Mn–La metal composite and studied fluoride ion uptake [39]. Xiang et al. reported synthesis of composite containing metal ions i.e. Ca, Al, La and further utilised it for the fluoride ion removal [35]. But still, the role of ternary metal hydroxide materials with Zn, Cu and Ni metal ions as adsorbent has not been explored.

Therefore keeping this in view efforts have been made to develop EG functionalized Zn–Cu–Ni ternary metal hydroxide composite. In addition the role of composite material as a highly efficient adsorbent was systematically investigated for CR sequestration by varying controlling parameters i.e. contact time, concentration, pH. Similar trials were also performed for methyl orange (MO) and rhodamine B (RhB); however, dye removal efficiency was found out be maximum for CR. The regeneration and reutilisation capability of synthesized adsorbent was also assessed. A detailed study of the adsorption kinetics and isotherm mathematical models was carried out to enhance information regarding the mechanistic pathway for adsorption of CR onto EG functionalized Zn–Cu–Ni ternary metal hydroxide for utilization in future. Schematic representation of this work is presented in Scheme 1.

Experimental Section

Materials

Reagents procured for EG functionalized Zn–Cu–Ni ternary metal hydroxide synthesis were used as obtained. Zinc

acetate dihydrate (98.0% purity), copper acetate monohydrate (98.0% purity) and Ethylene Glycol [$(CH_2OH)_2$] (< 99.0% purity) were purchased from MERCK. Nickel acetate tetrahydrate (98.00% purity) was obtained from CDH chemicals. CR, MO and RhB dyes used for adsorption study were procured from S.D. Fine chemicals, Sigma Aldrich and Loba Chemie respectively. All experiments were performed using deionized water.

Synthesis of EG Functionalized Zn–Cu–Ni Ternary Metal Hydroxide Composite

First, 1.2 g $Zn(CH_3COO)_2{\cdot}2H_2O$, 0.2 g $Cu(CH_3COO)_2{\cdot}H_2O$, and 0.2 g $Ni(CH_3COO)_2{\cdot}4H_2O$ were weighed and transferred to a beaker having 100 mL solution of ethylene glycol and deionised water (1:1). Later, the resultant mixture was ultrasonicated for 30 min. Finally, the resultant solution was placed in oven at 120 °C till the total volume of solution reduced to approximately 10 mL. The blue colored precipitate obtained were treated using deionised water and ethyl alcohol and finally dried at 60 °C in oven.

Characterization Techniques

PXRD pattern was collected by a high-resolution X-ray diffractometer model Bruker D8 advanced, with Cu Kα radiation (λ = 1.5418 Å) over the angle span of $2\theta = 5–70°$ using a gobel mirror. The data was recorded with a step size 0.02° and 1.0 s/step scan rate at 298 K. The FTIR spectrum was recorded using PerkinElmer Fourier-transform infrared spectrometer (version 10.5.3) with the KBr technique. Thermogravimetric analysis curve was recorded using the PerkinElmer TGA under flowing nitrogen in temperature range 50–900 °C with 10 °C min^{-1} uniform heating rate. SEM and EDX micrographs were captured with a Hitachi S-3700 M microscope to study morphology of composite surface. The XPS measurements were recorded on a PHI 5000 Versa Probe III with Al Kα for photoelectron excitation. The specific surface area (S_{BET}) and average pore size distribution of composite was recorded by BET-BJH (Barrett-Joyner-Halenda) nitrogen adsorption–desorption technique using Quanta chrome Nova-1000 instrument at liquid nitrogen temperature 77 k.

Adsorption Experiment

The batch equilibrium testing procedure was applied to study the adsorptive removal of dyes by EG functionalized Zn–Cu–Ni ternary metal hydroxide. The influence of controlling parameters was determined at ambient temperature. 100 mg L^{-1} CR stock was used for preparation of successive CR solutions. Experiments were conducted in sets of three and CR concentration was ascertained using WinCary UV/VIS spectrophotometer at CR $\lambda_{max} = 497$ nm. The pH of CR solutions was regulated using 0.1 M HCl and/or 0.1 M NaOH solutions. Adsorption isotherms were studied by adding 0.01 g adsorbent in 20 mL CR solution with concentrations over span of 20 to 100 mg L^{-1}. The dispersions were centrifuged at 3000 rpm after stirring in dark at room temperature. Equilibrium dye uptake capacity, q_e (mg g^{-1}) was computed by:

$$q_e = V \times \frac{(C_i - C_e)}{m} \tag{1}$$

where the initial and equilibrium CR concentrations are represented using C_i and C_e (mg L^{-1}), adsorbent mass (mg) using m while V represents volume of dye solution (mL).

Removal efficiency for CR was assessed using the expression:

$$RemovalEfficiency(\%) = \frac{(C_i - C_e)}{C_i} \times 100 \tag{2}$$

For the adsorption kinetics experiment, first 0.05 g adsorbent was dispersed in 100 mL of CR solution (20 mg L^{-1}) and stirred for a duration of 30 min in dark at room temperature. After pre-fixed time durations, 4–5 mL aliquots were retrieved and solid adsorbent was isolated from supernatant solution by centrifugation at 3000 rpm. Dye uptake capacity was calculated at time t using the expression:

$$q_t = V \times \frac{(C_i - C_t)}{m} \tag{3}$$

where q_t represents dye adsorbed at time t in mg g^{-1} while C_i and C_t (mg L^{-1}) represent the dye concentration at initial and final state *t* in mg L^{-1}, m and V represent mass of adsorbent (mg) and volume of solution (L) respectively. Removal efficacy of the adsorbent was computed using expression (2). Similar adsorption experiments were conducted for MO and RhB, and removal efficiency was computed. Solutions were used without pH adjustment unless otherwise mentioned.

Results and Discussion

Characterization of EG Functionalized Zn–Cu–Ni Ternary Metal Hydroxide

The PXRD pattern of EG functionalized Zn–Cu–Ni ternary metal hydroxide is presented in Fig. 1a, exhibiting prominent and intense reflection at lower 2θ angle (10.6°) and less sharp, asymmetric reflections at higher 2θ angles (21°–57°) typical of turbostratic materials [15]. The reflections appearing at 10.6°, 21.3° and 33.5° were assigned corresponding *hkl* values (003), (006) and (012), considering rhombohedral

symmetry having lattice parameters *a* and *c* equal to 3.14 Å and 25.04 Å respectively. These parameter values are concordant with those reported in previously published literature suggesting the successful formation of EG functionalized Zn–Cu–Ni ternary metal hydroxide having layered arrangement [12]. The *d*-value for *003* basal reflection was found to be 8.36 Å, and the gallery height, calculated to be 3.56 Å, is almost equal to the molecular diameter of the isolated EG molecule (4.2 Å), indicating the existence of EG molecules in a tilted monolayer arrangement in the composite lattice [34]. The PXRD pattern thus confirms the successful formation of EG functionalized Zn–Cu–Ni ternary metal hydroxide having layered arrangement.

Thermal stability of the sample analysed using TGA is presented in Fig. 1b and is indicative of the trends in the thermal behaviour of EG functionalized Zn-Cu-Ni ternary metal hydroxide. The TGA profile consists of three step mass loss. The preliminary stage involves a weight loss of 8.34% below 104 °C on account of the elimination of volatile water molecules physisorbed on composite surface followed by the dissipation of intercalated water molecules along with superficially adsorbed EG molecules below 260 °C corresponding to a weight loss of 10.44%. The rapid weight loss of nearly 19.46% in between 260 °C and 414 °C occurs due to combustion of the EG molecules intercalated between the layers of Zn–Cu–Ni ternary metal hydroxide. The complete dehydroxylation and thermally induced oxidative decomposition of EG functionalized Zn–Cu–Ni ternary metal hydroxide into their relative oxides occurs above 414 °C [24].

The formation of EG functionalized Zn–Cu–Ni ternary metal hydroxide was further confirmed using FTIR, presented in Fig. 1c. The broad and intense nature of band at 3428 cm^{-1} correspond to hydroxyl group ($\nu_{O\text{-}H}$) stretching of Zn–Cu–Ni ternary metal hydroxide and EG that are hydrogen bonded. The shift of O–H stretching vibration to 3428 cm^{-1} in EG functionalized Zn–Cu–Ni ternary metal hydroxide from 3310 cm^{-1} in pristine EG is a consequence of the organo-functionalization of Zn–Cu–Ni ternary metal hydroxide with ethylene glycol leading to interaction of EG molecules with the layers of synthesized lattice. Fingerprint bands ascribed to antisymmetric and symmetric stretching of $–CH_2$ occurs at 2881 cm^{-1} and 2836 cm^{-1} while the existence of bands at 1573 cm^{-1} and 1411 cm^{-1} indicate scissoring together with wagging & twisting modes of $–CH_2$ deformation [31]. Existence of ethyleneoxy ($O\text{-}CH_2\text{-}CH_2\text{-}O$) units in trans-conformation in sample were corroborated by band at 1342 cm^{-1}. In addition, the appearance of moderately intense bands at 1070 cm^{-1} and 907 cm^{-1} along with a weak shoulder at 868 cm^{-1} are designated to in-plane C–C–O/C–O–H bending and rocking vibrations of CH_2 [30]. The C–H stretching vibrations show a red shift along with weakening and

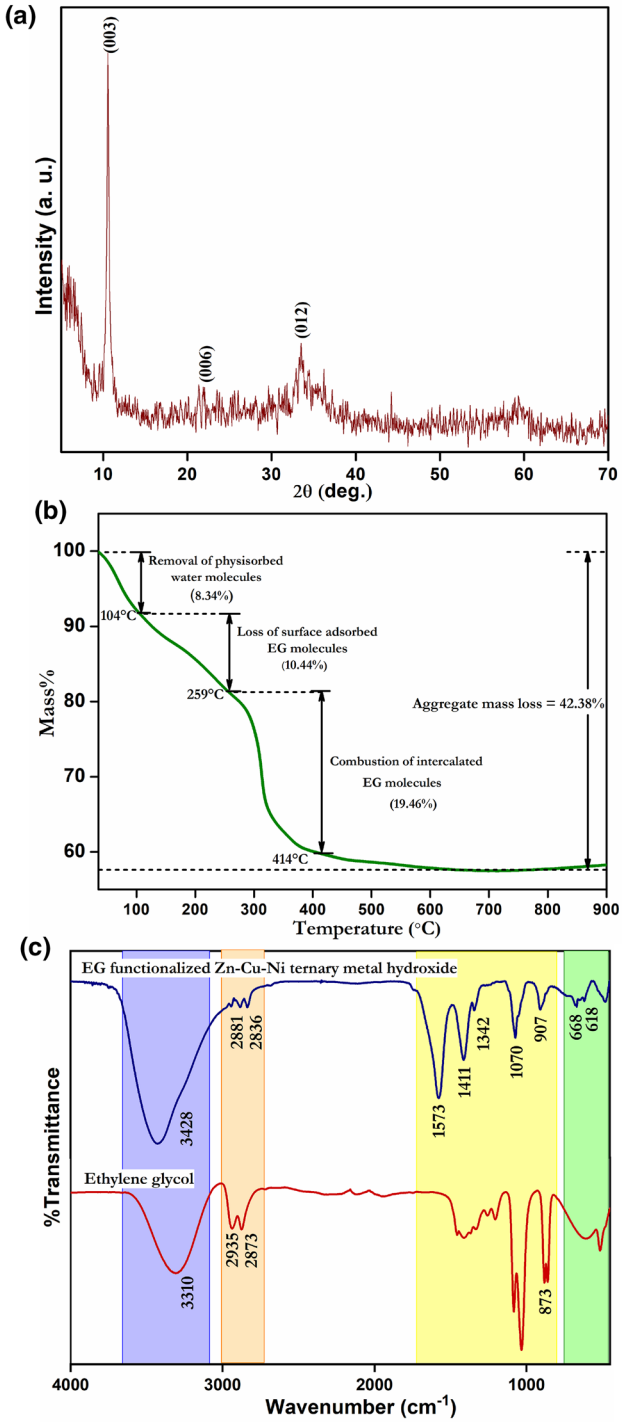


Fig. 1 **a** PXRD pattern **b** Thermogravimetric trace and (**c**) FTIR spectra of EG functionalized Zn–Cu–Ni ternary metal hydroxide and pristine EG

broadening of the band, while the CH_2 moiety rocking vibrations exhibit a blue shift as compared to pristine EG (2935, 2873, 873 and 859 cm^{-1}), occurring due to change in environment of EG molecules on grafting in the Zn–Cu–Ni ternary metal hydroxide lattice. The bands at

668 cm^{-1} and 618 cm^{-1} confirm metal–oxygen linkages (δ_{M-O-H}, ν_{M-O}) present in the lattice.

SEM imaging was done to identify the morphological features, as presented in Fig. 2a and it showed the scattered plate-like morphology in synthesized composite, typical of layered metal hydroxides [9]. This observation corroborates the formation of layered organofunctionalized Zn-Cu-Ni ternary metal hydroxide. EDX analysis and elemental mapping (Fig. 2a, b) were further conducted to determine the constituent elements present in EG functionalized Zn-Cu-Ni ternary metal hydroxide and it confirmed the homogenous distribution of carbon (C), oxygen (O), Zn, Cu and Ni elements throughout the sample.

The elemental composition on surface and the electronic states of the constituents of EG functionalized Zn–Cu–Ni ternary metal hydroxide were examined by XPS study (Fig. 3a–f) and reveals that peaks for C, O, Zn, Cu and Ni appear in survey spectrum of synthesized composite. The C *1 s* spectrum consists of dual peaks fitted at 284.8 eV and 287.9 eV that are associated with the C–C skeletal structure of intercalated ethylene glycol molecule and the C–O bonds [32]. O *1 s* core level spectrum was studied by best fitting the peak positioned at 531.0 eV and deconvolution to peaks at 529.5, 530.5 and 531.5 eV indicated the presence of metal-O, metal-OH linkages and adsorbed water molecules respectively [33]. XPS core spectrum of Zn *2p* consists of 2 fingerprint peaks, positioned at 1021.2 eV and 1044.3 eV, which can be attributed to the Zn^{2+} ion respectively [28]. The Cu *2p* core level HRXPS spectrum upon deconvolution exhibits 2 peaks positioned at 932.3 eV and 952.1 eV characteristic of Cu^{2+} ion [38]. The HRXPS of Ni *2p* was deconvoluted and it exhibited peaks at 855.8 eV and 873.2 eV without any satellites, indicating the presence of Ni^{2+} ion [17].

BET technique was applied to obtain N_2 adsorption/desorption isotherm and ascertain the specific surface area (S_{BET}) and pore size distribution. The S_{BET} value, calculated from BET isotherm, was 47.44 $m^2\ g^{-1}$ while the BJH (Barret-Joyner-Halenda) pore volume and pore diameter were computed to be 0.157 cc g^{-1} and 15.08 nm g^{-1} respectively for the synthesized lattice. The resemblance to IUPAC classification-Type IV isotherm (Fig. 4) and the pore diameter lying in the range of 2–50 nm indicated presence of aggregated mesoporous platelet like-particles in a layered arrangement in the synthesized composite [14]. The preliminary portion of the type IV- isotherm can be correlated with monolayer adsorption, and concurs with the Langmuir isotherm model. Moreover, existence of H3 hysteresis loop points to the intrinsic presence of disordered slit-like pores due to stacking of topotactic layers of composite, and validates the involvement of capillary condensation mechanism in their role as adsorbent for sequestration of organic dyes [10, 21]. Therefore, mesoporosity and considerably high S_{BET} provide multiple active adsorption sites, thereby favouring the faster diffusion of dye adsorbate into the porous composite lattice and therefore lend good adsorptive character to EG functionalized Zn-Cu-Ni ternary metal hydroxide.

Adsorption Behaviour

Impact of Contact Time

A key parameter in adsorption studies between adsorbent and dye solution is contact time, as it imparts information about the equilibrium state. The impact of the same on the dye uptake behaviour for CR using EG functionalized Zn–Cu–Ni ternary metal hydroxide as adsorbent is displayed in Fig. 5. Adsorbent demonstrated rapid uptake of CR in the initial 15 min of study with a removal efficiency of 90.3% that gradually increased to 93.3% after 20 min. Removal efficiency did not show any significant increase above 94.6% after 30 min, implying the equilibrium state was attained. Rapid preliminary adsorption may be a consequence of the physisorption of CR on the composite surface since initially a substantial amount of vacant adsorption sites are available to interact with the CR molecules. But with extended contact time, the availability of vacant adsorption sites reduces, leading to decreased dye uptake capacity, before ultimately plateauing to equilibrium state.

Impact of Dye Concentration

Adsorbate solution concentration also influences the adsorption behaviour of the adsorbent in a significant manner, and here it shows an upward trend, as presented in Fig. 6a. The curve elucidates that dye uptake capacity of the synthesized composite augmented from 38.93 mg g^{-1} to 125.51 mg g^{-1} with elevation in CR concentrations from 20 to 100 mg L^{-1}. This trend can be attributed to the increased molecular collisions in solution phase that are a consequence of the increased dye concentration. The elevated concentration gradient helps in surpassing the hindrance to transfer of mass of CR molecules between the adsorbent and the supernatant, leading to enhanced adsorption capacity of the synthesized composite [18].

Impact of pH

The pH of a solution is a significant operational variable governing the interactions at the interface of adsorbent and adsorbate by altering the charge on adsorbent surface as well as the ionizing power of CR molecules. For the synthesized lattice, uptake of CR dye by EG functionalized Zn–Cu–Ni ternary metal hydroxide was studied for a series of solutions having pH within the range of 3–11 (Fig. 6b). On raising the pH from 3 to 7, percentage dye removal of CR dye elevated slightly from 97 to 99%. At lower pH values, the majority

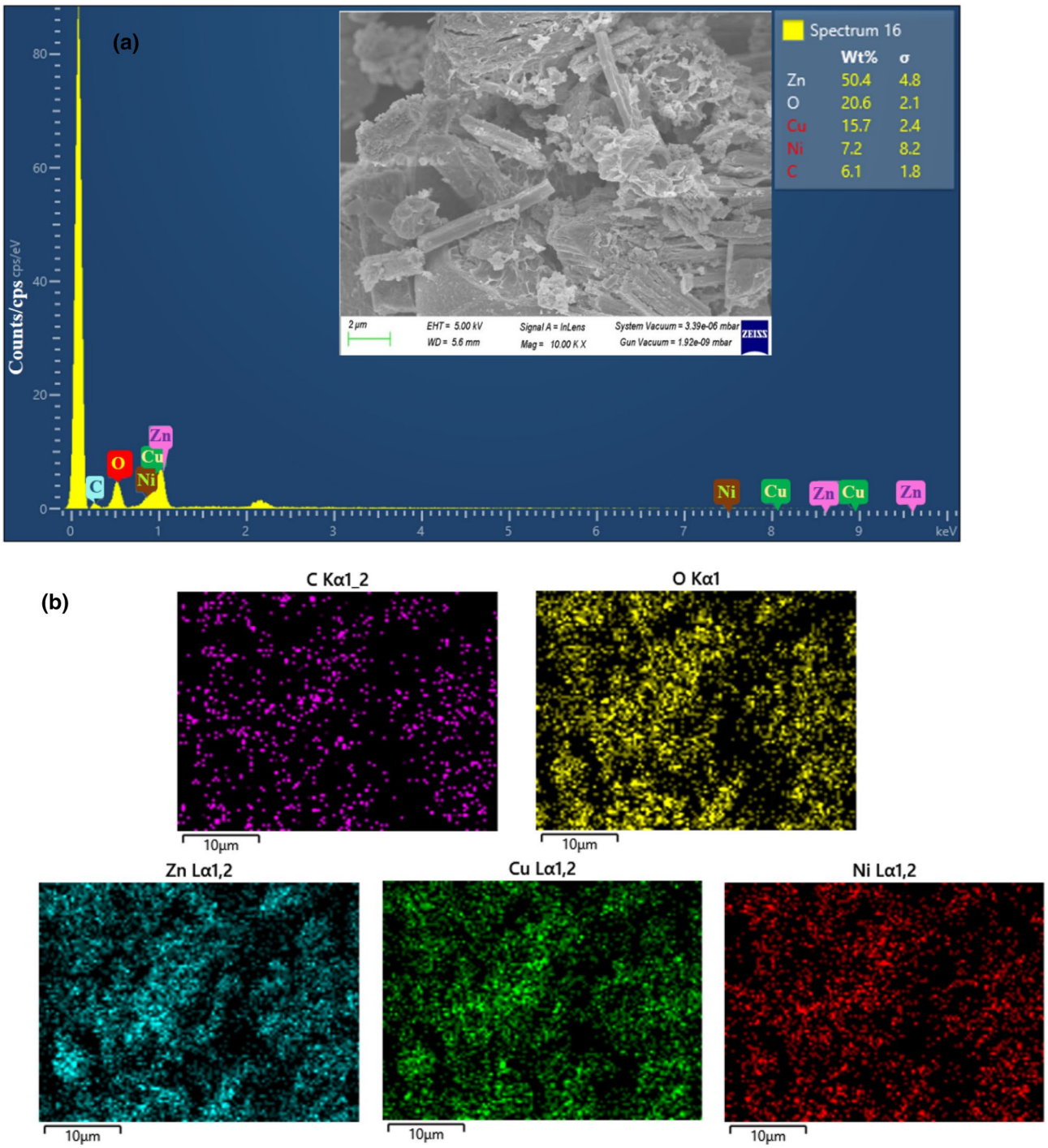


Fig. 2 **a** EDX with inset SEM micrograph and **b** elemental mapping of EG functionalized Zn–Cu–Ni ternary metal hydroxide

of CR molecules exist in their anionic form (CR^-) in the aqueous medium (pKa = 3.3), and the strong electrostatic attraction of SO_3^- group of CR molecules towards the protonated metal hydroxide surface leads to high adsorption capacity and good buffering ability in the pH range 3–7, rendering it a suitable candidate for use in real-life applications. Additional increase in pH to 9 leads to a slight decrease in removal efficiency while solution pH beyond 9 causes a drastic reduction in removal efficacy for dye, that may be a consequence of: (i) increased contention for unoccupied active sites between the CR molecules and OH^- molecules in basic solution or (ii) repulsive action of negatively charged metal hydroxide surface towards the anionic CR moieties present in solution. pHzc. The point of zero charge

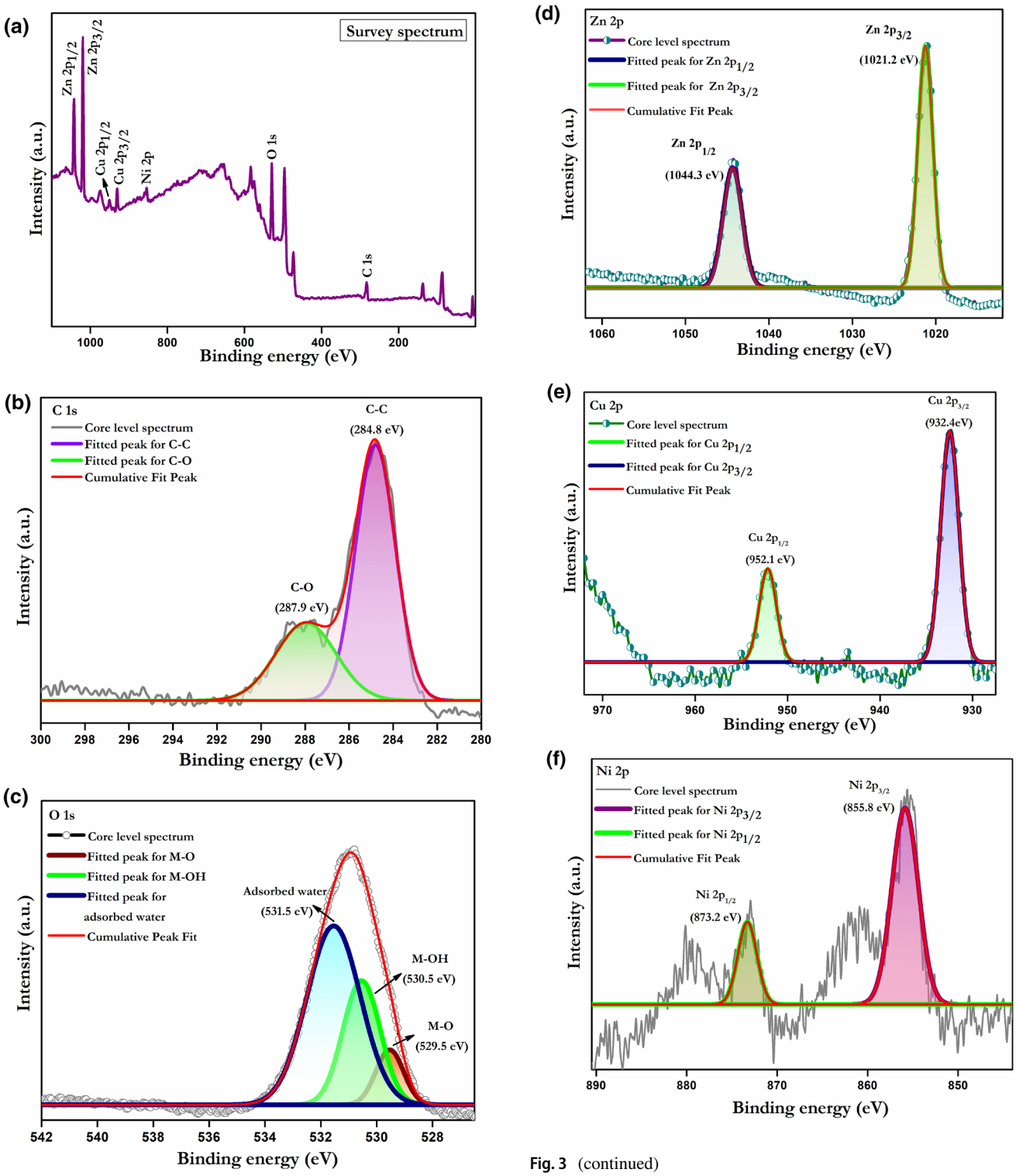


Fig. 3 (continued)

Fig. 3 XPS **a** survey spectrum and analysis of core level spectrums of **b** C *1 s*, **c** O *1 s*, **d** Zn *2p*, **e** Cu *2p* and **f** Ni *2p*

(pHzc) was analysed to determine the surface charge of synthesized adsorbent using the salt addition method. The procedure was as follows: 0.01 g adsorbent was added to a series of 20 mL 0.1 M NaCl solutions with initial pH (pH_i) value adjusted to 3, 5, 7, 9 and 11 respectively. The resulting suspensions were vigorously agitated for 24 h at room temperature. After agitation, the final pH (pH_f) value of the suspensions was measured using a pH meter and ΔpH was calculated using the expression $\Delta pH = pH_f - pH_i$. The pHzc

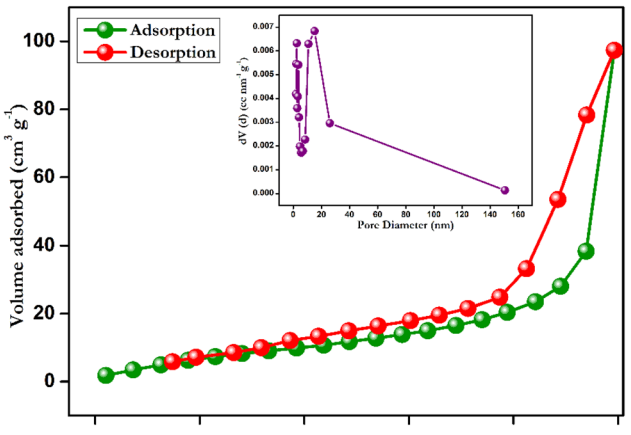


Fig. 4 N_2 adsorption–desorption isotherm with inset of pore size distribution plot from desorption isotherm for EG functionalized Zn-Cu-Ni ternary metal hydroxide

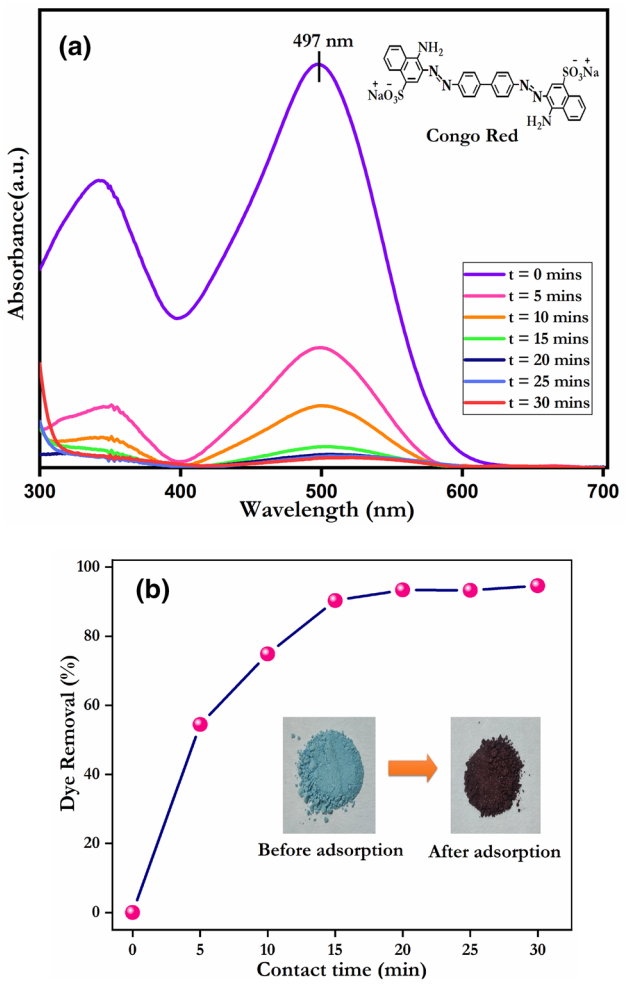


Fig. 5 **a** UV–Visible adsorption spectra and **b** Plot of removal efficiency v/s contact time for CR adsorption using EG functionalized Zn-Cu-Ni ternary metal hydroxide

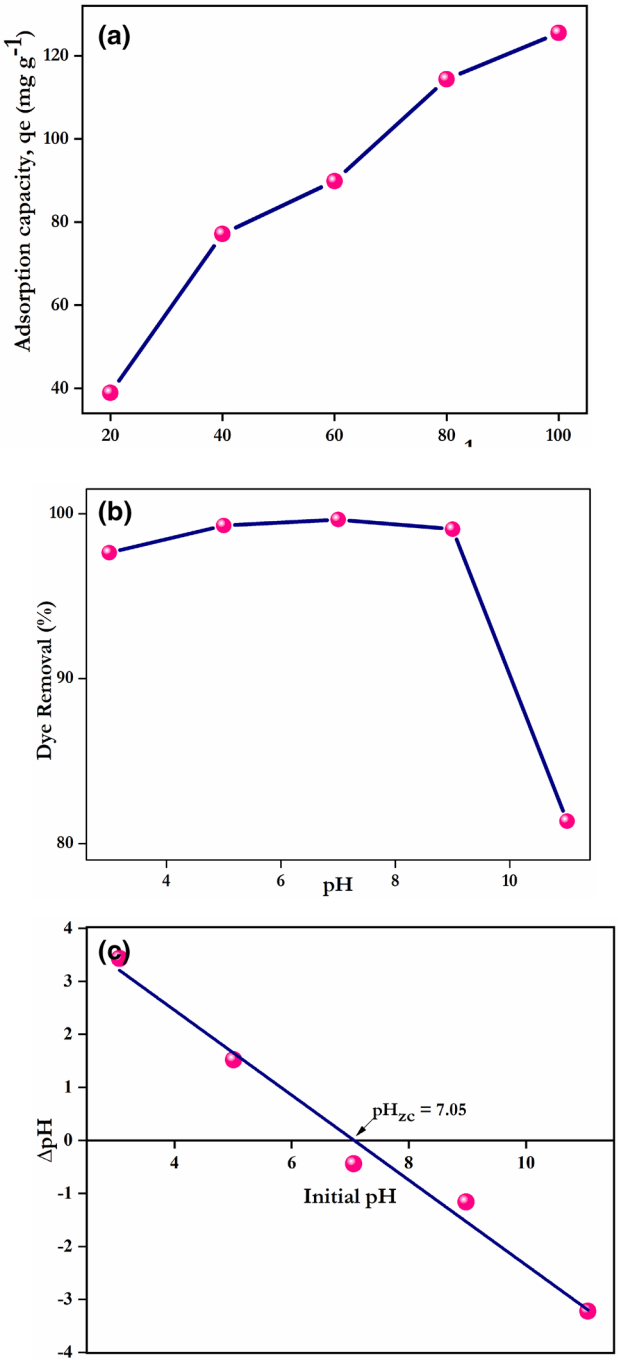


Fig. 6 **a** Impact of dye concentration on adsorption capacity (q_e) **b** Impact of pH on dye removal percentage of CR dye and **c** plot of pHzc for EG functionalized Zn–Cu–Ni ternary metal hydroxide as adsorbent

value was determined from the plot between ΔpH and pH_i at the point where $pH_f = pH_i$ presented in Fig. 6c. The pHzc value of EG functionalized Zn-Cu-Ni ternary metal hydroxide adsorbent was found to be 7.05, implying that the adsorbent surface was protonated at pH < 7.05, that consequently

promotes electrostatic attraction with the anionic CR molecules, enhancing their adsorption on the adsorbent surface. Conversely, when pH > 7.05, adsorbent surface is negatively charged, and thus electrostatically repels the CR anions, leading to reduced adsorption capability at high pH values.

Adsorption Isotherms

Equilibrium studies and isotherm modelling impart significant knowledge about the interaction mechanisms at the juncture of adsorbent-adsorbate phases. The experimental data was analysed using linear regressions of various mathematical models [5, 6, 11, 20] to compute the maximum dye uptake capacity and deduce the adsorption nature.

(i) Langmuir Adsorption Isotherm:

The model proposed by Langmuir attempts to explain the saturated adsorbate uptake by adsorbent in monolayer fashion with uniform binding sites and is empirically described as [20]:

$$q_e = \frac{\left(q_{max} C_e K_L\right)}{\left(1 + K_L C_e\right)} \tag{4}$$

The linear regression of Langmuir model is:

$$\frac{C_e}{q_e} = \frac{1}{(K_L q_{\max})} + \frac{C_e}{q_{max}} \tag{5}$$

where saturation adsorption capacity, maximum CR uptake ability and equilibrium CR concentration are represented by q_e (mg g^{-1}), q_{max} (mg g^{-1}) and C_e (mg L^{-1}) and Langmuir constant given by K_L (L mg^{-1}).

The separation factor R_L, a dimensionless constant, is used to determine linearity (R_L=1), irreversibility (R_L=0), favourability (0<R_L<1) or unfavourability (R_L>1) of adsorption and is calculated from formula:

$$R_L = \frac{1}{1 + K_L C_i} \tag{6}$$

where CR solution concentration is given in mg L^{-1} by C_i.

(ii) Freundlich Adsorption Isotherm

The Freundlich model describes the unideal adsorbate uptake on energetically heterogeneous binding sites in multilayer arrangement on adsorbent surfaces in low concentration systems [6]. The empirical and logarithmic linearized forms can be stated as:

$$q_e = K_f \times C_e^{1/n_f} \tag{7}$$

$$\log q_e = \log K_f + \frac{1}{n_f} \log C_e \tag{8}$$

where adsorption capacity (mg g^{-1}) and CR equilibrium concentration (mg L^{-1}) are represented as q_e and C_e. Adsorbate uptake capacity and heterogenic adsorption intensity are indicated by Freundlich coefficients K_f [(mg g^{-1}) (L $mg^{-1})^{1/n_f}$] and n_f.

Value of $1/n_f$ factor is a useful parameter in determining if the adsorption process is favourable ($0 < 1/n_f < 1$), unfavourable ($1/n_f = 1$) or irreversible ($1/n_f = 0$).

(iii) Dubinin-Radushkevich (D-R) Adsorption isotherm:

The D-R model explains the bonding adsorbent-adsorbate system interactions and is formulated on the theory of Polanyi potential [5]. The Gaussian distribution of average free energy relays information about nature of adsorption and involved mechanistic pathways. The non-linear and linear regressions are:

$$q_e = q_s e^{(-\beta\varepsilon^2)} \tag{9}$$

$$\ln q_e = \ln q_s - \beta\varepsilon^2 \tag{10}$$

where equilibrium and theoretical saturation capacity (mg g^{-1}) are given by q_e and q_s while is the activity coefficient for average free energy of adsorption (mol^2 kJ^2) is β. ε, the Polanyi potential, is described by:

$$\varepsilon = RT\ln\left(1 + \frac{1}{C_e}\right) \tag{11}$$

where universal gas constant (8.314 J K^{-1} mol^{-1}), temperature (K) and CR equilibrium concentration (mg L^{-1}) are presented by R, T and C_e respectively.

The average free energy E provides information about adsorption classification and is quantified by equation (12):

$$E = \frac{1}{\sqrt{2\beta}} \tag{12}$$

When E>8 kJ mol^{-1} it implies physisorption and when E<8 kJ mol^{-1}, it suggests involvement of chemical adsorption.

(iv) Harkins–Jura (H–J) adsorption isotherm

The H-J model is adopted for adsorption systems with multilayer adsorption and for adsorbents with non-uniform pore distribution on surface [11, 16]. The non-linear and linear regressions are:

$$q_e = \left(\frac{A}{B - \log C_e}\right)^{1/2} \quad (13)$$

$$\frac{1}{q_e^2} = \left[\frac{B}{A}\right] - \left[\frac{1}{A}\right] \log C_e \quad (14)$$

where equilibrium dye uptake capacity is q_e (mg g^{-1}), and B and A are Harkins-Jura constants while equilibrium CR concentration is given by C_e (mg L^{-1}).

Isotherm plots of the mathematical models analysed for the uptake of CR on EG functionalized Zn–Cu–Ni ternary metal hydroxide are presented in Fig. 7 and calculated values are presented in Table 1. The adsorptive removal of CR by EG functionalized Zn–Cu–Ni ternary metal hydroxide was found to be favourable according to Langmuir model as the correlation of experimental data was better with Langmuir isotherm model ($R^2 = 0.97$), giving evidence for homogenous monolayer adsorption. The $R_L = 0.106$ and high q_{max} value in Langmuir model (127.71 mg g^{-1}) establish the superior dye removal capability of synthesized composite owing to the immense attraction between anionic CR and EG functionalized Zn–Cu–Ni ternary metal hydroxide. Further, the lower R^2 value for Freundlich, D-R and H-J isotherm model indicates poor fit, further justifying the homogenous monolayer adsorption. Nonetheless, the value of E = 1.82 kJ mol^{-1}, calculated from Eq. 12, implied that the primary process involved in uptake of CR by EG functionalized Zn–Cu–Ni ternary metal hydroxide is physical adsorption [10].

Adsorption Kinetics

Kinetic modelling studies shed light onto the mechanistic pathways involved in adsorption process such as surface adsorption, capillary adsorption, surface and pore diffusion, and the adsorption rates in a system. Adsorption kinetics for uptake of CR on EG functionalized Zn–Cu–Ni ternary metal hydroxide were simulated by correlating the obtained experimental data to linear regressions of various kinetic models [2, 3, 13, 19].

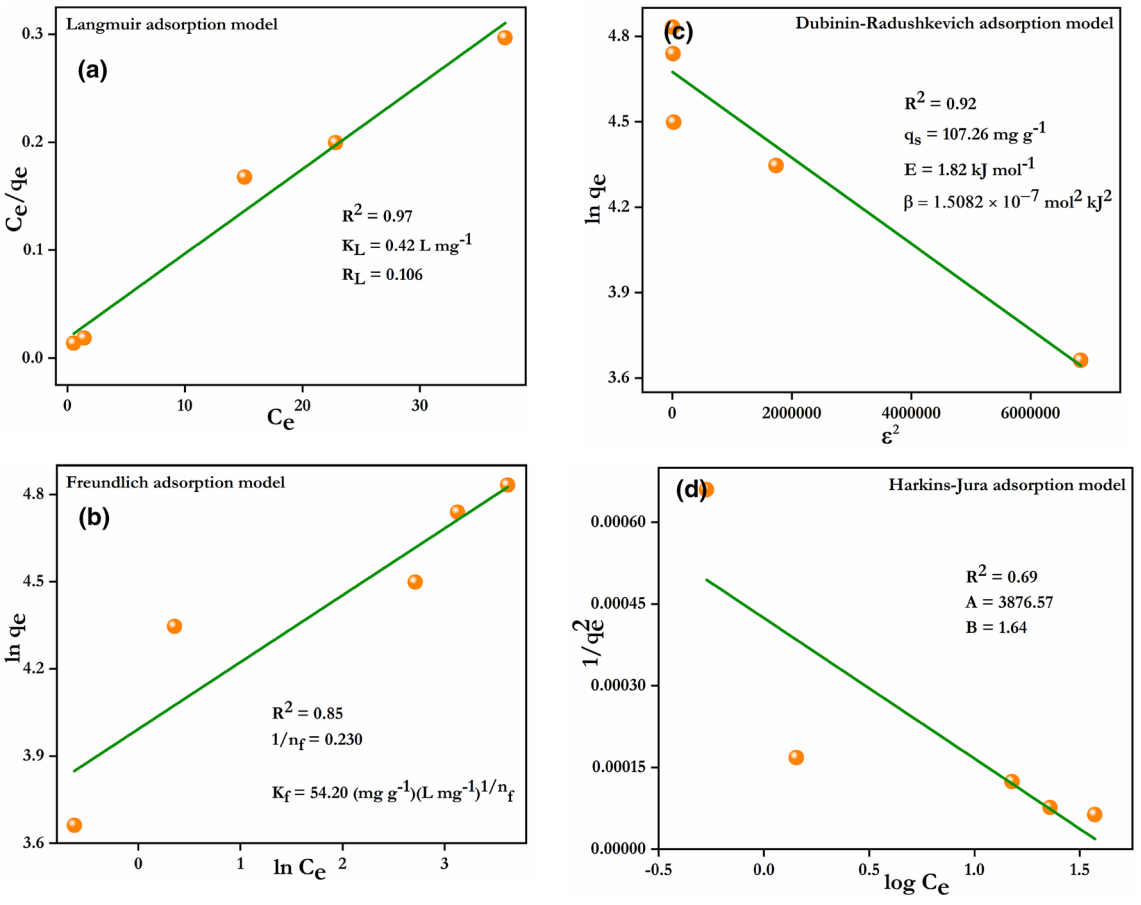


Fig. 7 Adsorption isotherm curves for adsorption of CR using EG functionalized Zn–Cu–Ni metal ternary hydroxide correlated to linear **a** Langmuir **b** Freundlich **c** Dubinin-Radushkevich and **d** Harkins–Jura model

Table 1 List of calculated adsorption isotherm parameters for various mathematical models

Model	Parameters	Values
Langmuir Model $\frac{C_e}{q_e} = \frac{1}{(K_L q_{max})} + \frac{C_e}{q_{max}}$	q_m	127.71 mg g^{-1}
	K_L	0.42 L mg^{-1}
	R_L	0.106
	R^2	0.97
Freundlich Model $\log q_e = \log K_f + \frac{1}{n_f} \log C_e$	$1/n_f$	0.23
	K_f	54.20 (mg g^{-1})(L mg^{-1})$^{1/n_f}$
	R^2	0.85
Dubinin-Radushkevich Model $\ln q_e = \ln q_s - \beta\varepsilon^2$	q_s	107.26 mg g^{-1}
	B	1.508×10^{-7} mol^2 kJ^2
	E	1.82 kJ mol^{-1}
	R^2	0.92
Harkins–Jura Model $\frac{1}{q_e^2} = \left[\frac{B}{A}\right] - \left[\frac{1}{A}\right] \log C_e$	A	3876.56
	B	1.64
	R^2	0.69

(i) The pseudo-first order model, dealing with physical adsorption, is stated in non-linear and linear expressions as:

$$q_t = q_e\left(1 - e^{-kt}\right) \tag{15}$$

$$\ln\left(q_e - q_t\right) = \ln q_e - k_1 t \tag{16}$$

where are dye uptake capacity (mg g^{-1}) of adsorbent at equilibrium and time t are represented as q_e and q_t and adsorption rate constant (min^{-1}) is given by k_1.

(ii) The pseudo-second order model explains rate equations for second order reactions and the adsorbent-adsorbate bonding, with the non-linear and linear regressions expressed as:

$$q_t = \frac{k_2 q_e^2 t}{1 + k_2 q_e t} \tag{17}$$

$$\frac{t}{q_t} = \frac{1}{k_2 q_e^2} + \frac{t}{q_e} \tag{18}$$

where dye uptake capacities of adsorbent at equilibrium and time t are represented as q_e and q_t (mg g^{-1}) , and k_2 (min^{-1}) is the rate constant.

Determination of initial adsorption rate h (mg g^{-1} min^{-1}) is done using the formula:

$$h = k_2 q_e^2 \tag{19}$$

(iii) The Elovich model, elucidating the adsorption–desorption kinetics, is given by the non-linear and linear equations:

$$q_t = \beta \ln(\alpha\beta t) \tag{20}$$

$$q_t = \frac{1}{\beta} \ln(\alpha\beta) + \frac{1}{\beta} \ln t \tag{21}$$

where is the uptake capacity of adsorbent at time t (mg g^{-1}), initial rate of adsorption (mg g^{-1} min^{-1}) and desorption constant (g mg^{-1}) are represented by q_t, α and β.

If value of α is greater than β, then chemical adsorption occurs between the adsorbate-adsorbent phases.

(iv) The Bangham model is employed to understand the diffusion activity in adsorbent pores and is expressed as:

$$\log\log\left(\frac{C_i}{C_i - q_t m)}\right) = log\left(\frac{k_o m}{2.303V}\right) + \alpha \log t \tag{22}$$

where initial adsorbate concentration (mg L^{-1}) and dye uptake capacity for dye at time t (mg g^{-1}) are C_i and q_t while m (g) is mass of adsorbent, V (L) is the volume of adsorbate solution and k_o (dm^3 g^{-1}) and α are Bangham model constants.

The kinetics plots for above models are presented in Fig. 8 and Table 2 contains parameteric values obtained from the linear regressions of employed mathematical models. Value of $R^2 = 0.99$ suggests the most effective correlation to the linear pseudo-second order model, implying that chemical adsorption was the rate-determining step in CR uptake by synthesised composite. Value of h was calculated from Eq. (19) and found to be 9.828 mg g^{-1} min^{-1}.

Further, the experimental data was fitted into the non-linear forms of adsorption models that showed best fit in the linear forms i.e. Langmuir and pseudo second order ($R^2 = 0.97$ and 0.99 respectively). The non-linear regression plots of Langmuir and pseudo 2nd order model, presented in Fig. 9, possess a Type-I convex upward isotherm shape. Correlation coefficients of linear and non-linear regressions of Langmuir and pseudo 2nd order models were found to be in good agreement.

Thus, the adsorption of CR on EG functionalized Zn–Cu–Ni ternary metal hydroxide may comprise of three steps: (i) initial physical adsorption on adsorbent surface, (ii) chemical adsorption and (iii) plausible saturation at equilibrium stage. The schematic representation of adsorption of CR on EG functionalized Zn–Cu–Ni ternary metal hydroxide is presented in Fig. 10a.

Selective Adsorption and Mechanistic Aspects

It is important to evaluate the versatility of EG functionalized Zn–Cu–Ni ternary metal hydroxide in the removal of CR over other dye pollutants. For this objective, an adsorption experiment was conducted to evaluate the selectivity of adsorbent towards CR over RhB. The adsorbent was dispersed in a mixture of 1:1 CR and RhB solutions with concentrations of 20 ppm. The visual examination of pink supernatant solution retrieved after 30 min of adsorption exhibited a clear distinction from the original pinkish-red solution (Fig. 10b), hinting to the selective uptake of CR by the synthesized material. To further ascertain the mechanistic pathway, removal efficiency of EG functionalized Zn–Cu–Ni ternary metal hydroxide towards MO and RhB was investigated in a similar manner to CR and is presented in Fig. 10c. For this study, 20 ppm MO and RhB dye solutions were prepared. Then, 0.01 g adsorbent was dispersed in 20 mL of each dye solution. The results show that EG functionalized Zn–Cu–Ni ternary metal hydroxide has good removal efficacy of 83.84% for MO and only 4.71% for RhB. Therefore, dye removal efficacy of EG functionalized Zn–Cu–Ni ternary metal hydroxide for the dye conforms to the order: CR > MO > RhB. Preferential adsorption of anionic dyes (CR, MO) over the cationic dye (RhB) may be accredited to electrostatic attractions and hydrogen bonding arising

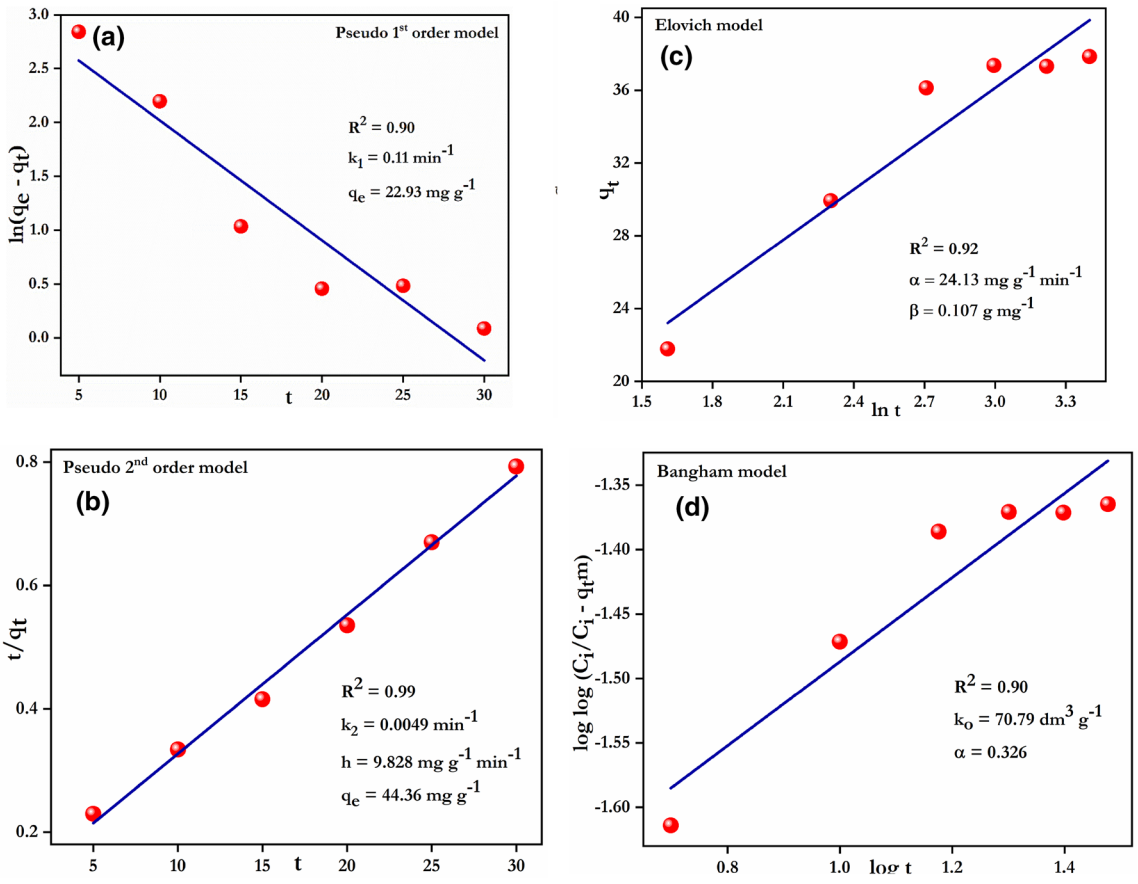


Fig. 8 Adsorption kinetics curves for CR uptake by EG functionalized Zn–Cu–Ni ternary metal hydroxide fitted to linear **a** pseudo-first order **b** pseudo-second order **c** Elovich and **d** Bangham model

Table 2 List of calculated adsorption parameters for kinetics models

Model	Parameters	Values
Pseudo 1st Order $\ln(q_e - q_t) = \ln q_e - k_1 t$	q_e	22.93 mg g^{-1}
	k_1	0.11 min^{-1}
	R^2	0.90
Pseudo 2nd Order $\frac{t}{q_t} = \frac{1}{k_2 q_e^2} + \frac{t}{q_e}$	q_e	44.36 mg g^{-1}
	k_2	0.0049 min^{-1}
	h	9.82 mg g^{-1} min^{-1}
	R^2	0.99
Elovich Model $q_t = \frac{1}{\beta}\ln(\alpha\beta) + \frac{1}{\beta}\ln t$	α	24.13 mg g^{-1} min^{-1}
	β	0.107 g mg^{-1}
	R^2	0.92
Bangham Model $\log\log\left(\frac{C_i}{C_i - q_t m)}\right) = log\left(\frac{k_o m}{2.303V}\right) + \alpha \log t$	α	0.326
	k_o	70.79 dm^3 g^{-1}
	R^2	0.90

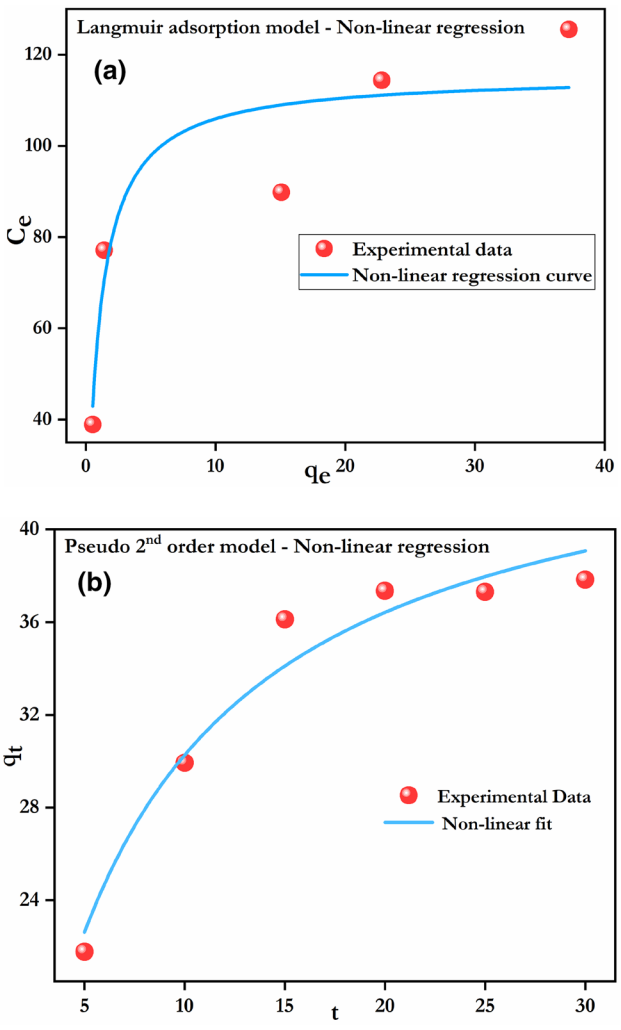


Fig. 9 Non-linear regression plot of (**a**) Langmuir and (**b**) pseudo-second order adsorption model for CR uptake by EG functionalized Zn–Cu–Ni ternary metal hydroxide

among the adsorbate-adsorbent interfaces. The electrostatic attractive forces and the hydrogen bonding between the positively charged surface of EG functionalized Zn–Cu–Ni ternary metal hydroxide and the R-SO_3^- and N–H/N–C moieties of anionic CR and MO and results in excellent adsorption capability and removal efficiency of adsorbent. Alternatively, the low removal efficiency for cationic RhB may be an outcome of the repulsion between the positive RhB–CH_3–NH^+ molecules and the positively charged adsorbent surface. Thus it was inferred that the synthesized EG functionalized Zn–Cu–Ni ternary metal hydroxide may be employed as an effective adsorbent for anionic dye uptake.

Regeneration and Reusability

Regeneration and reusability of an adsorbent are key parameters to assess its performance and cost effectiveness in real life applications. The regeneration ability was investigated using the batch experiment method. To regenerate and reuse EG functionalized Zn–Cu–Ni ternary metal hydroxide, the adsorbent was first saturated with CR dye solution for 30 min. After adsorption, the adsorbent was filtered, washed with deionised water and dried at 60 °C. Then the dried adsorbent was reused for another adsorption cycle in a similar manner. This procedure was repeated upto 5 cycles and the removal efficiency was calculated for each cycle. It was observed that in the first adsorption cycle, the removal efficiency was found to be 98.55% and it decreased to 75.43% after five cycles, as presented in Fig. 11. These results suggest the excellent regeneration and reutilization capability of EG functionalized Zn–Cu–Ni ternary metal hydroxide, thereby deeming it as an excellent prospective adsorbent for real life wastewater treatment.

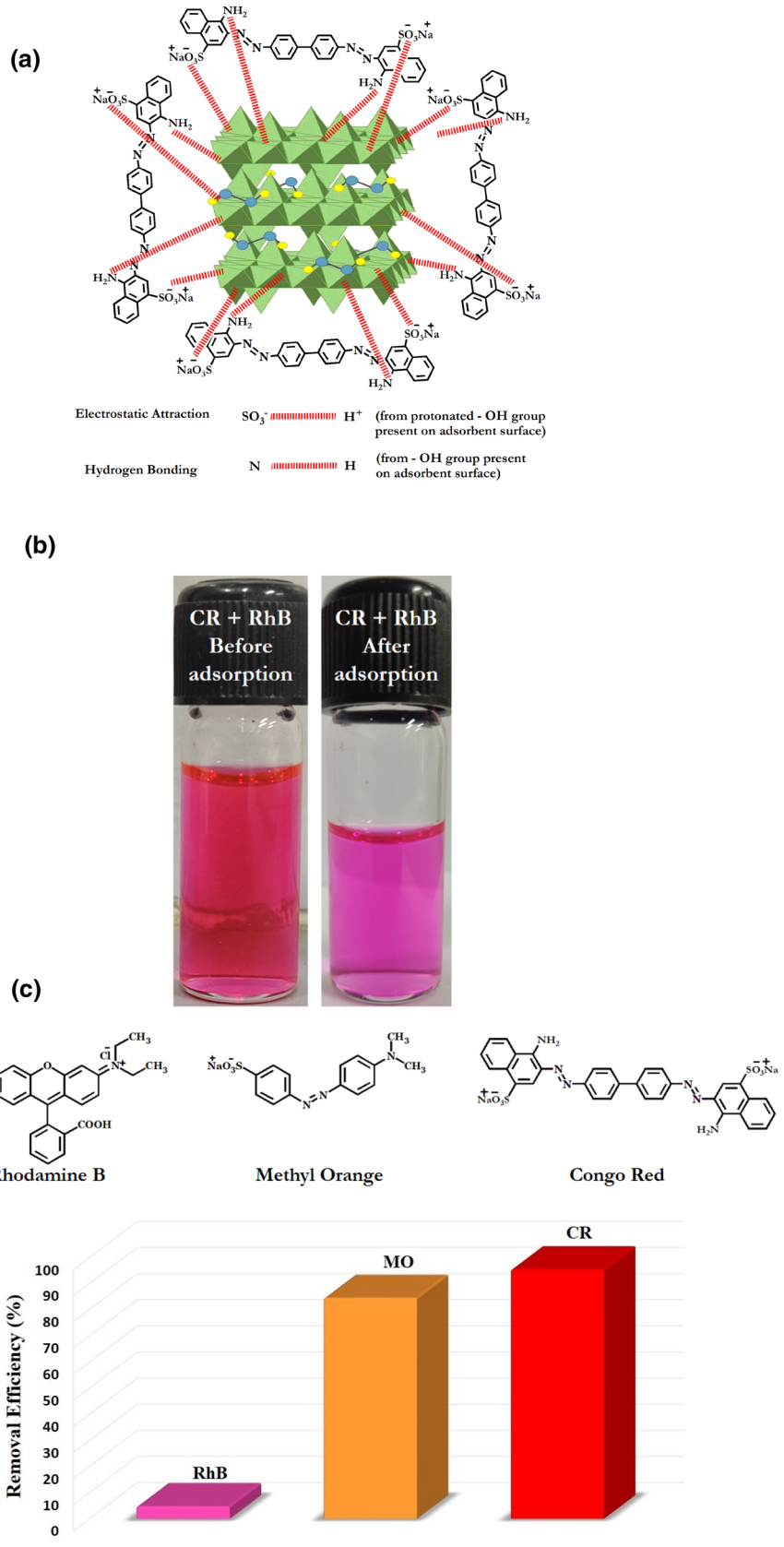


Fig. 10 **a** Schematic representation of interactions between CR molecules and EG functionalized Zn–Cu–Ni ternary metal hydroxide **b** Selective uptake of CR vs RhB and **c** Structure of CR, MO and RhB molecule and effect of adsorbate variation on removal efficiency of CR using EG functionalized Zn–Cu–Ni ternary metal hydroxide

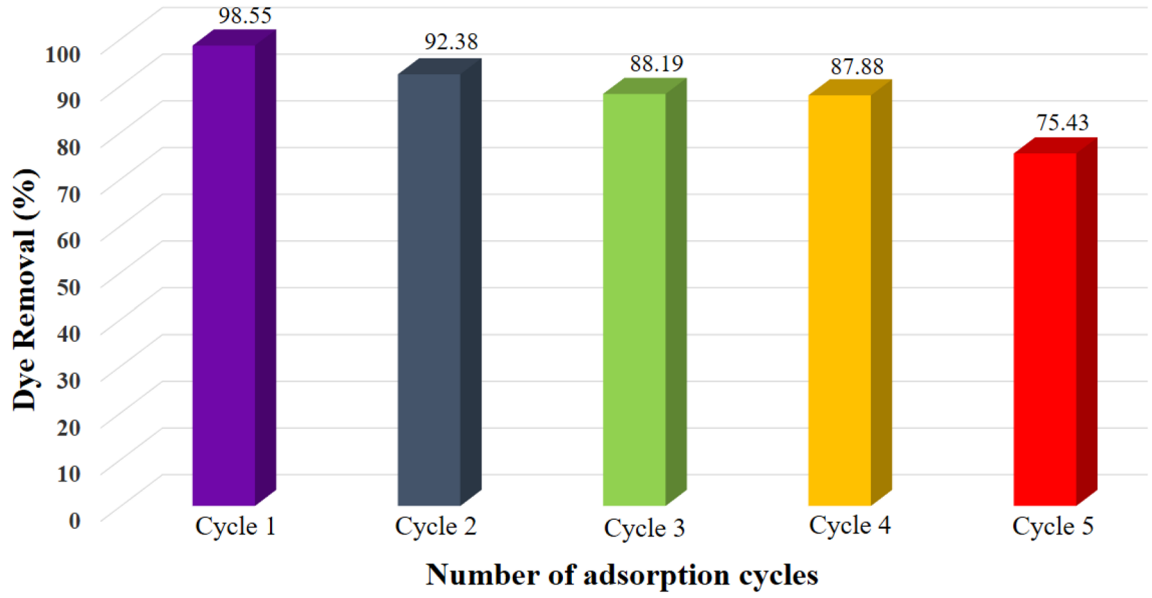


Fig. 11 Regeneration cycles of EG functionalized Zn–C Cu–Ni ternary metal hydroxide for CR dye

Conclusion

This study reports the successful fabrication of ethylene glycol functionalized Zn–Cu–Ni ternary metal hydroxide that can be used as an adsorbent for efficient CR sequestration from water system. In addition, the influence of controlling parameters to observe their effect on adsorption efficiency was also investigated. The Langmuir adsorption model correlated best with the experimental data, suggesting monolayer adsorption on homogenous vacant binding sites of adsorbent. The dye removal capacity of 127.71 mg g^{-1} was ascribed to strong electrostatic attraction and hydrogen bonds between EG functionalized Zn–Cu–Ni ternary metal hydroxide and CR molecules. The kinetics of adsorption were effectively correlated with the model of pseudo-second order, suggesting the occurrence of chemical adsorption. The synthesized adsorbent exhibited excellent dye removal efficiency even after five regeneration cycles. The above findings warrant the utilization of EG functionalized Zn–Cu–Ni ternary metal hydroxide as a feasible adsorbent material for dye pollutant removal in field-scale applications.

Acknowledgements The authors acknowledge Delhi Technological University for the immense support and resources to conduct this research. The author PS is highly thankful to UGC BSR for Research startup Grant No. F.30-416/2018 (BSR).

Authors Contributions JP: methodology, writing, review, editing. PS: conceptualization, writing, review, editing.

Funding Authors are grateful to UGC BSR Startup Grant No. F.30-416/2018 (BSR) for the funds to carry out this research.

Data Availability Data available upon request.

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.

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publication Not applicable.

References

1. Al-Salihi KJ, Alfatlawi WR (2021) Synthesis and characterization of low-cost adsorbent and used for alizarin yellow GG and alizarin red S dyes removal from aqueous solutions. IOP Conf Ser Mater Sci Eng 1094:012175. https://doi.org/10.1088/1757-899x/1094/1/012175
2. Austin Taylor H, Thon N (1952) Kinetics of Chemisorption. J Am Chem Soc 74:4169–4173. https://doi.org/10.1021/ja01136a063
3. Bangham DH (1946) The recognition of phase transitions in adsorbed films on solids. J Chem Phys 14:352. https://doi.org/10.1063/1.1724150
4. Deng L, Zeng H, Shi Z, Zhang W, Luo J (2018) Sodium dodecyl sulfate intercalated and acrylamide anchored layered double hydroxides: a multifunctional adsorbent for highly efficient removal of Congo red. J Colloid Interface Sci 521:172–182. https://doi.org/10.1016/j.jcis.2018.03.040
5. Dubinin MM (1959) The Potential Theory of Adsorption of Gases and Vapors for Adsorbents with Energetically Nonuniform Surfaces. Theory Adsorpt Gases Vap 235–241

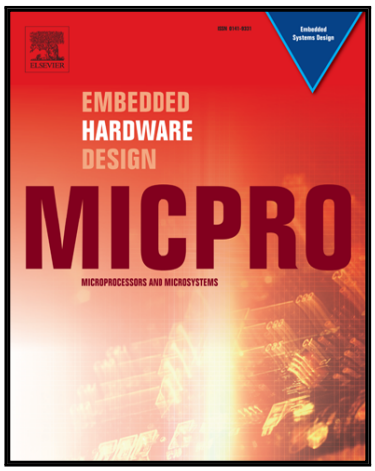
6. Freundlich H (1907) Über die Adsorption in Lösungen. Zeitschrift für Phys Chemie 57U:385–470. https://doi.org/10.1515/zpch-1907-5723
7. Garg VK, Kumar R, Gupta R (2004) Removal of malachite green dye from aqueous solution by adsorption using agro-industry waste: A case study of Prosopis cineraria. Dye Pigment 62:1–10. https://doi.org/10.1016/j.dyepig.2003.10.016
8. George G, Saravanakumar MP (2018) Facile synthesis of carbon-coated layered double hydroxide and its comparative characterisation with Zn–Al LDH: application on crystal violet and malachite green dye adsorption—isotherm, kinetics and Box-Behnken design. Environ Sci Pollut Res 25:30236–30254. https://doi.org/10.1007/s11356-018-3001-3
9. Greenwell HC, Jones W, Rugen-Hankey SL, Holliman PJ, Thompson RL (2010) Efficient synthesis of ordered organo-layered double hydroxides. Green Chem 12:688–769. https://doi.org/10.1039/b916301h
10. Guo Y, Zhu Z, Qiu Y, Zhao J (2013) Enhanced adsorption of acid brown 14 dye on calcined Mg/Fe layered double hydroxide with memory effect. Chem Eng J 219:69–77. https://doi.org/10.1016/j.cej.2012.12.084
11. Harkins WD, Jura G (1944) Surfaces of solids. XIII. A vapor adsorption of a monolayer area, and the areas occupied be nitrogen and other molecules on the surface of a solid. J Am Chem Soc 66:1366–1373
12. He J, Wei M, Li B, Kang Y, Evans DG, Duan X (2006) Preparation of layered double hydroxides. Interface Sci Technol 119:89–119. https://doi.org/10.1016/S1573-4285(04)80047-4
13. Ho YS, McKay G (1999) Pseudo-second order model for sorption processes. Process Biochem 34:451–465. https://doi.org/10.1021/acs.oprd.7b00090
14. Hu X, Liu L, Zeng HY, Xu S, Cao X, Cao XJ (2019) An advanced NiCoFeO/Polyaniline composite for high-performance supercapacitors. Chem An Asian J 14:977–985. https://doi.org/10.1002/asia.201801905
15. Jeevanandam P, Koltypin Y, Gedanken A (2001) Synthesis of nanosized α-Nickel hydroxide by a sonochemical Method. Nano Lett 1:263–266. https://doi.org/10.1021/nl010003p
16. Jura G, Harkins WD (1943) A new adsorption isotherm which is valid over a very wide range of pressure. J Chem Phys 11:430–431. https://doi.org/10.1063/1.1723870
17. Kosova NV, Devyatkina ET, Kaichev VV (2007) Mixed layered Ni–Mn–Co hydroxides: Crystal structure, electronic state of ions, and thermal decomposition. J Power Sources 174:735–740. https://doi.org/10.1016/j.jpowsour.2007.06.109
18. Lafi R, Charradi K, Djebbi MA, Ben Haj Amara A, Hafiane A (2016) Adsorption study of Congo red dye from aqueous solution to Mg-Al-layered double hydroxide. Adv Powder Technol 27:232–237. https://doi.org/10.1016/j.apt.2015.12.004
19. Lagergren S (1898) Zur Theorie der sogenannten Adsorption geloster Stoffeergren Kungliga Svenska Vetenskapsakademiens. Zeitschrift f{\"u}r Chemie und Ind der Kolloide 24:1–39. https://cir.nii.ac.jp/crid/1572824501080908544#citations_container
20. Langmuir I (1917) The constitution and fundamental properties of solids and liquids. Part II-Liquids J Franklin Inst 184:721. https://doi.org/10.1016/s0016-0032(17)90088-2
21. Liu L, Hu X, Zeng HY, Yi MY, Shen SG, Xu S, Cao X, Du JZ (2019) Preparation of NiCoFe-hydroxide/polyaniline composite for enhanced-performance supercapacitors. J Mater Sci Technol 35:1691–1699. https://doi.org/10.1016/j.jmst.2019.04.003
22. Liu X, Ma R, Bando Y, Sasaki T (2012) A general strategy to layered transition-metal hydroxide nanocones: tuning the composition for high electrochemical performance. Adv Mater 24:2148–2153. https://doi.org/10.1002/adma.201104753
23. Mallakpour S, Behranvand V (2020) Layered double hydroxide polymer nanocomposites for water purification. Elsevier, Amsterdam
24. Nagarajan R, Gupta P, Singh P, Chakraborty P (2016) An ethylene glycol intercalated monometallic layered double hydroxide based on iron as an efficient bifunctional catalyst. Dalt Trans 45:17508–17520. https://doi.org/10.1039/c6dt03129c
25. Pandey B, Singh P, Kumar V (2021) Photocatalytic-sorption processes for the removal of pollutants from wastewater using polymer metal oxide nanocomposites and associated environmental risks. Environ Nanotechnol Monit Manag 16:100596. https://doi.org/10.1016/j.enmm.2021.100596
26. Saroj SK, Pal S, Nagarajan R (2020) Polyol intercalation in copper substituted zinc hydroxide acetate and evaluation of its adsorptive role towards Congo red dye. Appl Clay Sci 185:105411. https://doi.org/10.1016/j.clay.2019.105411
27. Selvam T, Inayat A, Schwieger W (2014) Reactivity and applications of layered silicates and layered double hydroxides. Dalt Trans 43:10365–10387. https://doi.org/10.1039/c4dt00573b
28. Shen X, Zhu Z, Zhang H, Di G, Chen T, Qiu Y, Yin D (2020) Carbonaceous composite materials from calcination of azo dye-adsorbed layered double hydroxide with enhanced photocatalytic efficiency for removal of Ibuprofen in water. Environ Sci Eur. https://doi.org/10.1186/s12302-020-00351-4
29. Taviot-Guého C, Prévot V, Forano C, Renaudin G, Mousty C, Leroux F (2018) Tailoring hybrid layered double hydroxides for the development of innovative applications. Adv Funct Mater 28:1–33. https://doi.org/10.1002/adfm.201703868
30. Tunney JJ (1996) Chemically modified kaolinite. Grafting of methoxy groups on the interlamellar aluminol surface of kaolinite. J Mater Chem 6:1679–1685. https://doi.org/10.1039/jm9960601679
31. Wang C, Zhang X, Xu Z, Sun X, Ma Y (2015) Ethylene glycol intercalated cobalt/nickel layered double hydroxide nanosheet assemblies with ultrahigh specific capacitance: structural design and green synthesis for advanced electrochemical storage. ACS Appl Mater Interfaces 7:19601–19610. https://doi.org/10.1021/acsami.5b03176
32. Wang X, Zhou S, Xing W, Yu B, Feng X, Song L, Hu Y (2013) Self-assembly of Ni–Fe layered double hydroxide/graphene hybrids for reducing fire hazard in epoxy composites. J Mater Chem A 1:4383–4390. https://doi.org/10.1039/c3ta00035d
33. Wani AA, Khan AM, Manea YK, Salem MAS, Shahadat M (2021) Selective adsorption and ultrafast fluorescent detection of Cr(VI) in wastewater using neodymium doped polyaniline supported layered double hydroxide nanocomposite. J Hazard Mater 416:125754. https://doi.org/10.1016/j.jhazmat.2021.125754
34. Xi Y, Davis RJ (2010) Intercalation of ethylene glycol into yttrium hydroxide layered materials. Inorg Chem 49:3888–3895. https://doi.org/10.1021/ic1000478
35. Xiang W, Zhang G, Zhang Y, Tang D, Wang J (2014) Synthesis and characterization of cotton-like Ca–Al–La composite as an adsorbent for fluoride removal. Chem Eng J 250:423–430. https://doi.org/10.1016/j.cej.2014.03.118
36. Xu Z, Zhong Y, Wang Y, Song X, Huang W (2022) Removal performance and mechanism of phosphorus by different Fe-based layered double hydroxides. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-022-21047-7
37. Yan Q, Zhang Z, Zhang Y, Umar A, Guo Z, O'Hare D, Wang Q (2015) Hierarchical Fe3O4 core-shell layered double hydroxide composites as magnetic adsorbents for anionic dye removal from wastewater. Eur J Inorg Chem 2015:4182–4191. https://doi.org/10.1002/ejic.201500650
38. Zhu J, Zhu Z, Zhang H, Lu H, Qiu Y (2019) Efficient degradation of organic pollutants by peroxymonosulfate activated with

MgCuFe-layered double hydroxide. RSC Adv 9:2284–2291. https://doi.org/10.1039/C8RA09841G
39. Zhu Y, Yue X, Xie F (2020) Adsorptive removal of phosphate by a Fe–Mn–La tri-metal composite sorbent: adsorption capacity, influence factors, and mechanism. Adsorpt Sci Technol 38:254–270. https://doi.org/10.1177/0263617420942709

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 Pre-proof



ADVANCED ENERGY EFFICIENT PEGASIS BASED ROUTING PROTOCOL FOR IoT APPLICATIONS

Priyanka chugh , Meenu Gupta , Dr.S. Indu , Gopal Chaudhary , Manju Khari , Vimal Shanmuganathan

PII: S0141-9331(22)00255-1
DOI: https://doi.org/10.1016/j.micpro.2022.104727
Reference: MICPRO 104727

To appear in: *Microprocessors and Microsystems*

Received date: 16 November 2020
Revised date: 7 February 2021
Accepted date: 8 March 2021

Please cite this article as: Priyanka chugh , Meenu Gupta , Dr.S. Indu , Gopal Chaudhary , Manju Khari , Vimal Shanmuganathan , ADVANCED ENERGY EFFICIENT PEGASIS BASED ROUTING PROTOCOL FOR IoT APPLICATIONS, *Microprocessors and Microsystems* (2022), doi: https://doi.org/10.1016/j.micpro.2022.104727

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier B.V.

Journal Pre-proof

ADVANCED ENERGY EFFICIENT PEGASIS BASED ROUTING PROTOCOL FOR IoT APPLICATIONS

Priyanka chugh[1], Meenu Gupta[2], Dr. S. Indu[3], Gopal Chaudhary[4], Manju Khari[5,*], Vimal Shanmuganathan[6,*]

[1]Research Scholar, Ansal University, Gurugram
[2]Chandigarh University, Punjab
[3]Delhi Technological University, Delhi, India
[4]Bharati Vidyapeeth's College of Engineering, Delhi, India
[5]Netaji Subhas University of Technology, East Campus, Delhi
[6]National Engineering College, K.R.Nagar, Kovilpatti, Tamilnadu, India
{shivankachugh84@gmail.com, gupta.meenu5@gmail.com, gopal.chaudhary88@gmail.com, *manjukhari@yahoo.co.in,*svimalphd@gmail.com}* Corresponding authors

Journal Pre-proof

ABSTRACT

The Scientific betterments in Wireless Sensor Networks (WSNs) technology have paved the way for the advancement of lower sized, reduced-cost, and highly-powered sensor nodes. The sensor nodes are consistently expected to work with limited power batteries, and their energy is the most precious asset in a WSN, proper usage of the total energy to perpetuate the Network's lifetime is the primary constraint and objective of the majority of the research works. As energy conservation and prolonging of network-lifetime have become a critical issue, a novel protocol for WSN based on an Advanced Fixed Path and Mobile Base Station energy-efficient PEGASIS based routing protocol is proposed in this research work. The proposed method focuses on the base station's specified mobility with a multiple-chain model to attain small size chains and reduce the single leader node load that directly affect the Network's energy conservation and lifetime. The experimental results based on average energy, residual energy, consumed energy, dead nodes, and alive nodes reflect the proposed method's efficiency to fulfill the above-said objectives. The proposed method is also compared with three current routing protocols, i.e., Energy Efficient Routing Algorithm with Mobile Sink Support for WSNs (EERCCSM), Cluster-Chain Mobile Agent Routing (CCMAR), and Energy-efficient Cluster-based Dynamic Routing Algorithm (ECDRA). This novelty and advancement of the base station will open new horizons for getting it placed much closer to the cluster leader to prolong Network's lifetime to a great extent in the future.

INTRODUCTION

WSNs are adapted to sense distinct parameters of the environment by adopting wireless sensor nodes [1]. One of the noteworthy improvements in WSN-based smart healthcare centers is the infusion of IoTs (the Internet of Things). The universal Network and extraordinary intelligence of the IoTs with remote detecting technology turn into its main USPs. The IoT from a smart healthcare center's viewpoint can be portrayed as interfacing the instruments, mainly electronic and electrical apparatuses, and wearable and implanted sensors and actuators, to the Internet's full services [2]. The information gathered by these sensors and the user's instrument utilization

patterns permits the IoT to track its routine. The information can be adequately collected and further processed to uncover serious conditions by utilizing state of the art Machine/Deep Learning (ML/DL) & Artificial Intelligence (AI) based methods [3]. So, the performance of a perfect IoT system majorly depends on the working efficiency and accuracy of the WSN system [4].

Existing WSN routing protocols contend with ensuring fair data sharing [5]. The sensor nodes are consistently expected to work with limited power batteries. Their energy is the most precious asset in a WSN; proper usage of the total energy to perpetuate the Network's lifetime is the primary constraint and objective of most of the research works for WSNs [1]. Although sensor fabrication is on a progressive mode, including computing speed and processing, the advancement of battery mechanization still lags. This makes the advancement of energy resources a fundamental constraint [5]. Besides this, some existing researchers are working on the basic idea to elect some nodes as a specific area leader. These elected nodes have a leading responsibility as compared to other sensor nodes. This undergoes to set up a balanced relationship amongst all the nodes [6]. The overall focus of all the methods is to maximize the lifetime of the Network. So, energy conservation and efficient utilization of the energy is of paramount importance in WSNs [5].

As discussed above, energy conservation and prolonging of network lifetime have become a critical issue; clustering of the Network has been confirmed energy adequate in WSN particularly. The cluster leader transmits the collected data to the base station via multi-hop or single-hop [7]. Along with clustering and cluster leader deputation and base station mobility is also gaining importance day by day. Current developments i.e. Optimized routing in Multi-Layer Multi-Domain (MLMD) has shifted the spotlight focus on activities of base station instead of the sensor to advance the network lifetime. Improved Energy Efficient Power-Efficient (IEEPB) gathering in Sensor Information Systems PEGASIS based Protocol also efficiently saves the chain leader's energy when multi-head chain methodology is employed. IEEPB lesser down the development of long links in the chain by using a threshold [8]. As far as, the existing routing protocols Low-energy adaptive clustering hierarchy (LEACH), IEEPB, Distributed Energy Efficient Clustering (DEEC) and Stable Election Protocol (SEP) and other new routing protocols are a concern, they are not efficient to conquered the energy dissipation of nodes and to provide fixed path sink mobility competently.

The base station's specific path mobility and a multiple-chain model are undoubtedly adequate for boosting the Network's efficiency [9, 10, 11]. On the one hand, a mobile base station lowers down sensor nodes' energy usage by making data collection their location. On the other hand, it reduces the time taken in data packet delivery for all the chain's presented nodes. The multiple-chain model accounts to attain small size chains and reducing the load on a single leader.

Journal Pre-proof

ORGANIZATION OF THE PAPER

The present research is formulated as follows. Section 2 describes the various already existing studies connected to the significant multi-cluster based energy-efficient wireless routing protocols. Section 3 presents the proposed research methodology to resolve the current research problem. Section 4 figures out the outcomes of the simulations and comparison amongst the proposed method and existing methods. The last section of the paper concludes the overall research along with findings the future work.

LITERATURE REVIEW

Wendi Rabiner Heinzelman et al. [12, 13] introduced LEACH, a protocol that is based on clustering. This Protocol utilizes the arbitrary movement of cluster leaders to hand out the load amongst the node sensors. But the network lifetime was still low. Young Sang Yun et al. introduced research work to expand the network lifetime by applying a mobile base station [5, 6, 14] to prolong the network lifetime. For the same problem, Stephanie Lindsey et al. [10] introduced PEGASIS, a near-optimal chain-based protocol that is an improvement over LEACH. In continuation to the modification to PEGASIS, Feng Sen et al. proposed IEEPB, improved multi-chain PEGASIS, and mobility of a base station in enhanced energy-efficient PEGASIS-based Protocol, respectively to prolong the lifetime of the Network in WSNs [6, 11]. The multi-leader chain, multiple chains concept, and base station mobility largely prolong sensor nodes' lifetime.

Apart from the above three methods, Abo-Zahhad et al. [15] also explored a Mobile Sink-based adaptive Immune Energy-Efficient clustering Protocol (MSIEEP) to lighten the energy holes. An MSIEEP uses the flexible immune technique to pilot the mobile base station to lower overall energy consumption. On the other hand, Gharaei N. Kamalrulnizam et al. [16] proposed a joint mobile base station movement time optimization method. This method aims to optimize the mobile base station's movement time in each fixed area to get a stabilized network lifetime of cluster leaders that belong to various parts of the Network.

Zhou Haibo et al. [8] introduced a novel WSN model with energy diversity and computational diversity in contemporary work. Besides, the author suggested a new energy adequate routing algorithm named EDFM. The method for electing the cluster leader in EDFM is based on the energy forecasting technique's dissipation. This technique acknowledges the rate of energy consumption and residual energy in all the presented nodes. Ajay Kaushik et al. [9] explored a Mobile Sink Distributed load Routing algorithm Using Biogeography-Based Optimization (MSDR-BBO) that chooses the efficient cluster leader in mobile sink WSN according to the requirements of the sensor for data transfer and leader. Barani H. Jaradat et al. [17] designed two popular-location methods and a random method based on base station location. The authors

acknowledged a cell-subdivided network to handle signal intervention among the cells. The given results exhibit that there is an establishment between the dealy and capacity in both the methods. In comparison to the random scheme, Yung-Fa Huang et al. [18] introduced a Fixed Clustering Algorithm (FCA) to optimize the energy efficiency for WSNs. This technique reduces the energy utilization of the sensor nodes. It also enhances the network lifetime that is 80% greater than that of the random clustering (RC). Few authors worked mainly on the selection of chain leader and computing node energy. Marios Gatzianas et al. [19] proposed a distributed technique based on the sub-gradient algorithm and the base station as a leader. K. L. Ang et al. [20] presented detailed methods to resolve the sensor node's energy consumption for large-scale WSN mobile data collection techniques. It also provides the working models for calculating the maximum number of clusters to reduce energy utilization. Hidehisa Nakayama et al. [21] introduced a new technique based on the traveling salesman problem and set the packing method. The main focus of this novel technique is to access all sensor nodes in the presented networks.

Some researchers also gave significant comparison and analysis for existing routing protocols. Rajni Meelu et al. [7] analyzed the performance of distributed energy-efficient clustering methods concerned with the network lifetime, energy balancing, and energy consumption. Whereas Maya M. Warrier et al. [22] provided a comparative analysis among the Modified LEACH and Mobile Sink improved energy-efficient PEGASIS-based routing protocol (MIEEPB).

Apart from the basic routing protocols of a WSN, a comparative study and implementation of some problem-solving optimization techniques has been done. These are Ant Colony Optimization(ACO) [23], AI [24], ML/DL [3], and K-Means [23] clustering methods for energy optimization & jobs scheduling in IoT, given by some researchers [2,4].Gaurav Dhiman et al. introduced some Bio-Inspired optimization algorithms that are Rat Swarm Optimizer (RSO) [25], Emperor Penguin Optimizer (EPO) [26.27], which mimics the chasing and attacking behaviors of rats in nature and huddling behavior of emperor penguins (Aptenodytes forsteri), respectively. Some other authors also have given very significant other optimization methods Orientation Search Algorithm (OSA [28]), spring search algorithm (SSA)[29], Tunicate Swarm Algorithm (TSA) [30].

The description of routing protocols given by various researchers is shown below in table 1.

TABLE I: Description of Routing Protocol by Various Researchers

S. No	Author's Names	Year	Methodology used	Problem
[1]	Young Sang Yun et al.	2010	Mobile Sink in Delay Tolerant Applications	Maximize the lifetime of WSN

[2]	Hemant Ghayvat et al.	2015	Wellness sensor networks	IOT-Based Smart Homes and Their Extension to Smart Buildings
[3]	Iman Azimi et al.	2017	HiCH: Hierarchical Fog-Assisted Computing Architecture	IOT-Based Smart Healthcare centers
[5]	Weifa Liang et. al.	2010	Mixed Integer Linear Programming (MILP)	Enhancing the lifetime of WSN
[6]	Feng Sen et. al.	2011	Energy-Efficient PEGASIS-Based Protocol (EEPB)	Prolong the lifetime of WSN & balancing energy consumption
[7]	Rajni Meelu & Rohit Anand	2011	DEEC	Network lifetime, energy consumption and energy balancing
[8]	ZhouHaibo et. al.	2009	Energy Dissipation Forecast Method	Energy dissipation forecasting
[9]	Ajay Kaushik	2019	MSDR-BBO	Optimal distribution of the data relay load
[10]	Stephanie Lindsey and Cauligi S. Raghavendra	2001	PEGASIS	Enhancement of Energy Efficiency
[11]	Shailly Thakur1, Balraj Singh2	2016	MIEEPB	Energy Efficiency
[12]	Wendi Rabiner Heinzelman et. al.	2000	LEACH	Even distribution of the energy load among the sensors
[13]	Wendi Rabiner Heinzelman et. al.	2002	LEACH	Maximizing the Network's lifetime
[14]	JinWang et. al.	2019	Energy-efficient routing method combined with clustering and mobile base station technique	Energy-efficient routing
[15]	Abo-Zahhad et. al.	2015	MSIEEP	Data collection in WSN
[16]	Gharaei N. Kamalrulnizam et. al.	2018	Collaborative Mobile Sink Sojourn Time Optimization (CMS2TO)	Optimization of the sojourn time of the mobile base station in each cluster
[17]	Barani H. Jaradat et. al.	2018	Popular-location scheme and random scheme	Redundancy on the capacity of node and delay in data
[18]	Yung-Fa Huang et. al.	2007	FCA	Reduction of energy consumption
[19]	Marios Gatzianas and Leonidas Georgiadis	2008	A distributed technique for prolonging the network lifetime with the mobile base station	Maximizing the Network's lifetime

Journal Pre-proof

[20]	K. L. Ang et. al.	2018	Analytical approach to determine the node energy consumption	Node energy calculation
[21]	Hidehisa Nakayama et. al.	2011	Based on the Traveling Salesman problem and Set Packing algorithm	Efficient data gathering
[22]	Maya M. Warrier and Ajay Kumar	2016	Comparison analysis between Modified LEACH and MIEEPB	Energy conservation
[23]	Sumit Kumar et al.	2020	ACO and K-Means Clustering Approaches	Comparative Study for Jobs Scheduling and Energy Optimization IoT
[24]	Cristian González García et al.	2018	AI	Review of AI in the IoT

After going through the literature available, the existing protocols have still some deficiencies that are given as follows:

• When the Energy Efficient PEGASIS-based Protocol (EEPB) builds a chain, there is a considerable load on the single-chain leader. Also, a substantial energy loss happens due to a substantial distance between itself and the base station.

• Long single chains are also responsible for more significant delays in the delivery of the data. This fact is un-suited for delay intolerant or delay-susceptible operations.

• Due to these long inter-mutual distances, the presented sensor nodes affect badly in the instability period.

Due to the above imperfections, there is a strong need to utilize fixed path base station mobility and multiple cluster leaders in WSN to enhance the network lifetime. In the proposed method, fixed trajectory routing with the concept of multi-leader chains is exercised. A technique for the fixed path of the mobile base station has been developed. The proposed approach is also beneficial for delay-sensitive operations. It does not lead to the time delay in data packets delivery due to smaller sized chains having fewer sensor nodes. The contribution of the author of this research paper can be outlined below:

- The technique for fixing the mobile base station's trajectory to bound it for only a few specific locations is focused on this research work.
- The workload of collected data transfer on the single-chain leader is reduced by introducing the concept of multiple leaders in the chains.

Based on the problems mentioned above, this research work presents Advanced Fixed Path Mobile Sink Energy Efficient PEGASIS Based Protocol to provide mobility of the base station with a multiple-chain model, attain small size chains, and reduce the load on a single leader node.

Journal Pre-proof

IV. PROCESS FLOW FOR THE PROPOSED METHOD

The main steps for the process flow are given as follow:

A. The architecture of the Network

A field area of 100m x 100m has been taken for the sensor network. In this field, 100 sensor nodes have been deployed. The above-said field has been divided into four equal parts or clusters. These nodes have also been separated and deployed randomly in these clusters, i.e., 25 nodes per part, using uniform random distribution. Figure 1 shows equally distributed chained sensors network field through simulation.

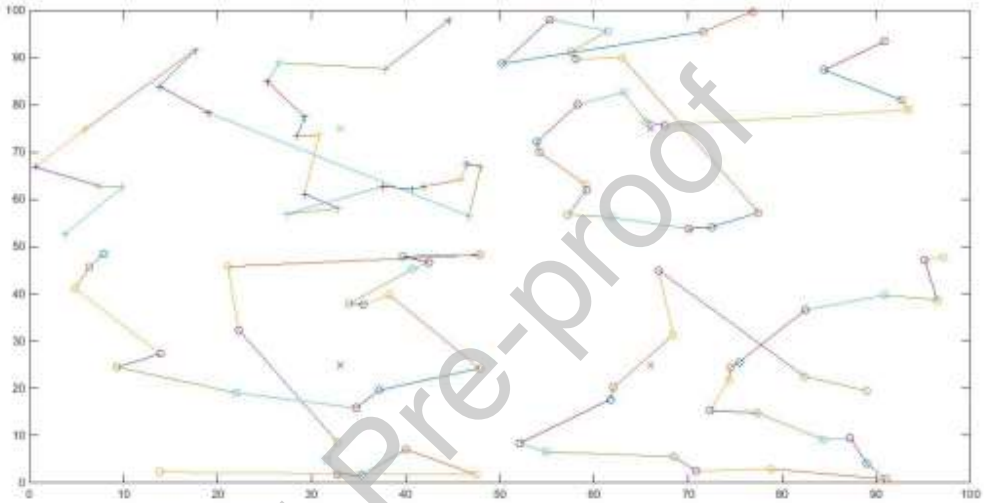


Figure 1. Equally distributed chained sensors network field through simulation

The process is designed so that the mobile base station moves at the fixed center of each equally spaced part and complete its 1st round with a particular controlled speed and so on for upcoming rounds. The base station stays at each fixed center for a specified time duration to accumulate the sensor nodes' full data packets. Sensor nodes also perform a data compression process on received bits with data aggregation factor (0.6). The energy consumed (e_{TX}) in the data transmission has been by the sensor nodes has been calculated by implementing the first order radio model as given in figure 2.

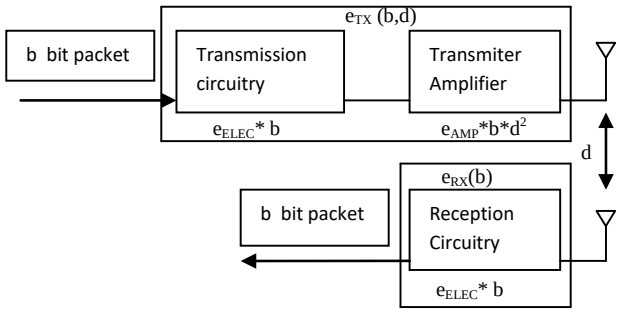


Figure 2. First-order Radio Model

$$e_{TX}(b,d) = e_{ELEC} * b + e_{AMP} * b * d^2 \quad (1)$$

$$e_{RX}(b) = e_{ELEC} * b \quad (2)$$

As per the first order radio model, some fixed parameters (used above) are to be used in the implementation. The value and description of these parameters are given below.

1. e_{ELEC} = 50 nJ/bit, The energy consumption by the radio for the overall running of transmitter or receiver system
2. e_{amp} = 100 pJ/bit/m^2, The energy needed for transmitter amplifier
3. b number of bits that are to be transmitted throughout the network lifetime
4. d is the distance between transmitter and receiver.
5. e_{DA}= 50 nJ/bit, The energy consumption by transmitter circuitry for the aggregation of the junior nodes' data.

Several other parameters have been used for simulation in the proposed method. The detail of these parameters is given in Table 2

Table 2 Different parameters used in simulations

	Network Parameter	Value
1	Size of the Network	100 x 100 m^2
2	The initial energy of sensor nodes	0.5 J
3	Size of the packet	2000 bits
4	Number of nodes	100 nodes
5	Number of rounds	4500 rounds
6	Energy consumption	50 nJ/bit
7	Fixed location of Base Station at each cluster	1st cluster (33 m Horizontal, 25 m Vertical) 2nd cluster (33 m Horizontal, 75 m Vertical) 3rd cluster (66 m Horizontal, 25 m Vertical) 4th cluster (66 m Horizontal, 75 m Vertical)
9	Data Aggregation factor	*0.6*
10	e_{DA}	$5\ nj/bit$
11	e_{TX}	$50\ nj/bit$
12	e_{RX}	$50\ nj/bit$
14	e_{amp}	$.0013\ pj/bit/m^4$
15	$e_{amp(fs)}$	$10\ pj/bit/m^2$

Journal Pre-proof

Journal Pre-proof

B. Construction of multiple chains

As in the case of PEGASIS, the chain building method of the proposed method is the same. In the proposed method, four chains have been prepared. The process of chain formation is as follows:

• Base station takes the initiative to start the data sharing by sending a hello packet to all the presented nodes in the particular cluster.

• Now, the base station measures the distance with each sensor node in the cluster and finds the sensor node having maximum distance from the base station by making a distance comparison with each node.

• The formation of the chain starts from the node (i^{th}) having maximum distance from the base station, also known as the end node. After this, the end node measures the distance with each node in the cluster, assesses the nearest sensor node from itself, and connects with it.

• This process is followed by each sensor node to find the nearest sensor node that is not already connected to make a full chain.

• In this process, every node (i^{th}) that gets the data from any other node (j^{th}), termed as a senior to the node (j^{th}), whereas node (j^{th}) acts as a junior to the node (i^{th}).

The above process is followed by the sensor nodes and the base station in each cluster for the chain formation.

C. Selection of Chain's leaders

The first leader's selection in the chain takes place based on the weight '*w'* that is assigned initially equally to each sensor node.

Later on, the value of weight (w_i) changes and depends on the residual energy (e_i) of the i^{th} node and its distance (d_i) from the base station.

$$w_i = e_i / d_i \quad (4)$$

Now it's the Network's turn to compare the weights of the entire presented sensor nodes. The Network finds the node with the largest weight. Based on this computation, the Network allows this node as the first chain leader.

As discussed in the last section for the chain formation, each junior node (j^{th}) evaluates its distance (dp) with its senior node (i^{th}). After this, the junior node also evaluates its distance (dbs) from the base station and compares. If its distance with the senior node is greater than that with the base station ($ds \geq dbs$), that particular node acts as a secondary chain leader. This secondary leader starts the collected data sending process with the base station instead of sending it to the senior node.

D. Mobility of the Base Station

The base station is the mobile participant in the whole Network, and its unlimited energy with mobility maximizes the lifetime of the Network. The base station is made bounded to roam in the

field with a fixed bath and stays to a fixed location. The base station's time duration to keep at a particular location (for data collection from the chain leader) at each cluster is also fixed. The fixed locations where the base station stays temporarily for data collection are 33 m Horizontal, 25 m Vertical (1st cluster), 33 m Horizontal, 75 m Vertical (2nd cluster), 66 m Horizontal, 25 m Vertical (3rd cluster), and 66 m Horizontal, 75 m Vertical (4th cluster).

1) Proposed algorithm for the base station's mobility: There are three stages of the base station mobility that is as follows.
   - First, the fixed time for the base station is calculated for all four clusters.
   - Based on these time profiles, the base station starts its journey and stays in between at the fixed locations mentioned
   - In the next step, the base station calculates the staying time in each round. Based on this, it calculates the total staying time of all four fixed locations. It can be calculated as follows

$$t_s = \sum_{i=1}^{4} (\tau i) \quad (5)$$

where τi is the time taken by the base station at one cluster. Ts is the total staying time of one round.

The main objectives of this research work are:

1. To enhance the network lifetime by improving the total staying time (*ts*).

   $$Maximize \sum_{i=1}^{4} (\tau i) \quad (6)$$

2. To ensure consistent data sharing among the base station and chain leaders.

   $$b_{ij} = \begin{cases} B & if\ i=j \\ 0 & otherwise \end{cases} \quad (7)$$

Here, bij is the number of data packet bits shared between the base station and chain leaders at horizontal locations i and vertical location j, $1 \leq i, j \geq 4$. B is the total data shared between the base station and chain leaders.

E. Data transmission

Multiple chain leaders use Time Division Multiple Access (TDMA) method to share the data with the base station in the proposed method. After data sharing with the base station, the end node receives a token from the chain leader and passes it to another node. Each node follows this process in the cluster. Therefore, the entire presented node in the cluster shares the data with the cluster leader, and the cluster leader transmits all the accumulated data to the base station. Now, the same process is followed in all four clusters. Therefore, the taken based methodology is used for data transmission in the proposed method. The token starts its journey from the end node to the chain leader. If there are multiple juniors of any senior node, data sharing is done using the TDMA mechanism.

F. Aggregation of the data

The senior node receives and compresses the data (obtained through a junior node) by a data aggregation factor of 0.6. Senior nodes use compressive sampling to combine their data with the received data.

Journal Pre-proof

The above process flow of the proposed method can be easily understood by the below-given block diagram in figure 3.

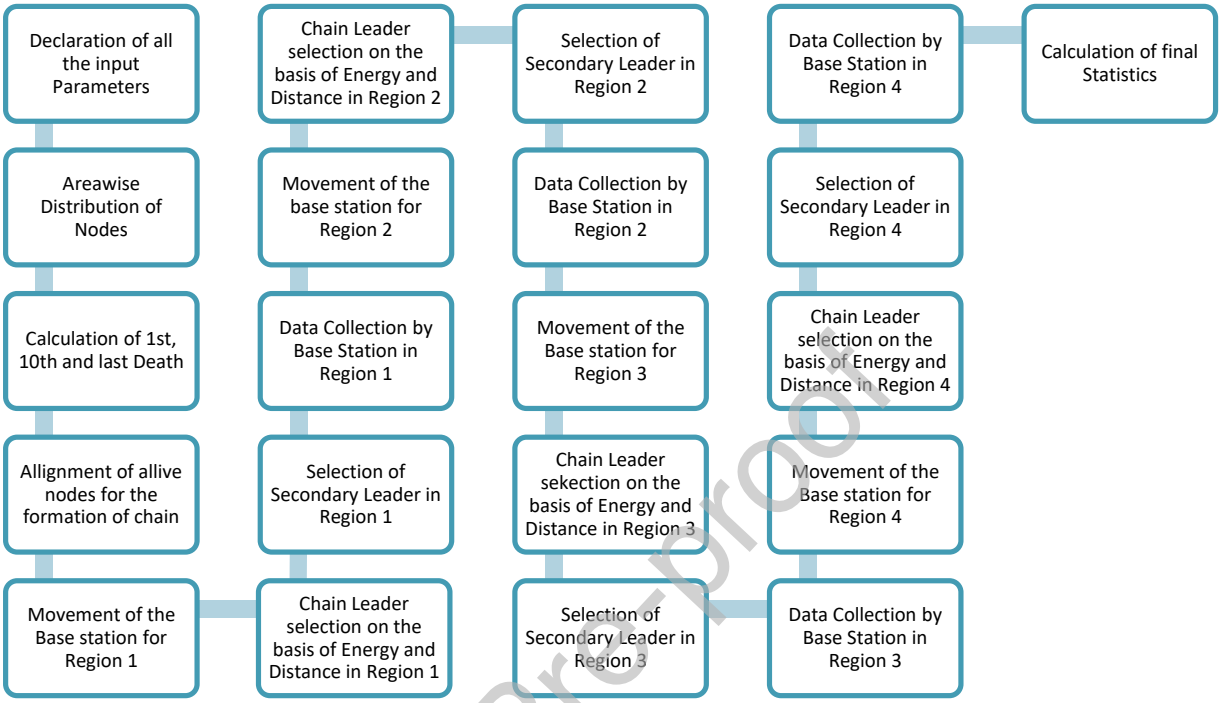


Figure 3 Block Diagram of the process flow of the proposed method

V. EXPERIMENTAL RESULTS

A novel method for WSN based on advanced fixed path mobile base station energy-efficient PEGASIS based routing protocol is proposed in this research work. In the proposed method, the mobile base station travels along with its fixed trajectory and stays there for a fixed time at a specified location to assure the full collection of data. Also, an algorithm for the fixed path of the mobile base station is developed. The base station's specific path mobility and a multiple-chain model are undoubtedly adequate for boosting the Network's efficiency [9, 10, 11]. On the one hand, a mobile base station lowers down sensor nodes' energy usage by making data collection their location. On the other hand, it reduces the time taken in data packet delivery for all the chain's presented nodes. The multiple-chain model accounts to attain small size chains and reducing the load on a single leader.

A wide range of experiments is performed for the assessment of the working functionality of the proposed method. These experiments are related to the mobile base station's speed and the size of the node's field. The base station's different speed is 10 m/s, 50 m/s, and 100 m/s, respectively. The different node’s field size is 100m X 100m, 200m X 200m and 300m X 300m, respectively. The proposed method is also compared with three existing methods: EERCCSM, CCMAR, and

Journal Pre-proof

ECDRA. The efficiency and outcome of the proposed method are shown and compared graphically and tabular manner. The parameters which have been taken for experimental purpose are Normalized average energy of the nodes, Residual energy of the nodes, Consumed energy of the nodes, total alive nodes and total dead nodes in the field.

The proposed method is implemented in MATLAB software, and the experimental results are given in the form of some graphs (taken from the MATLAB) and tables. All the parameters in the graph are changing concerning the increasing number of nodes, i.e., 0 – 4500. Figure 4 to 6 shows the normalized average energy plot, total residual energy and total consumed energy of the nodes for all three different sizes of fields, respectively.

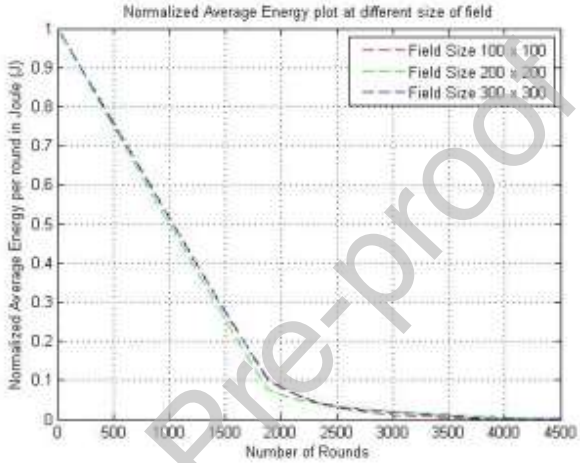


Figure 4 Normalized average energy plot of the nodes for all three different sizes of fields

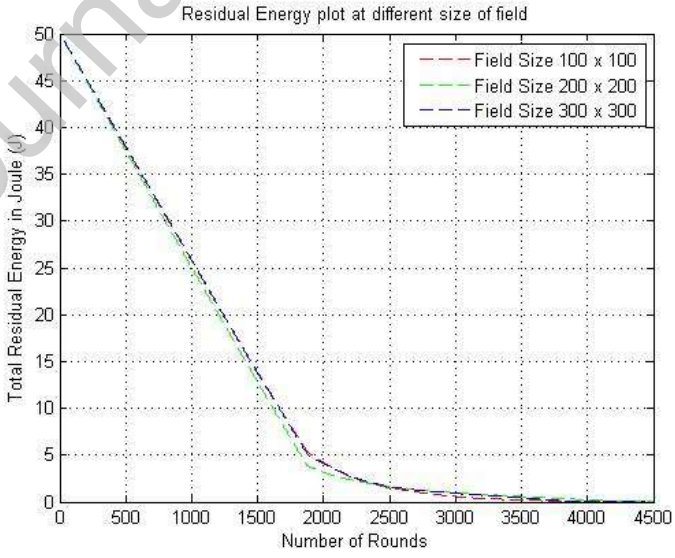


Figure 5 shows the total residual energy of the nodes for all three different sizes of fields

Journal Pre-proof

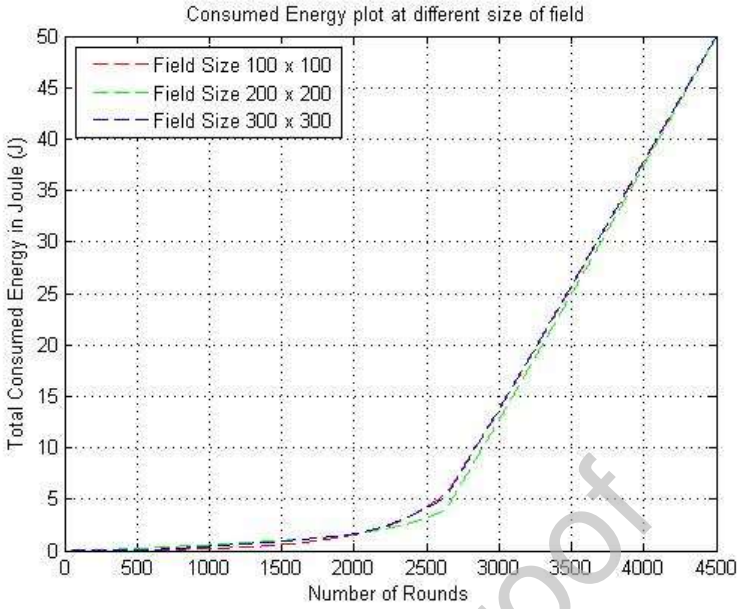


Figure 6 shows the total consumed energy of the nodes for all three different sizes of fields

Table 3 below gives the comparative analysis of the different percentage levels of consumption i.e., 1%, 10%, 50%, 90%, and 100%, at the respective round number for all three different sizes of fields.

Table 3 Comparative analysis of the different percentage level of consumption at the respective round number for all three different sizes of fields

S. No.	Parameter	Round Number		
		Field Size 100m x 100m	Field Size 200m x 200m	Field Size 300m x 300m
1.	1% Consumption	1425	887	1060
2.	10% Consumption	2585	2674	2610
3.	50% Consumption	3472	3501	3467
4.	90% Consumption	4297	4302	4294
5.	100% Consumption	4500	4500	4500

If we examine Table 3, one percent of energy is consumed in the 1425^{th} round in the 100 x 100 size field, which is very much late compared to the rest of the two sized fields. But if we look at 10 – 90 percent energy consumption, field size 200 X 200 is very much efficient and favorable. Also, From the overall analysis given in figure 4, figure 5, figure 6, and table 3, it can be concluded that field size 200 X 200 is more favorable for all the nodes in the field as far as energy consumption and residual energy and normalized energy are concerned.

Journal Pre-proof

Figures 7 and 8 show the total dead nodes and the total alive nodes for all three different sizes of fields, respectively. Table 4 gives the comparative analysis of the total node's death i.e., 1st death, 10th death, 50th death, 90th death, and 100th death, at the respective round number for all three different sizes fields.

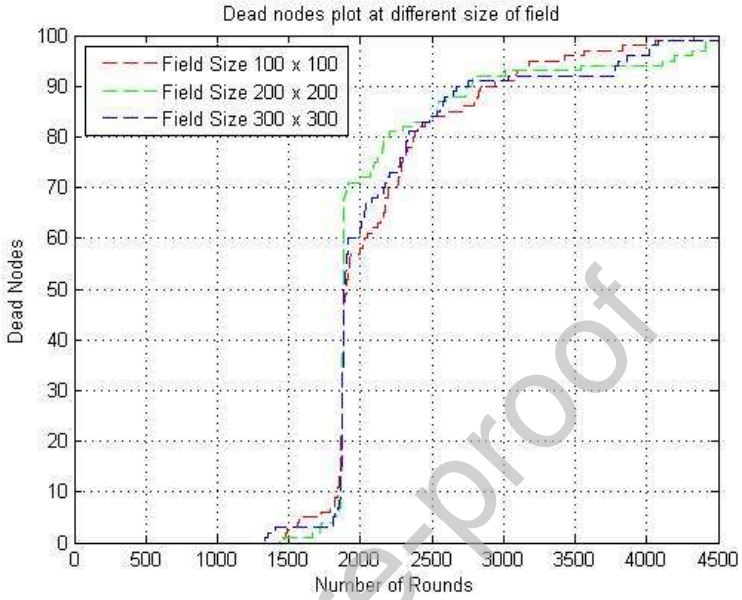


Figure 7 shows total dead nodes for all three different sizes of fields

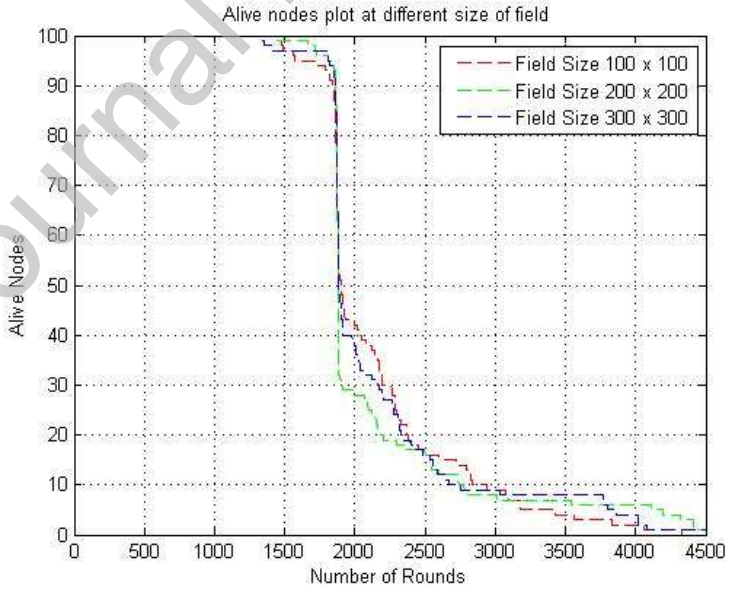


Figure 8 shows total alive nodes for all three different sizes of fields

Journal Pre-proof

Table 4 Comparative analysis of total node's death at the respective round number for all three different sizes of fields

S. No.	Parameter	Round Number		
		Field Size 100x100	Field Size 200x200	Field Size 300x300
1.	First Death	1463	1440	1331
2.	10^{th} Death	1823	1863	1856
3.	50^{th} Death	1891	1878	1884
4.	90^{th} Death	2827	2732	2674
5.	100^{th} Death	4334	4410	4083

If we examine Table 4, 1^{st}, 50^{th}, and $90^{th,}$ death occurs at round number 1463, 1891, and 2827, respectively, in the case of 100 x 100 size field, which is very much late compared to the rest two sized fields. But if we take a look at the 10th and 100^{th} death occurring at round number 1863 and 4410, respectively, field size 200 X 200 is very much late compared to the rest of the two field sizes. Also, from the overall analysis given in figure 7, figure 8, and table 4, it can be concluded that field size 100 X 100 is more favorable for all the nodes in the field as far as nodes lifetime is concerned.

Figure 9 to 11 shows the normalized average energy plot, total residual energy of the nodes, and nodes' total consumed energy for all three different base station speeds, respectively. Table 5 compares the different percentage levels of consumption, i.e., 1%, 10%, 50%, 90%, and 100%, at the respective round number for all three different base station speeds.

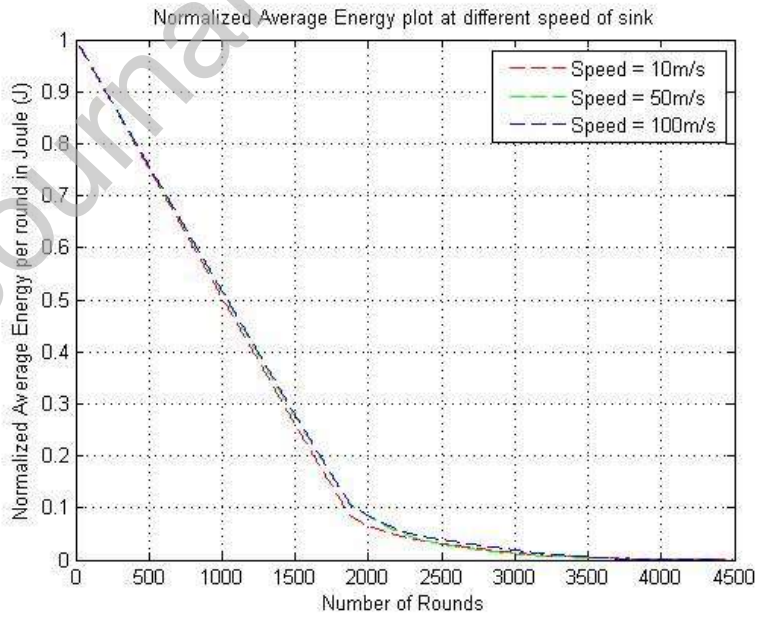


Figure 9 Normalized average energy plots of the nodes for all three different speed of the base station

Journal Pre-proof

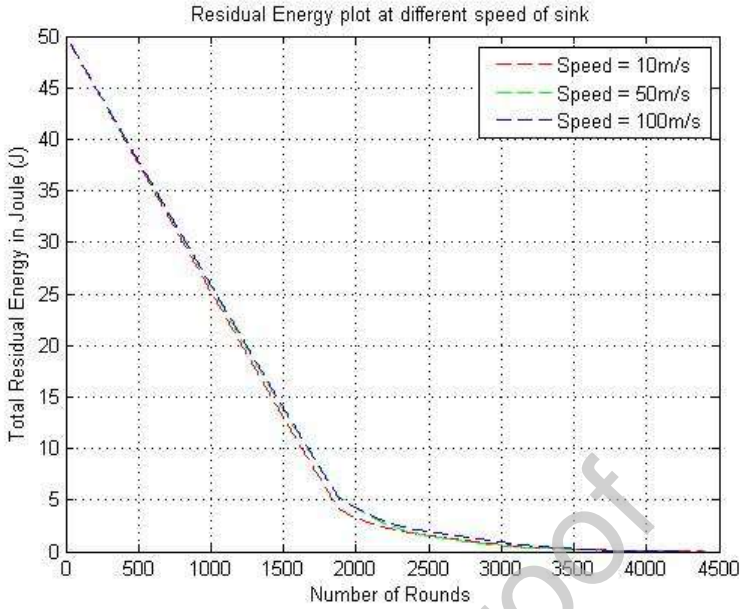


Figure 10 total residual energy of the nodes for all three different speed of the base station

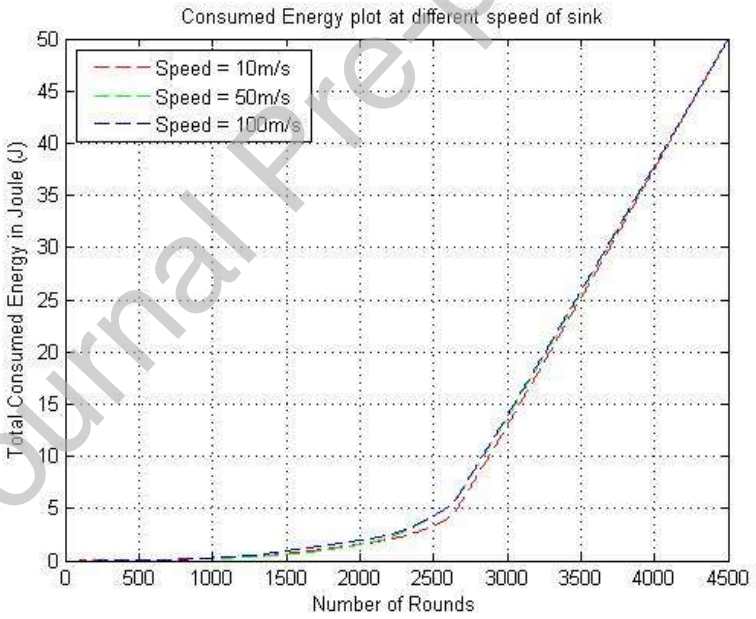


Figure 11 Total consumed energy of the nodes for all three different speed of the base station

Table 5 Comparative analysis of the different percentage level of consumption at the respective round number for all three different speeds of the base station

S. No.	Parameter	Round Number		
		Speed10 m/s	Speed 50 m/s	Speed 100 m/s
1.	1% Consumption	1363	1425	1238

Journal Pre-proof

2.	10% Consumption	2661	2585	2587
3.	50% Consumption	3494	3472	3466
4.	90% Consumption	4299	4297	4295
5.	100% Consumption	4500	4500	4500

If we examine Table 5, one percent energy is getting consumed in the 1425th round in base station speed 50m/s which are very much late compared to the rest two sized fields. But if we look at 10 – 90 percent energy consumption, base station speed 10m/s is very much efficient and favorable compared to the other two speeds. Also, from the overall analysis given in figure 9, figure 10, figure 11, and table 5, it can be concluded that base station speed of 10 m/s is more favorable for all the nodes in the field as far as energy consumption and residual energy and normalized energy are concerned.

Figure 12 and 13 shows the total dead nodes and total alive nodes for all three different base station speeds, respectively. Table 6 compares the total node's death, i.e., 1st death, 10th death, 50th death, 90th death, and 100th death, at the respective round number for all three different base station speeds.

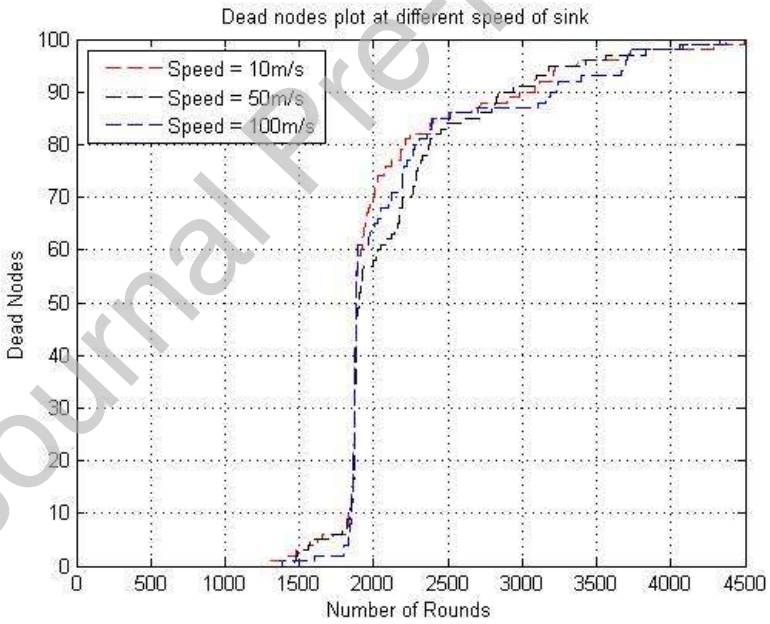


Figure 12 shows total dead nodes for all three different speed of the base station

Journal Pre-proof

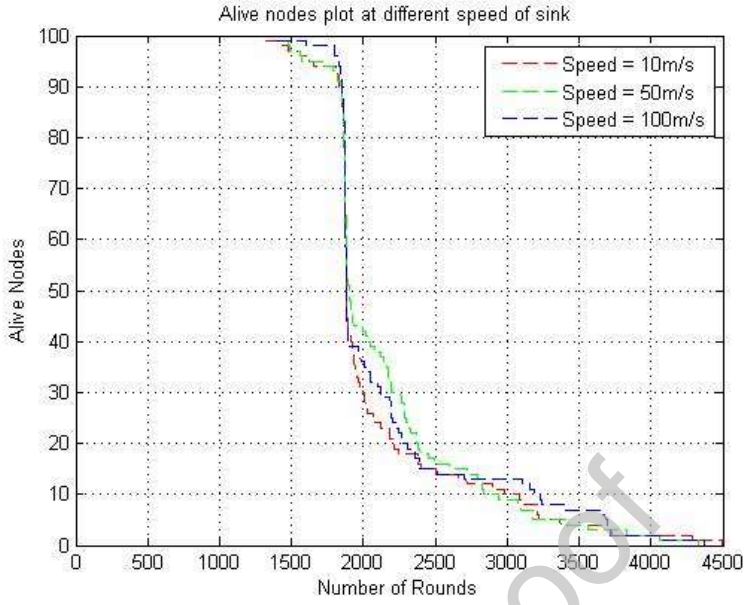


Figure 13 Total alive nodes for all three different speed of the base station

Table 6 Comparative analysis of total node's death at the respective round number for all three different speed of the base station

S. No.	Parameter	Round Number		
		Speed 10 m/s	Speed 50 m/s	Speed 100 m/s
1.	First Death	1309	1463	1385
2.	10^{th} Death	1833	1823	1847
3.	50^{th} Death	1880	1891	1879
4.	90^{th} Death	2896	2827	3158
5.	100^{th} Death	4497	4334	4076

If we examine Table 6, 1^{st} and 50^{th} death are occurring at round number 1463 and 1891, respectively, in the base station speed 50 m/s, which is very much late compared to the rest two sized fields. But if we take a look at 10^{th}, $90^{th,}$ and 100^{th} death occurring at round number 1833, 2896, and 4497, respectively, base station speed 10 m/s is very much late compared to the rest two base station speeds. Also, from the overall analysis given in figure 12, figure 13, and table 6, it can be concluded that base station speed 10 m/s is more favorable for all the nodes in the field as far as nodes lifetime is concerned.

A comparative graphical analysis is also given for the proposed method with three existing methods that are EERCCSM, CCMAR, and ECDRA. These comparisons are provided for the sensor nodes' total energy consumption and total alive nodes or lifetime concerning the increasing number of rounds. Figure 14 shows the total alive nodes for all four different

Journal Pre-proof

methods. Figure 15 shows the total energy consumption of nodes for all four different methods. Both of the proposed method plots are made with the single base station, field size 100m X 100m and base station speed 50 m/s.

Figure 14 Graphical comparative analysis of the proposed method with existing methods for Network lifetime / total alive nodes

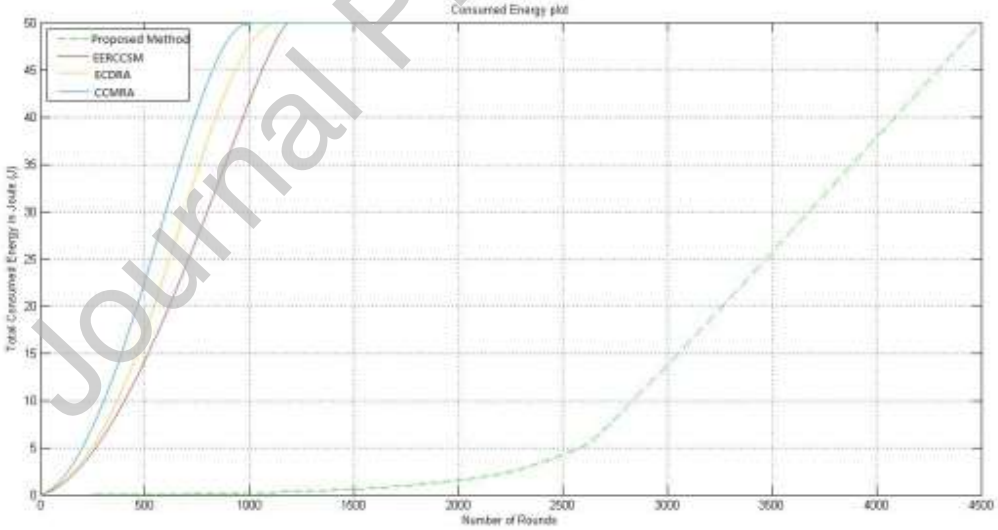


Figure 15 Graphical comparative analysis of the proposed method with existing methods for total energy consumption

After analyzing graphical figures 14 and 15, it is clear that the proposed method is much efficient as far as total energy consumption and network lifetime are concerned as existing methods lifetime and the total energy consumption is limited to 1300 rounds.

Journal Pre-proof

CONCLUSION AND FUTURE WORK

In the presented research work, we proposed a novel technique for WSN based on an advanced fixed path mobile sink energy-efficient PEGASIS based Protocol to maximize the network lifetime. The proposed method focuses on specified mobility of the base station to maintain the energy conservancy and to enhance the network lifetime. The proposed method implements a multiple-chain model to attain small size chains. This process decreases the overall load on the presented chain leaders at the initial stage and declines the sensor nodes' pressure at the end. The experimental work is very much supportive in reducing the hold in the delivery of the packets. It also declines the stretch between the linked nodes by using shorter chains. This base station's advancement has proven a significant superiority on static base station in prolonging the Network's lifetime. The proofs for the above statements are given in the experimental results section. The proposed method is also working efficiently and efficiently compared to existing routing protocols in total consumed energy and alive nodes of the Network. This novelty and advancement of the base station will open new horizons for getting it placed much closer to the cluster leader to prolong Network's lifetime to a great extent in the future. It is also necessary to apply the developed methodologies in several case studies to generalize the obtained results. In the present research, PEGASIS based routing protocol is used to implement the proposed method. The utilization of other routing protocols such as LEACH and DEEC can be considered to enhance the scope of the current work.

Journal Pre-proof

Conflict of interest

The authors declare that they do not have any conflict of interests. This research does not involve any human or animal participation. All authors have checked and agreed the submission.

REFERENCES

[1] YoungSang Yun and Ye Xia, "Maximizing the Lifetime of Wireless Sensor Networks with Mobile Sink in Delay-Tolerant Applications", IEEE Transactions on Mobile Computing, Vol. 9, No. 9, September 2010.

[2] Hemant Ghayvat, Subhas Mukhopadhyay , Xiang Gui and Nagender Suryadevara, "WSN- and IOT-Based Smart Homes and Their Extension to Smart Buildings" Sensors 2015, 15, 10350-10379; doi:10.3390/s150510350 sensors ISSN 1424-8220.

[3] Azimi I., Anzanpour A., Rahmani A.M., Pahikkala T., Levorato M., Liljeberg P., Dutt N. HiCH: hierarchical fog-assisted computing architecture for healthcare IoT. ACM Trans. Embed. Comput. Syst. (TECS) 2017;16(5s):174.

[4] Mustafa Kocakulak and Ismail Butun, "An overview of Wireless Sensor Networks towards internet of things", 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC).

[5] Weifa Liang, Jun Luo and Xu Xu, "Prolonging Network Lifetime via a Controlled Mobile Sink in Wireless Sensor Networks", IEEE Globecom 2010 proceedings.

[6] Feng Sen, Qi Bing and Tang Liangrui, "An Improved Energy-Efficient PEGASIS-Based Protocol in Wireless Sensor Networks", 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD).

[7] Rajni Meelu & Rohit Anand, "Performance Evaluation of Cluster-Based Routing Protocols used in Heterogeneous Wireless Sensor Networks", International Journal of Information Technology and Knowledge Management January-June 2011, Volume 4, No. 1, pp. 227-231.

[8] ZhouHaibo, WuYuanming and Xie Guangzhong, "EDFM：A Stable Election Protocol based on Energy Dissipation Forecast Method for Clustered Heterogeneous Wireless Sensor Networks", Conference Paper · October 2009, DOI: 10.1109/WICOM.2009.5304152 · Source: IEEE Xplore.

[9] Ajay Kaushik, S. Indu and Daya Gupta, "Adaptive Mobile Sink for Energy Efficient WSN Using Biogeography - Based Optimization", International Journal of Mobile Computing and Multimedia Communications Volume 10 • Issue 3 • July-September 2019.

[10] Stephanie Lmdsey and Cauligi S. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems", Updated Sept 29, 2001.

[11] Shailly Thakur and Balraj Singh, "Mobile Sink-Based Multi-Chain Pegasis Protocol for Improving the Lifetime of WSNs", International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering, Vol. 4, Issue 6, June 2016.

[12] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.

[13] Wendi B. Heinzelman, Member, IEEE, Anantha P. Chandrakasan, Senior Member, IEEE, and Hari Balakrishnan, "An Application-Specific Protocol Architecture for Wireless

Journal Pre-proof

Journal Pre-proof

Microsensor Networks", IEEE Transactions on Wireless Communications, VOL. 1, NO. 4, October 2002.

[14] JinWang, Yu Gao, Wei Liu, Arun Kumar Sangaiah and Hye-Jin Kim, "Energy Efficient Routing Algorithm with Mobile Sink Support for Wireless Sensor Networks", Sensors 2019, 19, 1494.

[15] Abo-Zahhad, M., Ahmed, S. M., Sabor, N. and Al-Ajlouni, A. F., (2015) "MS-Based Adaptive Immune Energy-Efficient Clustering Protocol for Improving the Lifetime and Stability Period of WSNs", IEEE Sensors Journal, vol. 15, no. 8.

[16] Gharaei N. Kamalrulnizam A. B., Hashim S. Z., Pourasl A. H., Butt S. A. (2018). "Collaborative Mobile Sink Sojourn Time Optimization Scheme for Cluster-Based WSN", IEEE Sensors Journal, Vol. 18, no. 16. Gonsalves, T. and Egashira, A., (2013). Parallel Swarms Oriented PSO, Applied Computational Intelligence and Soft Computing, vol. 2013.

[17] Barani H. Jaradat Y. Huang H. Li Z. Misra S. (2018). "Effect of sink location and redundancy on multi-sink wireless sensor networks: a capacity and delay analysis", IET Communications, vol. 12 no. 8, 941-947.

[18] Yung-Fa Huang, John Sum and Lin-huang Chang, "Lifetime Performance of an Energy Efficient Clustering Algorithm for Cluster-Based Wireless Sensor Networks", Conference Paper August 2007.

[19] Marios A. Gatzianas and Leonidas G. Georgiadis, "A Distributed Algorithm for Maximum Lifetime Routing in Sensor Networks with Mobile Sink", IEEE Transactions on Wireless Communications · April 2008.

[20] Kenneth Li-Minn Ang, Jasmine Kah Phooi Seng and Adamu Murtala Zungeru, "Optimizing Energy Consumption for Big Data Collection in Large-Scale Wireless Sensor Networks With Mobile Collectors" IEEE Systems Journal.

[21] Hidehisa Nakayama, Zubair Md. Fadlullah, Nirwan Ansari and Nei Kato, "A Novel Scheme for WSAN Sink Mobility Based on Clustering and Set Packing Techniques", IEEE Transactions on Automatic Control, Vol. 56, No. 10, October 2011.

[22] Maya M. Warriera, Ajay Kumar, "An energy-efficient approach for routing in Wireless Sensor Networks", RAEREST 2016.

[23] Kumar, S., Solanki, V. K., Choudhary, S. K., Selamat, A., & González Crespo, R. (2020). Comparative Study on Ant Colony Optimization (ACO) and K-Means Clustering Approaches for Jobs Scheduling and Energy Optimization Model in Internet of Things (IoT). International Journal of Interactive Multimedia & Artificial Intelligence, 6(1).

Journal Pre-proof

[24] González García, C., Núñez Valdéz, E. R., García Díaz, V., Pelayo García-Bustelo, B. C., & Cueva Lovelle, J. M. (2019). A review of artificial intelligence in the internet of things. International Journal Of Interactive Multimedia And Artificial Intelligence, 5.

[25] Dhiman, G., Garg, M., Nagar, A., Kumar, V., & Dehghani, M. (2020). A novel algorithm for global optimization: Rat swarm optimizer. Journal of Ambient Intelligence and Humanized Computing, 1-26.

[26] Dhiman, G., & Kumar, V. (2018). Emperor penguin optimizer: a bio-inspired algorithm for engineering problems. Knowledge-Based Systems, 159, 20-50.

[27] Dhiman, G., Oliva, D., Kaur, A., Singh, K. K., Vimal, S., Sharma, A., & Cengiz, K. (2021). BEPO: A novel binary emperor penguin optimizer for automatic feature selection. Knowledge-Based Systems, 211, 106560.

[28] Dehghani, M., Montazeri, Z., Malik, O. P., Dhiman, G., & Kumar, V. (2019). BOSA: Binary orientation search algorithm. Int. J. Innov. Technol. Explor. Eng.(IJITEE), 9, 5306-5310.

[29] Dehghani, M., Montazeri, Z., Dhiman, G., Malik, O. P., Morales-Menendez, R., Ramirez-Mendoza, R. A. & Parra-Arroyo, L. (2020). A spring search algorithm applied to engineering optimization problems. Applied Sciences, 10(18), 6173.

[30] Kaur, S., Awasthi, L. K., Sangal, A. L., & Dhiman, G. (2020). Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization. Engineering Applications of Artificial Intelligence, 90, 103541.

Journal Pre-proof

Journal Pre-proof

Author Profile

Ms Priyanka chugh(Shivanka) received her M.Tech(CSE) from Mullana University Ambala in 2010 and she has completed B.Tech(IT) from Kurukshetra University in 2005 . She has worked as Assistant Professor in Maharaja Surajmal Institute of Technology for 12 years (July 2007 to July 2019) in Information Technology department affiliated from GGSIPU New Delhi and She has also worked as Lecturer in M. M.E.C Mullana university for 6 month as well as lecturer in C.I.T.M college Faridabad for 6 month. She has received a letter of appreciation from SMES for excellent result in 2016 along with it she received 3rd prize for supervising the project for B.Tech Intercollege project competition in Nov 2013.Her research area in Wireless Sensor Network and she has published 3 research paper International reputed journal and 3research paper published in International conference in reputed conference and one book chapter in Springer .

Dr S. Indu received her Ph.d (ECE) in 2012 from Delhi University as well as she has completed her M.Tech (Control System) in 1990 from University of Kerala along with it She has completed her Bachelor of engineering(Electrical Engineering) degree in 1987 from University of Kerala. She has many achievement in her life like She was University topper for M.Tech andBranch Counselor - IEEE DTU Student Branch Aug 2012-Aug 2017 as well as Best Branch Counselor award 2014 IEEE USA and Outstanding Branch Counselor award for consecutive 5 years, IEEE Delhi Section along with it Best paper award ICCI-SEM-2017 (Computer session) and she has received a Commendable research award 2018 and Worked as General Chair ICSPVCE 2019

Dr.Meenu Gupta is an Assistant Professor (CSE-UIE) at the Department ofComputer Science Engineeringat Chandigarh University, Punjab, India. She has done her Ph.D. from Ansal University, Gurugram, in Computer Science in the year 2020. She also received an award of Best project display in 2006. She completed many certifications like a rational seed, Vb.Net, DBMS by oracle, DBA from Microsoft, and CDC from Govt. consultancy.

She also certified faculty of MTA and MOS. Her research interests are Cloud Computing, Blockchain, Big Data and IoT, Deep Learning and Machine Learning. A total of 12+ Years of Academic / Research Experience with more than 30+ Publications in various National, International Conferences and International Journals (Scopus/ESCI/SCIE/ SCI). She is a Lifetime member of educational society like ISTE, IAENG.

Dr. Gopal Chaudhary is currently working as an assistant professor in Bharati Vidyapeeth's College of Engineering, Guru Gobind Singh Indraprastha University, Delhi, India. He holds a Ph.D. in Biometrics at the division of Instrumentation and Control engineering, Netaji Subhas Institute of Technology, University of Delhi, India. He received the B.E. degree in electronics and communication engineering in 2009 and the M.Tech. degree in Microwave and optical communication from Delhi Technological University (formerly known as Delhi College of Engineering), New Delhi, India, in 2012. He has 30 publications in refereed National/International Journals & Conferences (Elsevier, Springer, Inderscience) in the area of Biometrics and its applications. His current research interests include soft computing, intelligent systems, information fusion and pattern recognition. He has organized many conferences and special issues.

Dr. Manju Khari an Assistant Professor in Ambedkar Institute of Advanced Communication Technology and Research, Under Govt. Of NCT Delhi affiliated with Guru Gobind Singh Indraprastha University, Delhi, India. She is also the Professor- In-charge of the IT Services of the Institute and has experience of more than twelve years in Network Planning & Management. She holds a Ph.D. in Computer Science & Engineering from National Institute Of Technology Patna and She received her master's degree in Information Security from Ambedkar Institute of Advanced Communication Technology and Research, formally this institute is known as Ambedkar Institute Of Technology affiliated with Guru Gobind Singh Indraprastha University, Delhi, India. Her research interests are software testing, information security, optimization, Image processing and machine learning. She has 70 published papers in refereed National/International Journals & Conferences (viz. IEEE, ACM, Springer, Inderscience, and Elsevier) and 10+ edited books from reputed publishers. She is also co-author of two books published by NCERT of Secondary and senior Secondary School.

Journal Pre-proof

Dr.S.Vimal is working in Department of Information Technology, National Engineering College, Kovilpatti, Tamilnadu, India. He has around Thirteen years of teaching experience, EMC certified Data science Associate and CCNA certified professional too. He holds a Ph.D in Information and Communication Engineering from Anna University Chennai and he received Masters Degree from Anna University Coimbatore. He is a member of various professional bodies and organized various funded workshops and seminars. He has wide publications in the highly impact journals in the area of Data Analytics, Networking and Security issues and published 04 book chapters. He has hosted two special session for IEEE sponsored conference in Osaka, Japan and Thailand. He has acted as Session chairs, organizing committee member, advisory committee and outreach committee member in various international conferences in IEEE and Other prestiguos Conferences. His areas of interest include Game Modelling, Artificial Intelligence, Cognitive radio networks, Network security, Machine Learning and Big data Analytics. He is a Senior member in IEEE and holds membership in various professional bodies. He has served as reviewer for Springer, Elsevier and Wiley journals. He has hosted 2 special issues in CMC tech science journals. He has served as Guest editor for SCI journals and editored 3 books in scopus indexed. He is a Senior Member in IEEE

Journal Pre-proof

arXiv:2211.12374v1 [cs.CL] 22 Nov 2022

AN EMOTION-AWARE MULTI-TASK APPROACH TO FAKE NEWS AND RUMOUR DETECTION USING TRANSFER LEARNING

Arjun Choudhry*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
choudhry.arjun@gmail.com

Inder Khatri*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
inderkhatri999@gmail.com

Minni Jain
Department of Computer Science
Delhi Technological University
New Delhi, India
minnijain@dtu.ac.in

Dinesh Kumar Vishwakarma
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
dinesh@dtu.ac.in

November 23, 2022

ABSTRACT

Social networking sites, blogs, and online articles are instant sources of news for internet users globally. However, in the absence of strict regulations mandating the genuineness of every text on social media, it is probable that some of these texts are fake news or rumours. Their deceptive nature and ability to propagate instantly can have an adverse effect on society. This necessitates the need for more effective detection of fake news and rumours on the web. In this work, we annotate four fake news detection and rumour detection datasets with their emotion class labels using transfer learning. We show the correlation between the legitimacy of a text with its intrinsic emotion for fake news and rumour detection, and prove that even within the same emotion class, fake and real news are often represented differently, which can be used for improved feature extraction. Based on this, we propose a multi-task framework for fake news and rumour detection, predicting both the emotion and legitimacy of the text. We train a variety of deep learning models in single-task and multi-task settings for a more comprehensive comparison. We further analyze the performance of our multi-task approach for fake news detection in cross-domain settings to verify its efficacy for better generalization across datasets, and to verify that emotions act as a domain-independent feature. Experimental results verify that our multi-task models consistently outperform their single-task counterparts in terms of accuracy, precision, recall, and F1 score, both for in-domain and cross-domain settings. We also qualitatively analyze the difference in performance in single-task and multi-task learning models.

***K*eywords** Multi-task Learning · Emotion Classification · Principal Component Analysis · Plutchik's emotions · Ekman's emotions · Cross-domain analysis.

1 Introduction

In recent years, we have witnessed a substantial increase in the usage of social networking sites, news services, and blogs. This increase in the user base, and thus, the usage of social media has brought along an increase in the spread of *fake news* and *rumours*. According to the work by Meel and Vishwakarma (2019), fake news and rumours both fall under *false information*. Fake news refers to the spread of false information under the pretext of being correct, while rumours are unverified pieces of information that may or may not be false. While there are noticeable differences in the

*Equal Contribution

semantics of a rumour and a fake news text, machine learning and deep learning approaches used for the detection of both remain largely similar. Further, both fake news and rumours have often been used interchangeably in research works. For example, Ajao, Bhowmik and Shahrzad. (2019) proposed a sentiment-aware approach to fake news detection, but tested their approach on the PHEME (Kochkina, Liakata and Zubiaga, 2018b) dataset, predominantly used for rumour detection. Thus, both tasks can be countered with common model architectures and solutions.

Due to the deceptive nature of fake news and rumours, it is difficult for people to understand which text is real, and which is fake (or a rumour). For example, Starbird, Maddock, Orand, Achterman and Mason (2014) found that people on social media were unable to distinguish between rumours and non-rumours for the 2013 Boston Marathon bombings. For every tweet tweeted debunking a rumour, as many as 44 tweets were tweeted in support of it. This proves the inability of users on social media to identify fake news and rumours. Hence, we can say that social media has been unable to keep fake news and rumours off their platforms. This calls into question the credibility of social media and the web as a source of information.

Significant research has been done in recent years towards improved detection of fake news and rumours. Researchers have focused on a variety of directions to implement models that can reliably detect fake news and rumours. Some works have focused on improved classifier architectures for better learning the semantic and syntactic features that differentiate fake news texts from real news texts. For example, Han, Karunasekera and Leckie (2020) used Graph Neural Networks with continual learning to differentiate between the propagation patterns of fake news and real news on social media. Other works have focused on multi-modal data to extract features from the text, as well as the accompanying multimedia. This has been shown to further improve the performance of classifiers for the detection of fake news and rumours. For example, Singhal, Shah, Chakraborty, Kumaraguru and Satoh (2019) proposed a multi-modal framework for the detection of fake news, and found that the performance of classifiers toward the detection of fake news was significantly improved when compared with models targeting a single modality. Singhal, Kabra, Sharma, Shah, Chakraborty and Kumaraguru (2020) further observed improved performance by using transfer learning for large transformer models.

Some research works have also focused on the extraction of specific features from the texts, which are found to aid them in the detection of fake news and rumours. For example, Kochkina, Liakata and Zubiaga (2018a) utilized veracity and stance detection features for the detection of rumours, and observed noticeably better performance from their classifiers, as compared to the classifiers trained without these features. Ajao et al. (2019) used the sentiment ratio extracted from the text using Latent Dirichlet Allocation and concatenated the extracted vector for the sentiment ratio with the feature matrix. They observed improved performance on the PHEME 6 dataset. Some recent works have also shown the correlation between the legitimacy of a text and its emotion (Vosoughi, Roy and Aral, 2018).

In this work, we verify the correlation between the legitimacy of a text and its intrinsic emotion and show that for the same emotion class, real news and fake news are often represented slightly differently. This work is the full paper for our abstract (Choudhry, Khatri and Jain, 2022) and contains suitable extensions for a more thorough evaluation. Based on the correlation between fake news and emotions, we further propose a multi-class emotion-aware multi-task approach for the detection of fake news and rumours, where classifiers implicitly learn features associated with fake news or rumour detection, as well as features associated with emotions portrayed by the text, with the help of transfer learning. We annotate datasets with their emotion class labels using transfer learning. We treat fake news detection and rumour detection as similar tasks and verify our approach's efficacy on the PHEME 9 dataset for rumour detection, FakeNews AMT dataset for multi-domain fake news detection, Celeb and Gossipcop datasets. Our experimental evaluations prove that implicit extraction of emotion-based features in a multi-task setting helps improve the model performance across a variety of model architectures and datasets, both in in-domain and cross-domain settings. Our contributions can be summarized as:

- We verify the correlation between the legitimacy of a text and its portrayed emotions by dimensionality reduction. We experimentally use this correlation for improved fake news and rumour detection by proposing a multi-task framework with fake news or rumour detection as the primary task, and multi-class emotion detection as the auxiliary task.

- We show that for the same emotion, rumours and non-rumours are often represented slightly differently in the feature space upon dimensionality reduction. These differences will vary from dataset to dataset due to differences in the domain and other biases. We suggest that these differences can aid the model to better incorporate this correlation for improved rumours and fake news detection.

- We further discover that emotion-guided implicit features are more domain-independent than features extracted by models without the use of emotions, due to improved performance observed in cross-domain fake news detection.

- We evaluate both Ekman's and Plutchik's emotion classes for fake news and rumour detection, showing that Ekman's emotion classes lead to similar or better performance than Plutchik's emotion classes. To the best of our knowledge, no previous work on fake news detection considers Ekman's emotion classes.
- We evaluate the performance of the single-task and multi-task models in in-domain and cross-domain settings for different deep learning architectures and datasets. Based on the results obtained, we verify that emotion-guided implicit features extracted by our multi-task models aid in cross-domain fake news and rumour detection.

The organization of our paper is as follows: Section 2 contains the Related Works and associated literature review, Section 3 contains the Proposed Methodology, Section 4 contains the Results and Discussion, Section 5 consists of the Broader Impact and Ethical Considerations regarding our work, Section 6 conveys the Current Limitation of Our Approach, and Section 7 contains the Conclusion.

2 Related Work

In recent years, significant work has been done to tackle the problem of *fake news* and *rumour detection*. Bhutani, Rastogi, Sehgal and Purwar (2019) used a combination of cosine similarity, Tf-Idf and sentiment scores as features for the detection of fake news. Pérez-Rosas, Kleinberg, Lefevre and Mihalcea (2018) presented two novel datasets, *FakeNews AMT* and *Celeb*, for fake news detection across multiple domains, and used hand-crafted linguistic features and an SVM model for fake news detection. Saikh, De, Ekbal and Bhattacharyya (2019) treated fake news detection as a text classification task and presented two deep learning models for fake news detection on FakeNews AMT and Celeb datasets.

Bhattacharya, Patel, Gupta, Tanwar and Rodrigues (2021) proposed a Blockchain and LSTM-based classification technique incorporating Q-GloVe word embeddings for the detection of fake news. They efficiently used Blockchain to protect real news from adversaries to prevent its alteration. Verma, Agrawal, Amorim and Prodan (2021) curated a new generalized dataset by combining four previously available benchmark datasets, to reduce bias associated with domains specific to each dataset. They generated embeddings from the linguistic features of the combined dataset and were used for classification using an ensemble of multiple machine learning classifiers. Shahid, Jamshidi, Hakak, Isah, Khan, Khan and Choo (2022) further presented a detailed survey of topics, challenges, and future areas of research associated with fake news detection, providing a detailed overview of dissemination, detection, and mitigation of fake news. They further focused on a specific potential future research agenda, i.e., using artificial intelligence explainable fake news credibility system.

Some contemporary works have studied rumour detection from the direction of network sciences, focusing on its evolution and impact on users. Cheng, Yin, Nazarian and Bogdan (2021b) modelled a misinformation network for the spread of rumours during the COVID-19 pandemic. Each node in the network represented a different rumour, while the edges indicated the similarity between the rumours. This work showed that information propagation could be controlled by deleting some central nodes in the misinformation graph, and further proposed a deep learning framework to predict Twitter posts that could act as central nodes. Verma, Bhardwaj, Aledavood, Choudhury and Kumar (2022), on the other hand, studied the impact of misinformation about the COVID-19 pandemic on the mental health of people and found that individuals who consumed a higher amount of misinformation suffered from high stress and anxiety, and exhibited an increased tendency for suicidal thoughts.

Recent research works have incorporated advanced mainstream deep learning approaches like Generative Adversarial Networks (GANs) and Transformers for fake news detection and rumour detection to train detectors in an unsupervised manner. Cheng, Li, Nazarian and Bogdan (2021a) proposed a GAN-based layered model for rumour detection, which was trained in a label-free manner without using any verified database. The Generator and the discriminator were trained adversarially in a sequential manner, where the generator produced rumours by fabricating the non-rumours, and the discriminator learnt to detect rumours and non-rumours with the help of the produced rumours. Roy, Bhanu, Saxena, Dandapat and Chandra (2022) proposed a transformer-based architecture with a modified multi-head attention mechanism used to learn multiple contextual correlations for enhancing feature representation learning.

Kochkina et al. (2018a) proposed a branch LSTM-based multi-task learning model trained using veracity and stance detection as auxiliary tasks to rumour detection. The model trained with both veracity and stance as auxiliary tasks outperformed single-task models, as well as multi-task models with either veracity or stance as an auxiliary task. Building upon this work, Cheng, Nazarian and Bogdan (2020) also used multi-task learning for the rumour classification, but unlike Kochkina et al. (2018a), they tackled all four component tasks related to rumour detection using the same model: rumour detection, rumour tracking, stance classification, and veracity classification. To improve the overall performance on the component tasks, they utilized an LSTM-based Variational-Autoencoder (VAE) to get

classifier-friendly latent representations. Our proposed approach also leverages multi-task learning for rumour and fake news detection. While the above works utilize multi-task learning to tackle the different component tasks relate to rumour detection simultaneously, we exploit multi-task learning to improve implicit feature representation using emotion classification as an auxiliary task.

Various works in fake news and rumour detection have shown the efficacy of using sentiment analysis and text polarity for fake news and rumour detection. Ajao et al. (2019) calculated *emoratio* (ratio of negative polarity to positive polarity in text) for the PHEME dataset, showing a significant difference in values for rumours and non-rumours. Augmentation of emoratio into the feature matrix showed noticeable improvements to their model's performance. Yang, Zheng, Zhang, Cui, Li and Yu (2018) showed that real news has higher median values and lower standard deviation for both positive and negative sentiments than fake news. The median values for negative sentiment for fake news were also higher than the median values for positive sentiment. Both indicate that more negative sentiment is associated with fake news.

Some recent works have started proving that there exists an affinity between certain emotions and the legitimacy of news, and this affinity can be leveraged for better detection of fake news. K, P and L (2020) introduced a technique to develop emotion-amplified (*emotionized*) text representations for the detection of fake health news articles, based on the assumption that fake and real health news articles portray a different affective character, and this can be used to aid in the detection of fake news. Vosoughi et al. (2018) analyzed all of the verified fake and real news stories propagated on Twitter from 2006 to 2017 and labeled them as real or fake using information extracted from six fact-checking organizations. Upon evaluating the replies on each of these stores using the lexicon curated by the National Research Council Canada (NRC) based on Plutchik's basic emotions, they detected that false stories typically garnered surprise, disgust and fear in replies, while real stories garnered anticipation, joy, trust and sadness.

Fersini, Armanini and D'Intorni (2020) proposed an approach incorporating the modeling of users based on four characteristics, i.e. stylometry, personality, emotions, and feed embeddings, for the detection of fake news spreaders for the Profiling Fake News Spreaders on Twitter shared task at PAN 2020 (Pardo, Giachanou, Ghanem and Rosso, 2020). They observed that stylometry, emotions, and embeddings contributed positively to the detection capability of the model. Kumari, Ashok, Ghosal and Ekbal (2021) proposed a common framework for novelty detection, sentiment analysis, emotion analysis, and fake news detection using a hybrid LSTM and BERT-based model. They map the emotions into two classes, *emotion true* and *emotion false*, which are used in their framework. Guo, Cao, Zhang, Shu and Yu (2019) proposed a novel Emotion-based Fake News Detection (EFN) framework utilizing the emotions extracted from both the content and the comments, i.e., the emotions portrayed by the publishers and the users respectively, apart from the content, for the detection of fake news. Zhang, Cao, Li, Sheng, Zhong and Shu (2021) further built upon the work by Guo et al. (2019), by extracting the *Emotion Gap* between the *publisher emotion* and the *social emotion*, and further proposed a *Dual Emotions Features* set for the detection of fake news. Most of the above emotions-based works present analytical studies about the correlation between fake news and emotions on different datasets. Only a few of them actually use emotions for detection. Further, most of these works which utilize emotions to aid in rumour or fake news detection, take the emotion labels as input or exploit the emotional inconsistencies between the main content and the comments. This makes them highly dependent on additional inputs like emotion labels, comments, and thread information for evaluation. Our work, in contrast, intends to align the feature representations according to the emotion labels in a multi-task learning setup, which does not require any additional inputs during inference time.

Some very recent works, done parallelly to our original abstract (Choudhry et al., 2022), incorporate multi-task learning and utilize emotion recognition as an auxiliary task to aid in their main classification tasks. Saha, Upadhyaya, Saha and Bhattacharyya (2022) proposed a combined sentiment and emotion-aided framework for tweet act classification. They utilized the multi-modal feature extractors to obtain the textual and emoticon features. Another work (Kamila, Hasanuzzaman, Ekbal and Bhattacharyya, 2022) investigated the impact of emotions on the temporal orientation of tweets, and further proposed a multi-task learning-based approach to leverage the influence of emotions on the temporal orientation of the text. It parallelly trained the CNN and Bi-GRU networks to learn the shared input embeddings for the primary task classifier, i.e., to detect the temporal orientation among three the classes (past, present, and future), and the auxiliary task classifier, i.e., to detect emotions among four labels (joy, sadness, anger or fear). Although these works utilize emotion classification as an auxiliary task in multi-task frameworks and are similar to our proposed approach in terms of architecture, their base problems (tweet act classification and temporal orientation of tweets) are significantly different from our task.

We observe that, contrary to the claim by Vosoughi et al. (2018), rumours and non-rumours cannot always be associated with specific types of emotions, as these may vary depending on the dataset. Thus, for smaller datasets, it may not be correct to generalize rumours being associated with specific emotion classes respectively. We incorporate multi-class emotion classification to aid in fake news and rumour detection on various datasets for in-domain and cross-domain evaluation. We further use the basic emotions considered by psychologists defined by Ekman and Plutchik respectively, which form the complete set of basic emotions, in relation to fake news detection. We portray the efficacy of using

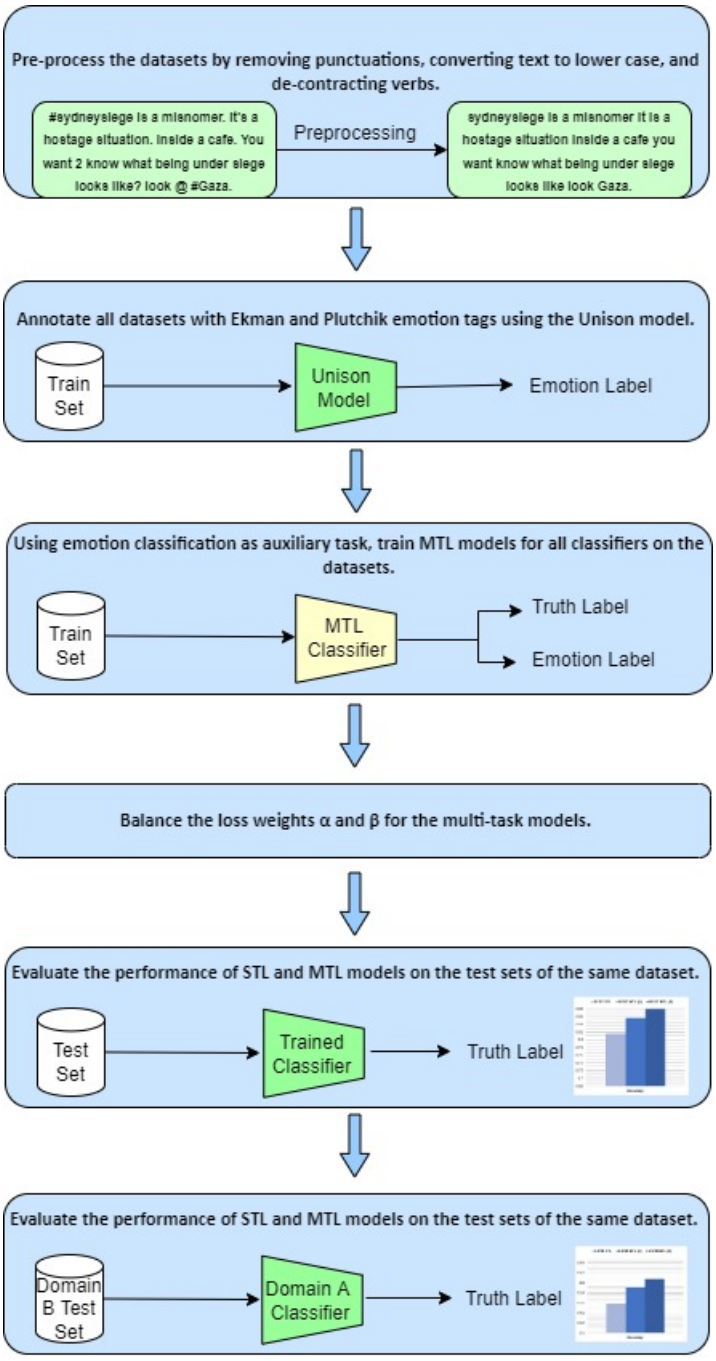


Figure 1: Flowchart representing our proposed methodology.

multi-class emotion classification as an auxiliary task, which aids in generalizing a model for use over a wider set of datasets.

3 Proposed Methodology

This section elaborates on our proposed methodology and is structured as follows: Subsection 3.1 elucidates the dataset preprocessing steps we used. Subsection 3.2 elaborates our approach for choosing the basic emotion classes for annotation and annotating the datasets. Subsection 3.3 explains the intuition behind the use of emotions as an

auxiliary feature for fake news and rumour detection and shows why they aid in improving the performance of classifiers. Subsection 3.4 elaborates upon the multi-task learning framework we used for various classifiers to incorporate emotion features into the model. Subsection 3.5 elucidates the process for cross-domain analysis, the reason why we evaluated cross-domain results, and their implication. Figure 1 shows a general overview of our work.

3.1 Preprocessing the Data

In this work, we use the PHEME 9 (Kochkina et al., 2018b), FakeNews AMT (Pérez-Rosas et al., 2018), Celeb (Pérez-Rosas et al., 2018) and Gossipcop datasets for evaluating our approach. We preprocess all the datasets using the same approach. We first convert all text to lowercase. We define a contracted verb's dictionary with general as well as specific cases that commonly occur in normal use. Using this contracted verb's dictionary, we de-contract verb forms to their root forms (eg. *"I'll"* to *"I will"*. We then remove all punctuation marks and numbers. Finally, we tokenize the text and store it in a list for further use.

3.2 Emotion Classes & Annotating the Datasets

Incorporating multi-class emotion classification as an auxiliary task in our multi-task framework poses two challenges: choosing a set of basic emotions, and annotating the datasets with the chosen emotion classes. There are two widely accepted theories on *basic* emotion classes, proposed by Plutchik (1982) and Ekman (1992) respectively. Plutchik (1982) designed a wheel with 8 basic emotions (*Joy*, *Surprise*, *Trust*, *Anger*, *Anticipation*, *Sadness*, *Disgust*, *Fear*) based on his *Ten Postulates*. Combinations of these 8 basic emotions can be used to form 56 other emotions at one intensity level. Previous psychological works (Vosoughi et al., 2018) on the correlation between emotions and fake news have made use of Plutchik's basic emotion classes using lexicon-based approaches. During a cross-cultural study, Ekman (1992) inferred that there are 6 basic emotions (*Joy*, *Surprise*, *Anger*, *Sadness*, *Disgust*, *Fear*). Each emotion was considered a discrete category, rather than an individual emotion. Bann and Bryson (2014) later found Ekman's basic emotion set to be the most semantically distinct.

We incorporate the use of Ekman's basic emotions in our work as well as Plutchik's basic emotions and see if they lead to significantly different results in the detection of fake news. Due to the unavailability of fake news datasets annotated with emotion classes, we use the *Unison model*[1] (Colneric and Demsar, 2018) to generate both Plutchik and Ekman emotion tags to annotate our datasets. The Unison model achieved state-of-the-art results for the classification of Twitter data into their respective emotions. Colneric and Demsar (2018) used a dataset of 73 billion tweets to train and test their approach. Annotation using the Unison model enables us to implicitly transfer some emotion-based features from the Unison model to our classifiers when trained in a multi-task setting. While the Unison model achieves an F1-score of around 0.70 for various basic emotion classes, it is a relatively robust model trained over 73 billion tweets from a variety of domains. It is also the only available model which can annotate for both Ekman's and Plutchik's basic emotion classes, an important factor for evaluating any differences between the two emotion sets. Further, as it is trained on an extremely large real-world dataset, it is less likely to show biases than smaller models trained on smaller corpora. However, in the future, as better and more robust emotion classification models become available, we recommend their use to minimize misannotations of emotion tags and further optimize performance. We compare the performance of various classifiers in multi-task settings on both emotion sets.

3.3 Intuition behind use of Emotions

We incorporate emotion-guided features in our fake news detection model by implementing a multi-task framework incorporating a multi-class emotion classification task in addition to the original fake news or rumour detection task. We further hypothesize that there is a prominent correlation between the legitimacy of a text and its intrinsic Ekman's and Plutchik's emotion, and within the same emotion, real and fake texts are sometimes represented differently. We further suggest that this correlation can positively affect a classifier's capability in the detection of fake news or rumours.

Figure 2 represents 3-dimensional Principal Component Analysis (PCA) graphs for the PHEME dataset, computed using embeddings generated from the Unison model for Ekman emotions. The 3-dimensional principal components are computed over the whole PHEME dataset but are plotted separately for rumours and non-rumours. If there wasn't a correlation between emotions and the legitimacy of text, then the two graphs obtained should have been nearly identical, except for the volume of sample points. However, we observe some stark differences between the plots for rumours and non-rumours.

Non-rumours show better-formed clusters and a higher percentage of *Joy* and *Fear* than rumours, which are more scattered and show a higher percentage of *Sadness* and *Surprise*. A similar conclusion can be drawn from Figure 3,

[1] https://github.com/nikicc/twitter-emotion-recognition

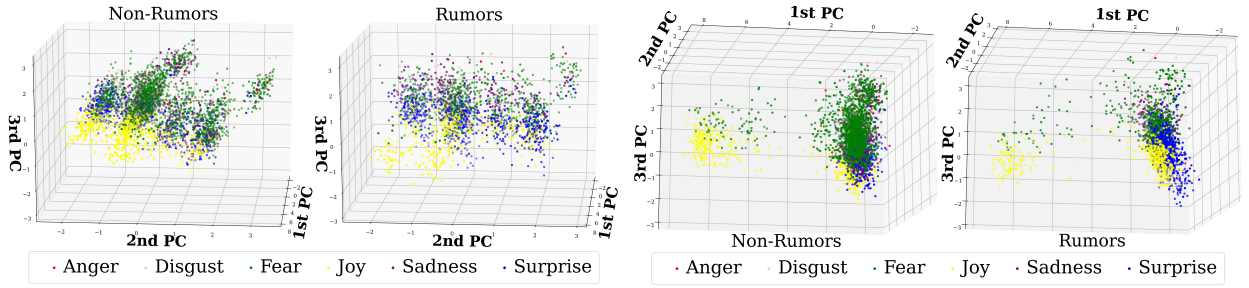


Figure 2: 3-dimensional Principal Component Analysis (PCA) plots for Rumours and Non-rumours from the PHEME 9 dataset, colored by Ekman emotions. ***(A)***: Upper View ***(B)***: Lower View

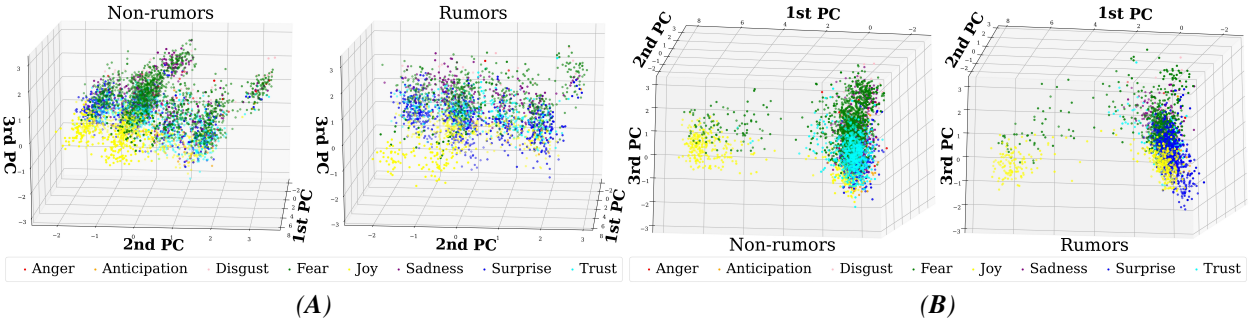


Figure 3: 3-dimensional Principal Component Analysis (PCA) plots for Rumours and Non-rumours from the PHEME 9 dataset, colored by Plutchik emotions. ***(A)***: Upper View ***(B)***: Lower View

which represents 3-dimensional Principal Component Analysis (PCA) graphs for the PHEME dataset, computed using embeddings generated from the Unison model for Plutchik emotions. Non-rumours show better-formed clusters and a higher percentage of *Trust* and *Fear* than rumours, which are again more scattered, with a higher percentage of *Sadness* and *Surprise*.

We further observe that for the same emotion, rumours and non-rumours clusters are plotted differently. Thus, even when two texts portray the same emotion, they are sometimes represented differently in the feature space for rumours and non-rumours. While the distribution of texts into emotion classes will vary from dataset to dataset, primarily due to differences in domains and various biases in each dataset, the fact that distributional and positional variations in the feature space occur between rumours and non-rumours for the same emotion should hold true for a variety of datasets, as these help form better decision boundaries for improved rumour or fake news detection in our multi-task models, leading to higher performance. The comparative non-emotion models, in comparison, perform worse and aren't able to correctly label texts as rumours or non-rumours (or fake and real news) as they are not guided by the variations in emotions between the classes. This indicates that a correlation exists between the legitimacy of a text and its underlying emotion, thereby supporting our proposition that emotion tags can aid in rumour detection and fake news detection.

3.4 Multi-task Learning for Fake News Detection

Multi-task learning (MTL) uses shared representations between the primary and auxiliary tasks for better extraction of common features, which can otherwise get ignored, and to generalise better on the primary task. We train our models to predict both the legitimacy and emotion of a text to verify the relationship between its legitimacy and emotion and evaluate them against their single-task (STL) counterparts. We further leverage the domain tags for texts in FakeNews AMT dataset and evaluate the performance of multi-task learning models predicting the legitimacy and domain of the text, to compare with multi-task learning models predicting the legitimacy and emotion of the text. Figure 4 shows the multi-task learning architecture we use, with fake news detection as the primary task, and emotion classification as the auxiliary task.

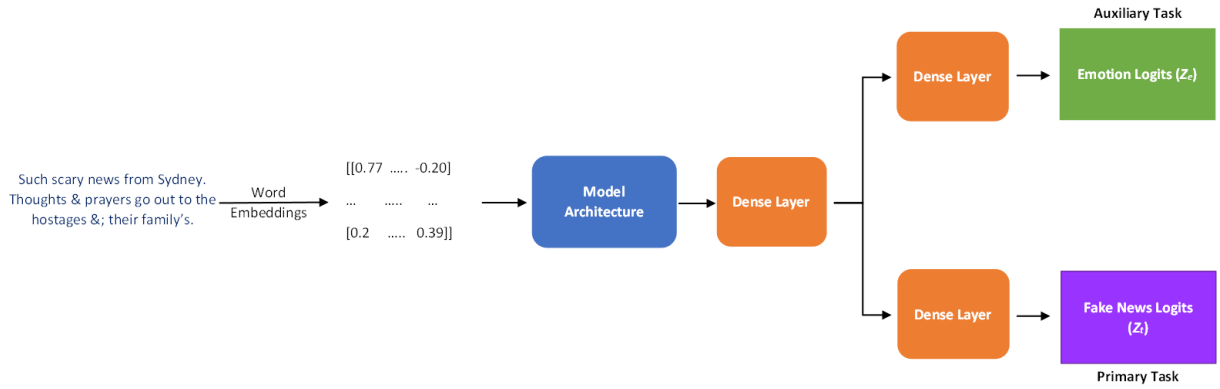


Figure 4: Generalized multi-task learning framework with fake news detection as the primary task and multi-class emotion classification as the auxiliary task.

3.5 Cross Domain Analysis

Previous works by Pérez-Rosas et al. (2018) and Saikh et al. (2019) show that classifiers exhibit a significant drop in performance when training the classifier on a dataset from one domain and testing it on a dataset from a different domain, as compared to its performance on the test set from its own domain. This is primarily due to the implicit extraction of domain-dependent features by the models, which reduces its efficacy in correctly detecting fake news from other domains. We evaluate the performance of LSTM and CNN-LSTM models in both single-task and multi-task settings for cross-domain analysis by training them on FakeNews AMT dataset and testing on Celeb dataset and vice versa. This helps us verify our hypothesis of the applicability of emotions for better generalization across domains and datasets for the detection of fake news.

4 Experimental Analysis & Results

This section describes the experimental analysis, results, and implications. Subsection 4.1 describes the datasets that we have used. Subsection 4.2 describes our α-β tuning approach applied to ensure optimal performance in multi-task settings. Subsection 4.3 elucidates the various standard classification models we have used with our framework to show the wide applicability of our approach. It further explains how we have balanced the loss weights of the two branches in the multi-task learning models for optimum performance. Subsection 4.4 contains the performance metrics we have used to evaluate the classifiers. Subsection 4.5 contains the results and the implications we draw from them.

4.1 Datasets

We have used the PHEME 9 (Kochkina et al., 2018b), FakeNews AMT (Pérez-Rosas et al., 2018), Celeb (Pérez-Rosas et al., 2018) and Gossipcop[2] datasets for evaluating our approach in this paper. The PHEME 9 dataset contains rumours and non-rumours posted on Twitter for 9 different events. Table 1 illustrates the distribution of the PHEME 9 dataset. The FakeNews AMT dataset is a multi-domain fake news dataset covering business, education, entertainment, politics, sports, and technology, while the Celeb dataset contains news about celebrities only. The Gossipcop dataset contains entertainment news articles scrapped from the web, containing over 12,000 samples. Hereon, the terms rumour and non-rumour are used specifically in the context of the PHEME 9 dataset or its evaluation, while the term fake news is used specifically in the context of the FakeNews AMT, Celeb and Gossipcop datasets, or their evaluation.

4.2 Parameter Tuning for Multi-task Models

We optimize the performance of the multi-task learning models by rebalancing the weights for the loss function for the primary and auxiliary branches of the models. We take the loss weight for the emotion classification branch as α, and the loss weight for the fake news or rumour detection branch as β. The relation between α and β is in Equation 1:

$$\beta = 1 - \alpha \tag{1}$$

[2]https://www.gossipcop.com/

Table 1: Distribution of the PHEME 9 dataset by the event. The total number of non-rumours are 67 percent more than the number of rumours.

Events	Threads	Tweets	Rumours	Non-rumours
Charlie Hebdo	2,079	38,268	458	1,621
Sydney Siege	1,221	23,996	522	699
Ferguson	1,143	24,175	284	859
Ottawa shooting	890	12,284	470	420
Germanwings-crash	469	4,489	238	231
Putin missing	238	835	126	112
Prince Toronto	233	902	229	4
Gurlitt	138	179	61	77
Ebola Essien	14	226	14	0
Total	6,425	105,354	2,402	4,023

The primary branch, tasked with the detection of the legitimacy of the text, gives a logits vector Z_t of size 2×1 as output, as shown in Equation 2. The auxiliary branch, which detects the Ekman or Plutchik emotion portrayed by the text, gives a logits vector Z_e of size 6×1 (for Ekman emotion classes) or 8×1 (for Plutchik emotion classes) as output, as shown in Equation 3.

$$Z_t = \begin{bmatrix} Z_t^0 \\ Z_t^1 \end{bmatrix} \tag{2}$$

$$Z_e = \begin{bmatrix} Z_e^0 \\ Z_e^1 \\ \vdots \\ Z_e^{n-1} \end{bmatrix} \tag{3}$$

Z_e and Z_t are further activated using the Softmax activation function to generate the probability distributions $\hat{y_t}$ and $\hat{y_e}$ for the legitimacy classes and emotion classes respectively. $\hat{y_t}$ is represented as shown in Equation 4.

$$\hat{y_t} = \begin{bmatrix} \hat{y_t^0} \\ \hat{y_t^1} \end{bmatrix} = Softmax(Z_t) \tag{4}$$

where $\hat{y_t^j}$ is defined as Equation 5.

$$\hat{y_t^j} = \frac{e^{z_t^j}}{\sum_{k=0}^{1} e^{z_t^k}} \tag{5}$$

Similarly, for the emotion branch, we get $\hat{y_e}$ as shown in Equation 6.

$$\hat{y_e} = \begin{bmatrix} \hat{y_e^0} \\ \hat{y_e^1} \\ \vdots \\ \hat{y_e^{n-1}} \end{bmatrix} = Softmax(Z_e) \tag{6}$$

where $\hat{y_e^j}$ is defined as in Equation 7.

$$\hat{y_e^j} = \frac{e^{z_e^j}}{\sum_{k=0}^{n} -1 e^{z_e^k}} \tag{7}$$

Further, we use cross entropy loss to penalize our model for both the fake news detection branch and the emotion detection branch, as shown in Equations 8 and 9.

$$Loss_t = -\sum_{i=0}^{1} \hat{y_t^i} * log(\hat{y_t^i}) \tag{8}$$

$$Loss_e = -\sum_{i=0}^{n-1} \hat{y_e^i} * log(\hat{y_e^i}) \tag{9}$$

The total loss for the multi-task learning models is defined in Equation 10.

$$Loss_(Total) = \alpha \times Loss_e + \beta \times Loss_t \tag{10}$$

Therefore, using Equations 1, 8, 9 and 10, we get 11.

$$\begin{aligned} Loss_{Total} = \alpha \times -\sum_{i=0}^{n-1} \hat{y_e^i} * log(\hat{y_e^i}) \\ +(1-\alpha) \times -\sum_{i=0}^{1} \hat{y_t^i} * log(\hat{y_t^i}) \end{aligned} \tag{11}$$

4.3 Classification Models and Hyperparameters

We have evaluated the performance of a variety of commonly used deep learning architectures in STL and MTL settings to show the wide applicability of our approach. Across all datasets, we have trained LSTM, CNN-LSTM, and BERT (Devlin, Chang, Lee and Toutanova, 2019) models in both MTL and STL settings. The CNN layer in the CNN-LSTM model aids in mirroring the local and position invariant features and preserving n-gram representations, while the LSTM layer helps to capture the long-term dependencies between word sequences better. For the PHEME 9 dataset, we have also evaluated the performance of CNN, CapsuleNet (Sabour, Frosst and Hinton, 2017) and HAN (Yang, Yang, Dyer, He, Smola and Hovy, 2016) models, which have previously been used for rumour detection (Guo, Cao, Zhang, Guo and Li, 2018). Comparison across multiple models enables us to better judge the efficacy of using emotion classification as an auxiliary task in a multi-task learning setup for fake news and rumour detection. This further ensures that our approach is applicable across a variety of architectures, and isn't just dependent on the model. We have made our code available online.[3]

We used the pre-trained *bert-base-uncased* model, available in the Huggingface Transformers (Wolf, Debut, Sanh, Chaumond, Delangue, Moi, Cistac, Rault, Louf, Funtowicz, Davison, Shleifer, von Platen, Ma, Jernite, Plu, Xu, Scao, Gugger, Drame, Lhoest and Rush, 2020) library, and further fine-tuned it for our task. Additional dense layers with softmax activation have been added to the ends of each branch, to predict the legitimacy of the text and its emotion. The single-task and multi-task BERT models were trained for 5 epochs with a batch size of 32 on the PHEME 9 dataset, while for the FakeNews AMT and Celeb datasets, the BERT models have a batch size of 64, and were trained for up to 50 epochs with early stopping to prevent over-fitting.

The CNN layer in the CNN and CNN-LSTM models consists of 32 filters with a filter size of 5, followed by a max pooling layer. The LSTM layer in the LSTM and CNN-LSTM models has 100 units. The CNN, LSTM and CNN-LSTM models were trained for 5 epochs, with a batch size of 64. The HAN and CapsuleNet models were trained for 5 epochs, with a batch size of 32. All models were trained using the Adam optimizer, with a learning rate of 2e-5 and cross-entropy loss function. The models trained for evaluation on the Gossipcop dataset follow the same parameters as those for PHEME 9.

During loss weights rebalancing, we took the initial value of α as 0.20, and after the training and testing of a model for a given set of values for α and β, we incremented α by 0.05, consequently decreasing β by 0.05. We trained each model for thirteen values of α and β, till α reaches 0.80, and consequently β reaches 0.20. Figure 5 illustrates the line graphs showing the variation in the accuracy of CNN, LSTM, and CNN-LSTM models on the PHEME 9 dataset with variations in α and β for Ekman's and Plutchik's emotions.

4.4 Performance Metrics

To compare the performance of the classifiers in single-task and multi-task settings, we calculated the metrics accuracy, precision, recall and F1-score for all models.

To calculate these metrics, we need to first determine the number of True Positives, True Negatives, False Positives, and False Negatives for our models using the confusion matrix. These metrics are defined as:

[3] https://github.com/Arjun7m/Emotion-Aware-MTL-FND-TCSS

- True Positive: Test samples whose labels were correctly predicted as Positive.
- True Negative: Test samples whose labels were correctly predicted as Negative.
- False Positive: Test samples whose labels were incorrectly predicted as Positive, but were in fact Negative.
- False Negative: Test samples whose labels were incorrectly predicted as Negative, but were in fact Positive.

Based on the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), accuracy, precision, recall and F1-score are defined as in Equations 12, 13, 14 and 15.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{15}$$

Accuracy is used to evaluate the model performance across all classes when all classes are given equal importance. Mathematically, it represents the ratio of the number of correct predictions by a model to the total number of samples in the testing set.

Precision exhibits a model's performance in classifying a sample as positive. It is a measure of *exactness* and tends to reduce if the number of samples in the minority classes is increased. Mathematically, it represents the ratio of the number of labels that are correctly predicted as positive to those that are labeled as positive.

Recall exhibits a model's performance in of correctly predicting true positives (TPs). It tends to increase if the number of samples in the minority classes is increased. Mathematically, it represents the ratio of the number of samples that are correctly predicted as positive to those that are actually positive.

F1-score acts as an alternative to accuracy, as it takes into account precision and recall scores of a model equally, but doesn't require the value of the total number of observations. Mathematically, F1-score is represented as the harmonic mean of precision and recall.

4.5 Results and Discussion

We evaluated the performance of our proposed framework on three datasets and several deep learning models, using the same train-test split for each dataset across all classifiers. Table 3 illustrates the results of our experiments on the PHEME 9 dataset, Table 2 compares the results of our MTL models with their STL counterparts on the FakeNews AMT and Celeb datasets, Table 5 shows the results of our experiments on the Gossipcop dataset, and Table 4 compares the performance of our multi-task learning models with their single-task counterparts in cross-domain settings. Some interesting findings observed are:

MTL models outperform their STL counterparts On comparing the performance of various classifiers in both single-task and multi-task settings, we observed an improvement in the performance of the classifiers in multi-task settings over the single-task setting across most metrics and on all datasets, verifying the correlation between the legitimacy of a text and its intrinsic emotion. The LSTM and CNN-LSTM models on the Celeb dataset unexpectedly showed slightly worse performance in the multi-task settings with Plutchik emotions compared to the single-task setting. However, the same models in the multi-task settings with Ekman's emotions outperform their single-task counterparts handily in terms of accuracy, recall, and F1 score. Further, the BERT model trained in the multi-task setting with Plutchik's emotions handily outperforms its multi-task counterpart with Ekman's emotions and its single-task counterpart across all metrics.

MTL models with Ekman's and Plutchik's emotions perform comparably We observed largely comparable performance between multi-task models trained on Ekman's and Plutchik's emotions across all datasets, except for LSTM and

CNN-LSTM models on the Celeb dataset, which unexpectedly performed worse with Plutchik's emotions. Multi-task models with Ekman's emotions generally performed slightly better on the FakeNews AMT, Celeb, and Gossipcop datasets while multi-task models with Plutchik's emotions performed slightly better on the PHEME 9 dataset. However, multi-task models trained with either emotion set generally outperformed the single-task models across most metrics on all datasets. Thus, we verify the correlation between the legitimacy of a text and its portrayed Ekman emotion and concur that Ekman's emotions can also act as an alternative to Plutchik's emotions as a feature for fake news detection, particularly for fake news detection datasets.

MTL models with domain classification as auxiliary task outperform their STL counterparts Classifiers trained to identify both domain and legitimacy of news (*MTL (Domain)* in Table 2) outperformed their single-task counterparts on the FakeNews AMT dataset, achieving performance comparable to those of emotion-based MTL models. This indicates a deeper correlation between the legitimacy of news and its domain. We intend to further build upon this correlation in future work for building a more generalizable framework.

Performance drops in cross-domain settings for all models, but MTL models outperform their STL counterparts We observed an all-around drop in the performance of classifiers in cross-domain settings, both when trained on the FakeNews AMT dataset and tested on the Celeb dataset, and vice versa. A similar drop in performance was also observed by Pérez-Rosas et al. (2018) and Saikh et al. (2019). However, our multi-task models still outperformed their single-task counterparts, implying improved extraction of shared features and better generalization. The results further indicate that intrinsic emotions aid in domain-independent implicit feature extraction for the detection of fake news.

Rebalancing weights of the loss function for fake news or rumour detection and emotion classification branches improves the performance of multi-task models We observed an improvement in the performance of all multi-task models on rebalancing the weights of the loss function for each branch's output. We observe from Figure 5 that the performance of the CNN, LSTM, and CNN-LSTM models on the PHEME 9 dataset reaches the maximum for α between 0.35 and 0.55. Upon increasing the value of α beyond that, we see a sharp decrease in the performance of the models. This can be attributed to the fact that the loss function penalizes the emotion classification branch significantly more than it penalizes the rumour detection branch (in the case of the PHEME 9 dataset). For $\alpha = 0.80$, the value of β is 0.20. This leads to a 4:1 ratio in the weights of the loss function, in favor of the emotion classification branch, which leads to the model focusing significantly more on the performance of the emotion classification branch.

Upon qualitatively analyzing the results of the single-task and multi-task learning models, we found many instances where by leveraging the emotion-specific features, the multi-task learning models were able to accurately detect the legitimacy of the text. Table 6 summarises some of the tweets found in the PHEME 9 dataset that was incorrectly

Table 2: Performance evaluation on FakeNews AMT and Celeb datasets for fake news detection, using accuracy, precision, recall, and F1-Score. Multi-task learning models outperform their single-task counterparts.

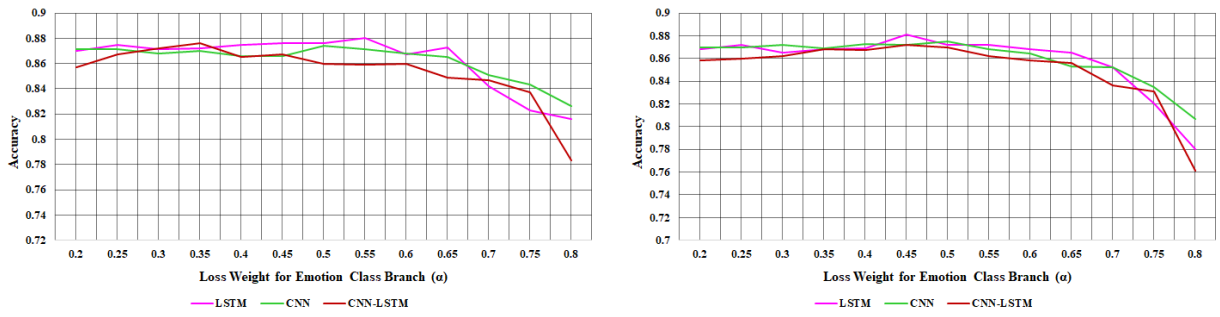
Dataset	Model	Setting	Accuracy	Precision	Recall	F1
FAMT	LSTM	STL	0.725	**0.829**	0.566	0.673
		MTL (Ekman)	**0.775**	0.753	0.816	**0.784**
		MTL (Plutchik)	0.758	0.712	**0.866**	0.781
		MTL (Domain)	**0.775**	0.770	0.783	0.776
	CNN-LSTM	STL	0.733	0.725	0.750	0.737
		MTL (Ekman)	0.758	0.754	0.766	0.760
		MTL (Plutchik)	**0.766**	0.735	**0.833**	**0.781**
		MTL (Domain)	0.758	**0.803**	0.683	0.738
	BERT	STL	0.816	0.851	0.766	0.806
		MTL (Ekman)	**0.875**	**0.868**	**0.883**	**0.875**
		MTL (Plutchik)	0.866	0.866	0.867	0.866
		MTL (Domain)	0.866	0.854	**0.883**	0.868
Celeb	LSTM	STL	0.736	**0.780**	0.639	0.702
		MTL (Ekman)	**0.752**	0.758	**0.721**	**0.739**
		MTL (Plutchik)	0.712	0.735	0.639	0.684
	CNN-LSTM	STL	0.704	0.740	0.606	0.666
		MTL (Ekman)	**0.712**	0.692	**0.737**	**0.714**
		MTL (Plutchik)	0.696	**0.744**	0.573	0.648
	BERT	STL	0.816	0.816	0.815	0.815
		MTL (Ekman)	0.856	0.859	0.854	0.855
		MTL (Plutchik)	**0.880**	**0.881**	**0.879**	**0.879**

Table 3: Performance evaluation on the PHEME 9 dataset for rumour detection, using accuracy, precision, recall, and F1-score. MTL (Ekman) and MTL (Plutchik) models outperform their STL counterparts across most metrics.

Dataset	Model	Setting	Accuracy	Precision	Recall	F1
PHEME 9	CNN	STL	0.870	0.819	0.826	0.822
		MTL (Ekman)	0.874	**0.843**	0.800	0.822
		MTL (Plutchik)	**0.875**	0.825	**0.834**	**0.830**
	LSTM	STL	0.857	0.807	0.800	0.804
		MTL (Ekman)	0.880	0.835	**0.837**	**0.836**
		MTL (Plutchik)	**0.881**	**0.847**	0.823	0.835
	CNN-LSTM	STL	0.860	0.786	**0.849**	0.816
		MTL (Ekman)	**0.876**	0.822	0.843	**0.832**
		MTL (Plutchik)	0.872	**0.835**	0.809	0.821
	CapsuleNet	STL	0.853	**0.818**	0.748	0.782
		MTL (Ekman)	0.858	0.804	**0.789**	0.797
		MTL (Plutchik)	**0.863**	0.817	0.784	**0.800**
	HAN	STL	0.847	0.790	0.769	0.779
		MTL (Ekman)	**0.867**	**0.859**	0.742	0.796
		MTL (Plutchik)	**0.867**	0.828	**0.784**	**0.805**
	BERT	STL	0.859	0.844	0.854	0.848
		MTL (Ekman)	**0.884**	**0.873**	**0.874**	**0.873**
		MTL (Plutchik)	0.874	0.859	0.870	0.864

Table 4: Cross-domain evaluation on FakeNews AMT and Celeb datasets using accuracy, precision, recall, and F1-score. We notice a drop in performance for all models compared to the in-domain setting. MTL models outperform their STL counterparts, indicating better generalization and domain-independent feature extraction.

Train Set	Test Set	Model	Setting	Accuracy	Precision	Recall	F1
FAMT	Celeb	LSTM	STL	0.560	0.557	0.475	0.513
			MTL (Ekman)	0.576	0.554	**0.672**	**0.607**
			MTL (Plutchik)	**0.632**	**0.641**	0.557	0.596
		CNN-LSTM	STL	0.568	0.559	0.540	0.550
			MTL (Ekman)	**0.632**	0.600	**0.737**	**0.661**
			MTL (Plutchik)	**0.632**	**0.608**	0.688	0.646
Celeb	FAMT	LSTM	STL	0.558	0.549	0.650	0.595
			MTL (Ekman)	0.591	0.557	**0.883**	**0.683**
			MTL (Plutchik)	**0.608**	**0.594**	0.683	0.635
		CNN-LSTM	STL	0.525	0.529	0.450	0.486
			MTL (Ekman)	0.575	**0.563**	0.666	0.610
			MTL (Plutchik)	**0.583**	0.556	**0.816**	**0.662**



(A) *(B)*

Figure 5: Line graph plot for variation in accuracy of CNN, LSTM and CNN-LSTM models in multi-task setting on the PHEME 9 dataset with variation in α (and consequently β) for: *(A)* Ekman's emotions *(B)* Plutchik's emotions.

Table 5: Performance evaluation on the Gossipcop dataset for fake news detection, using accuracy, precision, recall, and F1-score. MTL (Ekman) and MTL (Plutchik) models outperform their STL counterparts across most metrics.

Dataset	Model	Setting	Accuracy	Precision	Recall	F1
Gossipcop	LSTM	STL	0.806	0.845	0.932	0.886
		MTL (Ekman)	**0.835**	**0.847**	0.973	**0.906**
		MTL (Plutchik)	0.833	0.844	**0.975**	0.905
	CNN	STL	0.822	0.850	0.948	0.897
		MTL (Ekman)	0.835	0.850	**0.969**	**0.906**
		MTL (Plutchik)	**0.836**	**0.853**	0.964	0.905
	CNN-LSTM	STL	0.820	0.842	0.957	0.896
		MTL (Ekman)	**0.838**	**0.847**	0.978	**0.907**
		MTL (Plutchik)	0.837	0.845	**0.980**	**0.907**
	CapsuleNet	STL	0.824	0.828	0.977	0.897
		MTL (Ekman)	**0.834**	**0.838**	0.978	0.902
		MTL (Plutchik)	0.833	0.834	**0.983**	**0.903**
	HAN	STL	0.864	**0.903**	0.927	0.915
		MTL (Ekman)	0.869	0.882	**0.962**	**0.921**
		MTL (Plutchik)	**0.870**	0.888	0.955	0.920
	BERT	STL	0.851	0.803	0.716	0.745
		MTL (Ekman)	**0.866**	**0.823**	0.750	0.777
		MTL (Plutchik)	0.856	0.790	**0.768**	**0.778**
	RoBERTa	STL	0.841	0.764	**0.760**	0.762
		MTL (Ekman)	0.867	0.833	0.743	0.774
		MTL (Plutchik)	**0.872**	**0.836**	0.751	**0.782**

Table 6: Qualitative analysis of the results obtained from single-task and multi-task learning models on the PHEME 9 dataset, where multi-task models correctly predicted the label. Label 0 corresponds to Non-rumours, while label 1 corresponds to Rumours.

Text	Label	Ekman Emotion	Plutchik Emotion	STL	MTL (Ekman)	MTL (Plutchik)
Family claims #CorneliusGurlitt was mentally unfit. Read the news @Artlyst http://t.co/1tmcwtUYC8 http://t.co/tfH3caD1EP	1	Sadness	Surprise	0	1	1
Such scary news from Sydney. Thoughts & prayers go out to the hostages &; their family's.	0	Fear	Fear	1	0	0
Video footage shows police storming Sydney cafe http://t.co/3U4aWpVmdM #SydneySiege	0	Surprise	Surprise	1	0	0
This is trending on French twitter #MouradHamydInnocent started by classmates who say Mourad was in class when #CharlieHebdo attack happened	1	Fear	Fear	0	1	1

Table 7: Qualitative analysis of the results obtained from single-task and multi-task learning models on the PHEME 9 dataset, where multi-task models with Ekman's emotions correctly predicted the label, while multi-task models with Plutchik's emotions incorrectly predicted the label. Label 0 corresponds to Non-rumours, while label 1 corresponds to Rumours.

Text	Label	Ekman Emotion	Plutchik Emotion	STL	MTL (Ekman)	MTL (Plutchik)
Putin juggling enough instability. He would make a live appearance by now to squash death/coup jitters.	0	Fear	Trust	1	0	1
Australian Mufti condemns hostage taking in #Sydney - Statement #SydneyHostageCrisis #SydneySiege #SydneyCafe http://t.co/n38XZVq5iu	1	Surprise	Trust	0	1	0
Whenever you want to justify shooting an unarmed person you always say they reached for your gun. George Zimmerman taught us that #Ferguson	0	Fear	Fear	1	0	1

labelled by the majority of the single-task learning models, but using the emotion-based features, the multi-task learning

Table 8: Qualitative analysis of the results obtained from single-task and multi-task learning models on the PHEME 9 dataset, where multi-task models with Plutchik's emotions correctly predicted the label, while multi-task models with Ekman's emotions incorrectly predicted the label. Label 0 corresponds to Non-rumours, while label 1 corresponds to Rumours.

Text	Label	Ekman Emotion	Plutchik Emotion	STL	MTL (Ekman)	MTL (Plutchik)
Latest footage from #sydneysiege shows security forces throwing objects into the cafe http://t.co/HdTPHSMjUQ https://t.co/2pu3pWqHhz	0	Surprise	Trust	1	1	0
It just feels very odd.What'ss going on in #Kremlin? Military coup and putin removal?I'm afraid the next Russia's leader won't be any better	1	Fear	Trust	0	0	1
PM Tony Abbott will hold a press conference on the #SydneySiege in 20 minutes.	0	Surprise	Anticipation	0	1	0

Table 9: Qualitative analysis of the results obtained from single-task and multi-task learning models on the PHEME 9 dataset, where multi-task models with both Ekman's and Plutchik's emotions incorrectly predicted the label. Label 0 corresponds to Non-rumours, while label 1 corresponds to Rumours.

Text	Label	Ekman Emotion	Plutchik Emotion	STL	MTL (Ekman)	MTL (Plutchik)
#sydneysiege is a misnomer. It's a hostage situation. Inside a cafe. You want 2 know what being under siege looks like? look @ #Gaza.	1	Fear	Trust	1	0	0
Gun fire exchange in Parliament Hill building #Ottawa #parliament #shooting #Canada #BeSafeOttawa -https://t.co/9RDjNvVkJW via @martinvass2	1	Joy	Trust	1	0	0

models correctly predicted the labels. We observed that the multi-task learning models were able to correctly predict the legitimacy label for a sample even if the sample portrayed an emotion typically associated with the other class. This can be attributed to the fact that even within the same emotions, fake news and real news differ in their semantic and syntactic features, as shown in the PCA graphs in Figures 2 and 3.

We also found a few instances where multi-task learning models leveraging either emotion sets incorrectly predicted the labels, sometimes even when the single-task learning models correctly predicted the labels. These are summarised in Tables 7, 8 and 9. We can see from these results that multi-task learning models with Ekman's emotions can sometimes incorrectly predict rumours as non-rumours when the intrinsic emotions detected are *Fear* or *Joy*. Similarly, multi-task learning models with Plutchik's emotions can sometimes incorrectly predict rumours as non-rumours when the intrinsic emotions detected are *Fear* or *Trust*. However, the instances where single-task learning models were correct and multi-task learning models were incorrect were significantly lesser than the other way around.

5 Broader Impact and Ethical Considerations

The spread of misinformation has become a significant problem for society with adverse consequences. Our research further improves upon previous machine learning-based fake news and rumour detection approaches by incorporating the correlated features between the legitimacy of a text and its intrinsic emotion, to more accurately detect fake news or rumours. Instead of taking the sentiment or emotion features as input to the models like in previous works, we apply a multi-task learning approach to detect fake news and rumours. This mitigates the need to generate emotion-based features, either computationally or manually, to be used as input to the fake news detection or rumour detection models during testing or deployment. Instead, the models implicitly detect the emotion-guided features and use them for the detection of fake news or rumours. Evaluation across various datasets for rumour detection as well as fake news detection, as well as on multiple deep learning architectures proves the wide applicability of our approach, both for in-domain and cross-domain settings. Our approach can further be implemented with any state-of-the-art deep learning model used for fake news detection or rumour detection, to further improve the performance of the model. Hence, systems incorporating our approach can aid in curbing the spread of false information on social media with increased accuracy.

From an ethical point of view, our models may be affected by biases inherited from the training dataset. This may be biased toward gender, race, or a different set of people. We hope that further works lead to the creation of more robust

large-scale fake news detection and rumour detection datasets annotated with emotion classes, which can help minimize biases in the data, and hence the models. However, we do not foresee our proposed approach causing any negative impact on society, directly or indirectly.

6 The Current Limitations of Our Approach

One limitation of our proposed approach is dependence on the Unison model for annotating the datasets with Ekman's and Plutchik's emotion tags. Despite giving state-of-the-art results for emotion classification for both Ekman's and Plutchik's emotions, the Unison model achieves a micro F1-score of around 0.70 for both Ekman's and Plutchik's emotion classes. This can sometimes lead to incorrect annotations for some texts in the datasets. These incorrect annotations can lead to contradictory information being learned by the models, leading to lower performance. While we achieved positive performance gains in our experimental evaluations, future advancements in emotion classification models as well as manually annotated fake news detection and rumour detection datasets with emotion tags can help in further improving the performance of classifiers trained to detect fake news or rumours in our proposed multi-task setting.

7 Conclusion

In this work, we verified the correlation between the legitimacy of a text and its intrinsic emotion. We annotated the datasets with their emotion classes using the Unison model, to help our models learn the associated emotion features by transfer learning. We used Plutchik's basic emotions, based on previous psychological studies, as well as proposed the use of Ekman's emotion classes for improved feature extraction in our classifiers. For implicit utilization of features linked to the emotions portrayed by a text, we proposed a multi-task learning approach using multi-class emotion detection as an auxiliary task for the detection of fake news and rumours on FakeNews AMT, Celeb, Gossipcop, and PHEME 9 datasets. Our experimental evaluation revealed that most multi-task models consistently outperformed their single-task counterparts in terms of accuracy, precision, recall, and F1-score for both Ekman's and Plutchik's emotions for in-domain and cross-domain settings. Our cross-domain evaluations showed that emotion-guided features are more domain generalizable, as we observed similar gains in cross-domain settings in MTL settings over STL settings as in in-domain settings. Upon qualitative analysis, we observed that MTL models showed improved performance not only for fake news samples portraying emotions normally associated with fake news and for real news portraying emotions normally associated with real news, but also for fake samples portraying emotions associated with real news and real samples portraying emotions associated with fake news. We further showed the efficacy of the use of the domain of news as another plausible feature for fake news detection on the FakeNews AMT dataset.

References

Ajao, O., Bhowmik, D., Shahrzad., Z., 2019. Sentiment Aware Fake News Detection on Online Social Networks, in: IEEE 44th International Conference on Acoustics, Speech and Signal Processing, (ICASSP), IEEE, Brighton, United Kingdom. pp. 2507–2511.

Bann, E., Bryson, J., 2014. The conceptualisation of emotion qualia: semantic clustering of emotional tweets, in: Mayor, J., Gomez, P. (Eds.), Computational Models of Cognitive Processes, World Scientific. pp. 249–263. NCPW 13th Neural Computation and Psychology Workshop.

Bhattacharya, P., Patel, S.B., Gupta, R., Tanwar, S., Rodrigues, J.J.P.C., 2021. Satya: Trusted bi-lstm-based fake news classification scheme for smart community. IEEE Transactions on Computational Social Systems , 1–10doi:10.1109/TCSS.2021.3131945.

Bhutani, B., Rastogi, N., Sehgal, P., Purwar, A., 2019. Fake news detection using sentiment analysis, in: 2019 Twelfth International Conference on Contemporary Computing (IC3), pp. 1–5. doi:10.1109/IC3.2019.8844880.

Cheng, M., Li, Y., Nazarian, S., Bogdan, P., 2021a. From rumor to genetic mutation detection with explanations: a gan approach. Scientific Reports 11, 5861. doi:10.1038/s41598-021-84993-1.

Cheng, M., Nazarian, S., Bogdan, P., 2020. Vroc: Variational autoencoder-aided multi-task rumor classifier based on text. Proceedings of The Web Conference 2020 .

Cheng, M., Yin, C., Nazarian, S., Bogdan, P., 2021b. Deciphering the laws of social network-transcendent covid-19 misinformation dynamics and implications for combating misinformation phenomena. Scientific Reports 11. doi:10.1038/s41598-021-89202-7.

Choudhry, A., Khatri, I., Jain, M., 2022. An emotion-based multi-task approach to fake news detection (student abstract). Proceedings of the AAAI Conference on Artificial Intelligence 36, 12929–12930. URL: https://ojs.aaai.org/index.php/AAAI/article/view/21601, doi:10.1609/aaai.v36i11.21601.

Colneric, N., Demsar, J., 2018. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. IEEE Transactions on Affective Computing 11, 433–446. doi:10.1109/TAFFC.2018.2807817.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 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), Association for Computational Linguistics, Stroudsburg, PA, USA. pp. 4171–4186. doi:10.18653/v1/N19-1423, arXiv:1810.04805v1.

Ekman, P., 1992. An argument for basic emotions. Cognition & Emotion 6, 169–200.

Fersini, E., Armanini, J., D'Intorni, M., 2020. Profiling fake news spreaders: Stylometry, personality, emotions and embeddings, in: CLEF.

Guo, C., Cao, J., Zhang, X., Shu, K., Yu, M., 2019. Exploiting emotions for fake news detection on social media. ArXiv abs/1903.01728.

Guo, H., Cao, J., Zhang, Y., Guo, J., Li, J., 2018. Rumor detection with hierarchical social attention network, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Association for Computing Machinery, New York, NY, USA. p. 943–951. doi:10.1145/3269206.3271709.

Han, Y., Karunasekera, S., Leckie, C., 2020. Graph neural networks with continual learning for fake news detection from social media. CoRR abs/2007.03316. arXiv:2007.03316.

K, A., P, D., L, L.V., 2020. Emotion cognizance improves health fake news identification, in: Proceedings of the 24th Symposium on International Database Engineering & Applications, Association for Computing Machinery. doi:10.1145/3410566.3410595.

Kamila, S., Hasanuzzaman, M., Ekbal, A., Bhattacharyya, P., 2022. Investigating the impact of emotion on temporal orientation in a deep multitask setting. Scientific Reports 12. doi:10.1038/s41598-021-04331-3.

Kochkina, E., Liakata, M., Zubiaga, A., 2018a. All-in-one: Multi-task learning for rumour verification, in: COLING.

Kochkina, E., Liakata, M., Zubiaga, A., 2018b. Pheme dataset for rumour detection and veracity classification. doi:10.6084/m9.figshare.6392078.v1.

Kumari, R., Ashok, N., Ghosal, T., Ekbal, A., 2021. A multitask learning approach for fake news detection: Novelty, emotion, and sentiment lend a helping hand, in: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. doi:10.1109/IJCNN52387.2021.9534218.

Meel, P., Vishwakarma, D., 2019. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications 153, 112986. doi:10.1016/j.eswa.2019.112986.

Pardo, F.M.R., Giachanou, A., Ghanem, B., Rosso, P., 2020. Overview of the 8th author profiling task at pan 2020: Profiling fake news spreaders on twitter, in: Cappellato, L., Eickhoff, C., 0001, N.F., Névéol, A. (Eds.), Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020, CEUR-WS.org.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R., 2018. Automatic detection of fake news, in: Proceedings of the 27th International Conference on Computational Linguistics, ACL, Santa Fe, New Mexico, USA. pp. 3391–3401.

Plutchik, R., 1982. A psychoevolutionary theory of emotions. Social Science Information 21, 529–553.

Roy, S., Bhanu, M., Saxena, S., Dandapat, S., Chandra, J., 2022. gdart: Improving rumor verification in social media with discrete attention representations. Information Processing & Management 59, 102927. doi:https://doi.org/10.1016/j.ipm.2022.102927.

Sabour, S., Frosst, N., Hinton, G.E., 2017. Dynamic routing between capsules, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA. p. 3859–3869.

Saha, T., Upadhyaya, A., Saha, S., Bhattacharyya, P., 2022. A multitask multimodal ensemble model for sentiment- and emotion-aided tweet act classification. IEEE Transactions on Computational Social Systems 9, 508–517. doi:10.1109/TCSS.2021.3088714.

Saikh, T., De, A., Ekbal, A., Bhattacharyya, P., 2019. A deep learning approach for automatic detection of fake news, in: Proceedings of the 16th International Conference on Natural Language Processing, NLP Association of India, International Institute of Information Technology, Hyderabad, India. pp. 230–238.

Shahid, W., Jamshidi, B., Hakak, S., Isah, H., Khan, W.Z., Khan, M.K., Choo, K.K.R., 2022. Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities. IEEE Transactions on Computational Social Systems , 1–14doi:10.1109/TCSS.2022.3177359.

Singhal, S., Kabra, A., Sharma, M., Shah, R.R., Chakraborty, T., Kumaraguru, P., 2020. Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract). Proceedings of the AAAI Conference on Artificial Intelligence 34, 13915–13916. doi:10.1609/aaai.v34i10.7230.

Singhal, S., Shah, R., Chakraborty, T., Kumaraguru, P., Satoh, S., 2019. Spotfake: A multimodal framework for fake news detection, in: IEEE BigMM.

Starbird, K., Maddock, J., Orand, M., Achterman, P., Mason, R., 2014. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. IConference doi:10.9776/14308.

Verma, G., Bhardwaj, A., Aledavood, T., Choudhury, M., Kumar, S., 2022. Examining the impact of sharing covid-19 misinformation online on mental health. Scientific Reports 12, 8045. doi:10.1038/s41598-022-11488-y.

Verma, P.K., Agrawal, P., Amorim, I., Prodan, R., 2021. Welfake: Word embedding over linguistic features for fake news detection. IEEE Transactions on Computational Social Systems 8, 881–893. doi:10.1109/TCSS.2021.3068519.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359, 1146–1151. doi:10.1126/science.aap9559.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q., Rush, A.M., 2020. Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online. pp. 38–45.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., Yu, P.S., 2018. TI-CNN: convolutional neural networks for fake news detection. arXiv:1806.00749.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E., 2016. Hierarchical attention networks for document classification, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics. pp. 1480–1489. doi:10.18653/v1/N16-1174.

Zhang, X., Cao, J., Li, X., Sheng, Q., Zhong, L., Shu, K., 2021. Mining dual emotion for fake news detection, in: Proceedings of the Web Conference 2021, Association for Computing Machinery, New York, NY, USA. p. 3465–3476. doi:10.1145/3442381.3450004.

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.

An Iterative Analytic Approximation for a Class of Non-Linear Singularly Perturbed Parabolic Partial

KARTIKAY KHARI (kartikaybharadwaj@gmail.com)
Delhi Technological University Department of Applied Mathematics https://orcid.org/0000-0002-1019-4270

Vivek Kumar
Delhi Technological University Department of Applied Mathematics

Research Article

Keywords: Singular Perturbation, Non-Linear Parabolic, Reaction-Diffusion, Convection-Diffusion, Liouville-Green Transform, Lagrange Multiplier.

Posted Date: October 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2072461/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

An Iterative Analytic Approximation for a Class of Non-Linear Singularly Perturbed Parabolic Partial Differential Equations

Kartikay Khari*[1] and Vivek Kumar[1]

[1,2]Department of Applied Mathematics, Delhi Technological University, Delhi−110042, India.
[1]Email: kartikay_phd2k17@dtu.ac.in, kartikaybharadwaj@gmail.com
[2]Email: vivekkumar.ag@gmail.com

Abstract

In this article, a closed-form iterative analytic approximation to a class of nonlinear singularly perturbed parabolic partial differential equation is developed and analysed for convergence. We have considered both parabolic reaction diffusion and parabolic convection diffusion type of problems in this paper. The solution of this class of problem is polluted by a small dissipative parameter, due to which solution often shows boundary and interior layers. A sequence of approximate analytic solution for the above class of problems is constructed using Lagrange multiplier approach. Within a general frame work, the Lagrange multiplier is optimally obtained using variational theory and Liouville-Green's transformation. The sequence of approximate analytic solutions so obtained is proved to converge the exact solution of the problem. To demonstrate the proposed method's efficiency and accuracy, two linear and one nonlinear test problem have been taken into account.

Keywords- Singular Perturbation; Non-Linear Parabolic; Reaction-Diffusion; Convection-Diffusion; Liouville-Green Transform; Lagrange Multiplier.

1 Introduction

The nonlinear singularly perturbed parabolic partial differential equations (NSPPPDEs) plays an essential role in converting a real life phenomenon into a mathematical model. The dynamics of NSPPPDEs are utterly different from the conventional nonlinear partial differential equations. These types of problems depend on a small positive parameter, which makes the solution varies rapidly in narrow regions of the domain and changes slowly in rest parts of the domain. This behaviour of the solution in narrow regions is called layer phenomena and this class of problem is known as singularly perturbed problems. The problem is singular in such a way that when $\varepsilon \to 0$ changes the order of the differential equation, but the number boundary conditions remain the same, causing the problem ill-posed. This type of problems occur in oceanography, population dynamic, generic repression, size dependent cell growth, division modelling, ecology, quantum physics, chemistry, finance(Black-Schole Equation) and material science. One of the most remarkable examples of NSPPDES that occurs in fluid dynamics is the Navier-Stokes equation with a higher Reynolds number. Considering the following Navier-Stokes equation that governs unsteady incompressible viscous fluids flow problem.

$$\frac{\partial \upsilon_\varepsilon}{\partial t} + \upsilon_\varepsilon.\Delta\upsilon_\varepsilon + \Delta\rho = \frac{1}{\Re e}\Delta^2\upsilon_\varepsilon \tag{1.1}$$

$$\Delta \upsilon_\varepsilon = 0 \tag{1.2}$$

Where ρ signifies pressure and $\upsilon = (\upsilon_1, \upsilon_2)$ denotes the velocity, with velocity components υ_1 and υ_2 along with x and y directions. The parameter $\Re e$ represents Reynold number which is directly proportional to the velocity scale, length scale and inversely proportional to the kinematic viscosity of the fluid. The equation (1.1) behaves as a singularly perturbed PDE if the Reynold number $\Re e >> 1$ is large enough.

The designing of computational algorithms for such type of problems burden with difficulties because the solution of the proposed problem is contaminated by a small positive parameter ε and nonlinear term simultaneously. Since only few nonlinear systems can be solved explicitly, we rely on numerical techniques by linearizing the nonlinear problems. In recent years, the researchers have made efforts to provide consistent numerical techniques to solve NSPPPDEs. In [4] Igor Boglaev presented a monotone iterative algorithm for nonlinear singularly perturbed parabolic problem and then in [5] Igor Boglaev presented an inexact monotone iterative method based on method of upper and lower solution for a class of semi-linear parabolic singularly perturbed problem. In [18, 19] Kopteva and Linß used an elliptic reconstruction approach and provide the error estimate in maximum norm using Green's function. In [15] authors provide a consistent finite difference method for singularly perturbed semilinear parabolic differential difference equation. In [6] Yao Cheng , Chuanjing Song and Yanjie Mei used a discontinuous galerkin finite element method along with backward Euler method for our proposed problem. In [3] Mario Amrein and Thomas P. Wihler proposed an adaptive space-time Newton-Galerkin method which is consisted of a linearization technique using Newtons method along with finite element for spatial variable and backward-Euler for time variable and others [14, 17].
To best of our knowledge, till date there is no analytic method (exact or approximate) is developed to solve the NSPPPDEs. The novelty of this paper is the development of an iterative analytic method based on variation of iteration approach to find the approximate solution of the NSPPPDEs. The Variation of iteration method (VIM) was first developed by J. H. He [11] in 1999. The VIM is based on a Lagrange's multiplier developed by Inokuti *et al.* [13] in which they constructed adjoint operators and state that Lagrange multiplier λ is regarded as a Green's function rather then a constant. In [24] J. L. Ramos proves that the claim made by Inokuti *et al.* are indeed correct.

We can drive the VIM method by using Green's function, method of weighted residuals and using integration by parts. The best property of VIM method is to drive a closed form solution of a nonlinear problem without any need of linearization and discretization. Due to the flexibility and ability, the VIM has emerged as a promising tool to solve non-linear or complex systems. Consequently, this method was successfully implimented to find the approximate solution of Hilfer fractional advection diffusion equations with power law initial condition [2], of reaction diffusion system with fast reversible reaction [1], of wave-like and heat-like equations in large domains [9], of delay differential equations [10], of nonlinear singular boundary value problems [16], of the time fractional Fornberg-Whitham equation [25], of Burger's flow with fractional derivatives [28], of non-isothermal reaction diffusion model equations in a spherical catalyst [26], of Optical bright and dark soliton solutions for coupled nonlinear Schrdinger (CNLS) equations [27], to Bratu-like equation arising in electrospinningand [12], to free vibration analysis of a tapered beam mounted on two-degree of freedom subsystems [8] and to others [7, 20, 21, 22].
The paper is organized as follows: The continuous model problem is defined in Section 2. In Section 3, a brief sketch of the method is given. Section 4, contains the solution methodology and optimal value of Lagrange multiplier is evaluated. In Section 5 the convergence analysis is carried out . In Section 6, one linear and one nonlinear test problems are taken into account to validate theoretical results obtained in section 5. Finally, Section 7, contains a conclusion.

2 Statement of Problem

2.1 Parabolic Reaction Diffusion Problem

Consider the following class of nonlinear singularly perturbed parabolic problem of reaction diffusion type.

$$(v_\varepsilon)_t(x,t) - \varepsilon(v_\varepsilon)_{xx}(x,t) + v_\varepsilon(x,t) - f(x,t,v) = 0, \tag{2.1}$$

where $(x,t) \in (0,1)\times(0,1)$, subject to the initial condition and boundary conditions prescribed as:

$$\begin{cases} v_\varepsilon(x,0) = & \psi_b(x), \\ v_\varepsilon(0,t) = & \psi_l(t), \\ v_\varepsilon(1,t) = & \psi_r(t). \end{cases} \tag{2.2}$$

Where $0 < \varepsilon \ll 1$ is a small positive parameter known as singular perturbation parameter. Let $f(x,t,v)$ be sufficiently smooth and ϱ ,$\bar{\varrho}$ be some constant satisfies,

$$0 \le \varrho^2 \le f_v(x,t,v) \le \bar{\varrho}^2 \quad \forall (x,t,v) \in [0,1]\times[0,1]\times\mathbf{R} \tag{2.3}$$

2.2 Parabolic Convection Diffusion Problem

Consider the following class of singularly perturbed parabolic convection dominated problem.

$$(v_\varepsilon)_t(x,t) - \varepsilon(v_\varepsilon)_{xx}(x,t) + (v_\varepsilon)_x(x,t) - f(x,t,v) = 0, \tag{2.4}$$

where $(x,t) \in (0,1)\times(0,1)$, subject to the initial condition and boundary conditions prescribed as:

$$\begin{cases} v_\varepsilon(x,0) = & \psi_b(x), \\ v_\varepsilon(0,t) = & \psi_l(t), \\ v_\varepsilon(1,t) = & \psi_r(t). \end{cases} \tag{2.5}$$

Where $0 < \varepsilon \ll 1$ is a small positive parameter known as singular perturbation parameter. Let $f(x,t,v)$ be sufficiently smooth and ϱ ,$\bar{\varrho}$ be some constant satisfies,

$$0 \le \varrho^2 \le f_v(x,t,v) \le \bar{\varrho}^2 \quad \forall (x,t,v) \in [0,1]\times[0,1]\times\mathbf{R} \tag{2.6}$$

3 Variation of Iteration Method

In this section, we have given a brief sketch of variation of iteration method. Consider a general class non-linear parabolic partial differential equation.

$$Dv(x,t) = \mathcal{L}(v(x,t)) + \mathcal{N}(v(x,t)) = \mathcal{F}. \tag{3.1}$$

Where $\mathcal{L}$ and $\mathcal{N}$ are the linear and non-linear operator respectively and $\mathcal{F}(x,t)$ be the force term or inhomogeneous term. Then the correction functional associated with (3.1) is defined as:

$$v_{n+1}(x,t) = v_n(x,t) + \int_0^t \lambda(s)\{\mathcal{L}v_n(x,s) + \mathcal{N}\tilde{v_n}(x,s) - \mathcal{F}(x,s)\}ds \qquad n \ge 0. \tag{3.2}$$

Where λ be the general Lagrange's multiplier, which is evaluated optimally [11] by using variational theory and Liouville-Green transformation. Furthermore, v_n denotes the nth iterative approximation of v and $\tilde{v}_n$ denotes the restricted variation i.e $\delta\tilde{v}_n = 0$. After evaluating the Lagrange's multiplier λ and choosing an appropriate initial condition v_0, and using (3.2), one obtain the successive approximations v_n of solution $v(x,t)$. Hence the exact solution of the problems (3.1) is obtained as

$$v(x,t) = \lim_{n\to\infty} v_n(x,t).$$

4 Solution Methodology

4.1 Lagrange's Multiplier for Parabolic Reaction Diffusion Problem

Taking the correction functional in $t-$direction. The problem (2.1) can be expressed as:

$$v_{n+1}(x,t) = v_n(x,t) + \int_0^t \lambda(s)\left(\frac{\partial v_n(x,s)}{\partial s} - \varepsilon\frac{\partial^2 v_n(x,s)}{\partial x^2} + f_n(x,s,v)\right)ds, n \geq 0. \tag{4.1}$$

Now let $\delta\tilde{v_n}$ is the restricted variation and the restricted variation of nonlinear source term is denoted by $\tilde{f_n}(x,s,v)$.

$$v_{n+1}(x,t) = v_n(x,t) + \int_0^t \lambda(s)\left(\frac{\partial v_n(x,s)}{\partial s} - \varepsilon\frac{\partial^2 \tilde{v_n}(x,s)}{\partial x^2} + \tilde{f_n}(x,s,v)\right)ds. \tag{4.2}$$

Using integration by part and restricting the restricted variation term,

$$v_{n+1}(x,t) = v_n(x,t) + [\lambda(s)v_n(x,s)]_{s=0}^t - \left[\frac{d\lambda(s)}{ds}v_n(x,s)\right]_{s=0}^t. \tag{4.3}$$

We recall the variational theory for evaluating Lagrange's multiplier optimally and applying variation corresponding to v_n and making correction functional stationary i.e. $\delta v_{n+1} = 0$.

$$\delta v_{n+1}(x,t) = \delta v_n(x,t) + \lambda(s)\delta v_n(x,t) - \frac{d\lambda(s)}{ds}\delta v_n(x,t) = 0. \tag{4.4}$$

We obtain the following Euler-Lagrange equation and stationary conditions as:

$$\left[\frac{d\lambda(s)}{ds}\right]_{s=t} = 0; \qquad [1-\lambda(s)]_{s=t} = 0. \tag{4.5}$$

From stationary condition (4.5) we obtain the optimal value of Lagrange's multiplier

$$\lambda(s) = -1 \tag{4.6}$$

Now the iteration formulae is written as

$$v_{n+1}(x,t) = v_n(x,t) - \int_0^t \left(\frac{\partial v_n(x,s)}{\partial s} - \varepsilon\frac{\partial^2 v_n(x,s)}{\partial x^2} + f_n(x,s,v)\right)ds. \tag{4.7}$$

4.2 Lagrange's Multiplier for Parabolic Convection Diffusion Problem

Taking the correction functional in $t-$direction. The problem (2.4) can be expressed as:

$$v_{n+1}(x,t) = v_n(x,t) + \int_0^t \lambda(s)\left(\frac{\partial v_n(x,s)}{\partial s} - \varepsilon\frac{\partial^2 v_n(x,s)}{\partial x^2} + \frac{\partial v_n(x,s)}{\partial x} + f_n(x,s,v)\right)ds, n \geq 0. \tag{4.8}$$

Now let $\delta\tilde{v_n}$ is the restricted variation and the restricted variation of nonlinear source term is denoted by $\tilde{f_n}(x,s,v)$.

$$v_{n+1}(x,t) = v_n(x,t) + \int_0^t \lambda(s)\left(\frac{\partial v_n(x,s)}{\partial s} - \varepsilon\frac{\partial^2 \tilde{v_n}(x,s)}{\partial x^2} + \frac{\partial \tilde{v_n}(x,s)}{\partial x} + \tilde{f_n}(x,s,v)\right)ds. \tag{4.9}$$

Using integration by part and restricting the restricted variation term,

$$v_{n+1}(x,t) = v_n(x,t) + [\lambda(s)v_n(x,s)]_{s=0}^t - \left[\frac{d\lambda(s)}{ds}v_n(x,s)\right]_{s=0}^t. \tag{4.10}$$

We recall the variational theory for evaluating Lagrange's multiplier optimally and applying variation corresponding to v_n and making correction functional stationary i.e. $\delta v_{n+1} = 0$.

$$\delta v_{n+1}(x,t) = \delta v_n(x,t) + \lambda(s)\delta v_n(x,t) - \frac{d\lambda(s)}{ds}\delta v_n(x,t) = 0. \tag{4.11}$$

We obtain the following Euler-Lagrange equation and stationary conditions as:

$$\left[\frac{d\lambda(s)}{ds}\right]_{s=t} = 0; \qquad [1-\lambda(s)]_{s=t} = 0. \tag{4.12}$$

From stationary condition (4.12) we obtain the optimal value of Lagrange's multiplier

$$\lambda(s) = -1 \tag{4.13}$$

Now the iteration formulae is written as

$$v_{n+1}(x,t) = v_n(x,t) - \int_0^t \left(\frac{\partial v_n(x,s)}{\partial s} - \varepsilon\frac{\partial^2 v_n(x,s)}{\partial x^2} + \frac{\partial v_n(x,s)}{\partial x} + f_n(x,s,v)\right) ds. \tag{4.14}$$

5 Convergence

Theorem 1. *Let us suppose that χ be a Banach space such that $\boldsymbol{A} : \chi \to \chi$ be the non-linear mapping and suppose*

$$\|\boldsymbol{A}[v] - \boldsymbol{A}[\bar{v}]\| \leq \varrho\|v - \bar{v}\|, \qquad v, \bar{v} \in \chi \tag{5.1}$$

then $\boldsymbol{A}$ has a unique fixed point for $\varrho < 1$. Moreover

$$v_{n+1} = \boldsymbol{A}[v_n] \tag{5.2}$$

Sequence converges to a fixed point $\boldsymbol{A}$ for any arbitrary choice of $v_0 \in \chi$,

$$\|v_k - v_l\| \leq \|v_1 - v_0\| \sum_{j=l-1}^{k-2} \varrho^j \tag{5.3}$$

Proof. See [23] ■

The above iteration formulae (4.7) generates a sequence $< v_n(x,t) >$. Now we construct the series from variational iteration formulae (4.7) as,

$$v_0(x,t) + [v_1(x,t) - v_0(x,t)] + [v_2(x,t) - v_1(x,t)] + \cdots + [v_n(x,t) - v_{n-1}(x,t)] + \cdots . \tag{5.4}$$

Now let

$$S_{n+1}(x,t) = v_0(x,t) + [v_1(x,t) - v_0(x,t)] + [v_2(x,t) - v_1(x,t)] + \cdots + [v_n(x,t) - v_{n-1}(x,t)] = v_n(x,t), \tag{5.5}$$

then we have the following bounds.

Theorem 2. *Let $v_n(x,t) \in [0,1]$ be the sequence generated by (5.4) with $v_0 = q(x)c$, where $q(x) \in C^2[0,1]$ bounded in interval $[0,1]$ and c is an arbitrary constant. Let the optimal value of Lagrange's multiplier corresponding to the stationary conditions (4.5) belongs to $C^\infty[0,1]$. Then the sequence v_n convergences to the exact solution to $v(x,t)$ of the problem.*

Proof.

$$|v_1(x,t)-v_0(x,t)| = |\int_0^t \lambda(s)\left(\frac{\partial v_0(x,s)}{\partial s}-\varepsilon\frac{\partial^2 v_0(x,s)}{\partial x^2}+\tilde{f}_n(x,s,v_0)\right)ds|, \quad (5.6)$$

$$\leq \int_0^t |\lambda(s)|\left(\|\frac{\partial v_0(x,s)}{\partial s}-\varepsilon\frac{\partial^2 v_0(x,s)}{\partial x^2}\|+\|\tilde{f}_n(x,s,v_0)\|\right)ds, \quad (5.7)$$

$$\leq \int_0^t |A_i|\left(\|B_i\|_\infty+\|C_i\|_\infty\right)ds \quad (5.8)$$

$$\leq M\int_0^t ds, \quad (5.9)$$

$$\leq Mt. \quad (5.10)$$

Where $|A_i|=\max_{s\in[0,1]}|\lambda(s)|$; $\|B_i\|_\infty\leq\max_i|B_i|$ and $\|C_i\|_\infty\leq\max_i|C_i|$,

$$M=\max\{|A_i|\left(\|B_i\|_\infty+\|C_i\|_\infty\right),\|B_i\|_\infty\}. \quad (5.11)$$

Similarly ,

$$|v_2(x,t)-v_1(x,t)| = |\int_0^t \lambda(s)\left(\frac{\partial v_1(x,s)}{\partial s}-\varepsilon\frac{\partial^2 v_1(x,s)}{\partial x^2}+\tilde{f}(x,s,v_1)\right)ds|, \quad (5.12)$$

$$\leq \int_0^t |\lambda(s)|\left(\|\frac{\partial v_1(x,s)}{\partial s}-\varepsilon\frac{\partial^2 v_1(x,s)}{\partial x^2}\|+\|\tilde{f}(x,s,v_1)\|\right)ds, \quad (5.13)$$

$$\leq \int_0^t |A_i|\left(\|B_i\|_\infty+\|C_i\|_\infty\right)ds, \quad (5.14)$$

$$\leq M^2\int_0^t sds, \quad (5.15)$$

$$\leq \frac{M^2t^2}{2!}. \quad (5.16)$$

By mathematical induction,

$$|v_{n+1}(x,t)-v_n(x,t)| = |\int_0^t \lambda(s)\left(\frac{\partial v_n(x,s)}{\partial s}-\varepsilon\frac{\partial^2 v_n(x,s)}{\partial x^2}+\tilde{f}(x,s,v_n)\right)ds|, \quad (5.17)$$

$$\leq \int_0^t |\lambda(s)|\left(\|\frac{\partial v_n(x,s)}{\partial s}-\varepsilon\frac{\partial^2 v_n(x,s)}{\partial x^2}\|+\|\tilde{f}(x,s,v_n)\|\right)ds, \quad (5.18)$$

$$\leq \int_0^t |A_i|\left(\|B_i\|_\infty+\|C_i\|_\infty\right)ds, \quad (5.19)$$

$$\leq M^{n+1}\int_0^t s^n ds, \quad (5.20)$$

$$\leq \frac{M^{n+1}t^{n+1}}{(n+1)!}. \quad (5.21)$$

Hence, for $t\in[0,1]$ the series $\frac{M^nt^n}{n!}$ is uniformly convergent and converges to 0 as $n\to\infty$. Therefore we conclude that the sequence $<v_n>$ is uniformly convergent and converges to the exact solution $v(x,t)$. From above convergence analysis it is concluded that the proposed method is parameter uniform convergent i.e independent of singular perturbation parameter ε. ■

Theorem 3. *Let $v_n(x,t)\in[0,1]$ be the sequence generated by (4.14) with $v_0=q_1(x)c_1$, where $q_1(x)\in C^2[0,1]$ bounded in interval $[0,1]$ and c_1 is an arbitrary constant. Let the optimal value of Lagrange's multiplier corresponding to the stationary conditions (4.12) belongs to $C^\infty[0,1]$. Then the sequence v_n convergences to the exact solution to $v(x,t)$ of the problem.*

Proof. The proof is similar to the theorem (2). ■

6 Numerical Discussion

In this section, two linear and one non-linear reaction-diffusion test problems are considered, as well as one linear convection reaction diffusion kind of problem to validate the theoretical finding and evaluate the mentioned method's efficiency. The maximum absolute error is calculated by using the following:

$$\eta = \|v - v_n\| \tag{6.1}$$

The approximate solution at the nth iteration is symbolized by vn. We define

$$\hat{\eta} = \max_{x=0.1,0.2\cdots,0.9} \|\eta\| \tag{6.2}$$

The approximate solution obtained by proposed method for example (6.1) and example (6.2) with $\varepsilon = 2^{-4}$ and 2^{-6} at 10^{th} and 4^{th} iteration are plotted in Fig. (1) and Fig. (5) respectively, which coincide with the exact solutions plotted in Fig. (2) and Fig. (6), respectively. The approximate solution computed by proposed method for example (6.1) and example (6.2) with $\varepsilon = 2^{-10}$ at 10^{th} and 4^{th} iteration are plotted in Fig. (3) and Fig. (7) respectively, which coincide with the exact solutions plotted in Fig. (4) and Fig. (8), respectively. From figure (1) and (3), it is clear that the solution exhibits the boundary layers at the both boundaries, which corroborate the theoretical property of the considered test problem (6.1). The example (6.2) is a time depend second order differential equation, therefore the basic theory of singularly perturbed problem says that there should be two boundary layers at both the boundaries of spatial domain but contrary the graphs of the solution of the examples (6.2) in figure (5) and (8) shows one boundary layer at $x = 0$. This is happened because the exact solution of the example (6.2) $v(x,t) = e^{-t-\frac{x}{\sqrt{\varepsilon}}}$ satisfies the boundary condition $v(1,t) = e^{\left(-t-\frac{1}{\sqrt{\varepsilon}}\right)}$ at $x = 1$.

The computed results are tabulated in table (1) for example (6.1) and in table (2) for example (6.2). Therefore, from table (1) and (2) we observe that after further increases the number of iteration the error term decreases. Hence it is concluded that the proposed method is convergent and converges to the exact solution.

6.1 Reaction-diffusion problem

Example 6.1. *"Consider the following linear parabolic singularly perturbed reaction diffusion problem.*

$$\begin{cases} \frac{\partial v(x,t)}{\partial t} - \varepsilon \frac{\partial^2 v(x,t)}{\partial x^2} + v(x,t) = f(x,t), \\ v(x,0) = \left(\dfrac{e^{-\frac{x}{\sqrt{\epsilon}}} + e^{-\frac{1-x}{\sqrt{\epsilon}}}}{e^{-\frac{1}{\sqrt{\epsilon}}} + 1} - \cos^2(\pi x) \right), \quad x \in [0,1], \\ v(0,t) = 0, \quad v(1,t) = 0, \quad t \in [0,1] \end{cases} \tag{6.3}$$

with exact solution

$$v(x,t) = e^{-t} \left(\frac{e^{-\frac{x}{\sqrt{\epsilon}}} + e^{-\frac{1-x}{\sqrt{\epsilon}}}}{e^{-\frac{1}{\sqrt{\epsilon}}} + 1} - \cos^2(\pi x) \right)." \tag{6.4}$$

Example 6.2. *"Consider the following non-linear parabolic singularly perturbed reaction diffusion equation.*

$$\begin{cases} \frac{\partial v(x,t)}{\partial t} - \varepsilon \frac{\partial^2 v(x,t)}{\partial x^2} + v(x,t) + (v(x,t))^2 = e^{-2t-\frac{2x}{\sqrt{\varepsilon}}} - e^{-t-\frac{x}{\sqrt{\varepsilon}}}, \\ v(x,0) = e^{\left(-\frac{x}{\sqrt{\varepsilon}}\right)}, \qquad x \in [0,1], \\ v(0,t) = e^{(-t)}, \quad v(1,t) = e^{\left(-t-\frac{1}{\sqrt{\varepsilon}}\right)}, \quad t \in [0,1] \end{cases} \tag{6.5}$$

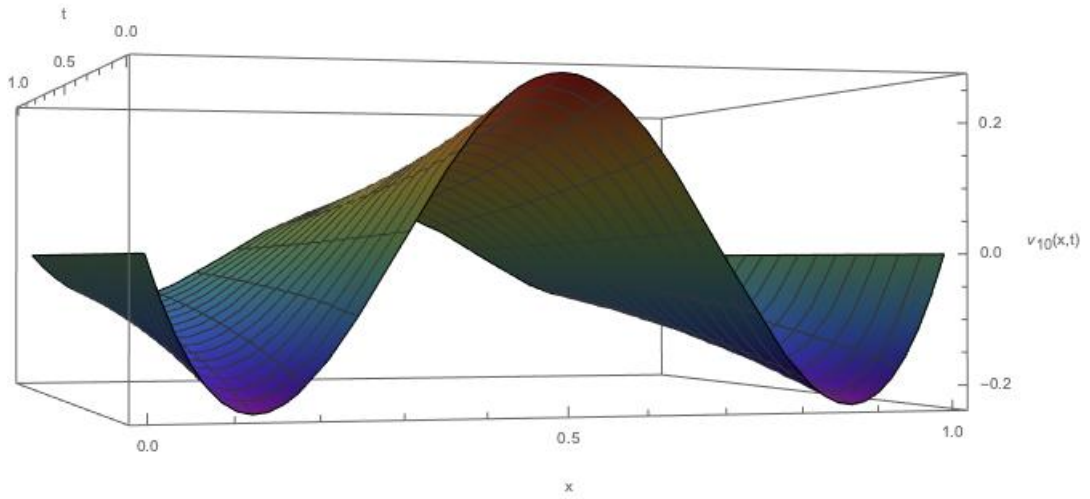


Fig. 1: "Solution Obtained by VIM Method at 10^{th} iteration for $\varepsilon = 2^{-4}$ for Example (6.1)."

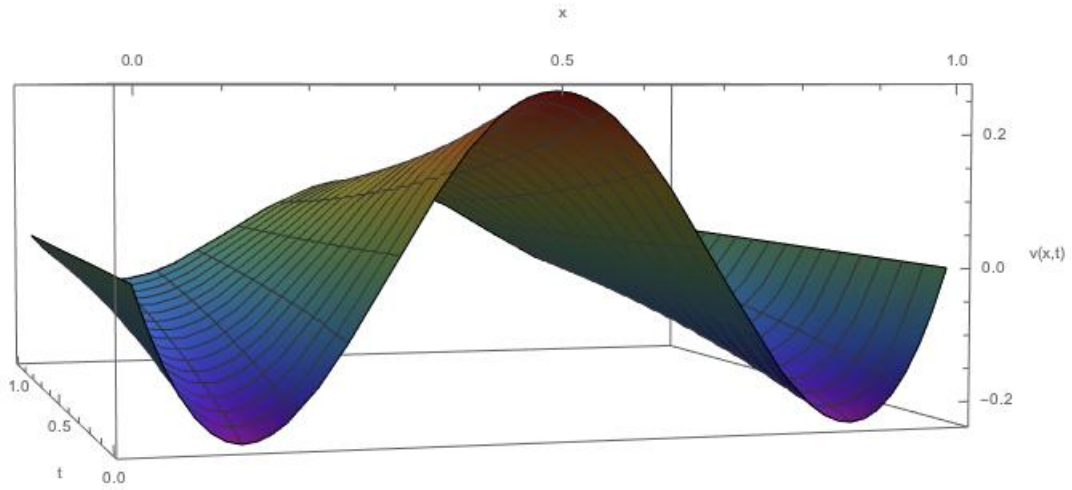


Fig. 2: "Exact Solution for $\varepsilon = 2^{-4}$ for Example (6.1)."

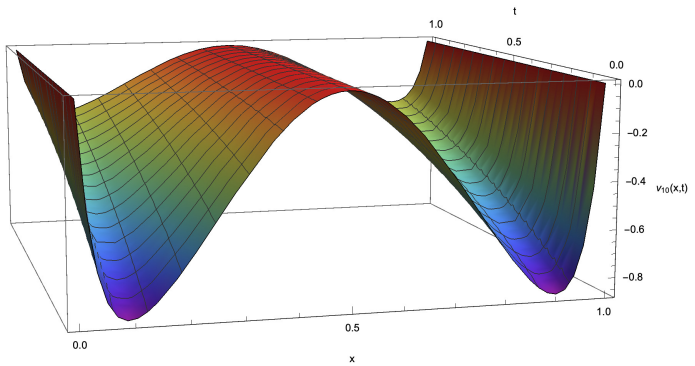


Fig. 3: "Solution Obtained by VIM Method at 10^{th} iteration for $\varepsilon = 2^{-10}$ for Example (6.1)."

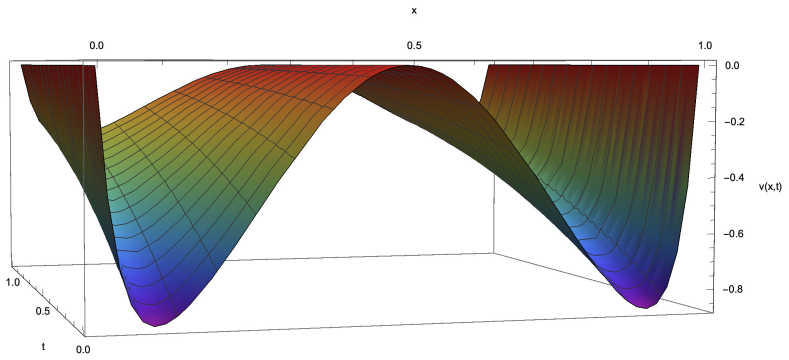


Fig. 4: "Exact Solution for $\varepsilon = 2^{-10}$ for Example (6.1)."

Table 1: "Maximum absolute error $\hat{\eta}$ of proposed method for Example (6.1) solved for fifth iteration v_5."

t	$\varepsilon = 2^{-4}$	$\varepsilon = 2^{-6}$	$\varepsilon = 2^{-8}$	$\varepsilon = 2^{-12}$	$\varepsilon = 2^{-16}$
0.2	2.6148×10^{-12}	2.1399×10^{-13}	1.3250×10^{-13}	3.1572×10^{-15}	1.2101×10^{-14}
0.4	6.8904×10^{-10}	4.6791×10^{-10}	2.0960×10^{-10}	8.5420×10^{-13}	9.40137×10^{-13}
0.6	5.9053×10^{-8}	3.9830×10^{-8}	1.7842×10^{-8}	7.2197×10^{-11}	7.8050×10^{-11}
0.8	1.3761×10^{-6}	9.2819×10^{-7}	4.1580×10^{-7}	1.6825×10^{-9}	1.8193×10^{-9}

with exact solution

$$v(x,t) = e^{-t-\frac{x}{\sqrt{\varepsilon}}}." \tag{6.6}$$

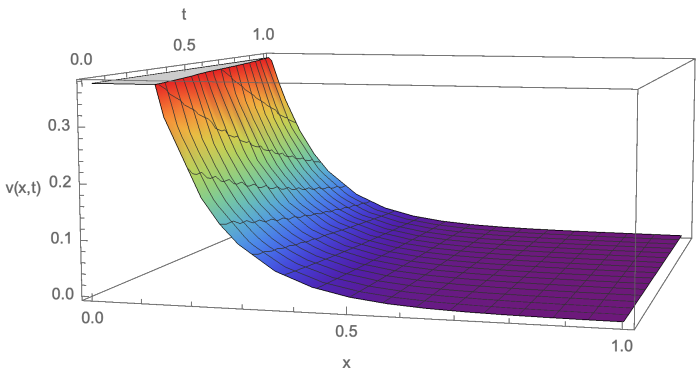


Fig. 5: "Exact Solution $v(x,t)$ for $\varepsilon = 2^{-6}$ for Example (6.2)."

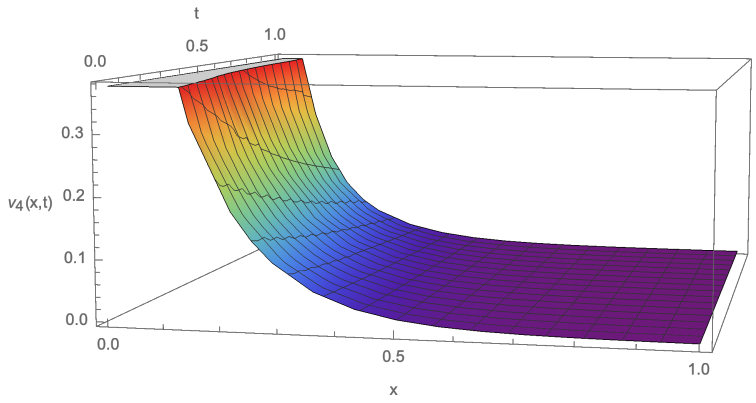


Fig. 6: "Solution Obtained by VIM Method at 4^{th} iteration for $\varepsilon = 2^{-6}$ for Example (6.2)."

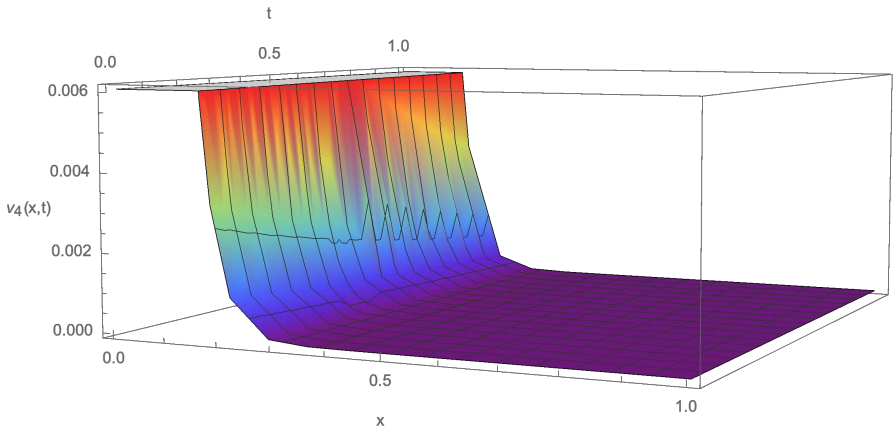


Fig. 7: "Solution Obtained by VIM Method at 4^{th} iteration for $\varepsilon = 2^{-10}$ for Example (6.2)."

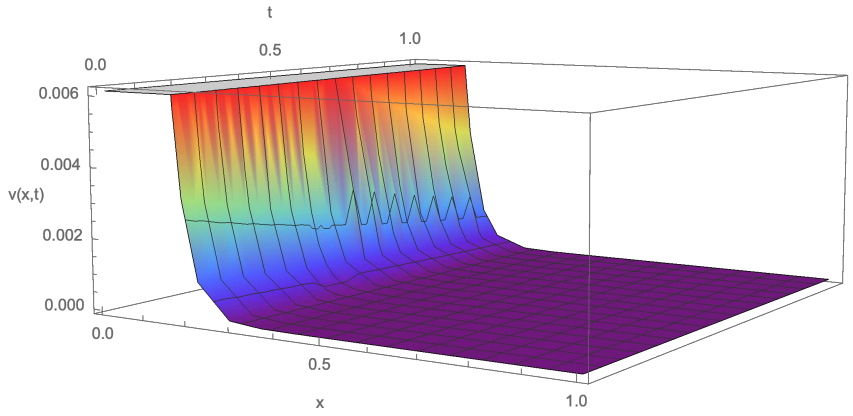


Fig. 8: "Exact Solution for $\varepsilon = 2^{-10}$ for Example (6.2)."

Table 2: "Maximum absolute error $\hat{\eta}$ of proposed method for Example (6.2) solved for third iteration v_3."

t	$\varepsilon = 2^{-2}$	$\varepsilon = 2^{-4}$	$\varepsilon = 2^{-6}$	$\varepsilon = 2^{-8}$	$\varepsilon = 2^{-10}$
0.2	3.85497×10^{-4}	1.3886×10^{-4}	2.40926×10^{-5}	2.62891×10^{-5}	1.68401×10^{-6}
0.4	3.56725×10^{-3}	8.5865×10^{-4}	7.28079×10^{-4}	4.24775×10^{-4}	2.53068×10^{-5}
0.6	5.11157×10^{-3}	5.11257×10^{-3}	5.11157×10^{-3}	2.15044×10^{-3}	1.20588×10^{-4}
0.8	2.26075×10^{-2}	2.19989×10^{-2}	1.98954×10^{-2}	6.7509×10^{-3}	3.59495×10^{-4}

6.2 Convection Reaction-diffusion problem

Example 6.3. *"Consider the following non-linear parabolic singularly perturbed convection-reaction diffusion problem.*

$$\begin{cases} \frac{\partial v(x,t)}{\partial t} - \varepsilon \frac{\partial^2 v(x,t)}{\partial x^2} + \frac{\partial v(x,t)}{\partial x} + v(x,t) + (v(x,t))^2 = f(x,t), \\ v(x,0) = e^{\left(-\frac{x}{\sqrt{\varepsilon}}\right)}, \qquad x \in [0,1], \\ v(0,t) = e^{(-t)}, \quad v(1,t) = e^{\left(-t-\frac{1}{\sqrt{\varepsilon}}\right)}, \quad t \in [0,1] \end{cases} \tag{6.7}$$

with exact solution

$$v(x,t) = e^{-t}\left(x\left(1 - e^{-1/\epsilon}\right) - e^{-\frac{1-x}{\epsilon}} + e^{-1/\epsilon}\right)." \tag{6.8}$$

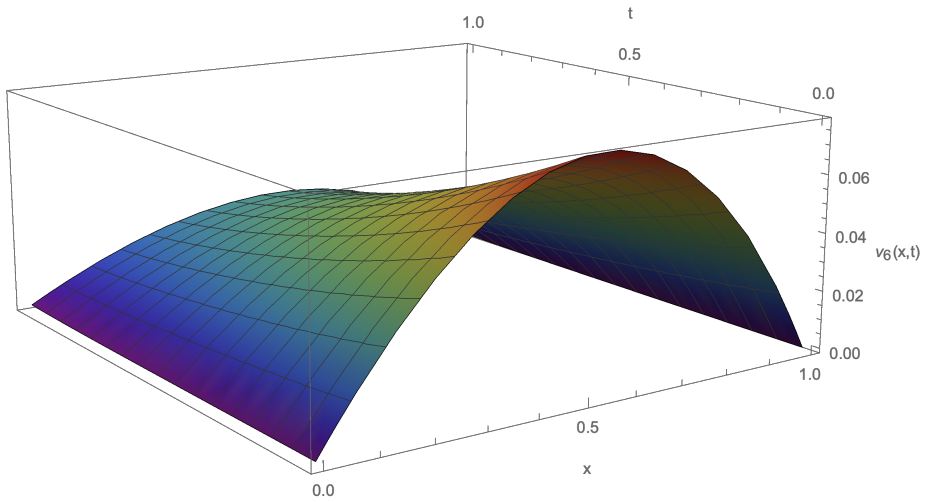


Fig. 9: "Solution Obtained by VIM Method at 6^{th} iteration for $\varepsilon = 2^0$ for Example (6.3)."

7 Conclusion

In this article, we have successfully developed the variational iteration method for solving NSPP-PDEs. In nonlinear problems, one often uses a quasi-linearization technique to linearize the problem and then solve the linearized problem by numerical or other existing techniques. Due to the linearization of the nonlinear problem, the approximated solution's accuracy somehow degenerates, which leads to deceptive solutions. In this paper, an attempt has been made to address such types of issues associated with NSPPPDEs. The method presented in this paper seems to be robust, as it maintains high accuracy concerning the perturbation parameter ε.

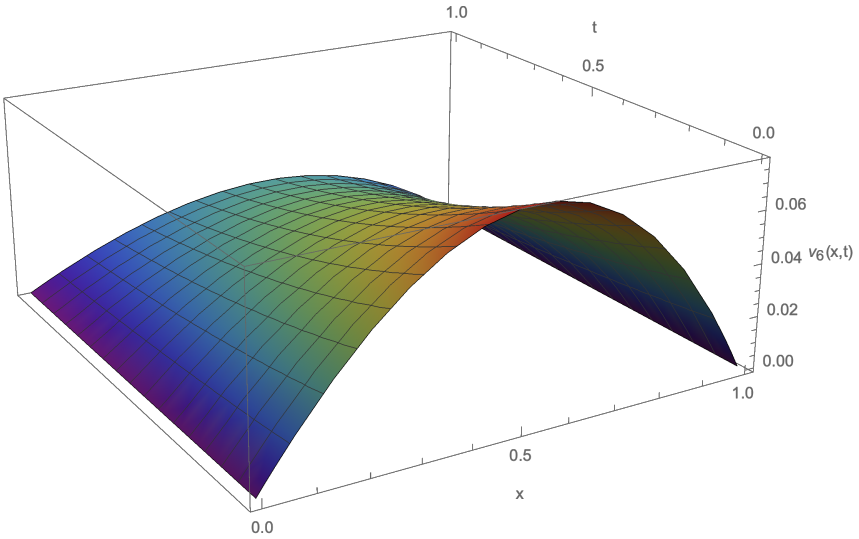


Fig. 10: "Exact Solution for $\varepsilon = 2^0$ for Example (6.3)."

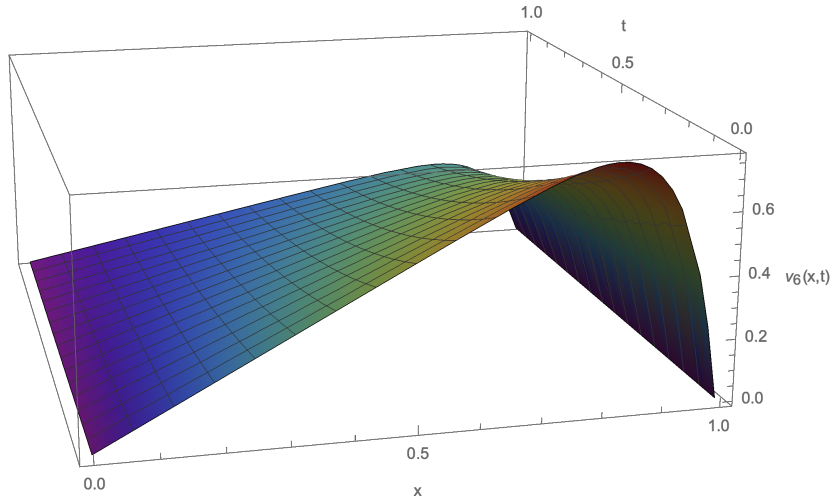


Fig. 11: "Solution Obtained by VIM Method at 6^{th} iteration for $\varepsilon = 2^{-4}$ for Example (6.3)."

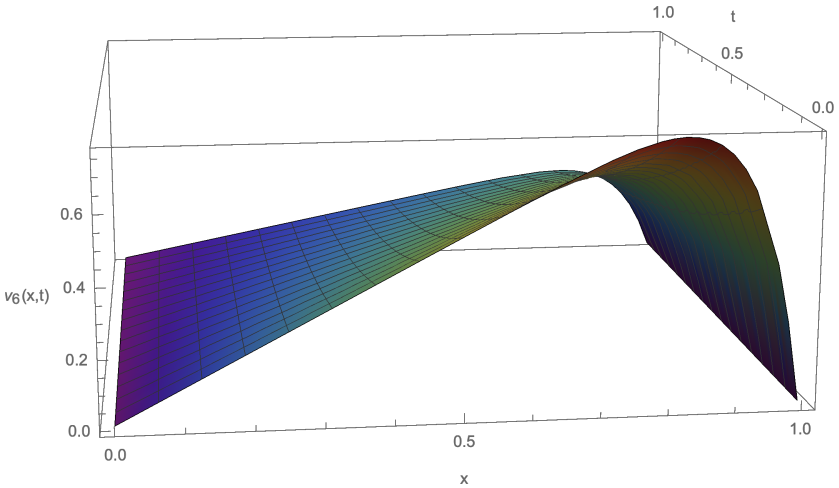


Fig. 12: "Exact Solution for $\varepsilon = 2^{-4}$ for Example (6.3)."

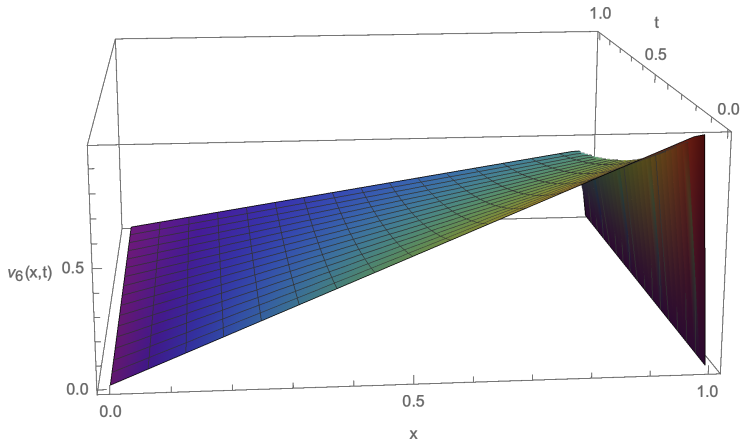


Fig. 13: "Solution Obtained by VIM Method at 6^{th} iteration for $\varepsilon = 2^{-8}$ for Example (6.3)."

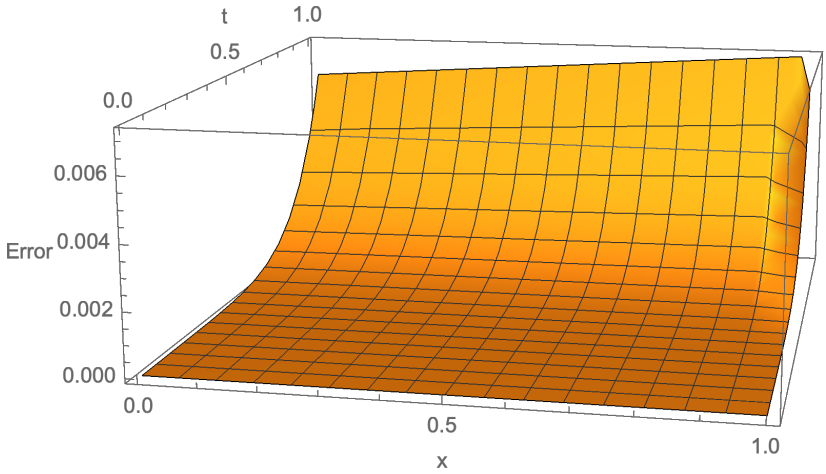


Fig. 14: Error at v_5 for $\varepsilon = 2^{-8}$ for Example (6.3).

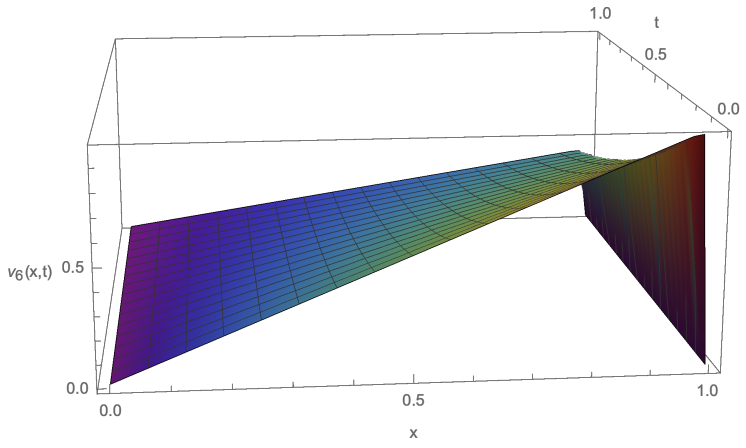


Fig. 15: "Solution Obtained by VIM Method at 6^{th} iteration for $\varepsilon = 2^{-8}$ for Example (6.3)."

Unlike numerical approaches, the proposed method is simple to implement because it does not require either linear or nonlinear systems of equations. It eliminates the need for time-consuming algebraic equations, prior simplification, discretion, or linearization, allowing it to greatly reduce the size of the calculation while maintaining excellent accuracy. We obtain a higher level of precision just in few iterations without any limiting assumption. The obtained numerical results appear to be promising and justify the efficiency of the proposed method.

Author Statement

"Kartikay Khari: Conceptualization, Methodology, Writing-Original draft, Supervision;

Data Availability Statement

"This manuscript has no associated data."

Conflicts of Interest

"The authors declare no conflict of interest".

References

[1] A. Al-Sawoor and M. Al-Amr, A new modification of variational iteration method for solving reaction diffusion system with fast reversible reaction, *J. Egyptian Math. Soc.* **22** (2014), 396–401. Available at https://doi.org/10.1016/j.joems.2013.12.011.

[2] I. Ali and N. Malik, Hilfer fractional advection diffusion equations with power law initial condition: A numerical study using variational iteration method, *Comput. Math. Appl.* **68** (2014), 1161–1179. Available at https://doi.org/10.1016/j.camwa.2014.08.021.

[3] M. Amrein and T. P. Wihler, An adaptive space-time newton–galerkin approach for semilinear singularly perturbed parabolic evolution equations, *IMA J. Numer. Anal.* **37** no. 4 (2017), 2004–2019. Available at https://doi.org/10.1093/imanum/drw049.

[4] I. Boglaev, Monotone iterative algorithms for a nonlinear singularly perturbed parabolic problem, *J. Comput. Appl. Math.* **172** no. 2 (2004), 313–335. Available at https://doi.org/10.1016/j.cam.2004.02.010.

[5] I. Boglaev, An inexact monotone method for solving semilinear parabolic problems, *Appl. Math. Comp.* **219** no. 6 (2012), 3253–3263. Available at https://doi.org/10.1016/j.amc.2012.09.067.

[6] Y. Cheng, C. Song, and Y. Mei, Local discontinuous galerkin method for time-dependent singularly perturbed semilinear reaction-diffusion problems, *Comput. Methods Appl. Math.* **21** no. 1 (2021), 31–52. doi:https://doi.org/10.1515/cmam-2019-0185.

[7] O. Doeva, P. K. Masjedi, and P. M. Weaver, A semi-analytical approach based on the variational iteration method for static analysis of composite beams, *Compos. Struct.* **257** (2021), 113110. Available at https://doi.org/10.1016/j.compstruct.2020.113110.

[8] T. El-Sayed and H. El-Mongy, Application of variational iteration method to free vibration analysis of a tapered beam mounted on two-degree of freedom subsystems, *Appl. Math. Model.* **58** (2018), 349–364. Available at https://doi.org/10.1016/j.apm.2018.02.005.

[9] H. Ghaneai and M. Hosseini, Variational iteration method with an auxiliary parameter for solving wave-like and heat-like equations in large domains, *Comput. Math. Appl.* **69** (2015), 363–373. Available at https://doi.org/10.1016/j.camwa.2014.11.007.

[10] J. He, Variational iteration method for delay differential equations, *Commun. Nonlinear Sci. Numer. Simul.* **2** (1997), 235–236. Available at https://doi.org/10.1016/S1007-5704(97)90008-3.

[11] J. He, Variational iteration method - a kind of non-linear analytical technique: some examples, *Int. J. Non Linear Mech.* **34** (1999), 699–708. Available at https://doi.org/10.1016/S0020-7462(98)00048-1.

[12] J.-H. He, H.-Y. Kong, R.-X. Chen, M.-s. Hu, and Q.-l. Chen, Variational iteration method for bratu-like equation arising in electrospinning, *Carbohydr. Polym.* **105** (2014), 229–230. Available at https://doi.org/10.1016/j.carbpol.2014.01.044.

[13] M. Inokuti, H. Sekine, and T. Mura, *General use of the Lagrange multiplier in nonlinear mathematical physics*, Pergamon Press, Oxford, 1978.

[14] K. K. Ali, A. Hadhoud, and M. Shaalan, Numerical study of self-adjoint singularly perturbed two-point boundary value problems using collocation method with error estimation, *J. Ocean Eng. Sci.* **3** (2018), 237–243. doi:https://doi.org/10.1016/j.joes.2018.07.001.

[15] M. J. Kabeto and G. F. Duressa, Robust numerical method for singularly perturbed semilinear parabolic differential difference equations, *Math. Comput. Simul.* **188** (2021), 537–547. doi:https://doi.org/10.1016/j.matcom.2021.05.005.

[16] A. R. Kanth and K. Aruna, He's variational iterational method for treating nonlinear singular boundary value problems, *Comput. Math. Appl.* **60** (2010), 821–829. Available at https://doi.org/10.1016/j.camwa.2010.05.029.

[17] G. G. Kiltu, G. F. Duressa, and T. A. Bullo, Computational method for singularly perturbed delay differential equations of the reaction-diffusion type with negative shift, *J. Ocean Eng. Sci.* **6** (2021), 285–291. doi:https://doi.org/10.1016/j.joes.2021.02.001.

[18] N. Kopteva and T. Linss, Maximum norm a posteriori error estimation for a time-dependent reaction-diffusion problem, *Comput. Methods Appl. Math.* **12** no. 2 (2012), 189–205. Available at https://doi.org/10.2478/cmam-2012-0013.

[19] N. Kopteva and T. Linss, Maximum norm a posteriori error estimation for parabolic problems using elliptic reconstructions, *SIAM J. Numer. Anal.* **51** no. 3 (2013), 1494–1524. Available at https://doi.org/10.1137/110830563.

[20] F. Motaharifar, M. Ghassabi, and R. Talebitooti, A variational iteration method (vim) for nonlinear dynamic response of a cracked plate interacting with a fluid media, *Eng. Comput.* (2020), 1–20. Available at https://doi.org/10.1007/s00366-020-00998-w.

[21] S. Mungkasi, Variational iteration and successive approximation methods for a sir epidemic model with constant vaccination strategy, *Appl. Math. Model.* **90** (2021), 1–10. Available at https://doi.org/10.1016/j.apm.2020.08.058.

[22] M. Nadeem and J.-H. He, He–laplace variational iteration method for solving the nonlinear equations arising in chemical kinetics and population dynamics, *J. Math. Chem.* **59** no. 5 (2021), 1234–1245. Available at https://doi.org/10.1007/s10910-021-01236-4.

[23] Z. M. Odibat, A study on the convergence of variational iteration method, *Math. Comput. Model.* **51** no. 9 (2010), 1181–1192. doi:https://doi.org/10.1016/j.mcm.2009.12.034.

[24] J. Ramos, On the variational iteration method and other iterative techniques for nonlinear differential equations, *Appl. Math.Comp.* **199** (2008), 39–69. Available at https://doi.org/10.1016/j.amc.2007.09.024.

[25] M. Sakar and H. Ergören, Alternative variational iteration method for solving the time fractional Fornberg-Whitham equation, *Appl. Math. Model.* **14** (2015), 3972–3979. Available at https://doi.org/10.1016/j.apm.2014.11.048.

[26] A.-M. Wazwaz, Solving the non-isothermal reaction-diffusion model equations in a spherical catalyst by the variational iteration method, *Chem. Phys. Lett.* **679** (2017), 132–136. Available at https://doi.org/10.1016/j.cplett.2017.04.077.

[27] A.-M. Wazwaz, Optical bright and dark soliton solutions for coupled nonlinear schrödinger (cnls) equations by the variational iteration method, *Optik* **207** (2020), 164457. Available at https://doi.org/10.1016/j.ijleo.2020.164457.

[28] G.-C. Wu and D. Baleanu, Variational iteration method for the Burger's flow with fractional derivatives: New Lagrange multipliers, *Appl. Math. Model.* **37** (2013), 6183–6190. Available at https://doi.org/10.1016/j.apm.2012.12.018.

An Overview of Machine Learning

Wazid Ansari

B.TECH, Delhi Technological University

ABSTRACT

When we talk about solving a highly complex problem which is selective in nature, AI (artificial Intelligence) plays a very important and exponential role over the growth in the past few years processing the Machine Learning for performing the human sounded learning. Nowadays if we talk about a car which drives on its own to be very specific we say it driverless car to the auto generated recommendations on Netflix or on any other online show platform, we are ought to be surrounded by artificial intelligence, even if it is not noticed by us. Moreover there are certain companies that also have recently adopted some new kind of frameworks to upgrade their routines which consists of mainly by algorithms which are able to solve much more complex problems in a very short period of time.

Artificial intelligence, a technology which has been unconditionally stunning. This all is only possible due to a sub category, or we can say a sub-field of Artificial intelligence known as Machine Learning. Its growing comparatively faster than its subsidiaries.

Talking on a small scale, machine learning can also be yielded as a quite simple and sorted system which is able to catch patterns on the data and also it learns from doing so. Moreover, it is more surely stated that if the learning process occurs in a large scale, it will enhance and allow the machine to copy the human behavior. By acquiring such process it will also perfume tasks that would eventually require human intelligence. Certainly , Machine learning (Arel, I., Rose, D. C., & Karnowski, T. P.,2010) has already become a blissful and profitable scientific area which strains continuous process of learning as there is always something new which is being discovered.

From the last decades, a enormous number of algorithms have also been planned which would individually help to understand more and remove the confusion of how to use them.

INTRODUCTION

During the early 2017, in the most famous paper journal, the author compared data as the new variety of an oil as the most valuable asset as of now. If we talk about the likeness, it's pretty clear as if the data is in raw format it has no such value if not prepared and analysed, similarly in the case of crude oil which had to be refined before using it as a fuel. Furthermore on looking to the possessions, it can be stated that when we talk about its importance, data is in the position of oil in 21st century as the oil was at the same in 18th century. This perception on Big Data had now been around for a while of years, but unfortunately the taking companies never had the proper development of their infrastructures which can store all the data and also the technology which could further analyse it. These days, companies are storing huge amounts of data every minute and they have specialists such as Data Scientists to extract value from it. Globally speaking, storing the data is a never-end exponential curve, every digital interaction leaves a footprint of data. According to a study of the World Economic Forum, in 2020 the digital universe will reach 44 Zettabytes (that is 44 with 21 zeros after it). People usually fall in the mistake of thinking that data is only increasing in terms of Volume, but in fact Big Data is described in the 4v's: Volume (the amount of collected data), Variety (different forms of data), Velocity (the speed of new generated data) and Veracity (the trustworthiness of the data). A fifth "v" may be added which refers to "Value", this may seem quite obvious but the data is only useful if we can convert huge amounts of data into value to the companies, so the data must have some intrinsic value. Big data has taken the main stage, it is a buzzword that everybody wants to speak about (World Economic Forum, 2020).

Though, we have understood the fact that big data enhances and also provides a tremendous source of information and also it becomes useless when not done well or correctly understood, which will gradually lead us to the automation in the era. In order to handle such a huge amount of data it will first require all the necessary infrastructures and also the resources to run that data (Andrieu, C., De Freitas,2003).

But even after presuming all the information can be processed by us, how will be the value extracted from it? Or in other words, how can we convert all that data into knowledge? And what sort of algos can be used to achieve that? All the answers related to the above questions are all about machine learning which is to be achieved with the help of these thesis. Machine Learning (ML) has been one of the most innovative breakthroughs of computer science by allowing computers to learn by themselves as they process large datasets. Additionally, Machine Learning has been already enacted by varieties of companies around the globe as they start to be much more aware of its potential.

This was first introduced as the concept of Business Intelligence as it was just the first step of a long journey where companies start to incorporate complex data analysis in their routines. The study of Machine Leaning indulges to increase the advancement of analytics from which many scientists , engineers and analysts who performs the same get their own prospect to further add anything new specially in algos, innovation and optimization. Some enhancements of the systems performance

have been absolutely impressive allowing the Machine Learning systems to exceed humans in areas of patterns detection and image recognition.

LITERATURE REVIEW

Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence. It is the learning and building of algorithms that can learn from and make predictions on data sets. These procedures operate by construction of a model from example inputs in order to make data-driven predictions or choices rather than following firm static program instructions. "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." – (Tom Mitchell, Carnegie Mellon University).

It's the opportunity of this research paper is to make concentration amid upcoming scholars about fresh developments in expertise, precisely profound knowledge a part of machine learning which finds requests in big data analytics and Machine Learning, Machine Learning, Big Data, Artificial Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and intelligence (Kumar, R, nd). It is the knowledge and construction of algorithms that can also enhance learning from and also to make predictions on data sets. These procedures operate by rom example inputs in order to make driven predictions or choices rather than following firm "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its mince on T, as measured by P, improves with Tom Mitchell, Carnegie Mellon University. So if we want our program to foresee, for example, traffic forms at a busy node (task T), we can run it through a machine out previous traffic patterns (experience E) and, if it has successfully "learned", it will then do better at predicting upcoming traffic patterns (performance We need machine learning in the following cases:- navigating on Mars. Humans are unable to explain their expertise. E. g. Speech Solution changes with time E. g. Temperature Control.

A new area of machine learning research, which has been introduced with the objective of moving machine learning closer to one of its original goals: Artificial Intelligence. Machine learning draws its roots from Noncognition; an Artificial Neuron Network (ANN) introduced by Kuniko Fukushima in 1980. An ANN is an interconnected network of processing units emulating the network of neurons in the brain. The idea behind ANN was to develop a learning method by modelling the human brain. However, this method lost machine learning community owing to the fact that it required an impractical amount of time as well as a humungous amount of data to train the network parameters for any decent application. Machine learning is a method to train multi(and hence the word "Machine") ANN using little data. This is the reason why ANN is back in the game. Using an example to compare Machine Learning with Machine Learning, we can say that if a machine learning algorithm learns parts of a face like eyes and nose for face detection tasks, a Machine learning algorithm will learn extra features like the distance between eyes and the length of the nose. Hence Machine major step away from Shallow Learning Algorithms. The term Machine learning gained traction in mid"vanishing

gradient problem" responsible reduction in speed was solved in a publication Hinton and Ruslan Salahuddin. They showed layered feed forward neural network could retrained at a time, treating each layer unsupervised restricted Boltzmann's machine, supervised back-propagation for fine tuning. A Machine Neural Network (DNN) is defined to Neural Network (ANN) with at least one hidden An Overview of Machine Learning and its Applications Dayananda Sagar College of Engineering, Bengaluru-78 A new area of machine learning research, which has been introduced with the objective of moving machine learning closer to one of its original goals: Artificial Intelligence(Simon et al., 2016)). Machine learning draws its roots from Noncognition; an Artificial NN) introduced by Kuniko Fukushima in 1980. An ANN is an interconnected network of processing units emulating the network of neurons in the brain. The idea behind ANN was to develop a learning method by modelling favour within the machine learning community owing to the fact that it required an impractical amount of time as well as a humungous amount of data to train the network parameters for any decent application. Machine learning is a method to train multi-layer and hence the word "Machine") ANN using little data (Kumar, nd.). This is the reason why ANN is back in the game. Using an example to compare Machine Learning with Machine Learning, we can say that if a machine learning algorithm learns parts of a face like face detection tasks, a Machine learning algorithm will learn extra features like the distance between Machine Learning is a Algorithms. mid 2000s after the for causing a publication by Geoffrey showed how a multi-could be effectively in turn as an machine, then using to be an Artificial hidden layer of units between the input and output layers. Added levels of abstraction, thus capability. The most popular kinds are known as Convolutional NeuralConvNets. These area type of feed network, extensively used in computer individual neurons are tiled in such overlapping regions in the visual field have also been successfully applied recognition (ASR). Machine Belief Networks Machine Belief Networks are some otharchitectures in use.

Types of Learning

The primary categories of machine learning are supervised, unsupervised, and semi-supervised learning. We will focus on the first two in this article. In supervised learning, the data contains the response variable (label) being modelled, and with the goal being that you would like to predict the value or class of the unseen data. Unsupervised learning involves learning from a dataset that has no label or response variable, and is therefore more about finding patterns than prediction.

Imagine that each row of the data is essentially a team snapshot (or observation) of relevant statistics for every game since 1920. The columns in this case, and the data contained in each, represent the features (values) of the data, and may include feature data such as game date, game opponent, season wins, season losses, season ending divisional position, post-season berth (Y/N), post-season stats, and perhaps stats specific to the three phases of the game: offense, defense, and special teams.

In the supervised case, your goal may be to use this data to predict if the Bears will win or lose against a certain team during a given game, and at a given field (home or away). Keep in mind that anything can happen in football in terms of pre and game-time injuries, weather conditions, bad

referee calls, and so on, so take this simply as an example of an application of supervised learning with a yes or no response (prediction), as opposed to determining the probability or likelihood of 'Da Bears' getting the win (Chao, W. L.,2011).

Since you have historic data of wins and losses (the response) against certain teams at certain football fields, you can leverage supervised learning to create a model to make that prediction. Now suppose that your goal is to find patterns in the historic data and learn something that you don't already know, or group the team in certain ways throughout history. To do so, you run an unsupervised machine learning algorithm that clusters (groups) the data automatically, and then analyze the clustering results.

Machine Learning Goals and Outputs

- Machine learning algorithms are used primarily for the following types of output:
- Clustering (Unsupervised)
- Two-class and multi-class classification (Supervised)
- Regression: Univariate, Multivariate, etc. (Supervised)
- Anomaly detection (Unsupervised and Supervised)
- Recommendation systems (aka recommendation engine)

Specific algorithms that are used for each output type are discussed in the next section, but first, let's give a general overview of each of the above output, or problem types. There are two disadvantages with DNNs, And computation time. Overfitting very specific details on the training layers. As a result, the DNN performs is given as input, but poorly when This problem is solved by a regularization where some units are the hidden layers during training. Computations required here are well we could speed up the computations enormous processing power. The figure below illustrates how images can be achieved using a Machine every layer learns a single feature can learn the different edges; in slightly more complex features like such as ears, noses, and eyes. In the even more complex features like the face shapes (McCrea, N.,2014). The final representations applications of categorization. Applications of Machine learning are as• Optical Character Recognition extracting text from it. layers. The extra layers give it thus enhancing its modelling kinds of Machine Learning models, Neural Nets (CNN), or simply feed-forward artificial neural computer vision, where the such a way that they respond to field. In recent times, CNNs applied to automatic speech Networks and Convolutional other popular Machine learning DNNs (Simon et al., 2015). They are overfitting Overfitting is when the DNN learns training data using its hidden performs well if the training data when the input data is different. method called "dropout" are randomly removed from training. The matrix and vector well suited for GPUs. Hence, computations by harnessing their how categorizing of different Machine learning model where at a time. Hence we could only speed up the computation by harnessing their enormous processing power. Machine Learning and Big Data are two high-focus areas of data science. Machine learning algorithms extract complex data patterns, through a hierarchical

learning process by analyzing and learning massive amounts of unsupervised data (Big Data). This makes it an extremely valuable tool for big data analysers (Arel, I., Rose, D. C., & Karnowski, T. P.,2010).

Big Data has 4 important characteristics, namely, Volume, Variety, Velocity, and Veracity. They are Learning algorithms are mainly concerned with issues related to Volume and Variety. Machine Learning algorithms deal with massive amounts of data, i. e. Volume whereas shallow learning algorithms fail to understand complex data patterns which are inevitably present in large data sets. Moreover, Machine Learning deals with analyzing raw data presented in different formats from different sources, i. e. Variety in Big Data. This minimizes the need for input from human experts to retrieve features from all new data typesfound in Big Data.

Big Data has 4 important characteristics, namely, Volume, Variety, Velocity and Veracity. They are Learning algorithms are mainly concerned with issues related to Volume and Variety. Machine Learning algorithms deal with massive amounts of data, i. e. Volume whereas shallow learning algorithms fail to understand complex data patterns which are inevitably present in large data sets. Moreover, Machine Learning deals with analyzing raw data presented in different formats from different sources, i. e. Variety in Big Data. This minimizes the need for input from human experts to retrieve features from all new data types found in Big Data (Patterson, J., & Gibson, A.,2017).

Semantic Indexing, Data Tagging and Fast Information Retrieval are the main objectives of Machine Learning in Big Data. Consider data that is unstructured and unorganized. Haphazard storage of massive amounts of data cannot be used as a source of knowledge because looking through such data for specific topics of interest and retrieving all relevant and related information would be a tedious task. Using Semantic Indexing and Data Tagging, we identify patterns in the relationships between terms and concepts based on the principle that words used in the same context have similar meanings. The related words can then be stored close to each other in the memory. T s helps us prese data in a more comprehensive manner and helps in improving efficiency. A direct result of such a form of storage would be that search engines would work more quickly and efficiently (Najafabadi,2015).

Artificial Intelligence is the theory and development of computers which are capable of performing tasks which humans can. Machine learning represents the rudimentary level of attempts towards achieving this task. It is utilized in visual perception, speech recognition, game playing, expert systems, decision-making, medicine, aviation and translation between languages (Economic times).

RESEARCH METHODOLOGY

The elaboration of this expert speculation was basically settled on ahead of time existing calculations which found a hard and fast theorical structure. The progress of this system was a long participation

formed by a few phases. The first of those strategies was to perceive the fundamental intends to talk about during this work, then again to depict which appraisal questions ought to be responded to with this work. The pondering was not exclusively to shape the speculation about the Machine Learning models, yet likewise portraying in which conditions each and every one of them ought to be utilized in basically the same manner as depicting their benefits and weaknesses. Since it was an astoundingly theoretical subject it was basic to complete an elevated and cautious assessment of creating outline, which was the second step of the correspondence. In the third step it was time to pick which assessments would be explored in basically the same manner as getting sorted out all the speculation structure. From the most reliable reference point that the improvement of this idea was considered as a secluding variable from other comparative works. To accomplish this, it was contemplated that each part would be committed to a family of assessments (that share near attributes) and in each sub-section a particular calculation would be clarify. The fourth and last advancement of the cycle was to dispense with decision about the calculations and to respond to the fundamental evaluation solicitations of the suggestion. This was made in two intriguing minutes, toward the finishing of each part with a system of the calculation and around the fulfilment of the idea with the outcomes acquired.

CONCLUSION

As mentioned, machine learning leverages algorithms to automatically model and find patterns in data, usually with the goal of predicting some target output or response. These algorithms are heavily based on statistics and mathematical optimization. Optimization is the process of finding the smallest or largest value (minima or maxima) of a function, often referred to as a loss, or cost function in the minimization case. One of the most popular optimization algorithms used in machine learning is called gradient descent, and another is known as the normal equation.

In a nutshell, machine learning is all about automatically learning a highly accurate predictive or classifier model, or finding unknown patterns in data, by leveraging learning algorithms and optimization techniques. Machine learning techniques have been criticized because there is no way of representing causal relationships (such as between diseases and their symptoms), and the algorithms fail to acquire abstract ideas like “sibling” or “identical to.” Not much theory is available for most of the methods which is disadvantageous to beginners. Machine Learning is only a small step towards building machines which have human-like intelligence. Further advancements must be made in order to achieve our ultimate goal. Organizations like Google, Facebook, Microsoft and Baidu (a Chinese search engine) are buying into this technology and exploring various avenues available. For example, Facebook is using Machine learning to automatically tag uploaded pictures. Google’s Machine Mind focusses on exploring new techniques in this area. Recent trends show that the interest in machine learning has only been

growing with time and has sparked an interest in countries like India and Singapore. Thus it has emerged as one of the most promising fields of technology in recent times.

BJAR

References

1. E. Alpaydin, Introduction to Machine Learning,MIT Press, 2014.
2. C.M. Bishop, Pattern Recognition and MachineLearning, Springer, 2006.
3. An introduction to machine learning theoryand its applications,A visual tutorial by NickMccrea.
4. Inza, I., Calvo, B., Armañanzas, R., Bengoetxea,E., Larrañaga, P., & Lozano, J. Machinelearning: an indispensable tool inbioinformatics. In Bioinformatics methods inclinical research (pp. 25-48), Humana Press,2010.
5. K.P. Murphy, Machine Learning: AProbabilistic Perspective, MIT Press, 2012.
6. Khattree, R., & Naik, D, Machine LearningTechniques for Bioinformatics InComputational Methods in Biomedical Research (pp. 57-88). Chapman and Hall/CRC,2007.
7. Kumar, R. MACHINE LEARNING: CONCEPT, DEEP LEARNING AND APPLICATIONS. *WIRELESS COMMUNICATION AND MATHEMATICS*, 49.
8. Chao, W. L. (2011). Machine learning tutorial. *Digital Image and Signal Processing*.
9. Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE computational intelligence magazine*, *5*(4), 13-18.
10. Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of big data*, *2*(1), 1-21.

BJAR

PAPER • **OPEN ACCESS**

Analog Performance Analysis of a Novel 5nm Stacked Oxide Top Bottom Gated Junctionless FinFET

To cite this article: L Gangwani and S Hajela 2022 *IOP Conf. Ser.: Mater. Sci. Eng.* **1258** 012046

View the article online for updates and enhancements.

You may also like

- Strategic Review on Different Materials for FinFET Structure Performance Optimization
K Bindu Madhavi and Suman Lata Tripathi
- FinFET Devices and Integration
Huiming Bu
- Heated Ion Implantation Technology for High Performance Metal-Gate/High-k CMOS SOI Finfets
Wataru Mizubayashi, Hiroshi Onoda, Yoshiki Nakashima et al.

This content was downloaded from IP address 14.139.251.98 on 03/11/2022 at 06:09

Analog Performance Analysis of a Novel 5nm Stacked Oxide Top Bottom Gated Junctionless FinFET

L Gangwani[1] and S Hajela[1]

[1]Applied Physics Department, Delhi Technological University, New Delhi, India

E-mail: lakshyagangwani1@gmail.com

Abstract: This paper depicts the analog investigation of a Novel Stacked Oxide Top Bottom Gated Junctionless (TBG-JL) Fin-shaped Field Effect Transistor (FinFET) structure. The structure is designed in this way to enhance the switching performance and curtail the Short Channel Effects (SCEs). The study is done on Silvaco Atlas TCAD tools for the Novel Stacked Oxide FinFET device and analyzed against Top-Bottom Gated Junctionless FinFET and Traditional FinFET. The Novel device shows 280 times improvement in switching ratio when stacked against the Traditional JL FinFET at identical gate length at room temperature (300K). The results from device simulation affirm that the Novel device has better analog performance over Traditional Junctionless FinFET and diminishes the Short Channel Effects (SCEs) a cut above the Traditional Junctionless FinFET.

Keywords*: FinFET, Top-Bottom Gated, Junctionless, Stacked Oxide, Switching Ratio*

1. INTRODUCTION

New technologies are being developed every day. These progressions and types of developments are intended to serve the larger part and make daily tasks more efficient. Amidst such progressions it ends up being extensively more fundamental to work on the usefulness and resource use of our devices [1]. Various headways are being designed to achieve precisely that. This trend of downsizing the devices to achieve more portability and efficiency starting right from MOSFET devices till SOI and FinFET helps in developing far more capable devices which are handheld at the same time [2]. But reducing the device dimension comes at a cost. As we scale down, the dominance of certain undesirable attributes becomes more prominent. These undesirable attributes comprises of various short channel effects such as Velocity Saturation, Subthreshold Swing and DIBL. These undesirable attributes can cause delays and inefficient power use [3]. The new gate configurations for FinFET technology is a step in the right direction. These progressions are being embraced by different goliaths in the Chip and Semiconductor fabricating industry. Thusly in this research article, we have investigated the analog performance parameters of an Innovative Stacked Oxide Top-Bottom Gated (TBG) Junctionless (JL) FinFET [4]. The structure has been contrasted with the TBG-JL FinFET and a Traditional Junctionless FinFET structure to better glean the merits of the novel structure in terms of improved switching and reduced short channel effects.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. DEVICE DESIGN

Figures 1, 2 and 3 outlines the Innovative FinFET structure. The oxide layer has been stacked i.e. one layer of Hafnium Oxide (HfO_2) placed over layer of Silicon Dioxide (SiO_2). Stacking is done to further develop the Ion/Ioff ratio and productivity of the gadget.

Gate is present only above the substrate in case for the Traditional Junctionless FinFET, wherein the Novel Stacked Oxide TBG JL FinFET has gates both above and below the substrate, subsequently working on the control of gate over the channel. This aides in improving the gadget efficiency [1].

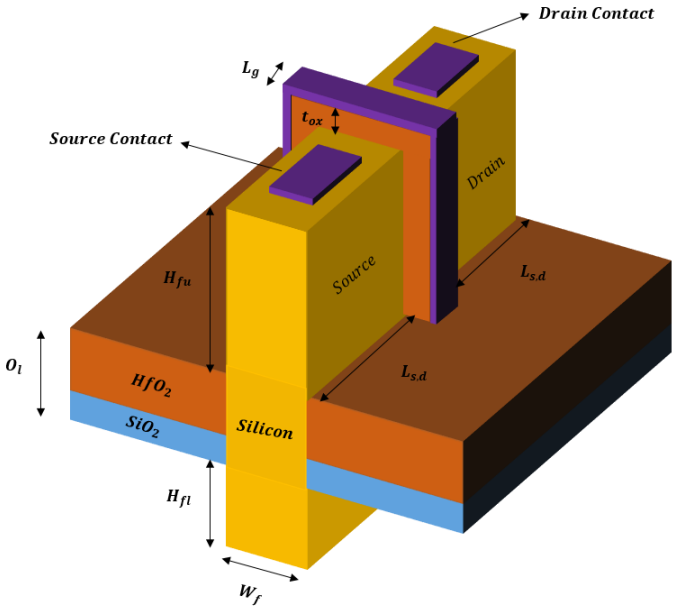


Figure 1. Systematic 3D structure of the Novel FinFET Structure

The Novel structure has two layers of oxide. 7nm HfO_2 oxide layer is placed on 3nm SiO_2 oxide layer. Hafnium Oxide (HfO_2) has dielectric constant in the range of 20 to 25 whereas for Silicon Dioxide (SiO_2) it is around 3.7 [5]. The dielectric constant of Hafnium Oxide (HfO_2) is taken as 25 for the simulation of this device structure using the TCAD tools [6]. This stacking of oxide layer alongside the fins reaching out on both the sides of the oxide helps in improving the analog performance and switching ratio of the device when stacked against the Conventional JL FinFET. Additionally it upgrades the gate controllability of the gadget when stacked against the TBG JL FinFET [2].

The Hafnium Oxide (HfO_2) has a band gap of 5.3- 5.7 eV and can behave as a dielectric insulator. HfO2 has high dielectric constant, great thermal stability and also lower values of leakage current [7].

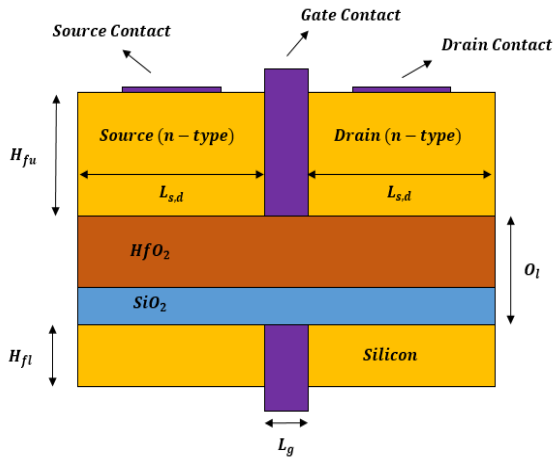


Figure 2. Horizontal View of the Novel FinFET Structure

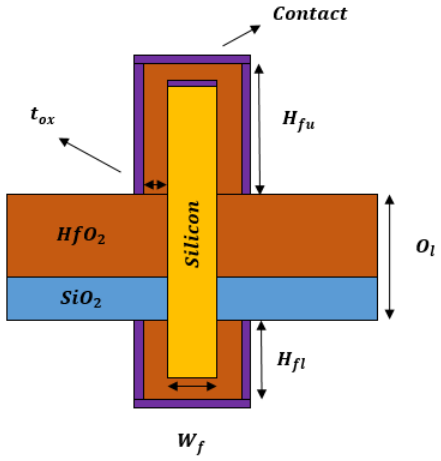


Figure 3. Vertical View of the Novel FinFET Structure

The Novel device structure simulated is of n-type with 5.0×10^{18} (cm^{-3}) doping concentration for all the regions of source, drain and channel.

So as we increase the doping from 5.0×10^{18} (cm^{-3}) to 5.0×10^{19} (cm^{-3}), the threshold voltage dips from 0.42V to 0.39V but the Switching ratio decreases by 16 times.

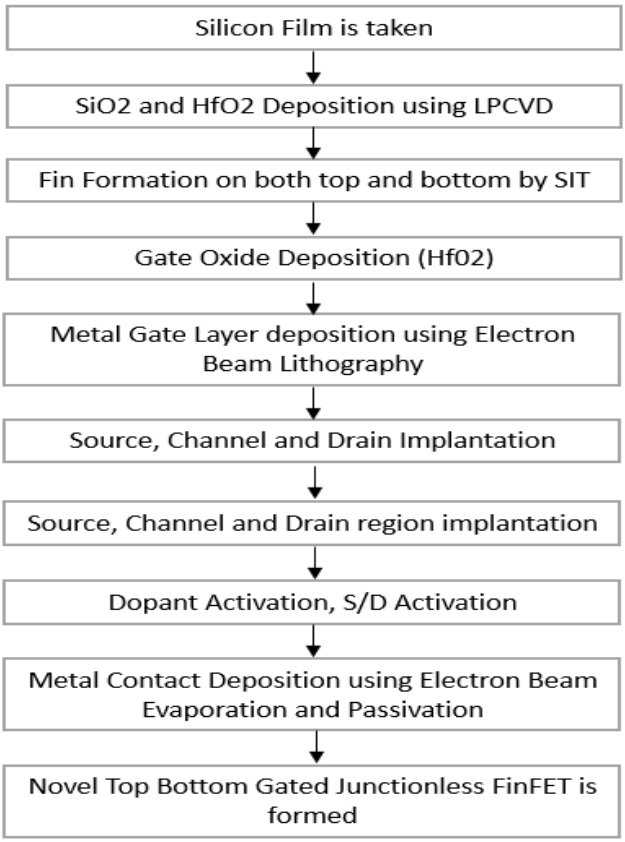


Figure 4. Fabrication Flowchart for Proposed TBG JL FinFET

Furthermore increasing the doping concentration shows improvement as far as Transfer Characteristics and Transconductance are concerned as depicted in the **figures 5 and 6**.

As we increase the doping concentration of the device, more charge carriers are available and hence the device exhibits better output current for same gate voltage as shown in **Figure 5** and similarly the Transconductance which varies directly with Output current increases as well as portrayed in **Figure 6**.

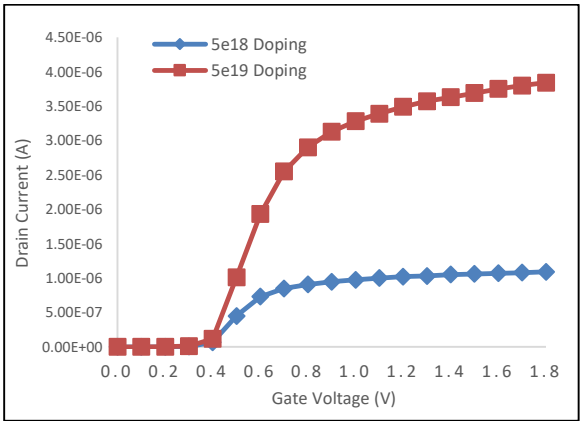


Figure 5. Variance of Transfer Characteristics with doping

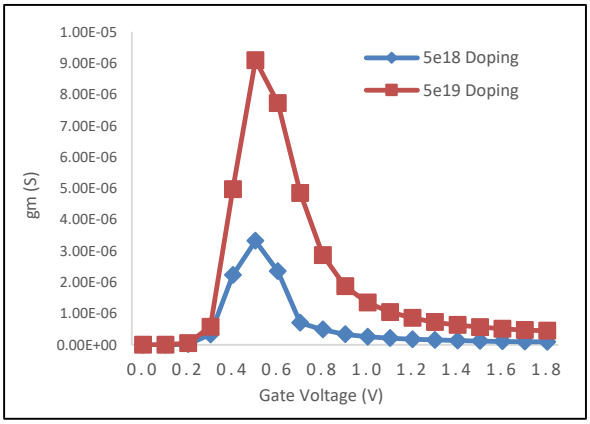


Figure 6. Variance of Transconductance with doping

The Top-Bottom Gated FinFET devices have fins on both sides of the oxide layer and thus have more device heights than the Traditional JL FinFET structure [2]. The Novel FinFET structure and TBG JL FinFET have device heights of 33nm due to presence of lower fin.

Introduction of High K HfO_2 gate dielectrics in FinFET devices prompts improvement in the analog efficiency and reduces power consumption [8]. The device parameters for the Novel structure are portrayed with the help of **Table 1**.

Table 1. Parameter Specification for Novel Stacked Oxide Top Bottom Gated FinFET

Parameter	Dimensions
Gate Length (L_g)	5nm
Upper Fin Height (H_{fu})	12nm
Lower Fin Height (H_{fl})	5nm
Fin Width (W_f)	5nm
Oxide Layer Thickness (O_l)	7nm HfO_2 + 3nm SiO_2
Substrate Width (W_s)	5nm
Source, Drain Length ($L_{s,d}$)	20nm

3. SIMULATION METHODOLOGY

The innovative gadgets have been simulated using the Silvaco Atlas TCAD tools [6]. A certain procedure is followed for developing FinFET devices using Silvaco tools. Above all the Construction Specification for the device is finished i.e. definition of the mesh grid structure for the FinFET lattice, definition of the doping concentration of regions and elucidating the regions itself. Secondly Definition of the Materials is done and physical models for the device are selected [6]. Finally the

Specification of Solution and Result Analysis is done for visualization of the results and analysis of various parameters [6].

4. RESULTS AND ANALYSIS

When a switch is on it has I_{ON} which is the ON state current and similarly it has I_{OFF} which is the off state current [9]. In an optimal circumstance, ON current is decided by the outside circuit while the OFF current should be zero [10].

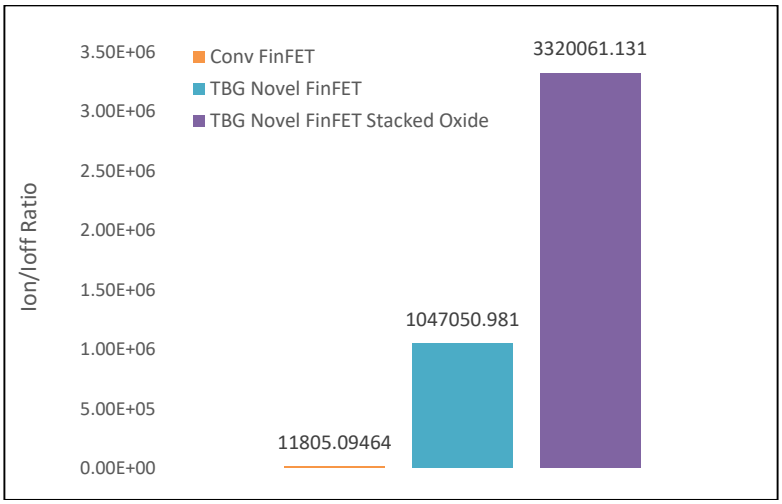


Figure 7. Switching Ratio for the three devices at 300K

For the common solid state technologies like MOSFET or FinFET that we use to emulate switches, the off current veers off from ideal zero and has a negligible value [11].

Figure 7 depicts the I_{ON}/I_{OFF} (Ion to Ioff) ratio of Traditional Junctionless FinFET, Novel TBG FinFET and the Innovative Stacked Oxide FinFET structure. It reveals that there is a great improvement in switching ratio for Innovative Stacked Oxide FinFET structure from that of Traditional JL FinFET and we observe that it improves by almost 280 times for Innovative Stacked Oxide FinFET structure when stacked against the Traditional JL FinFET and 3 times improvement as compared to TBGJL FinFET structure. Thus the proposed Stacked Oxide device has a higher on state current and better performance with reduced discharging ability due to the addition of high dielectric constant oxide. [10].

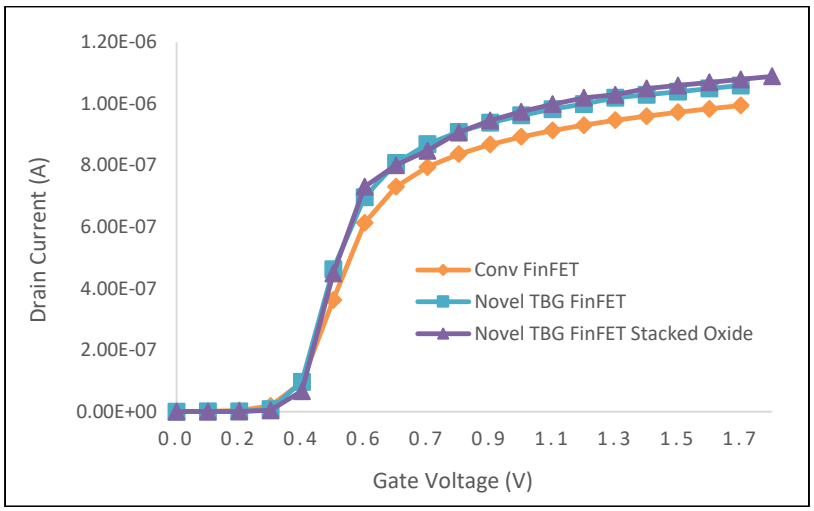


Figure 8. Transfer Characteristics Comparison for the devices at 300K

Figure 8 portrays the transfer characteristics of the Innovative Stacked Oxide FinFET structure and the same is shown for the other two devices for comparison [11]. It is clearly observed from the figure that the Innovative Stacked Oxide FinFET structure has improved yield current when stacked against the other two devices at the same given gate voltage at 300K. Thusly it can be inferred that the Innovative Stacked Oxide FinFET structure has incremented transfer characteristics [12].

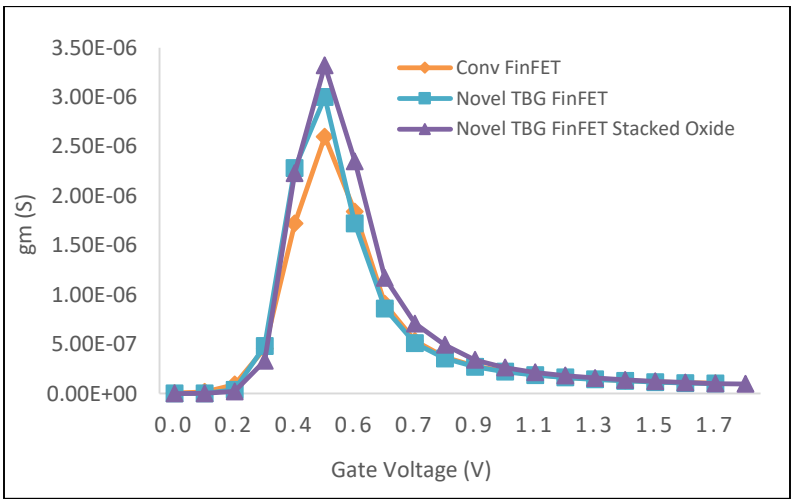


Figure 9. Transconductance Comparison for the devices at 300K

The transconductance as given by equation (1) (g_m- Input conductance) increments relatively with the linear gain in a gadget [3]. It connotes the variety in yield current with relating changes in input voltage V_{gs} (Gate Voltage) when V_{ds} (Drain Voltage) is kept unaltered [4].

The equation for Transconductance is:

$$g_m = \frac{\partial I_d}{\partial V_g} \tag{1}$$

Figure 9 gives the transconductance (g_m) of the three device. The Innovative Stacked Oxide FinFET structure gives incremented values of g_m for similar input voltage when stacked against the other two devices [13]. In this way the proposed gadget has better qualities with reference to transconductance. It moreover shows additionally created control of gate voltage over channel current [14].

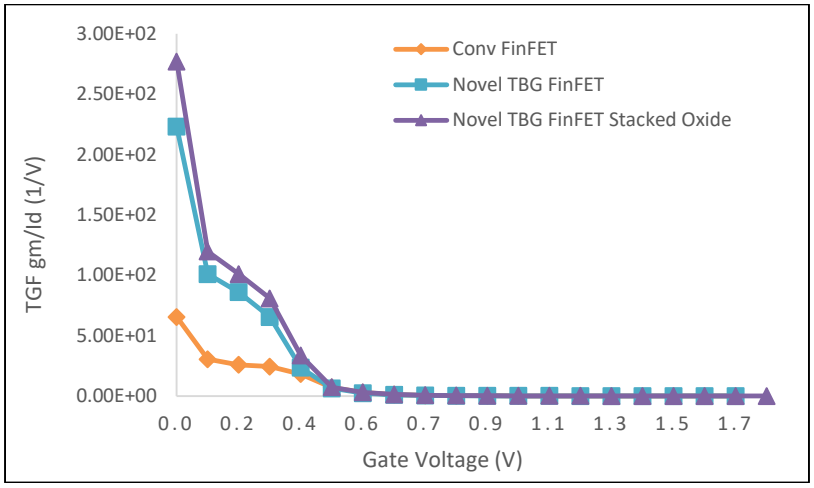


Figure 10. Transconductance Generation Factor Comparison for the devices at 300K

The ratio of g_m (Transconductance) to I_d (Drain current) is Transconductance Generation Factor (TGF) as depicted by equation (2) [15].

$$TGF = \frac{g_m}{I_d} \tag{2}$$

TGF demonstrates how successfully and deliberately the current is used to get a specific worth of transconductance [16]. The upside of high TGF is in the activity of circuits working at low supply voltage. And so we have superior performance [13]. **Figure 10** represents the variance in TGF as for changes in V_{gs} for the three devices. It tends to be seen that the Innovative Stacked Oxide FinFET structure has higher values of TGF when stacked against the other two devices and thus has more efficiency and performance values [16].

5. CONCLUSION

This work depicts the superiority of the Innovative Stacked Oxide TBG FinFET structure in terms of Analog Performance over the other two devices. 280 times improvement in switching ratio is observed for the Innovative Stacked Oxide FinFET structure when stacked against Traditional JL FinFET at 300K [1]. Further, enhanced analog performance is observed for the Innovative Stacked Oxide FinFET structure in terms of Transconductance and Transconductance Generation Factor when stacked against the other two devices [2]. Thus the device is an excellent choice for Analog Performance at Room Temperature (300K) and helps in diminishing Short Channel Effects a notch better.

References

[1] Gangwani L et al 2021 *J. Phys.: Conf.* Ser. **1921** 012100
[2] Gangwani L, Chakravarti R and Chaujar R 2021 *Devices for Integrated Circuit (DevIC) IEEE* 2021 pp. 1-5

[3] Kim S Y, Seo J, Yoon Y J, Yoo G, Kim Y J et al. 2014 *Journal of Semiconductor Technology and Science.* **14**. pp. 508-517
[4] Bughio A, Guerrieri S D, Bonani F and Ghione G 2017 *IEEE Electron Device Letters*, vol. **38**, no. 8, pp. 1004-07
[5] Kailasam M and Govindasamy M 2020 *International Journal Of Scientific & Technology*
[6] SILVACO I 2011 *ATLAS User's Manual* Santa Clara, CA, Ver., vol. **5**
[7] Boukortt N, Patanè S, and Crupi G 2019 *Silicon*, pp. 19–20, 2019
[8] Misra S, Mohan S B, Bara B, Kumar S S and Kumar S P *2018 IEEE Electron Devices Kolkata Conference (EDKCON)*, pp. 234-41
[9] Kaur N, Rattan M and Gill S S 2019 *Arab J Sci Eng* **44**, pp. 3101–16
[10] Singh M, Mishra S, Mohanty S S and Mishra G P 2016 *Advances in Natural Sciences: Nanoscience and Nanotechnology*, vol. **7**, no. 1, 2016
[11] Tripathi S, Gupta N, Tripathi P and Chaujar R 2019 *IEEE* 1-4
[12] Madan J and Chaujar R 2016 *Appl Phys A Material Science Process* **122**(11):973
[13] Kumar A, Gupta N and Chaujar R 2016 *Microsyst Technol* **22**, 2665–71
[14] Rai S, Sahu J, Dattatray W, Mishra R and Tiwari S 2012 *ISRN Electronics* 10.5402/2012/827452.
[15] Kranti A and Armstrong G 2007 *IEEE Trans. Electron Devices*, vol. **54**, no. 12, pp. 3308–16, 2007
[16] Kumar B, Kumar A and Chaujar R 2020 *IEEE VLSI Device, Circuit and System Conference* pp. 159–163

Advanced Production and Industrial Engineering
R.M. Singari and P.K. Kankar (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE220720

Analysing the Impact of Covid-19 on Automobile Sales in India Using ARIMA Modelling

Suyash SRIVASTAVA[a,1], Vasu SOOD[a,1], N. YUVRAJ[b]

[a]*Department of Production and Industrial Engineering, Delhi Technological University, Delhi, India*

[b]*Department of Mechanical Engineering, Delhi Technological University, Delhi, India*

Abstract. The Covid-19 pandemic has caused severe economic depression and has disrupted the supply chains of various industries. The automobile industry which contributes significantly to the Indian economy was gravely hit due to the lockdowns, semiconductor shortage and the uncertainty associated with the pandemic. This research paper analyses the effect of Covid-19 on the automobile sales in India using the time series modelling approach. The data recorded by SIAM from 2012 to 2019 was used to develop the Autoregressive Integrated Moving Average (ARIMA) model following the Box-Jenkins methodology. ARIMA model (2, 1, 3) was chosen as it had the lowest AIC and BIC criteria. This model was used to forecast the sales from 2020 to 2021 to give a picture of the expected automobile sales had the pandemic not occurred. The forecasted data from the model developed has then been compared with actual automobile sales data during the pandemic to gauge the level of impact Covid-19 had on the Indian automobile industry. The paper also explores the associated challenges that the automobile industry had to face due to the pandemic.

Keywords. Covid -19, Indian Automobile Industry, Time series, ARIMA, Forecasting

1. Introduction

The Covid-19 pandemic has been one of the most lethal epidemics in the history of mankind. Covid-19 is a contagious disease caused by the SARS-CoV-2 virus [1]. It was declared as a pandemic by the World Health Organisation on 11th March 2020 [2]. The pandemic has caused severe economic depression and disrupted the supply chains across numerous sectors. Many countries imposed lockdowns to help prevent the transmission of the virus. On 30th January, the first case of Covid-19 was reported in India [3]. Soon, a nationwide lockdown was enforced to stop the spread of the pandemic [4]. This pandemic exposed the vulnerability in the global supply chains and severely affected most of the industries across the world. The automotive industry can be considered as the barometer of the economy as it is effectively linked to several other key industries such as semiconductors, plastics, base metals and steel. In general, the automobile industry in India had witnessed significant growth over the past few

decades before the onset of the pandemic. However, since FY19 the industry had been witnessing a rapid fall in sales. This fall was fuelled by weak demand, regulatory issues, unfavourable exchange rates, weak investment scenario and the liquidity crisis [5]. The Covid-19 pandemic worsened the already precarious state of the Indian automobile industry.

2. Research Methodology

2.1. Data Collection

The data for this analysis has been collected from the Society of Indian Automobile Manufacturers (SIAM) from 2012 to 2021 [6]. The automobile sales data represents the number of cars sold in India during the given period. Further, the data has been sorted and the entries till the end of 2019 will be used to forecast the sales for 2020 and 2021. These forecast values will finally be compared to the actual sales data for our analysis. Figure 1 shows the graph of automobile sales across the years from 2012 to 2019.

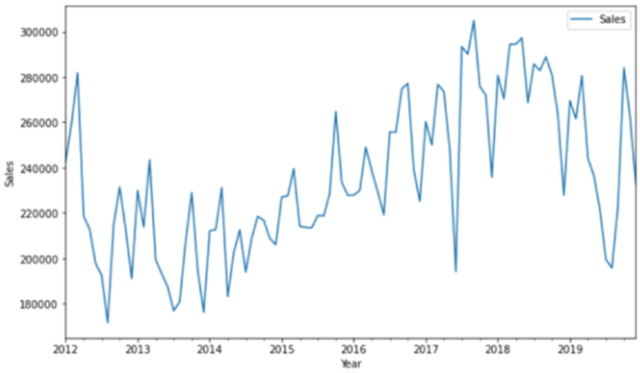


Figure 1. Automobile Sales in India from 2012 to 2019

Here, we will carry out the time series analysis using the Autoregressive Integrated Moving Average (ARIMA) model. It is a combination of Autoregressive (AR) Integrated (I) and Moving Average (MA) terms. In this model, 'AR' signifies that the data is regressed on past values, 'I' shows differencing of observations & 'MA' indicates that the forecast is linearly dependent on past error terms [7]. ARIMA model is denoted as ARIMA (p, d, q) where parameters (p, d, q) specify the type of ARIMA model used. The parameters for the model are as follows:

- p: the number of lag observations or lag order
- d: the number of times the data has been subtracted with past value or the degree of differencing.
- q: number of forecast errors or order of the moving average.

In order to build an ARIMA model, we will follow the Box-Jenkins methodology.

2.2. Model Identification

The second stage of the process is the identification of whether the variable to be forecast shows stationarity or not. The mean, variance, and covariance remain constant over the time in the stationary time series, whereas these values may change with time in non-stationary time series. ARIMA model can only be applied to stationary series. Therefore, we apply the Augmented Dickey-Fuller test to check for stationarity in the data [8]. This test was developed by statisticians David Dickey and Wayne Fuller in 1979. It has the following assumptions:

Null hypothesis (H_0) - series has a unit root. It is non-stationary.

Alternative Hypothesis (H_1) - series does not have a unit root. It is stationary.

For the p-value > 0.05 we fail to reject the null hypothesis (H_0) and for p-value <= 0.05 we accept the alternative hypothesis (H_1).

After applying the ADF test, the p-value is 0.62 which is greater than 0.05, therefore we fail to reject the null hypothesis (H_0) i.e., the series is non-stationary. Thus, we proceed with differencing the time series. After applying the first-order differencing (d=1) and then again using ADF test the p-value comes out to be 0.0025 which is less than 0.05. Hence, we accept the alternative hypothesis (H_1). Since this hypothesis states that the series is stationary, we do not need further differencing and can proceed with applying the ARIMA (p, d, q) model with d=1.

2.3. Estimation and Selection of Parameters

Further, we find the apt 'p' and 'q' values for our model. For this, we plot and examine the Autocorrelation (ACF) and Partial Autocorrelation (PACF) for our stationary time series. The ACF plot is a bar chart of coefficients of correlation between a time series and its lagged values. In other words, it explains how the present value is correlated with a past value in the given time series. The PACF demonstrates the partial correlation between the time series and the lags.

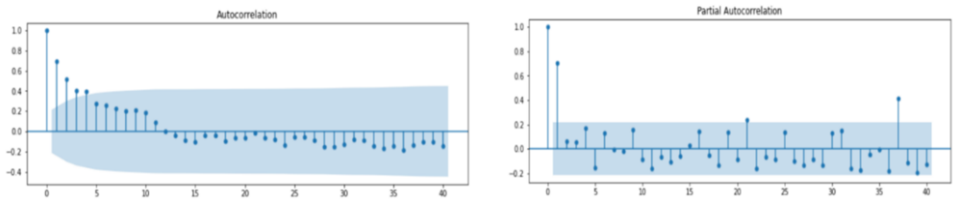


Figure 2. The plot of correlogram (ACF) and partial correlogram (PACF) for lags 1 to 40 of the first order differenced time series

From Figure 2, it can be seen that the Autocorrelation Function gradually decreases while the Partial Auto-Correlation Function has a sharp drop after 2 significant lags so we can conclude that the model is AR (2) i.e., the value of 'p' is 2.

By iterations, several ARIMA models with suitable 'q' values were fitted, the model with minimum Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) values was chosen [9]. The model ARIMA (2,1,3) has the lowest AIC and BIC criteria, hence this model will accurately predict the forecast values for our time series data.

2.4. Forecasting

Figure 3 gives a picture of the automobile sales in India in an ideal scenario had the Covid-19 pandemic not occurred. The blue lines depict the actual sales values till the end of 2019. The orange lines represent the forecast values for years 2020 and 2021 generated by the ARIMA (2, 1, 3) model.

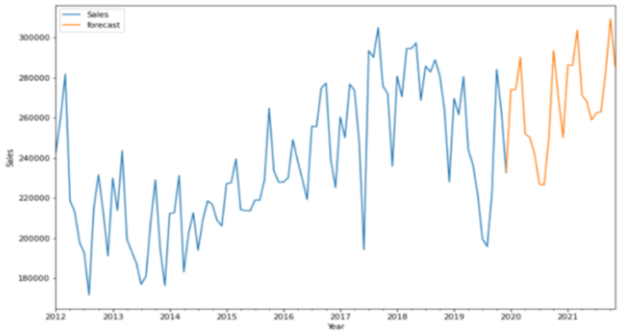


Figure 3. Forecast for the Automobile Sales for the 2020 and 2021 using ARIMA (2, 1, 3) model

3. Results and Discussions

Figure 4 shows the actual sales from January 2012 to December 2021 and the forecasted values for January 2020 to December 2021. Most notably, we can observe three significant negative deviations in the actual sales as compared to the forecasted values. These can be attributed to the first wave (March 2020 – June 2020), second wave (April 2021 – June 2021) and the subsequent semiconductor shortage in the latter part of 2021.

Figure 4. The graph shows actual sales in comparison to the forecasted sales by using ARIMA (2,1,3) model.

The outbreak of Covid-19 in India impacted the already struggling industry. According to SIAM, the drop in sales was 51% y-o-y for March and it was the worst ever decline recorded by the organization. It also estimated a revenue loss of over INR 23 billion per day for the industry due to the shut-down of plants and the piling up of unsold inventory in distribution channels [10]. The complete lockdowns also drove small-scale auto-component manufacturers to bankruptcy along with the 25,000 dealerships that run on very small margins. The pandemic had the greatest impact on the service sector in India which led to losses in employment and pay-cuts. These factors decreased the purchasing power of the Indian customer and thus led to further weakening of automobile sales.

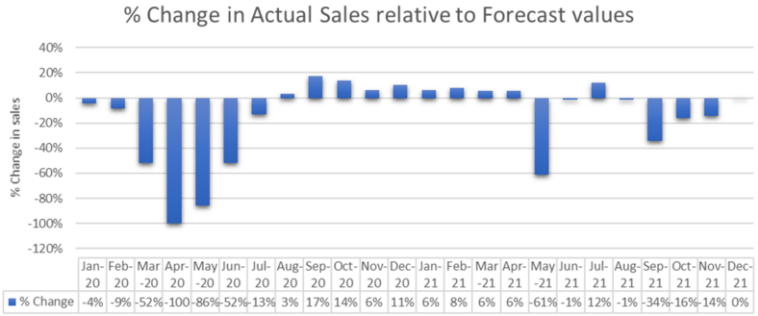


Figure 5. The graph shows percentage increase/decrease in actual sales relative to forecasted sales values for different months

From Figure 5, we can observe that there is an evident decrease in the actual sales since the start of 2020 as compared to our forecasted values through the ARIMA model. The initial decrease in January 2020 and February 2020 may be attributed to the disruption in the supply chain due to the spread of Covid-19 in China. In March 2020, the sales declined significantly (52%) and in April 2020 the sales were absolutely zero. The reason behind this observation is the implementation of lockdowns in mid-March and throughout April 2020. This trend continues throughout July 2020. From August 2020 to April 2021, however, the sales show a slight improvement (3%-17%). This boost is observed due to the gradual removal of restrictions by late 2020 and the festive season in India [11]. The 2nd wave of the pandemic emerged in Mid-April 2021 and it peaked by May 2021, therefore we can observe a steep 61% drop versus the forecast for the month. Going further, the sales rebounded in July but again dipped in comparison to our forecast. This decline was caused due to the growing semiconductor shortage across the globe which was boosted by the pandemic [12].

4. Conclusion

The Indian Automobile industry which is connected to various other important industries such as semiconductors, plastics, base metals and steel was witnessing a decline in sales due to weak demand, regulatory issues, unfavourable exchange rates, and the liquidity crisis in FY19. In March 2020, just as the automotive industry was beginning to recover, the pandemic struck and nationwide lockdowns were imposed which significantly affected the global supply chains. Moreover, the semiconductor shortage in the first quarter of 2021 also had a huge impact on the production capacity

of the automotive industry. This research paper utilises the Box Jenkins' ARIMA model to forecast the Automobile sales in India for the years 2020 to 2021. The model having the lowest AIC and BIC values was selected which was having the order (2, 1, 3). The forecasted data from the ARIMA model was compared with real data of sales of automobiles in India for the years 2020 to 2021 which was drastically affected due to the pandemic. The comparison showed that during the first wave there was a steep decline in actual sales relative to forecasted values for the months of March to June in the year 2020. The months from August 2020 to April 2021 showed a slight increase (3%-17%) in actual sales relative to forecasted values. However, there was a notable decline in the month of May 2021 when the second wave was at its peak. This trend improved slightly but continued till the end of the year. Despite all of these challenges, the sector is set to benefit from the global supply chain rebalancing, incentives from the government and new developmental opportunities which came to light after the tough challenges faced with the Covid-19 pandemic.

5. References

[1] Gan S, Xu J. Study on the Risk Identification Model and Countermeasures of Cruise under Epidemic Situation. Journal of Service Science and Management. 2020 Nov 19;13(6):824-40.

[2] Foster CE, Marquez L, Davis AL, Tocco E, Koy TH, Dunn J, Revell PA, Arrington AS, Campbell JR. A surge in pediatric coronavirus disease 2019 cases: the experience of Texas Children's Hospital from March to June 2020. Journal of the Pediatric Infectious Diseases Society. 2021 May;10(5):593-8.

[3] Andrews MA, Areekal B, Rajesh KR, Krishnan J, Suryakala R, Krishnan B, Muraly CP, Santhosh PV. First confirmed case of COVID-19 infection in India: A case report. The Indian journal of medical research. 2020 May;151(5):490.

[4] India will be under complete lockdown for 21 days. Economic Times [Internet]. 2020 Mar 25 [cited 2022 Mar 22]

[5] Taneja G. Managing the impact of COVID-19 on India's supply chains [Internet]. Assets.ey.com. 2020 [cited 2022 Mar 20]. Available from: https://assets.ey.com/content/dam/ey-sites/ey-com/en_in/topics/government-and-public-sector/2020/09/managing-the-impact-of-covid-19-on-india-supply-chains.pdf

[6] Society of Indian automobile manufactures [Internet]. Siam.in. [cited 2022 Mar 2]. Available from: https://www.siam.in/press-release.aspx?mpgid=48&pgidtrail=50

[7] Chatfield C. The analysis of time series: an introduction. Chapman and hall/CRC; 2003 Jul 29.

[8] Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association. 1979 Jun 1;74(366a):427-31.

[9] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018 May 8.

[10] Covid 19 impact on automotive business [Internet]. Deloitte India. 2020 [cited 2022 Feb 25]. Available from: https://www2.deloitte.com/in/en/pages/consumer-business/articles/covid19-response-for-automotive-companies.html

[11] Chibber B, Gupta N. The Indian automotive industry: From resilience to resurgence? [Internet]. Mckinsey.com. 2021 [cited 2022 Jan 22]. Available from: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-indian-automotive-industry-from-resilience-to-resurgence.

[12] Burkacky O, Lingemann S, Pototzky K. Coping with the auto-semiconductor shortage: Strategies for success [Internet]. Mckinsey.com. McKinsey & Company; 2021 [cited 2022 Feb 14]. Available from: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/coping-with-the-auto-semiconductor-shortage-strategies-for-success

Application to reduce distraction caused by phone calls while driving

Avinash Kumar
Department of Information Technology
Delhi Technological University
New Delhi, India
avi8654340@gmail.com

Mohak Gupta
Department of Information Technology
Delhi Technological University
New Delhi, India
mohakgupta_2k18it071@dtu.ac.in

Omkar Singh
Department of Information Technology
Delhi Technological University
New Delhi, India
omkarsingh_2k18it079@dtu.ac.in

Anamika Chauhan
Department of Information Technology
Delhi Technological University
New Delhi, India anamika@dtu.ac.in

2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA) | 978-1-6654-5834-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCSEA54677.2022.9936529

***Abstract*— The paper talks about the growing problem of accidents due to mobile phones. Mobile phone distraction is one of the major causes of road accidents. The objective of the paper is to reduce such a risky situation proposing an auto answer application while driving. It eliminates the physical engagement of the hand while talking on the phone. The driver may use the phone only if he has a Bluetooth device connected with the phone. The application has used android telephony manager and broadcast receiver for automatically answering phone calls. Details of this application forms the basis of this paper. The paper also discusses future scope of such apps and their usability.**

Keywords— road accident, Bluetooth, SMS, android, driving

I. INTRODUCTION

The impact of mobile phone usage while driving can be hazardous. If a driver is on a call and distracted, it creates a dangerous situation. In India especially, where there are no laws prohibiting the use of cell phones (without a hands-free device) on a four-wheeler and implementation is even less, the danger is aggravated. It causes active distraction and can lead to various kinds of accidents on the road. Distraction of any kind must be avoided while driving, but it is found that handheld usage of mobile phones while driving causes a lot of accidents. Drivers who engage in calling while driving are four times more likely to collide according to the World Health Organization.

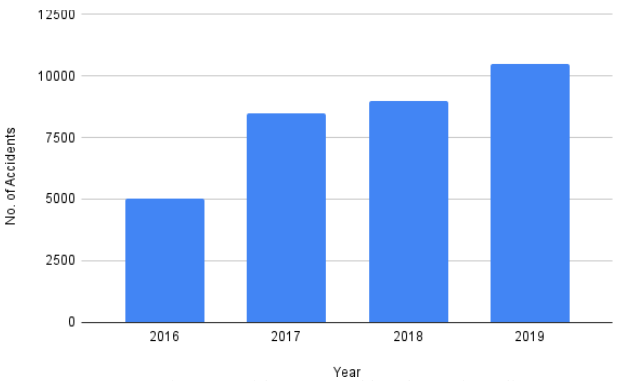


Fig. 1. Accidents caused by phones in India

Nearly 1.4 million people get killed because they are in wireless communication and their overbearing cell phones. Many accidents caused by mobile phones are due to the involvement of one hand in calling and only one hand in driving [1]. The consequences of road accidents can be as follows: -

- Physical-Some severe physical injuries and damage linked with an accident which may also lead to death. Physical destruction is obvious for any individual. These include head injuries, coma, memory loss, loss of limb, fractures etc.
- Emotional-Emotionally also, the victim is traumatized sometimes. Mental and emotional damages of an accident may include regret and remorse, emotional affliction, frustration, fear, humiliation, anxiety, embarrassment, shock, disturbed mental health, loss of appetite, weight fluctuations, deficit in energy, difficult mood, sexual dysfunction and sleep disturbances.
- Social effect- Social consequences of an accident include the loss of productivity of human resource, the cost of the legal system, nuisance and disturbance in nearby vehicles, the pain and suffering, of quality of life of the victim and their family. The loss of valuable human resources in terms of productivity and impact towards society constitutes a significant proportion of the total social impact.
- Economic impact- The accident may also cause a lot of damage not only to the driver but also to the public property. For example, it may cause damage to electric polls causing loss of electricity and many more.

A study proves the adverse effect of distraction caused by mobile phone calls on speed selection, vehicle spacing, and time headways. The study was based on experiments in a virtual environment. Overall, drivers drove slowly and with greater vehicle spacing and time headways when involved in phone conversations compared to default case- without phone conversations [2].

The handheld condition increases additional manual load on the driver. This coupled with the mental demand and exhaustion, leads to a greater distraction effect on drivers during the two phone conversations. Higher workload in handheld phone condition was also reported in [3] because physical engagement was greater for the handheld condition. No difference was noticed in mental demand between hands-free and handheld phone conversation. As a result, the car-following performance was adversely affected and

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:33:03 UTC from IEEE Xplore. Restrictions apply.

deteriorated. As a result of excessive mobile use, either for texting or calling, accidents on the roads have increased. The purpose of this paper is to build an android application which allows users to answer phone calls only if a hands-free Bluetooth device is connected. If the Bluetooth device is not connected the application does not pick up the call, thus keeping driver distraction due to mobile phone calls to a minimum. The application must work automatically without having to do a lot of things manually.

II. Literature Survey

There have been many prior researches on reducing cell phone distraction during driving automobiles or road accident prevention via mobile phones in general. In October 2020, the paper proposed an automated detection and alerting system for vehicular accidents. This system helped in detecting the accidents in a very short time typically in a few seconds and sending the basic information of the driver and vehicle to a first aid center in a message including the time and location of the accident. If there are no casualties or serious injuries, then the driver can terminate the message service. The message is sent with the aid of the GSM module, and the location of the accident is determined using the GPS module. The built-in accelerometer sensor in the mobile phone helps in precisely detecting the accident [4]. These accelerations, along with their location, are sent to a web service through the Internet. It is then delivered to other application users.

A research paper in 2019 [5] proposed a methodology and mobile application for driver monitoring, analysis, and recommendations based on unusual and unsafe driving behavior detection for accident prevention using a personal smartphone. The unusual behavior could be detected by the smartphone's cameras and built-in sensors (accelerometer, GPS, gyroscope and microphone). In this mobile application use of deep learning and computer vision was done which includes dangerous state classification, dangerous state detection and a reference model.

In 2019, a mobile application, based on GPS, was developed to track the user's current location. Initially, all blackspot and their speed limit were stored in the cloud database [6]. The mobile application extracted information about the next blackspot from the database and gave a voice notification 500 meters before the blackspot.

Another research for an accident prevention mobile application, eCall implementation was used [7]. eCall is an automatic accident detector which also requests help to the EMS, at the time of an accident. The application's working algorithm, Accident Detection Algorithm (ADA) uses built in sensors in mobiles like Accelerometer, gyroscope and the magnetometer. In addition to these sensors, the algorithm also takes into account vehicle sensor data primarily the airbag activation signal.

A survey of both Android and iOS application repositories for applications promoting road safety shows that only few applications have high ratings and positive user feedback. [8] Examples of such applications are -

1) Avertino- Advertino alerts the driver with audio and visual alerts. These alerts can be on a temporary basis or permanent.

2) iOnRoad Augmented Driving: This application detects when the driver does not keep minimum safety distance while in wheels and warns the driver in such a situation.

In both these applications, smartphones are used to detect accidents by use of accelerometer, gyroscope, and compass already present in a smartphone.

.

III. Methodology

The objective of the application is to reduce driver's distraction caused by mobile phone calls. The application mutes all notifications and puts the phone on silent. The application functions when a call arrives on the phone of the driver, it reads its phone state. To do this the application takes the permission of ReadPhoneState. Important classes used in this process are "PhoneStateListener" and "Telephony Manager".

- PhoneStateListener: - A listener class for monitoring changes in specific telephony states on the device, Methods are called when the state changes, as well as once on initial registration [9].
- PhoneCallBack: - The 3 states corresponding to idle, ringing and busy are TelephonyManager.CallStateIdle, TelephonyManager.CallStateRinging and Telephony-Manager.CallStateOffhook. A class is made named " PhoneCallBack " which inherits PhoneStateListener and displays phone state in the device. This method works only for devices with android version greater than Marshmallow.Abbreviations and Acronyms

Class/Method	Description
Check permission()	A function in main activity that checks all permissions
Phone State Listener()	A listener class for monitoring changes in specific telephony states on the device
TelephonyManager	Provides information about the telephony services
PhoneCallBack(_)	A class is made named "PhoneCallBack" which inherits PhoneStateListner and displays phone state in the device
CallPhoneManager()	A function in main activity that make instant of PhoneCallBack and listens phone state by LISTEN_CALL_STATE
onCallStateChange()	Function of PhoneCallBAck() class which changes the current state of the phone to string value

Fig. 2. Important Functions

A. Block Diagram

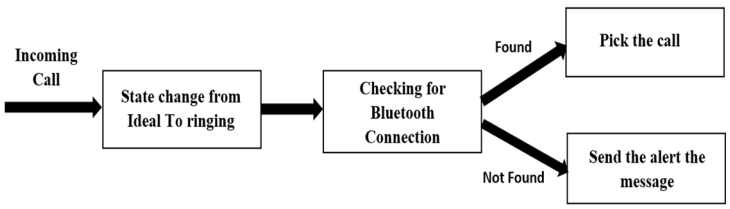


Fig. 3. Structure of App

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:33:03 UTC from IEEE Xplore. Restrictions apply.

The Overall Process Involve The following steps

- Step 1 -Incoming of call on phone
- Step 2-Changing the State of Phone
- Step 3- Method Invoke for the Bluetooth check
- Step 4- One of two cases will take place-
  - Case 1-If Bluetooth connect found to paired device, pick up the call
  - Case 2- Else extract the details of the caller and send the alert message

B. Bluetooth Check

The After that the application checks if any Bluetooth device is connected to the phone or not. Bluetooth is used because it avoids interference from other wireless devices. Also, it is present in maximum devices, has good short range, less danger to health, does not use the Internet and consumes low battery.

In order to work with bluetooth, the application needs the permission "android.permission.BLUETOOTH" and "Manifest. permission.BluetoothConnect" for the latest version. Checking for Bluetooth can be done by the Broadcast Receivers. Broadcast Receivers are responsible for the reception of system wide broadcast and take response aiming at the information that a broadcast transmits [10]. Broadcast receiver class is used to listen to Bluetooth broadcasts i.e., if a device just got connected or disconnected to Bluetooth. Then intent filters are used to listen to the ACTION_ACL_CONNECTED,ACTION_ACL_DISCONNECT_REQUESTED and ACTION_ACL_DISCONNECTED broadcasts.

Another way to know if Bluetooth is connected or not is by BluetoothAdapter. The BluetoothAdapter helps us to perform the fundamental Bluetooth tasks, such as initiate device discovery, query a list of bonded (paired) devices, instantiate a BluetoothDevice using a known MAC address. It then creates a BluetoothServerSocket which listens to the connection requests from other devices, and starts a scan for Bluetooth LE devices [11]. Further, to know if a paired device is connected to a phone, the method isConnected(Boolean) in BluetoothDevice system API can be used. If a Bluetooth device is connected, then the incoming call is diverted to the off-hook state but if Bluetooth is not connected, the application would send an SMS to the Caller via the send SMS method.

C. SMS sending

In The SMS sending feature is a more customizable feature than the two before [11]. The application provides provision to write a short message to numbers, which would be delivered to the numbers if the call is not answered or the driver is not using Bluetooth. Android.telephony module is used here. SmsManager module and the permission used is "android.permission.SENDSMS". The application sends SMS only if a message is saved prior for a given phone number.

D. Flowchart

- Call incoming: - Event of occurrence of a call.
- Permission: - Functions which check for permissions such as Bluetooth, SMS and telephony to access android default functions.
- CallPhoneManager: - Function inside main class used to listen to the call state [13].
- Phone State Listener: - Class/module to find call states [11],
- PhoneCallBack: - Class that extends PhoneStateListener.
- OnCallStateChange: - Function used to convert phone state to string and retrieve the incoming phone number.
- Bluetooth getState (): - Function to get state of bluetooth i.e., connected or free.; table

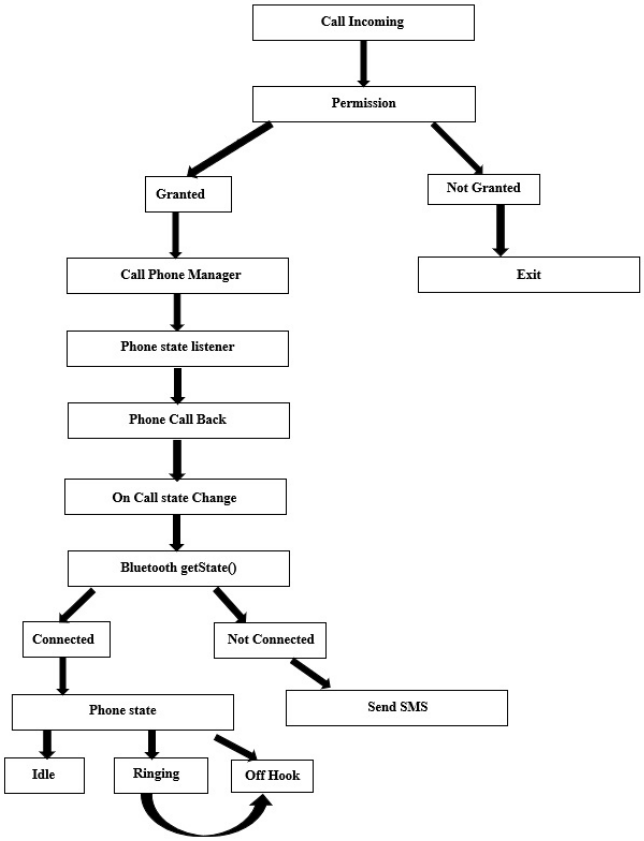


Fig. 4. Flowchart: description of program

E. Proposed Algorithm

Main objective of the proposed application is to automatically pick phone calls or send SMS if no bluetooth device is connected. The algorithm is based on Flowchart given in Fig 4. If all permissions are given by Android, then the application determines Call State.

Algorithm:

- All permissions such as "android.permission.Bluetooth", "android.permission.ReadPhoneState","Manifest.permission.BluetoothConnect" are checked by checkpermission() method.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:33:03 UTC from IEEE Xplore. Restrictions apply.

- The application exits if permissions are not granted, otherwise proceeds to the next step.
- CallPhoneManager () method is called which is responsible for listening to the status of the phone at that moment of time.
- An object of the PhoneCallBack class is instantiated by overriding the onCallStateChanged() method of the PhoneStateListener which finds the current call state.
- After finding the state, if the state is found to be ringing, then the BluetoothGetState () is called.
- If the bluetooth is found to be connected then the answer to the call method is called.
- If the state is found to be not connected then sendsms() method will be called by extracting the number of the caller and sending the alert message.

Fig. 5. Result 1: Showing call state change when a call arrives

```
D/In check permission: Call for callPhoneManager completed
D/in CAll State: Event triggered
D/in CAll State to string: converting to string
D/My debug: IDLE::
D/in CAll State to string: converting to string
D/My debug: IDLE::
D/in CAll State: Event triggered
D/in CAll State to string: converting to string
D/My debug: RINGING::
D/in CAll State to string: converting to string
D/My debug: RINGING::
D/in CAll State: Event triggered
D/in CAll State to string: converting to string
D/My debug: OFFHOOK::
D/in CAll State to string: converting to string
D/My debug: OFFHOOK::
D/in CAll State: Event triggered
D/in CAll State to string: converting to string
D/My debug: IDLE::
```

Fig. 6. Result-2: - Output on CLI showing possible state of phone and event triggers

IV. Conclusion

The problem of road accidents is immense. Using cell phones for calling or texting while driving creates a risky situation for everyone on the road. The paper proposes a software approach in the form of a mobile phone application which would automatically pick up incoming calls, provided a Bluetooth device is connected with the phone. In case of absence of a hands-free device, the application would deliver SMS text automatically. This android application can help to reduce the number of deaths that are occurring due to the accident because it limits the use of phones while driving. Reliability and maintainability of the system is strong. The system is available and functional all the time till the phone is dead. The reason behind its strongness is the continuous alert and call picking function of the application which lets it function all the time with accuracy.

V. Future Scope

Th The future vision is to enhance the system so that road accidents can be prevented before they actually happen. GPS functionality can also be added to the system which will improve alert systems by providing the real time location. The GPS module can be used to send the location of the mobile to the caller. GPS can be easily integrated in the application. Other services such as reporting the accident to the nearest emergency help like hospital, police and rescue team.

Another future addition can be an easy-to-use user interface for SMS and auto calling options. The application must consume minimal battery. In these ways, the application can be modified and enhanced

References

[1] Gorea, R. K. ”Financial impact of road traffic accidents on the society.”international Journal of Ethics, Trauma Victimology 2.01 (2016): 6-9.

[2] Mohammad Saifuzzaman, Md. Mazharul Haque, Zuduo Zheng, SimonWashington, Impact of mobile phone use on car-following behavior of young drivers, Accident Analysis Prevention, Volume 82, 2015, Pages 10-19, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2015.05.001.

[3] Roland Matthews, Stephen Legg, Samuel Charlton,The effect of cellphone type on drivers subjective workload during concurrent driving and conversing,Accident Analysis Prevention, Volume 35, Issue 4,2003,Pages 451-457,ISSN 0001-4575.

[4] Parveen, Nazia, Ashif Ali, and Aleem Ali. ”IOT Based Automatic Vehicle Accident Alert System.” 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA). IEEE, 2020.

[5] Kashevnik, Alexey, Igor Lashkov, and Andrei Gurtov. ”Methodology and mobile application for driver behavior analysis and accident prevention.” IEEE transactions on intelligent transportation systems 21.6 (2019):2427-2436.

[6] kumar Thangavel, Rajesh, et al. ”Blackspot Alert and Accident Prevention System.” 2019 10th International Conference on Computing,Communication and Networking Technologies (ICCCNT). IEEE, 2019.

[7] Fernandes, Bruno, et al. ”Mobile application for automatic accident detection and multimodal alert.” 2015 IEEE 81st vehicular technology conference (VTC spring). IEEE, 2015.

[8] Khalil, Usman, Tariq Javid, and Adnan Nasir. ”Automatic road accident detection techniques: A brief survey.” 2017 International Symposium on Wireless Systems and Networks (ISWSN). IEEE, 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:33:03 UTC from IEEE Xplore. Restrictions apply.

[9] Singh, A. J., and Akshay Bhardwaj. ”Android internals and telephony.”Int. J. Emerg. Technol. Adv. Eng 4 (2014): 51-59.

[10] Liu, Jianye, and Jiankun Yu. ”Research on development of android applications.” 2011 4th International Conference on Intelligent Networks and Intelligent Systems. IEEE, 2011.

[11] Yan, Ming, and Hao Shi. ”Smart living using Bluetooth-based Android smartphone.” International journal of wireless mobile networks 5.1 (2013): 65.

[12] Arzt, Steven, Siegfried Rasthofer, and Eric Bodden. ”Instrumenting android and java applications as easy as abc.” International Conference on Runtime Verification. Springer, Berlin, Heidelberg, 2013.

[13] Komatineni, Satya, and Dave MacLean. ”Using the Telephony APIs.” Expert Android. Apress, Berkeley, CA, 2013. 135-152

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:33:03 UTC from IEEE Xplore. Restrictions apply.

Biomass Conversion and Biorefinery
https://doi.org/10.1007/s13399-022-03433-w

ORIGINAL ARTICLE

Biosynthesis, characterization, and evaluation of antibacterial and photocatalytic dye degradation activities of silver nanoparticles biosynthesized by *Chlorella sorokiniana*

Lakhan Kumar[1] · **Navneeta Bharadvaja**[1]

Received: 9 July 2022 / Revised: 5 October 2022 / Accepted: 15 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Nanoparticle synthesis by physiochemical methods is highly expensive and uses hazardous chemicals, and leaves behind toxic by-products. Thus, green or biological synthesis has emerged as an alternative for the same. The present study reports the synthesis of silver nanoparticles (SNPs) using *Chlorella sorokiniana* microalgae extract. The influence of certain parameters (pH, cell disruption methods, temperature, extract to precursor metal salt solution ratio, salt strength, biomass concentration, and incubation conditions) on SNP synthesis was studied. A strong surface plasmon response band occurred with a maximum between 400 and 440 nm at different physiochemical parameters tested. Synthesized SNPs were spherical and were 79 nm in size with a zeta potential of − 23.7 mV at optimized conditions. XRD analysis confirmed the crystalline structure of nanoparticles. The synthesized SNPs were found to be an effective bactericide, with zones of inhibition up to 20 ± 1 and 15 ± 1 mm for *E. coli* and *S. aureus*, respectively. SNPs were observed to catalyze the degradation of crystal violet by 97.04%, followed by methylene blue (95.75%) and eosin Y (94.9%), and the least efficiency was observed for rhodamine B (56.05%), at the end of 18 h under direct exposure to sunlight. Thus, the study concludes that SNPs synthesized using *Chlorella sorokiniana* are potent antibacterial agents and catalysts for photocatalytic dye degradation.

Keywords Green synthesis · Silver nanoparticles · *Chlorella* · Antibacterial activity · Photocatalytic dye degradation · Wastewater treatment

1 Introduction

Due to increased industrialization and urbanization by anthropogenic activities, several pollutants have been gradually detected and quantified in and around the human being and their environment. Being xenobiotic and persistent, dyes are one of the prominent pollutants having proven harmful effects on public and environmental health [1]. Out of the 60,000 tons of dye being discharged into waterbodies worldwide each year, almost 80% comprise the azo-dyes. The printing and dyeing industry run-offs are complex, difficult to bio-degrade, and highly concentrated in color containing a variety of organics. Such wastes are considered to be mutagenic, carcinogenic, and teratogenic. The principal organic components of dye industry wastewater are benzene, anthraquinone, and so on. The chromatic pollution of dyes is attributed to the chromophore-molecular group in its structure. Hence, individual treatments of these wastewaters are not suggested. The wastewater containing such a high "color index" requires dedicated yet comprehensive treatments to match the stringent standards of discharge [2]. Health hazards of dyes on flora and fauna health, the microbial community, and water bodies are a rising concern [3].

Likewise, antibiotic-resistant bacterial infections are creating substantial distress to human health and the economy. The indiscriminate and overuse of antibiotics and other possible genetic causes, including mutations, genetic transfer, and acquisition leading to resistant genes, resulted in the development of microbial populations adapting to tolerate, exist, and propagate in the presence of a range of antibiotic agents [4, 5]. The antimicrobial resistance led to the increase

✉ Navneeta Bharadvaja
navneetab@dtu.co.in

Lakhan Kumar
adarsh.lakhan@gmail.com

[1] Plant Biotechnology Laboratory, Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India

Published online: 27 October 2022

of death case numbers which is a growing concern worldwide today and it has facilitated the exploration of novel antimicrobial agents [6].

In recent times, nano-dimensional particles are under investigation for a range of applications, including environmental remediation, antimicrobial agents, drug delivery, biosensors, and cosmetics [7]. Nanoparticles, due to their nanoscale dimension, increased surface area, shape, and inherent properties, have been proven as an alternative to conventional antibiotics [8, 9]. Metal nanoparticles, including silver, zinc, copper, titanium, gold, iron, magnesium, and gold, metal oxide nanoparticles, and several other organic and inorganic nanoscale agents have been reported to have the ability to combat different pathogens [4, 10–16]. Silver nanoparticles, in particular, are tested against a range of bacterial infections due to their potent antimicrobial activities via cell wall membrane defragmentation, apoptosis, alteration in the enzymatic pathway and cellular pathways, and synergy of several known and unknown mechanisms. Therefore, economical and ecofriendly routes of silver nanoparticle synthesis for diverse applications have received continued interest [17].

The synthesis of nanomaterials via conventional physicochemical methods is not economic and ecofriendly. In recent times, green synthesis or biological synthesis route of nanomaterials has gained momentum. Plants and microbes, including bacteria, algae, and fungi, have been explored for the synthesis of nanomaterials. The bioactive compounds present in plant materials and microbial biomass extracts act as reducing, capping, and stabilizing agents during synthesis of nanomaterials. In this direction, algae-based silver nanoparticle synthesis is a novel and sustainable approach [18]. The biogenic synthesis is considered to occur in phases, namely the activation, growth, and termination phases. The metal ion from the precursor salt gets reduced, leading to a colorimetric change in the activation phase. In the growth phase, the nucleation products fuse, forming nanoparticles of diverse morphologies. The nanoparticles attain their final morphologies in the termination phase [19, 20].

In the present study, *Chlorella sorokiniana* extract has been used to synthesize silver nanoparticles (SNPs). *Chlorella sorokiniana* belongs to *Chlorella*, an important genus of the Chlorophyta phylum, which can efficiently grow in a wide range of temperatures and light intensities and has been extensively used in application for wastewater treatment [21, 22]. It is a fast-growing green microalga. It can grow autotrophically, heterotrophically, and mixotrophically depending upon the water used, the nutrients provided, and the surrounding conditions. It is an essential source of biomass as well as biological components, including proteins, carbohydrates, lipids, pigments, and vitamins [23]. The cellular extract of the microalgae contains different enzymes and metabolites like carotenoids, polysaccharides, amino acids, polyphenols, aromatic compounds, and minerals, which function as reducing, capping, and stabilizing agents during nanoparticle synthesis [18]. *Chlorella sorokiniana*–mediated synthesized nanoparticles were characterized and tested for their photocatalytic degradation potential against four different dyes, namely methylene blue, crystal violet, eosin Y, and rhodamine B, and bactericidal effect against two bacteria, *Escherichia coli* and *Staphylococcus aureus*. Also, the influence of certain physicochemical parameters, including pH, cell disruption methods, temperature, extract to precursor metal salt solution ratio, salt strength, biomass concentration, and incubation conditions on SNP synthesis, has been studied.

2 Materials and methods

2.1 Microalgae extract preparation and silver nanoparticle synthesis

Microalgae culture was cultivated in BG-11 medium under 4000 Lux light intensity at 27 ± 1 °C and 16:08 h photoperiod in an algal growth chamber. Biomass harvested at the stationary phase was mixed into Milli-Q water and boiled for 20 min at 80 °C to prepare the microalgae extract. This extract contains a range of bioactive compounds, including protein, carbohydrates, fatty acids, and pigments like chlorophylls, carotenoids, phycobilin, and antioxidants like terpenoids or polyphenol. For biosynthesis of SNPs, a certain volume of microalgae extract was mixed with freshly prepared silver nitrate solution and kept at room temperature for the reaction to progress. SNP formation was indicated by a gradual change in the color, from light green to dark brown of the aqueous solution. This preparation was kept under observation for a period of 3 days, and nanoparticle synthesis was monitored using a UV–Vis spectrophotometer.

Silver nitrate ($AgNO_3$) was purchased from SRL, India. All other reagents were of analytical grade and used as received. All the solutions were freshly prepared with double distilled water for the experimental procedure and were kept in the dark to avoid any photochemical reaction.

2.2 Optimization of parameters influencing green synthesis of silver nanoparticles

Microalgae-mediated synthesis of SNPs depends on several conditions, including the strength of metal precursors used, ratio of salt to microalgae extract, pH of the microalgae extract, temperature, incubation condition, and period. Here in this study, we have optimized the abovementioned parameters to get the optimum synthesis of the silver nanoparticles. One condition was varied at a time, and the rest of the conditions were kept constant. This method is called "One Variable At a Time" or the OVAT method. Different

ratios pertaining to extract to salt ratio (20:1, 10:1, 1:1, and 1:10), metal precursor salt strength (1 mM, 2 mM, 5 mM, 10 mM, and 20 mM), pH (6–10), biomass concentration (0.5 g/100 mL, 1 g/100 mL, 2 g/100 mL, and 4 g/100 mL), cell disruption techniques (autoclaving, boiling, and sonication), temperature (room temperature (30 °C and 60 °C)), incubation conditions (light and dark), and incubation period (up to 72 h) were optimized for efficient green synthesis of SNPs using microalgae extract. Three cell disruption methods, namely boiling for 20 min at 80 °C, autoclaving at 121 °C for 20 min, and sonication (0.5 cycle, 50% amplitude) for 20 min, were tested for the selection of suitable cell disruption methods.

2.3 Characterization of silver nanoparticles synthesized using algal extract

The formation of SNPs is time-dependent and confirmed by the development of brownish color solution as the reaction progresses. Initial characterization was performed using UV–Vis spectrophotometry (Lambda 365, Perkin Elmer) by measuring the absorbance (200–700 nm) of the samples taken at regular intervals. The morphology of the synthesized SNPs was obtained by scanning electron microscopy FESEM (ZEISS microscope). Dynamic light scattering (DLS) technique (Zetasizer nano series ZS-Malvern PANalytical) and X-ray diffraction analysis (Bruker D8 Advance) were performed to find the average size and stability of nanoparticles and the crystalline nature of the nanoparticles formed respectively. For SEM analysis, an aliquot of SNP suspension was placed on a quartz thin film and allowed to settle, followed by obtaining the SEM images for morphological analysis.

2.4 Antibacterial activity of silver nanoparticles

To determine the antibacterial activity, the synthesized SNPs were tested against *Escherichia coli* and *Staphylococcus aureus* by well diffusion method. Overnight-grown bacterial cultures were spread on the nutrient agar plate. Wells were created using a well borer. At a concentration of 40 µg/mL, a colloidal solution of silver nanoparticles and a certain volume of synthesized SNP solution (10 µL, 20 µL, and 25 µL) were added to the wells. The algal extract and silver nitrate ($AgNO_3$) solutions were also placed in the well as control. Then, the plates were kept in incubation overnight. The antibacterial activity was determined by observing and measuring the presence or absence of a zone of inhibition (millimeters or mm) around the well compared to the control. The diameter of such zones of inhibition was measured using a meter ruler, and the mean value for each microorganism was recorded and expressed in millimeters. The experiment was performed in triplicates, and the zone of inhibition was measured after overnight incubation.

2.5 Photocatalytic degradation of select dyes

The photocatalytic degradation of four different dyes eosin yellow (EY), methylene blue (MB), crystal violet (CV), and rhodamine B (RhB), using SNPs as a catalyst, synthesized in this study, was carried out under natural sunlight. A 5 ppm dye solution was prepared by mixing the required amount of dye in water. A total of 0.5 mL of biogenic nanoparticle solution (40 µg/mL) was added into the dye solutions (5 ppm, 20 mL) and kept under sunlight for degradation to proceed. The degradation of dyes was monitored using UV–Vis spectrophotometry by taking samples at regular intervals. The percentage of decolorization is calculated by the following formula:

$$\% \text{ Degradation} = \left(C_0 - C_t / C_0\right) \times 100$$

where C_0 = absorbance of dye at time 0 min, and C_t = absorbance of the dye after the photocatalytic degradation.

3 Results and discussion

This study aimed to determine the antibacterial activity and photocatalytic dye degradation potential of silver nanoparticles (SNPs) synthesized using *Chlorella sorokiniana* extract. The green synthesized SNPs used in this study were characterized by UV–Vis spectroscopy, scanning electron microscopy, X-ray diffraction pattern, and a zeta potential analyzer. Certain parameters affecting the SNP synthesis have been optimized and presented in the following section.

3.1 Optimization of SNP synthesis parameters using *Chlorella sorokiniana* biomass

In the current study, different parameters for the synthesis of SNPs were considered and optimized. These included pH conditions, extract to salt ratios, the concentration of precursor salt, cell disruption methods, biomass concentration, incubation conditions, and method of extract preparation from *Chlorella sorokiniana*. Screening of optimum conditions for synthesis was done by keeping only one condition dynamic at a time, and the rest of the conditions were kept constant throughout the study (OVAT method). The final synthesis was performed at the combination of all the optimum parameters. The UV–Vis spectrum representing stability and synthesis of SNP synthesis by *C. sorokiniana* at different conditions is illustrated in Fig. 1. It was observed that the synthesis of SNPs took place at a slower pace in the initial hours; the synthesis rate increased and attained a

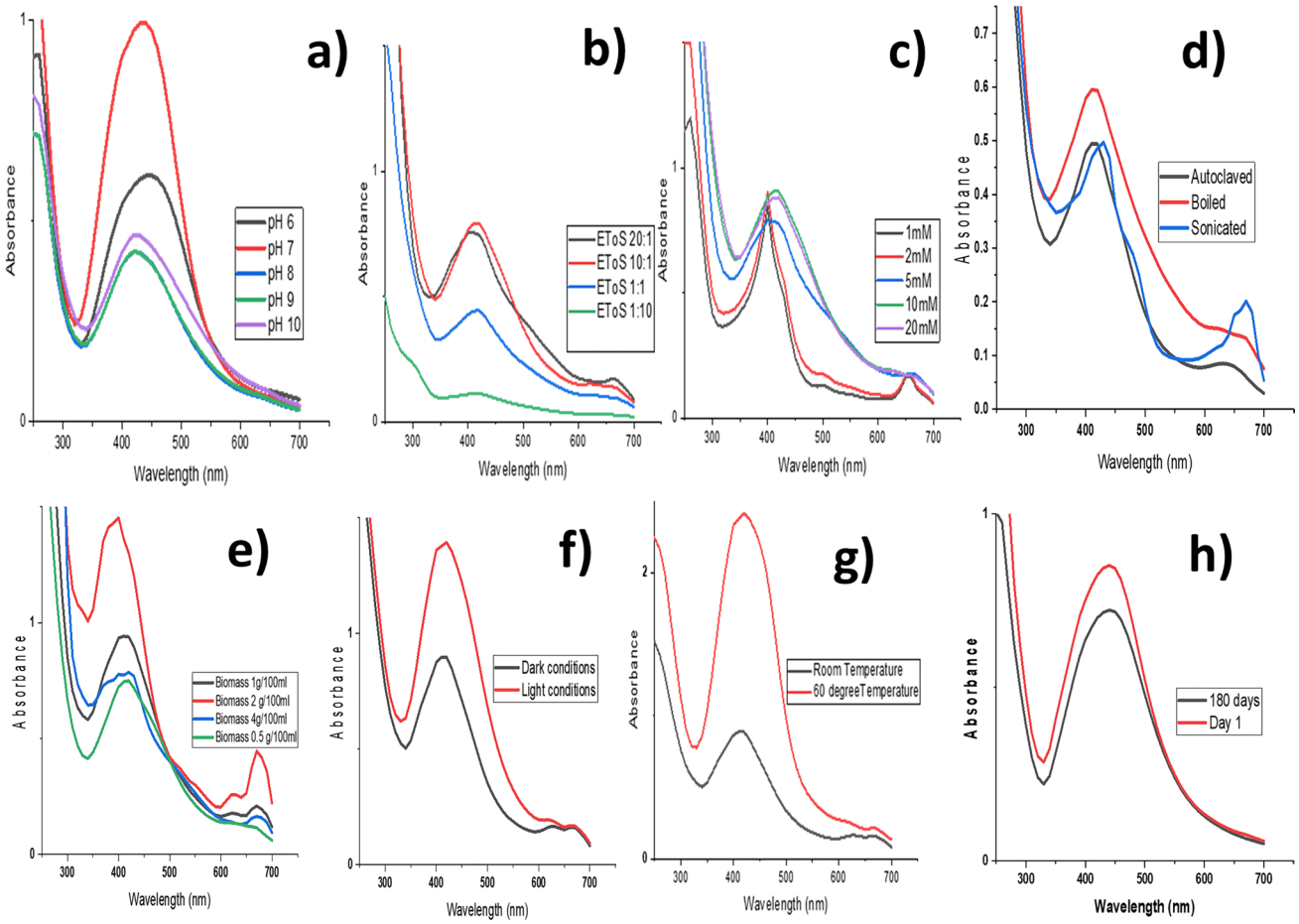


Fig. 1 UV–Vis spectrum representing stability and synthesis of SNP synthesis by *C. sorokiniana* at different conditions: **a** pH range from 6 to 10; **b** different extract to precursor metal salt ratios; **c** different strengths of silver nitrate solution; **d** different methods of cell disruption; **e** different concentrations of algal biomass; **f** incubation in light and dark; **g** incubation temperature; and **h** synthesis of SNPs at optimized conditions and stability of SNPs at different periods of incubation

maximum at 24 h. The brownish color of the solution was found to be increasing upon increase in the incubation time up to 24 h. This increase in color intensity with time indicates SNP synthesis. Similar results were reported for the SNP synthesis using the marine algae *Caulerpa racemosa* [24]. The environmental factors, including pH, reaction time and temperature, and precursor metal salt concentration, have a significant role to play during synthesis and morphology of silver nanoparticles. The spectrophotometric analysis recorded a strong surface plasmon response band between 400 and 440 nm at different physiochemical parameters tested in this study, as shown in Fig. 1. Literature has also confirmed that the occurrence of the peak in the range of 410–450 nm in the formulation indicates the presence of SNPs in the solution [25–29].

pH has a profound effect on the morphology and structural build of the synthesized nanoparticles [30]. The considered range of pH for the optimization of synthesis of biogenic NPs was chosen from 6 to 10. SNPs formed at this pH range were stable while those formed at pH (5, 11, 12) were highly unstable. Figure 1a shows the spectrophotometric spectra of synthesized SNPs at different pH values. The most stable SNPs were synthesized, with a strong SPR, at a pH of 7 among all other tested pH values. This indicates the higher activity of reducing, capping, and stabilizing agents present in the extracts of *Chlorella sorokiniana* [31]. Lower pH favors large-sized SNP formation while high pH facilitates a comparatively faster formation of small-sized and highly dispersed SNPs in the synthesis medium [32].

At higher concentrations of metal precursor salts, the formation of SNPs increases. Along with this, the aggregation of synthesized nanoparticles also increases. In order to get nanoparticles stable for a longer duration, optimization of precursor metal salt strength is essential [33]. In this study, the ratio of algal extract to precursor salt solution was varied at 20:1, 10:1, 1:1, and 1:10. The best synthesis occurred when the algal extract and precursor metal salt solution were mixed in the ratio of 10:1. The SNPs synthesized using this

ratio traced a broader peak around 430–440 nm as evident in Fig. 1b. This can be attributed to a rapid rate of reduction of Ag + ions with the optimal presence of secondary algal metabolites having high electron density [30, 34]. Ahmed and Mustafa reported that the particle size decreases with the increasing amount of extract [32]. Different concentrations of precursor metal salt solutions ($AgNO_3$) (1 mM, 2 mM, 5 mM, 10 mM, and 20 mM) were prepared. It was confirmed using UV–Vis spectroscopy that a 10 mM concentration of $AgNO_3$ was optimum for the synthesis of SNPs using *Chlorella sorokiniana* extract. A prominent surface plasmon response peak was observed at 10 mM silver nitrate solution, as illustrated in Fig. 1c. A recent study reported similar findings for synthesizing SNPs using different molar strengths of metal precursor salts [31].

For the extraction of algal bioactive compounds, different cell disruption techniques such as sonication, boiling, and autoclave were performed. The prominent SPR was observed for SNPs prepared using algal extract obtained after boiling algal biomass as compared to the algal extract obtained after autoclaving and sonicating the biomass, as indicated in Fig. 1d. Autoclaving of the algal biomass leads to degradation of cellular metabolites due to very high operational temperature and pressure, whereas in the case of sonication, complete algal cell wall disruption could not be achieved. Thus, the concentration of metabolites in the extract was low, which could not enable a rapid or stable synthesis of SNPs.

A 100-mL volume with varying algal biomass concentrations (0.5 g/100 mL, 1 g/100 mL, 2 g/100 mL, and 4 g/100 m) was tested for the synthesis of SNPs. The UV–Vis spectrophotometric analysis was recorded and is presented in Fig. 1e. The extract prepared using 2 g biomass of *Chlorella sorokiniana* per 100 mL distilled water facilitated optimum concentration of reducing, capping, and stabilizing phyco-compounds. These agents contributed in efficient bio-reduction of Ag + ions present in the reaction mixture leading to the synthesis of SNPs.

The mixture of algal extract and precursor metal salt was incubated in two different conditions: light and dark. The results from visual and UV–Vis analysis indicated that the SNP synthesis in light conditions is more favorable than the rate of synthesis in dark conditions. Figure 1f presents the UV–Vis spectrophotometric analysis of SNP formation under both light and dark conditions. This observation can be attributed to the activation of various compounds under light conditions, which aid in the synthesis process. A similar observation was reported by a recent study on silver nanoparticle synthesis using *Kalanchoe* leaf extract [33].

The favorable incubation temperature for the synthesis of SNPs was found to be 30 °C. A recent study on *Chlorella vulgaris*–mediated green synthesis of silver nanoparticles reported a similar observation which corroborates our results [30]. The algal extract and silver salt mixture was incubated at 30 °C (RT) and 60 °C. It was observed that the rate of synthesis was much higher when incubated at 60 °C as compared to the samples incubated at 30 °C, as illustrated in Fig. 1g. High temperature accelerated the synthesis of the SNPs, but the particles were highly unstable. It was further noted that the SNPs synthesized at higher temperatures were not stable for a longer time than those synthesized at lower temperatures. The optimum incubation temperature for precursor metal salt solution and algal extract for the synthesis of SNPs was observed at 30 °C (RT). Furthermore, any increase or decrease in the incubation temperature led to the broadening of peak values, indicating an increase in the size of SNPs as observed in [35] and [36].

SNPs were synthesized at optimum conditions, and the stability of SNPs was checked using UV–Vis at 24 h, 48 h, 72 h, 96 h, 168 h, and after 180 days of synthesis and it was observed that the SNPs prepared at optimum conditions are highly shelf stable as illustrated in Fig. 1h. The optimized condition for *Chlorella sorokiniana*–mediated synthesis of SNPs was extract to salt ratio 10:1 with 10 mM strength of precursor metal salt $AgNO_3$ at pH 7 under light conditions at room temperature (30 °C) as shown in Supplementary Fig. 1. Algal extract prepared by boiling algal biomass (2 g/100 mL) at 100 °C for 20 min gave better green synthesis of biogenic SNPs.

3.2 Characterization of SNPs synthesized from *Chlorella sorokiniana*

Silver nanoparticles synthesized at optimum conditions using *Chlorella sorokiniana* extract were characterized primarily by a UV–Vis spectrophotometer. A change in color from light green to brown was observed with time as the reaction progressed. The UV–Vis absorption spectrum of the synthesized SNPs at optimized conditions showed a broad peak at 441 nm, a characteristic band for SNPs. Furthermore, the synthesized particles were characterized using SEM, XRD, and Zetasizer, as illustrated in Fig. 2 (i, ii & iii). The XRD patterns for silver nanoparticles indicated the crystalline nature of the synthesized nanoparticles. The XRD showed four prominent diffraction peaks observed at 2θ values of 38.23°, 44.37°, 64.57°, and 77.51°, which correspond to the (111), (200), (220), and (311) crystallographic planes of face-centered cubic silver (Ag) crystals respectively. It was similar to the planes of the cubic structure of SNPs presented in JCPDS, File No. 04–0783) [37, 38].

The zeta potential of synthesized SNPs in aqueous solution was − 23.7 mV, indicating higher stability of SNPs in colloidal solution. It is known that zeta potential indicates the stability of SNPs containing aqueous suspension and particle morphology as well [39]. This parameter shows the degree of repulsion between the charged particles in

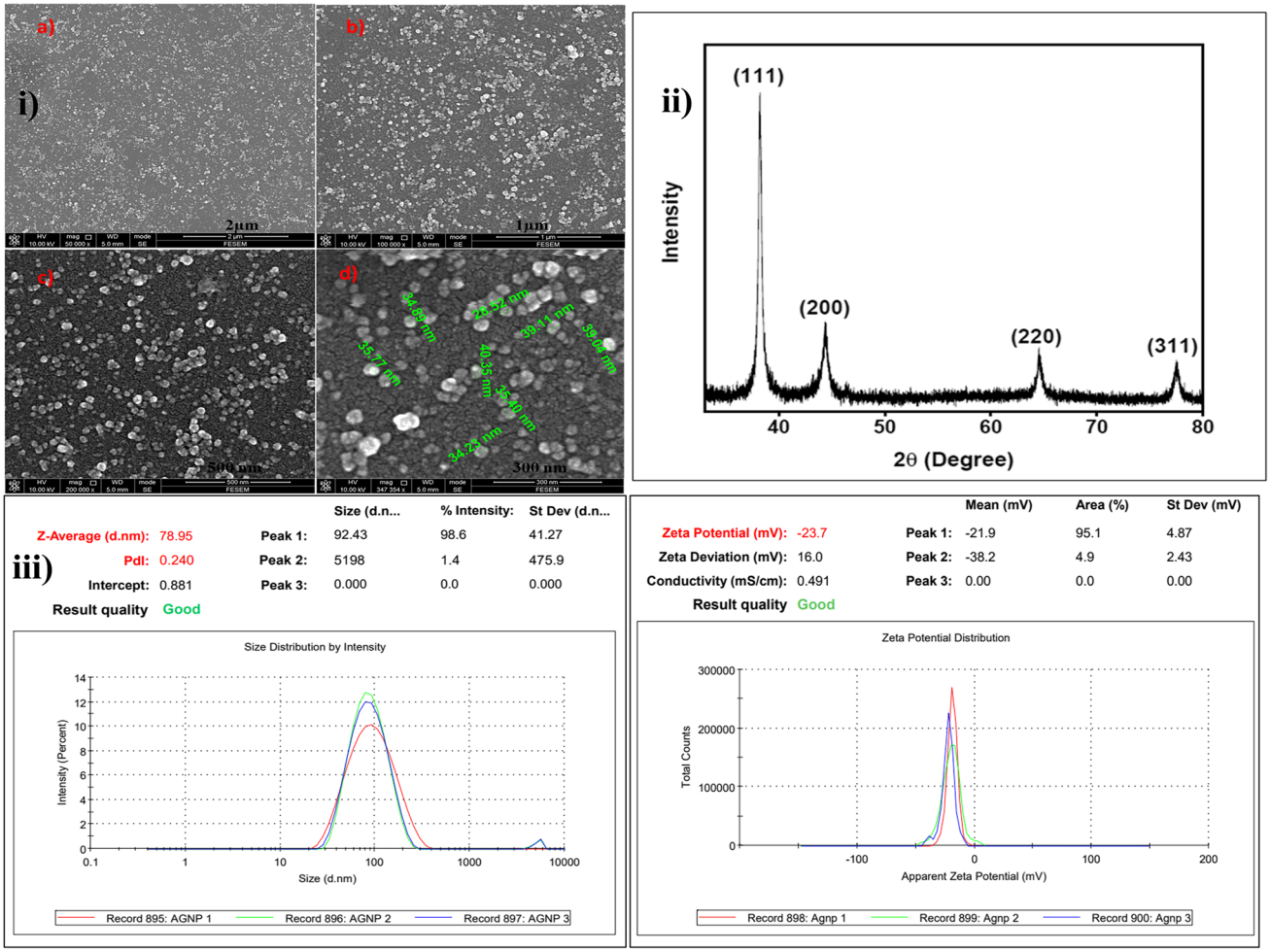


Fig. 2 i–iii Characterization of green synthesized silver nanoparticles using FESEM, XRD, and Zetasizer

the dispersion. ZP values are typically in the range of + 100 to − 100 mV. High ZP implies highly charged particles, which prevents aggregation of the particles due to electric repulsion. If the ZP is low, attraction overcomes repulsion, and it is likely that the mixture would form coagulate [40]. It is the measurement of immensity and type of surface charges corresponding to double layers all over the particles with high positive/negative values of zeta potential. Accordingly, ± 30 mV is considered a more stable colloidal solution due to electrostatic attraction/repulsion among the particles [41]. It is related to particle size and environmental factors like pH, ionic strength, and the type of ions present in suspension [42].

A scanning electron microscope (SEM) was used to determine the shape and size of the synthesized SNPs in this study. An aqueous solution of SNPs was added dropwise to a quartz thin film and allowed to settle, followed by SEM images being captured. The shape of SNPs observed from SEM was spherical, as shown in Fig. 2i. A broad size distribution of silver nanoparticles was observed. A recent study agreed with the observed absorption and Zetasizer [43]. A comparatively lower value of zeta potential for the silver nanoparticles suggests that SNPs have a tendency to aggregate to form bigger sized particles [44].

3.3 Antibacterial effect of SNPs produced using *Chlorella sorokiniana*

In the present study, the antibacterial activity of SNPs synthesized using *Chlorella sorokiniana* extract was tested against *E. coli* (gram-negative) and *S. aureus* (gram-positive) using the agar well diffusion technique. *E. coli* and *S. aureus* are considered potential human pathogens [45]. The result of the study indicated that the maximum zone of inhibition was observed at higher concentrations (25 µL) of SNPs as compared to lower concentrations (Fig. 3). The algal extract exhibited poor or no killing activity against any of the tested bacteria, thus indicating that the killing mechanism is solely due to the synthesized SNPs and not due to any of the bioactive compounds present in the algal extracts.

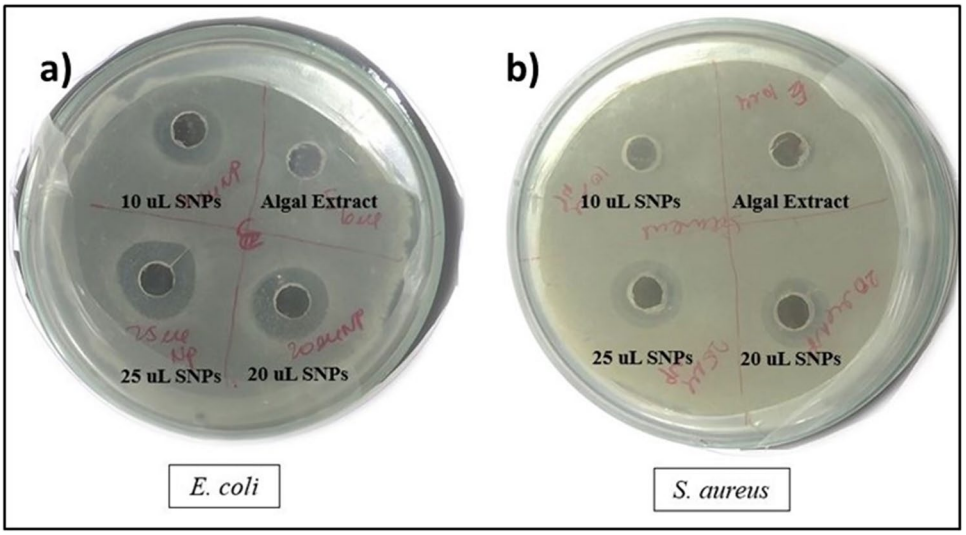


Fig. 3 Nutrient agar plates depicting zone of inhibition caused due to the activity of SNPs. **a** Zone of inhibition caused due to varied concentrations of SNPs against *E. coli*. **b** Zone of inhibition caused due to varied concentrations of SNPs against *S. aureus*

Table 1 Zone of inhibition (mm) depicted by synthesized SNPs against *E. coli* and *S. aureus*

Test solution (in µL)	Zone of inhibition (in millimeters)	
	E. coli	S. aureus
Algal extract	0 mm	0 mm
SNPs (10 µL)	11 ± 1 mm	7 ± 1 mm
SNPs (20 µL)	17 ± 1 mm	11 ± 1 mm
SNPs (25 µL)	20 ± 1 mm	15 ± 1 mm

As shown in Table 1, SNPs exhibited profound antibacterial activity against *E. coli* (20 ± 1 mm) and were found to be moderately effective against *S. aureus* (15 ± 1 mm). This difference in the effectiveness of SNPs can be attributed to the differential composition of cell walls of gram-positive and gram-negative bacterial species. SNPs exhibited a lower antibacterial activity against *S. aureus* due to the thick peptidoglycan layer in the cell wall of gram-positive bacteria, which hinders the absorption of SNPs in the cell cytoplasm, leading to a lower or ineffective killing activity [45, 46]. The profound killing activity against *E. coli* was due to a thin peptidoglycan layer present in the cell wall of gram-negative bacterial species, allowing the absorption of SNPs into the cell cytoplasm [47].

The antibacterial activity of SNP is associated with the nanoscale hydrodynamic diameters of the particles, which enable easy uptake by the cells for acting against bacterial cells [45]. It has been reported that the smaller the size of the nanoparticles, the larger is the surface area, enabling a greater contact with the cells, thus enhancing the killing efficacy [48]. The antibacterial properties of nanoparticles can be attributed to one or a combination of the following mechanisms: induction of oxidative stress among the bacterial cells, the release of metal ions that interact with cells and their components through various non-oxidative mechanisms [49, 50]. Some studies have reported that SNPs are responsible for neutralizing the charge on the bacterial cell wall leading to changes in permeability and ultimately causing cell death [48].

3.4 Photocatalytic dye degradation using SNPs produced using *Chlorella sorokiniana*

The photocatalytic dye degradation was observed due to the excitation of surface plasmon response in SNPs. The electrons in the valance band move towards the conduction band, causing a hole in the lower band, and shifting the electrons in the upper band when exposed to sunlight. Excited electrons at nanoparticle surfaces interact with oxygen molecules and produce superoxide ions and hydroxyl radicals. The nanoparticles and hydroxyl radicals react with target dyes and break them into smaller fragments, thus facilitating their degradation [33, 51–54]. In the current study, the photocatalytic degradation of 5 PPM solution of eosin yellow (EY), methylene blue (MB), crystal violet (CV), and rhodamine B (RhB) dye was carried out using *Chlorella sorokiniana* mediated SNPs (Fig. 4). MB is one of the most prominent water pollutants as it finds its use in the textile industry for various purposes. CV belongs to the group of triarylmethane dye and finds its application in the paper industry, textile dying, and other purposes [55]. EY finds its use in histology for cell staining and visualization. RhB is commonly used for industrial purposes such as printing, dyeing, paints, and leather and is known to cause various health problems such as eye and skin irritation [56]. It was noted that after 18 h of incubation in the sunlight, 0.5 mL of the biogenic SNPs per 20 mL of dye solutions was able to photo-catalytically degrade 97.04% of crystal violet, followed by 95.75% of methylene blue and 94.9% of eosin yellow. The degradation of EY reached 92.97% in 4 h, after which there

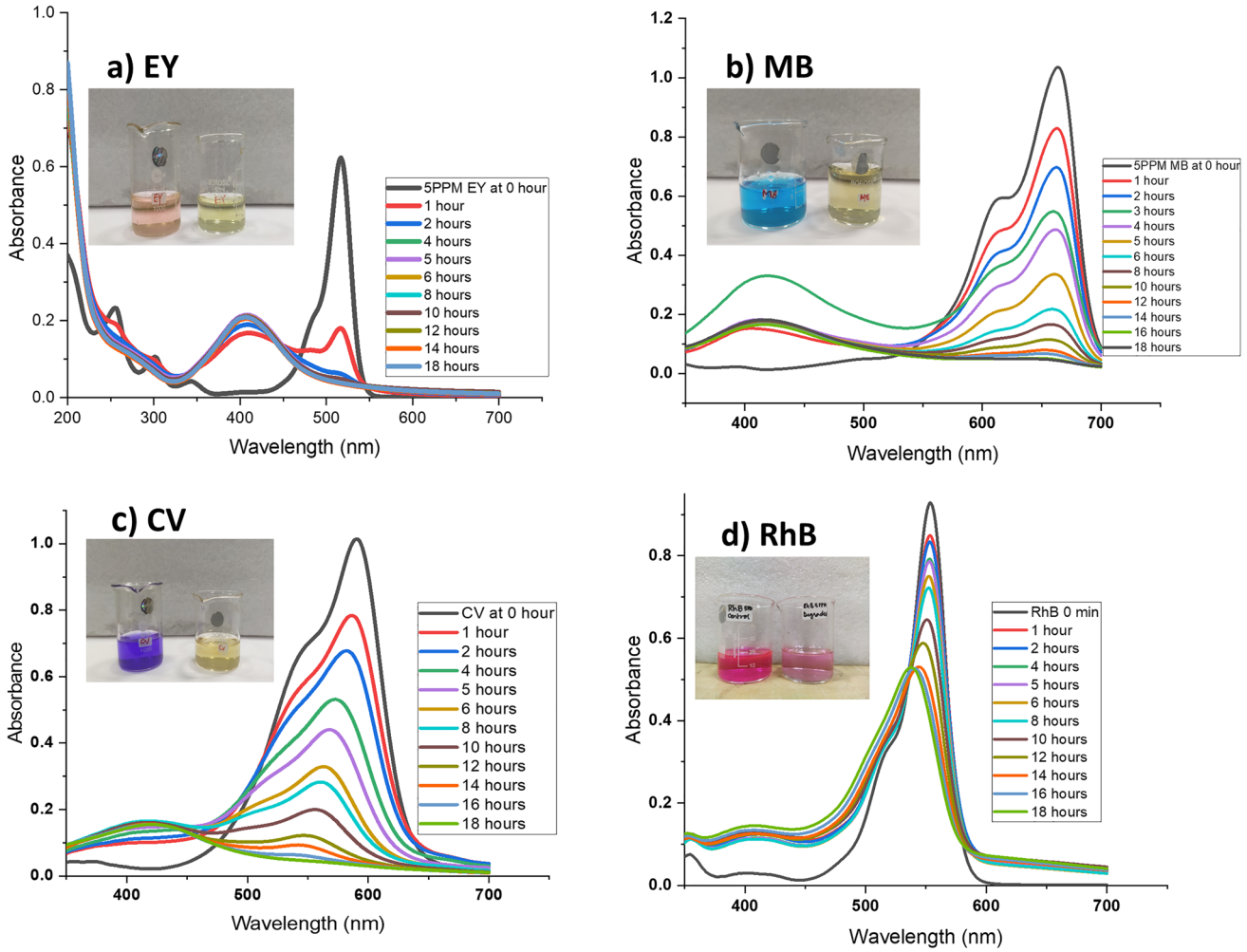


Fig. 4 Absorption spectrum and visual color change after dye degradation over a period of 18 h of incubation in light condition. **a** EY. **b** MB. **c** CV. **d** RhB

was only a little increment in degradation efficiency until 18 h (Fig. 5). The least degradation efficiency was found for rhodamine blue (56.05%). Vanaja et al. reported 95% degradation of MB using biogenic SNPs after 72 h of incubation in sunlight [57]. A study conducted by Karthik et al. reported that biogenic SNPs were able to degrade about 97% of EY after 1 h of irradiation [52]. Studies performed by Azeem et al. reported 89% degradation of crystal violet after 6 h of incubation [55]. The degradation efficiency of SNPs under UV irradiation was found to be 86.51% after 90 min of incubation in the case of RhB, as demonstrated by Shaikh et al. in their study [58]. In a different study conducted by Aryan et al. on SNP-based RhB degradation, 83% degradation within 45 min under UV irradiation was reported [33]. In this study, *C. sorokiniana*–mediated SNPs were able to degrade 56.05% of RhB in the presence of sunlight, making this experimental setup cost-effective due to non-requirement of UV light. In another study, photocatalytic degradation of rhodamine B by different metal nanoparticles, i.e., zinc oxide nanoparticles synthesized using the leaf extract of *Cynometra ramiflora*, was reported with 98% degradation efficiency under direct sunlight [59]. The photocatalytic ability refers to the creation of electron–hole pairs in a material when exposed to sunlight. SNPs have been widely studied as a photocatalytic agent for the degradation of dyes for the treatment of wastewater. The use of solar energy for photocatalytic degradation is gaining importance. The photocatalytic activity in the presence of sunlight is highly dependent upon the size of SNPs, morphology, and crystalline structure of SNPs [60].

4 Conclusion

The present study investigated the prospect of using the green microalga *Chlorella sorokiniana* extract for silver nanoparticle synthesis. Optimum SNP yield was achieved

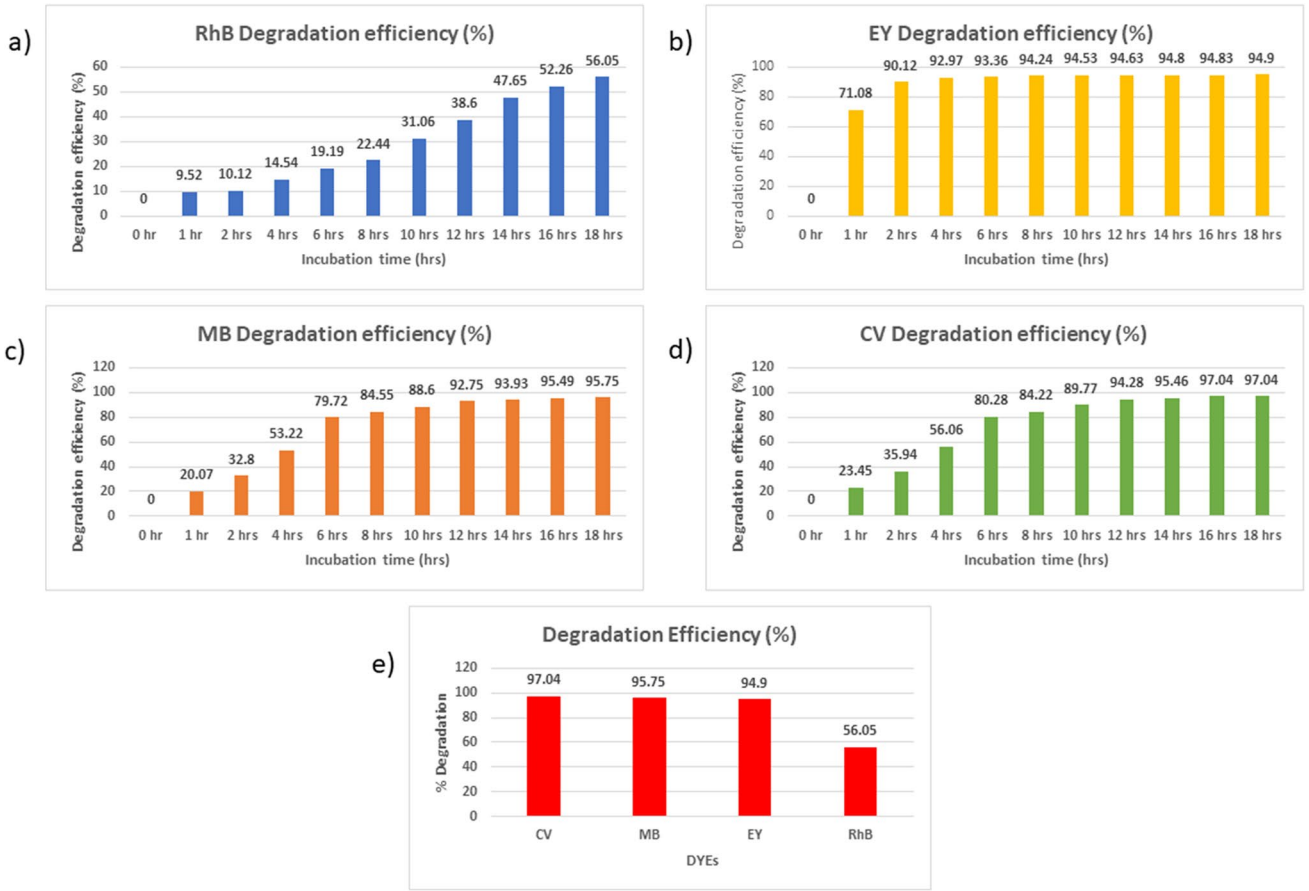


Fig. 5 Photocatalytic degradation efficiency using SNPs as catalysts against **a** rhodamine B (RhB) dye; **b** eosin yellow (EY); **c** methylene blue (MB); and **d** crystal violet (CV); and **e** comparative degradation efficiency of four dyes after 18 h of exposure to sunlight

at pH 7, with a metal precursor strength of 10 mM with extract to the salt ratio of 10:1. Boiling biomass at 80 °C for 20 min to prepare the extract is a better strategy than sonication and autoclave for cell disruption. SNPs were synthesized optimally at room temperature under light conditions. It was observed that the SNP synthesis rate was quite higher at 60 °C temperature compared to room temperature, but the SNP synthesis at the former condition was highly unstable. A color change from light green to dark brown initially confirmed the formation of SNPs. The UV–Vis absorption spectrum of the synthesized SNPs at optimized conditions showed a broad peak at 441 nm, a characteristic SNP band. FESEM and XRD analysis showed that the SNPs were spherical and crystalline in nature, whereas the significant value of zeta potential indicates their high stability over a longer period of time. SNPs thus prepared at optimum conditions exhibited potency against two bacteria, gram-negative *Escherichia coli* and gram-positive *Staphylococcus aureus*. Also, these SNPs were able to degrade four different dyes, namely crystal violet, methylene blue, eosin Y, and rhodamine B, under natural sunlight. It is evident from the study that *Chlorella sorokiniana* extract–mediated green SNPs can be considered a promising agent for the degradation of dyes in an aqueous solution as well as an adjuvant for the treatment of microbial infections**.**

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13399-022-03433-w.

Author contribution Both authors, Lakhan Kumar and Navneeta Bharadvaja, contributed to the study's conception and design. Lakhan Kumar performed material preparation, investigation, data collection, and analysis. Lakhan Kumar performed material preparation, investigation, data collection, and analysis. Lakhan Kumar wrote the draft of the manuscript under the supervision and guidance of Navneeta Bharadvaja. Both authors read and approved the final manuscript.

Data availability The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval Not applicable.

Competing interests The authors declare no competing interests.

References

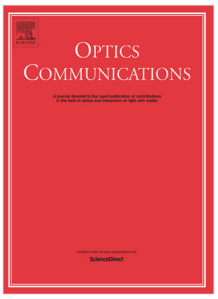
1. El-Sheekh MM, Gharieb MM, Abou-El-Souod GW (2009) Biodegradation of dyes by some green algae and cyanobacteria. Int Biodeterior Biodegrad 63:699–704. https://doi.org/10.1016/j.ibiod.2009.04.010
2. Liu Q (2020) Pollution and treatment of dye waste-water. IOP Conference Series: Earth and Environmental Science 514(5):052001
3. Kumar L, Bharadvaja N (2020) Microorganisms: a remedial source for dye pollution. In: Removal of toxic pollutants through microbiological and tertiary treatment. Elsevier, pp 309–333
4. Fatima F, Siddiqui S, Khan WA (2021) Nanoparticles as novel emerging therapeutic antibacterial agents in the antibiotics resistant era. Biol Trace Elem Res 199:2552–2564
5. Murray CJ, Ikuta KS, Sharara F et al (2022) Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399:629–655. https://doi.org/10.1016/S0140-6736(21)02724-0
6. Serwecińska L (2020) Antimicrobials and antibiotic-resistant bacteria: a risk to the environment and to public health. Water (Switzerland) 12:3313. https://doi.org/10.3390/w12123313
7. Kumar L, Ragunathan V, Chugh M, Bharadvaja N (2021) Nanomaterials for remediation of contaminants: a review. Environ Chem Lett 19:3139–3163
8. Gupta A, Mumtaz S, Li CH et al (2019) Combatting antibiotic-resistant bacteria using nanomaterials. Chem Soc Rev 48:415–427
9. Makabenta JMV, Nabawy A, Li CH et al (2021) Nanomaterial-based therapeutics for antibiotic-resistant bacterial infections. Nat Rev Microbiol 19:23–36
10. Truu M, Ligi T, Nõlvak H et al (2022) Impact of synthetic silver nanoparticles on the biofilm microbial communities and wastewater treatment efficiency in experimental hybrid filter system treating municipal wastewater. J Hazard Mater 440:129721. https://doi.org/10.1016/j.jhazmat.2022.129721
11. Abdolalian S, Taghavijeloudar M (2022) Performance evaluation and optimization of ZnO-PVP nanoparticles for photocatalytic wastewater treatment: interactions between UV light intensity and nanoparticles dosage. J Clean Prod 365:132833. https://doi.org/10.1016/j.jclepro.2022.132833
12. Chen H-f, Wu J-j, Wu M-y, Jia H (2019) Preparation and antibacterial activities of copper nanoparticles encapsulated by carbon. New Carbon Mater 34:382–389. https://doi.org/10.1016/S1872-5805(19)30023-X
13. Azizi-Lalabadi M, Ehsani A, Divband B, Alizadeh-Sani M (2019) Antimicrobial activity of titanium dioxide and zinc oxide nanoparticles supported in 4A zeolite and evaluation the morphological characteristic. Sci Rep 9:1–10. https://doi.org/10.1038/s41598-019-54025-0
14. Soo JZ, Chai LC, Ang BC, Ong BH (2020) Enhancing the antibacterial performance of titanium dioxide nanofibers by coating with silver nanoparticles. ACS Appl Nano Mater 3:5743–5751. https://doi.org/10.1021/acsanm.0c00925
15. Arakha M, Pal S, Samantarrai D et al (2015) Antimicrobial activity of iron oxide nanoparticle upon modulation of nanoparticle-bacteria interface. Sci Rep 5:1–12. https://doi.org/10.1038/srep14813
16. Shamaila S, Zafar N, Riaz S et al (2016) Gold nanoparticles: an efficient antimicrobial agent against enteric bacterial human pathogen. Nanomaterials 6:71. https://doi.org/10.3390/nano6040071
17. Yun JE, Lee DG (2017) Silver nanoparticles: a novel antimicrobial agent. In: Antimicrobial nanoarchitectonics: from synthesis to applications. Elsevier, pp 139–166
18. Kumar L, Mohan L, Anand S et al (2022) Phyconanoremediation: a sustainable approach to deal with environmental pollutants heavy metals and dyes. Vegetos 2022:1–16. https://doi.org/10.1007/S42535-022-00399-Y
19. El-Sheekh MM, El-Kassas HY (2016) Algal production of nano-silver and gold: their antimicrobial and cytotoxic activities: a review. J Genet Eng Biotechnol 14:299–310
20. Chaudhary R, Nawaz K, Khan AK et al (2020) An overview of the algae-mediated biosynthesis of nanoparticles and their biomedical applications. Biomolecules 10:1–35
21. Ziganshina EE, Bulynina SS, Ziganshin AM (2022) Growth characteristics of Chlorella sorokiniana in a photobioreactor during the utilization of different forms of nitrogen at various temperatures. Plants 11. https://doi.org/10.3390/plants11081086
22. Taghavijeloudar M, Yaqoubnejad P, Amini-Rad H, Park J (2021) Optimization of cultivation condition of newly isolated strain Chlorella sorokiniana pa.91 for CO2 bio-fixation and nutrients removal from wastewater: Impact of temperature and light intensity. Clean Technol Environ Policy 1:1–13. https://doi.org/10.1007/s10098-021-02199-5
23. Cazzaniga S, Dall'Osto L, Szaub J et al (2014) Domestication of the green alga Chlorella sorokiniana: reduction of antenna size improves light-use efficiency in a photobioreactor. Biotechnol Biofuels 7:1–13. https://doi.org/10.1186/s13068-014-0157-z
24. Kathiraven T, Sundaramanickam A, Shanmugam N, Balasubramanian T (2015) Green synthesis of silver nanoparticles using marine algae Caulerpa racemosa and their antibacterial activity against some human pathogens. Appl Nanosci 5:499–504. https://doi.org/10.1007/s13204-014-0341-2
25. Bhuyar P, Rahim MHA, Sundararaju S et al (2020) Synthesis of silver nanoparticles using marine macroalgae Padina sp. and its antibacterial activity towards pathogenic bacteria. Beni-Suef Univ J Basic Appl Sci 9. https://doi.org/10.1186/s43088-019-0031-y
26. Ibraheem IBM, Abd Elaziz BEE, Saad WF, Fathy WA (2016) Green biosynthesis of silver nanoparticles using marine red algae Acanthophora specifera and its antibacterial activity. J Nanomed Nanotechnol 7:7–10. https://doi.org/10.4172/2157-7439.1000409
27. Vivek M, Kumar PS, Steffi S, Sudha S (2011) Biogenic silver nanoparticles by Gelidiella acerosa extract and their antifungal effects. Avicenna J Med Biotechnol 3:143–148
28. Rani P, Kumar V, Singh PP et al (2020) Highly stable AgNPs prepared via a novel green approach for catalytic and photocatalytic removal of biological and non-biological pollutants. Environ Int 143:105924. https://doi.org/10.1016/j.envint.2020.105924
29. Kannan RRR, Arumugam R, Ramya D et al (2013) Green synthesis of silver nanoparticles using marine macroalga Chaetomorpha linum. Appl Nanosci 3:229–233. https://doi.org/10.1007/s13204-012-0125-5
30. Rajkumar R, Ezhumalai G, Gnanadesigan M (2021) A green approach for the synthesis of silver nanoparticles by Chlorella vulgaris and its application in photocatalytic dye degradation activity. Environ Technol Innov 21.https://doi.org/10.1016/j.eti.2020.101282
31. Anigol LB, Charantimath JS, Gurubasavaraj PM (2017) Effect of concentration and pH on the size of silver nanoparticles synthesized by green chemistry. Org Med Chem Int J 3:1–5. https://doi.org/10.19080/OMCIJ.2017.03.555622
32. Ahmed RH, Mustafa DE (2020) Green synthesis of silver nanoparticles mediated by traditionally used medicinal plants in Sudan. Int Nano Lett 10:1–14. https://doi.org/10.1007/s40089-019-00291-9
33. Aryan R, Mehata MS (2021) Green synthesis of silver nanoparticles using Kalanchoe pinnata leaves (life plant) and their antibacterial and photocatalytic activities. Chem Phys Lett 778:138760. https://doi.org/10.1016/j.cplett.2021.138760
34. Shaik MR, Khan M, Kuniyil M et al (2018) Plant-extract-assisted green synthesis of silver nanoparticles using Origanum vulgare L. extract and their microbicidal activities. Sustainability 10:1–14. https://doi.org/10.3390/su10040913

35. Amin M, Anwar F, Janjua MRSA et al (2012) Green synthesis of silver nanoparticles through reduction with Solanum xanthocarpum L. berry extract: characterization, antimicrobial and urease inhibitory activities against Helicobacter pylori. Int J Mol Sci 13:9923–9941. https://doi.org/10.3390/ijms13089923
36. Pastoriza-Santos I, Liz-Marzán LM (2002) Formation of PVP-protected metal nanoparticles in DMF. Langmuir 18:2888–2894. https://doi.org/10.1021/la015578g
37. Meng Y (2015) A sustainable approach to fabricating ag nanoparticles/PVA hybrid nanofiber and its catalytic activity. Nanomaterials 5:1124–1135. https://doi.org/10.3390/nano5021124
38. Priyadharshini RI, Prasannaraj G, Geetha N, Venkatachalam P (2014) Microwave-mediated extracellular synthesis of metallic silver and zinc oxide nanoparticles using macro-algae (Gracilaria edulis) extracts and its anticancer activity against human PC3 cell lines. Appl Biochem Biotechnol 174:2777–2790. https://doi.org/10.1007/s12010-014-1225-3
39. Sankhla A, Sharma R, Yadav RS et al (2016) Biosynthesis and characterization of cadmium sulfide nanoparticles - an emphasis of zeta potential behavior due to capping. Mater Chem Phys 170:44–51. https://doi.org/10.1016/j.matchemphys.2015.12.017
40. Shnoudeh AJ, Hamad I, Abdo RW et al (2019) Synthesis, characterization, and applications of metal nanoparticles. In: Biomaterials and bionanotechnology. Academic Press, pp 527–612
41. Xu R (2008) Progress in nanoparticles characterization: sizing and zeta potential measurement. Particuology 6:112–115. https://doi.org/10.1016/j.partic.2007.12.002
42. El Badawy AM, Luxton TP, Silva RG et al (2010) Impact of environmental conditions (pH, ionic strength, and electrolyte type) on the surface charge and aggregation of silver nanoparticles suspensions. Environ Sci Technol 44:1260–1266. https://doi.org/10.1021/es902240k
43. Pesika NS, Stebe KJ, Searson PC (2003) Relationship between absorbance spectra and particle size distributions for quantum-sized nanocrystals. J Phys Chem B 107:10412–10415. https://doi.org/10.1021/jp0303218
44. Satapathy S, Shukla SP, Sandeep KP et al (2015) Evaluation of the performance of an algal bioreactor for silver nanoparticle production. J Appl Phycol 27:285–291. https://doi.org/10.1007/s10811-014-0311-9
45. de Aragão AP, de Oliveira TM, Quelemes PV et al (2016) Green synthesis of silver nanoparticles using the seaweed Gracilaria birdiae and their antibacterial activity. Arab J Chem 12:4182–4188. https://doi.org/10.1016/j.arabjc.2016.04.014
46. Mohandass C, Vijayaraj AS, Rajasabapathy R et al (2013) Biosynthesis of silver nanoparticles from marine seaweed Sargassum cinereum and their antibacterial activity. Indian J Pharm Sci 75:606–610
47. Peiris MMK, Fernando SSN, Jayaweera PM et al (2018) Comparison of antimicrobial properties of silver nanoparticles synthesized from selected bacteria. Indian J Microbiol 58:301–311. https://doi.org/10.1007/s12088-018-0723-3
48. Wang L, Hu C, Shao L (2017) The antimicrobial activity of nanoparticles: present situation and prospects for the future. Int J Nanomed 12:1227–1249. https://doi.org/10.2147/IJN.S121956
49. Nagy A, Harrison A, Sabbani S et al (2011) Silver nanoparticles embedded in zeolite membranes: release of silver ions and mechanism of antibacterial action. Int J Nanomed 6:1833–1852. https://doi.org/10.2147/ijn.s24019
50. Leung YH, Ng AMC, Xu X et al (2014) Mechanisms of antibacterial activity of MgO: non-ROS mediated toxicity of MgO nanoparticles towards Escherichia coli. Small 10:1171–1183. https://doi.org/10.1002/smll.201302434
51. Kang SF, Liao CH, Po ST (2000) Decolorization of textile wastewater by photo-Fenton oxidation technology. Chemosphere 41:1287–1294. https://doi.org/10.1016/S0045-6535(99)00524-X
52. Karthik R, Govindasamy M, Chen SM et al (2017) Biosynthesis of silver nanoparticles by using Camellia japonica leaf extract for the electrocatalytic reduction of nitrobenzene and photocatalytic degradation of eosin-Y. J Photochem Photobiol B Biol 170:164–172. https://doi.org/10.1016/j.jphotobiol.2017.03.018
53. Khare P, Singh A, Verma S et al (2018) Sunlight-induced selective photocatalytic degradation of methylene blue in bacterial culture by pollutant soot derived nontoxic graphene nanosheets. ACS Sustain Chem Eng 6:579–589. https://doi.org/10.1021/acssuschemeng.7b02929
54. Singh MK, Mehata MS (2019) Phase-dependent optical and photocatalytic performance of synthesized titanium dioxide (TiO2) nanoparticles. Optik (Stuttg) 193:163011. https://doi.org/10.1016/j.ijleo.2019.163011
55. Abdel Azeem MN, Hassaballa S, Ahmed OM et al (2021) Photocatalytic activity of revolutionary Galaxaura elongata, Turbinaria ornata, and Enteromorpha flexuosa's bio-capped silver nanoparticles for industrial wastewater treatment. Nanomaterials 11.https://doi.org/10.3390/nano11123241
56. Thakur A, Kaur H (2016) Removal of hazardous Rhodamine B dye by using chemically activated low cost adsorbent: pine cone charcoal. Int J Chem Phys Sci 5:17–28
57. Vanaja M, Paulkumar K, Baburaja M et al (2014) Degradation of methylene blue using biologically synthesized silver nanoparticles. Bioinorg Chem Appl 2014.https://doi.org/10.1155/2014/742346
58. Shaikh WA, Chakraborty S, Islam RU (2020) Photocatalytic degradation of rhodamine B under UV irradiation using Shorea robusta leaf extract-mediated bio-synthesized silver nanoparticles. Int J Environ Sci Technol 17:2059–2072. https://doi.org/10.1007/s13762-019-02473-6
59. Varadavenkatesan T, Lyubchik E, Pai S et al (2019) Photocatalytic degradation of Rhodamine B by zinc oxide nanoparticles synthesized using the leaf extract of Cyanometra ramiflora. J Photochem Photobiol B Biol 199:111621. https://doi.org/10.1016/j.jphotobiol.2019.111621
60. Kumar B (2021) Green synthesis of gold, silver, and iron nanoparticles for the degradation of organic pollutants in wastewater. J Compos Sci 5.https://doi.org/10.3390/jcs5080219

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 Pre-proof



Broadband plasmonic switches based on nanodisc-dimers with progressively increasing diameters on a plasmonic film with a VO_2 spacer

Kirti Dalal, Yashna Sharma

PII: S0030-4018(22)00768-4
DOI: https://doi.org/10.1016/j.optcom.2022.129121
Reference: OPTICS 129121

To appear in: *Optics Communications*

Received date : 13 July 2022
Revised date : 16 October 2022
Accepted date : 6 November 2022

Please cite this article as: K. Dalal and Y. Sharma, Broadband plasmonic switches based on nanodisc-dimers with progressively increasing diameters on a plasmonic film with a VO_2 spacer, *Optics Communications* (2022), doi: https://doi.org/10.1016/j.optcom.2022.129121.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Elsevier B.V. All rights reserved.

Journal Pre-proof

Broadband Plasmonic Switches based on Nanodisc-Dimers with Progressively Increasing Diameters on a Plasmonic Film with a VO_2 spacer

Kirti Dalal and Yashna Sharma*

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

*yashnasharma@dtu.ac.in

ABSTRACT

In this paper, we propose active broadband plasmonic switches based on a periodic array of metallic dimers of nanodiscs with progressively increasing diameters on a gold coated SiO_2 substrate with a VO_2 thin film as a spacer between the nanodiscs and the underlying plasmonic substrate. Periodic arrays based on unit cells with plasmonic nanodisc dimers of varying diameters are employed to couple multiple wavelengths of the incident radiation into the plasmonic modes of the nanostructure, which results in a broadband plasmonic response. We employ a vanadium dioxide (VO_2) thin film as a spacer between these periodic sets of plasmonic nanodisc-dimers and the underlying gold coated substrate to achieve broadband switching. On exposure to voltage, infrared light or heat, the VO_2 spacer layer changes from its semiconducting state to its metallic state. This transformation leads to significant changes in its optical properties leading to changes in the reflectance spectra of the proposed nanostructure. This results in an efficient switching action over wide range of wavelengths spanning the C, L and U bands of optical communication. We demonstrate a broadband extinction ratio of 5 dB over an operating wavelength range of 650 nm spanning from 1460 nm to 2110 nm, and a broadband extinction ratio of 4 dB over an operating wavelength range of 702 nm spanning from 1432 nm to 2134 nm. In addition, we demonstrate the trade-off between the extinction ratio and the operating wavelength range of these switches. An exhaustive analysis of these switches has been carried out using finite difference time domain (FDTD) modelling to demonstrate that the spectral range over which switching can be achieved by employing these switches can be varied by varying the geometrical parameters of the proposed switches and thus these switches offer the flexibility of being employed for different wavelength regimes. These switches can be potentially employed in optical communication networks or as switching elements in plasmonic circuits.

Keywords: Plasmonics, Phase change materials, VO_2, Nanodiscs, Broadband switches

Journal Pre-proof

Journal Pre-proof

1. INTRODUCTION

Photonic switching, which is the control or switching of optical signals as they pass through an optical system, has become increasingly important for optical communication networks [1] and optical logic circuits [2]. Due to the recent advancements in high-speed optical communication systems, there is an increasing demand for ultrasmall and ultrafast switches with high extinction ratios [3]. In the last few decades, plasmonics [4] has been extensively employed for the realization of ultrasmall switches for telecommunication applications [5]. Plasmonics involves the interaction of light with metallic nanostructures and thin films. These metallic nanostructures and thin films support localized surface plasmon resonances and surface plasmon resonances [6-7]. Surface plasmons are the coherent collective oscillations of the conduction band electrons, which get excited when light of a specific wavelength (plasmon resonance wavelength) is incident on a metallic nanostructure [6]. As these plasmons can be confined to sub-wavelength volumes, the size of optical switches based on plasmonics can be reduced below the diffraction limit which consequently leads to an ultrasmall footprint and higher integration density of such devices [8].

Active plasmonics, which is extensively employed for plasmonic switching, is a sub-field of plasmonics where external stimuli are used to achieve a dynamically tunable and reversible plasmonic response from the metallic nanostructures [9]. In addition to being employed for telecommunication applications, plasmonic switches based on active plasmonics are also the potential components for the realization of plasmonic circuits with electro-optical [10] or all-optical control [11]. Over the past few decades, many designs for the active switching of plasmonic devices have been proposed. These include active plasmonic switches based on nanostructures coated with polymer layers doped with photochromic molecules [12], thermo-optical materials [13], electro-optic films based on liquid crystals molecules [14], electrically-controlled materials such as graphene [15] or electromechanically controlled conductive polymer layers [16]. In addition, active switching can also be achieved by employing plasmonic nanostructures with tunable gaps by incorporating plasmonic nanoparticles into freestanding stimulus-responsive hydrogels [17] or by using the non-linearity of materials for active switching [18].

One of the most promising methodologies to achieve active plasmonic switching is based on the integration of phase change materials [19] with the plasmonic nanostructures to demonstrate ultrafast and ultrasmall plasmonic switches. Over the years, many phase change materials have been employed for achieving active plasmonic switching such as oxides of vanadium [20], germanium antimony telluride (GST) [21] and others [22, 23]. Of special interest is vanadium dioxide (VO_2), which changes from its monoclinic semiconducting phase to its tetragonal metallic phase on being heated to 68ºC, or being exposed to an electric field [24] or by using optical radiation [25]. This phase transformation - which happens on a femtosecond scale [26]- leads to a change in the real and the imaginary parts of the refractive index of VO_2 which results in a significant change in its optical as well as electrical properties [27]. Moreover, the transition temperature of VO_2 can be reduced to near-room temperature by using dopants such as tungsten or molybdenum [28]. Due to the low transition temperature, low thermal hysteresis, and the ability of VO_2 to withstand millions of switching cycles without degradation, it has been employed extensively for plasmonic switching [29, 30]. Although, there have been numerous efforts to increase the extinction ratio of VO_2 based plasmonic switches [31] over the last few decades, the operating wavelength range of these switches remains quite narrow. There are limited reports of broadband plasmonic switching by employing

Journal Pre-proof

VO_2 as the phase change material [32-37]. These include broadband switches based on a hybrid SPP-VO_2 geometry [32], hybrid plasmonic switches based on aluminium nanohole arrays and vanadium oxide [33], metal-VO_2 hybrid dagger like plasmonic nanostructure exhibiting switching based on circular dichroism [34], VO_2 coated plasmonic indium tin oxide nanorod arrays for switching in the visible and the infrared region [35], and others [36, 37]. For most of these switches, the operating range of wavelengths over which switching is demonstrated is lower than 500 nm and they do not offer the flexibility of tuning the spectral range of operation. It must be noted that broadband switching, with the flexibility of controlling the operating wavelength range, offers enormous advantage for the design of optical communication networks [38] and optical computing systems [39].

In this paper, we propose broadband plasmonic switches based on a periodic array of metallic nanodisc-dimers with progressively increasing diameters. The nanodisc-dimers are placed on a gold coated silicon dioxide (SiO_2) substrate with a VO_2 thin film as a spacer between the discs and the underlying gold-coated substrate (see Fig. 1). It is well known that — for a periodic array of circular metallic discs on a substrate, the incident optical radiation is coupled into the plasmonic modes of the nanostructure at specific plasmon resonance wavelengths, thus resulting in narrow and strong dips in the reflection spectrum of the nanostructure [40, 41]. Further, there is a red-shift in the plasmon resonance wavelength due to the retardation of the depolarization field when the diameter of the nanodiscs is increased [6]. It can therefore be inferred that by employing an array of dimers of metallic nanodiscs with progressively increasing diameters as shown in Fig. 1, multiple wavelengths can be coupled into the plasmonic modes of the nanostructure such that each nanodisc dimer with a specific value of diameter results in the coupling of incident radiation into the plasmonic modes of the nanostructure at a specific plasmon resonance wavelength. This coupling of incident radiation to plasmonic modes at multiple wavelengths results in an overall broadband resonance dip in the reflection spectrum of the nanostructure. We, therefore, employ periodic arrays of metallic gold nanodisc-dimers with progressively increasing diameters — that are capable of exhibiting a broadband plasmonic behaviour — in combination with a phase change material, i.e., vanadium dioxide to achieve broadband switching.

We employ finite difference time domain (FDTD) modelling to demonstrate the broadband switching behaviour of the proposed plasmonic switches. We comprehensively analyze the performance of these switches with two, three, four and five sets of nanodisc dimers of progressively increasing diameters periodically arranged on the VO_2-coated gold film. We show that the operating wavelength range of these switches widens with the inclusion of more sets of nanodisc dimers of different diameters in a unit cell of the periodic array. This is because multiple wavelengths in the incident radiation get coupled into the plasmonic modes of this composite nanostructure, with a metallic nanodisc dimer of a specific diameter corresponding to a specific plasmon resonance wavelength. We demonstrate that with the proper optimization of the values of the diameters of the nanodisc dimers which are periodically arranged on the VO_2-coated gold film, these broadband switches can be appropriately designed to span the C, L and U bands of optical communication.

The performance of these switches is measured in terms of extinction ratio (ER) which is the ratio of the reflectance from the switch with VO_2 in its metallic phase and the reflectance from the switch with VO_2 in its semiconducting phase expressed in dB. When a periodic array of five sets of nanodisc-dimers with progressively increasing diameters is placed on the VO_2-thin film on top of the gold coated SiO_2 substrate, the proposed switch exhibits a broadband extinction ratio of 5 dB over an operating wavelength range of

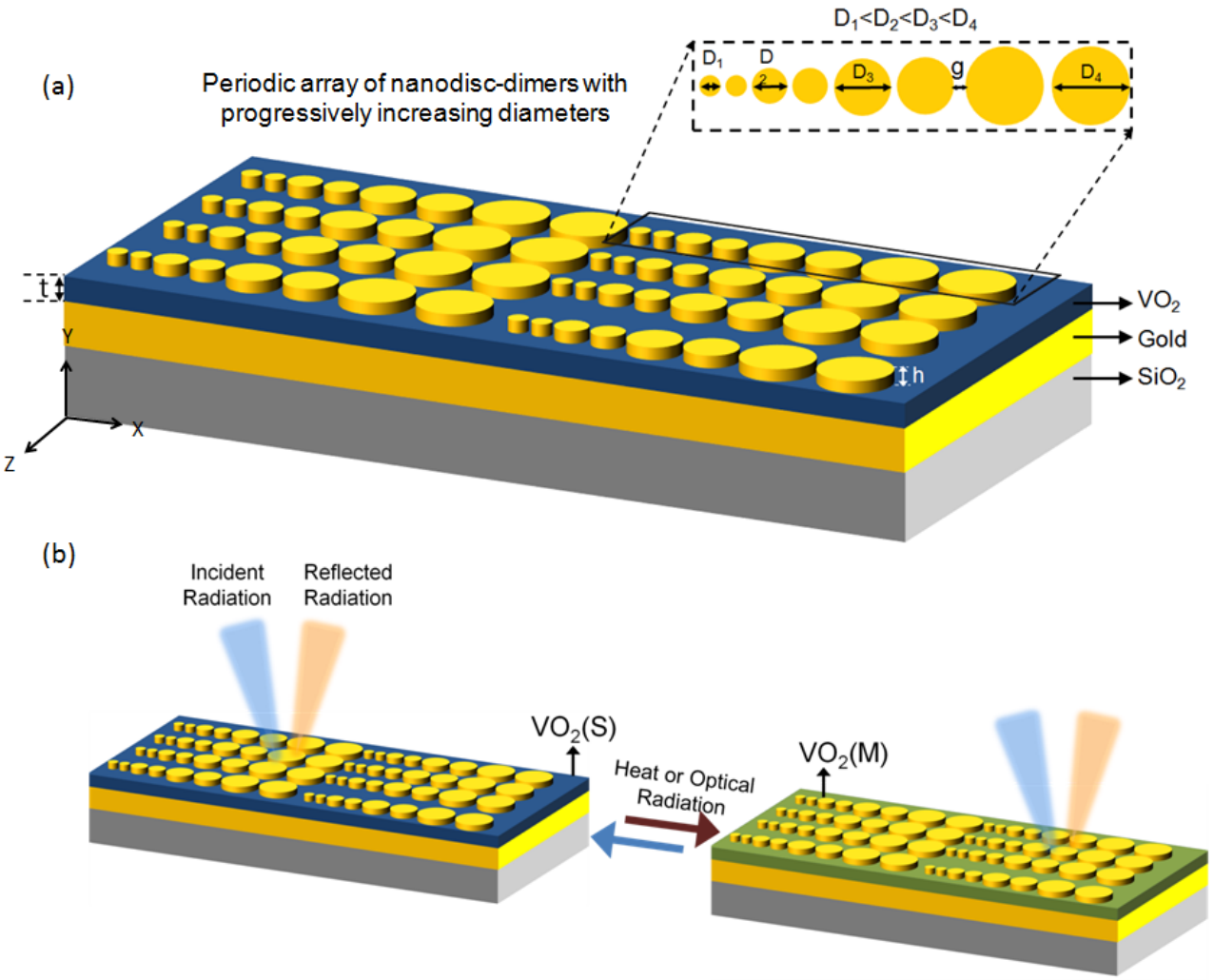


Fig. 1. (a) Schematic for the proposed broadband plasmonic switch based on a periodic array of nanodisc-dimers with progressively increasing disc sizes on a plasmonic film with a VO_2 spacer. The diameters of the dimers of nanodiscs are taken such that $D_1<D_2<D_3<D_4$, where D_1, D_2, D_3 and D_4 are the diameters of the first, second, third and fourth set of nanodisc-dimer, respectively. The periodic array of dimers of gold nanodiscs are placed on a gold coated SiO_2 substrate with a thin film of VO_2 as a spacer. (b) Schematic showing the switching action of the proposed switch. The phase of the VO_2 spacer layer changes from its semiconducting state to its metallic state by application of voltage, heat or optical radiation, thus leading to change in the reflectance spectra.

650 nm spanning from 1460 nm to 2110 nm. Previously , Joushanghani *et al.* [32] proposed a hybrid SPP-VO_2 waveguide switch which could provide a maximum extinction ratio of 16.4 dB but the operating wavelength range was only ~100 nm. Mandal *et al.* [34] designed a metal-VO_2 hybrid dagger like geometry to exhibit switching based on circular dichroism where the broadband operating wavelength was limited to ~200 nm. In addition, Sun *et al.* [33] reported a broadband operating wavelength of ~ 650 nm for a photonic switch based on the hybrid combination of hexagonal array of aluminium nanoholes and VO_2 with 5 dB as the broadband extinction ratio. However, the maximum extinction ratio for these switches was also limited to 5 dB. Therefore, the broadband plasmonic switch proposed by us that is capable of being operated over a

broad operating wavelength range of 650 nm with a broadband extinction ratio of 5 dB and highest extinction ratio >20 dB has not been demonstrated so far. In addition, we show a trade-off between the extinction ratio and the operating wavelength range of these switches. We demonstrate that the proposed switch can even provide a extinction ratio > 8 dB over a broadband operating wavelength range > 530 nm. As VO_2 exhibits an ultrafast switching [42], the proposed switches can be potentially employed for high-speed optical communication networks or as switching components in plasmonic circuits. Additionally, it should be noted that the wavelength range over which switching can be achieved by employing these switches can be tuned by changing the diameters of the nanodisc-dimers. Thus, these switches are capable of being employed for different wavelength regimes. Moreover, these switches can be employed with both angled and normally incident light, and specific arrangements such as prism coupling are not needed to couple incident light into plasmonic modes of the metallic nanostructures.

These switches can be fabricated by employing a thin film deposition technique like atomic layer deposition (ALD) [43] or E-beam evaporation [44] to deposit a layer of gold on a SiO_2 substrate. Thereafter, a thin layer of VO_2 can be deposited by using a thin film deposition technique like pulsed laser deposition (PLD) [45]. Subsequently, the gold nanodisc dimers can be fabricated on top of the VO_2 layer by employing E-beam lithography [46] or helium ion milling [47] to fabricate the proposed switches. (See Supplementary Material, Note 1 for proposed fabrication methodology)

2. NUMERICAL METHODS

We have used a commercial finite difference time domain (FDTD) solver by Lumerical Inc. to carry out the 3D FDTD modelling of the proposed switches. A plane wave source with a 1 V/m incident electric field polarized in the X-direction (See Fig. 1) was used to illuminate the structure. After careful convergence testing, a uniform mesh size of 1 nm was employed in the X-, Y- and Z- directions of the whole simulation region. The structure was taken to be periodic in the X- and Z- directions, and perfectly matched layer (PML) boundaries were used in the Y-direction. Fig. 1 shows the gold nanodiscs with progressively increasing diameters placed on a gold coated SiO_2 substrate with a VO_2 thin film as a spacer between the nanodiscs and the underlying gold-coated substrate. The calculation and subsequent fitting of the dielectric constants for metallic VO_2 and semiconductor VO_2 employed in this paper was based on the Lorentz-Drude Model [48, 49] as follows:

$$\varepsilon(\omega) = \varepsilon_\infty + \sum_{k=1}^{6} \frac{\Delta\varepsilon_k}{-\alpha_k\omega^2 - b_k(i\omega) + c_k} \quad (1)$$

The values of the constants for metallic VO_2 were taken as: ε_∞: 3.95; $\Delta\varepsilon_1$: 284.7835992; $\Delta\varepsilon_2$: 34.49365809; $\Delta\varepsilon_3$: 195.7081155; $\Delta\varepsilon_4$: 323.4583261; $\Delta\varepsilon_5$: 570.5996093; $\Delta\varepsilon_6$: 0; α_1: 1; α_2: 1; α_3: 1; α_4: 1; α_5: 1; α_6: 0; b_1: 3.344700147; b_2: 7.914574094; b_3: 13.79232715; b_4: 18.34111935; b_5: 24.4771238; b_6: 0; c_1: 0; c_2: 18.99430511; c_3: 201.3457979; c_4: 311.0176212; c_5: 543.4281993; c_6:0 [50]. The values of the constants for semiconductor VO_2 were taken as: ε_∞: 4.26; $\Delta\varepsilon_1$: 21.10833326; $\Delta\varepsilon_2$: 20.57271235; $\Delta\varepsilon_3$: 27.9097635; $\Delta\varepsilon_4$: 104.1014275; $\Delta\varepsilon_5$: 411.6548635; $\Delta\varepsilon_6$: 384.8646853; α_1: 1; α_2: 1; α_3: 1; α_4: 1; α_5: 1; α_6: 1; b_1: 4.083574816; b_2: 3.122733683; b_3: 3.671568571; b_4: 7.469830329; b_5: 23.27526155; b_6: 20.19793471; c_1: 26.7194092; c_2: 43.40234673; c_3: 57.78418944; c_4: 194.2190812; c_5: 312.807647; c_6: 363.0798918 [50].

(See Supplementary Material, Note 2 for dielectric constants) The diameter of the nanodiscs (D), the thickness of VO_2 layer (t), and the gap between the nanodiscs (g) were varied to analyze the effect of these geometrical parameters on the extinction ratio spectra of the proposed broadband plasmonic switch.

3. RESULTS AND DISCUSSIONS

We employ finite difference time domain (FDTD) modelling to illustrate the broadband switching behaviour of our proposed switches based on a periodic array of metallic dimers of nanodiscs with progressively increasing diameters on a gold coated SiO_2 substrate with a VO_2 thin film as a spacer between the discs and the underlying substrate. In this section, we comprehensively analyze the performance of these switches with two, three, four and five sets of nanodisc dimers of different diameters periodically arranged on the underlying VO_2 thin film and compare them with an equivalent periodic array of uniform-sized nanodiscs arranged on a VO_2 coated SiO_2 substrate. The height of the nanodisc-dimers is considered to be *h*, the thickness of VO_2 layer is *t*, the gap between the nanodiscs is *g* and the diameters of the nanodisc-dimers are kept to be variable as shown in Fig. 1.

To elaborate the utility of the proposed switches for broadband switching, an analysis of the switching behaviour of a periodic array of metallic nanodiscs with uniform diameters (UNDs) placed on a VO_2 coated SiO_2 substrate is carried out first. Figs. 2(a)-(b) show the reflectance spectra for such a periodic array of uniform metallic nanodiscs with VO_2 in its semiconductor and metallic state, respectively.

Fig. 2 (a) shows the reflectance spectra from the periodic array of uniform metallic nanodiscs (UNDs) for different values of diameters placed on a gold coated SiO_2 substrate with the VO_2 spacer in its semiconductor state. It must be noted that for this case of uniform nanodiscs, for each value of the nanodisc diameter, the incident optical radiation is coupled into a plasmonic waveguide mode (M1) around 720 nm and a geometry-dependent plasmonic mode (M2) of the nanostructure at a specific plasmon resonance wavelength resulting in a strong dip in the reflection spectra of the nanostructure. As the underlying layer of VO_2 is switched from its semiconducting state to its metallic state (See Fig. 2(b)), the real part of the refractive index of VO_2 decreases leading to a blue shift of the plasmonic dip in the reflectance spectra. Moreover, the dip in the reflectance spectra corresponding to the mode M2 also exhibits a broadening for the case when VO_2 is in its metallic state due to the damping of the plasmon resonance by the oscillations of mobile electrons in metallic VO_2 [35]. The overall broadening can also be attributed to the overlapping between the plasmonic mode and the plasmonic waveguide mode (at 720 nm). This change in the reflectance spectra as the underlying VO_2 is switched from its semiconducting state to its metallic state, can be utilized for switching.

The switching efficiency of all switches discussed in this paper is measured in terms of the extinction ratio (ER) which is the ratio of the reflectance from the switch with VO_2 in its metallic phase (R_M) and the reflectance from the switch with VO_2 in its semiconducting phase (R_S) as shown in Eq. (2).

$$ER = 10 * log\left(\frac{R_M(\lambda)}{R_S(\lambda)}\right) \quad dB \tag{2}$$

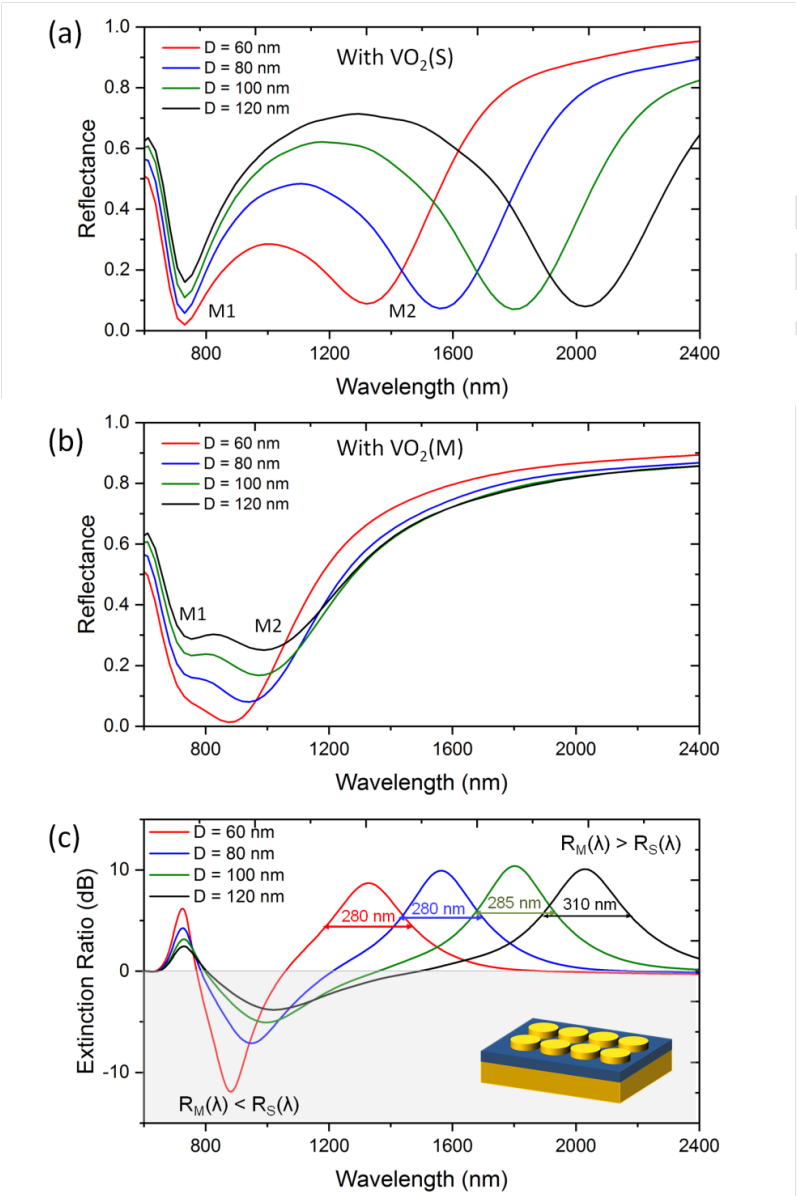


Fig. 2. Reflectance spectra of a periodic array of metallic nanodiscs with uniform diameters placed on a gold coated SiO_2 substrate with a thin film of VO_2 as a spacer. The reflectance spectra is shown for the values of the nanodisc diameters varying from 60 nm to 120 nm, when VO_2 is in its (a) semiconductor phase and (b) metallic phase. (c) Extinction ratio (ER) spectra for the periodic array of uniform metallic nanodiscs for different values of diameters placed on a gold coated SiO_2 substrate with a thin film of VO_2 as the spacer. The operating wavelength range at the full width half maximum (FWHM) of the maximum ER is shown for all cases. The VO_2 layer thickness, 't', the height of the nanodiscs, 'h', and the gap between the nanodiscs, 'g', were taken to be 10 nm, 50 nm, and 5 nm, respectively.

It can be observed from Figs. 2(a)-(b) that as the diameters of the uniform metallic nanodiscs (UNDs) are varied from 60 nm to 120 nm, the reflectance spectra for the switch for both the semiconducting and metallic states of the underlying VO_2 exhibit a red-shift of the plasmon resonance wavelengths. With VO_2 in its semiconducting state, the switch demonstrates a significant red-shift in the plasmon resonance wavelength from 1320 nm to 2030 nm as the diameter of the nanodiscs is varied from 60 nm to 120 nm. With VO_2 in its metallic state, there is a red-shift in the plasmon resonance wavelength from 880 nm to 990

nm as the diameter of the nanodiscs is varied from 60 nm to 120 nm. This red-shift in the plasmon resonance wavelength is expected due to the retardation of the depolarization field as the diameter of the nanodiscs is increased. As the diameter of the nanodiscs increases, the distance between the charges at the opposite interfaces of the nanodisc increases, thus leading to a lower restoring force causing a red-shift in the plasmon resonance wavelength [6]. Additionally, it must be noted that for both the semiconducting and the metallic state of VO_2, a plasmonic waveguide mode can also be observed around 720 nm, and this dip does not exhibit any red-shift on changing the diameter of the nanodiscs. Fig. 2(c) shows the extinction ratio (ER) of the periodic array of uniform metallic nanodiscs (UNDs) for different values of nanodisc diameters. It can be noticed from Fig. 2(c) that when the reflectance of the metallic phase is greater than the reflectance of the semiconductor phase (i.e., $R_M(\lambda) > R_S(\lambda)$), the ER is positive. However, when the reflectance of the metallic phase is less than the reflectance of the semiconductor phase (i.e., $R_M(\lambda) < R_S(\lambda)$), the ER becomes negative. It must be noted that although, the switch can be employed for switching in both wavelength regimes, i.e., with positive and negative values of the ER, we will be carrying out an extensive analysis and comparison of the switching characteristics of all switches in the wavelength regime where the ER is positive. It can be seen from Fig. 2(c), that, for all values of diameters for the uniform nanodiscs, the operating wavelength range — for the full width at half maxima (FWHM) of the extinction ratio spectrum — remains below 350 nm. It is well known that such narrow bandwidths are not suitable to be employed for applications such as wavelength division multiplexed (WDM) networks [51], free space optical (FSO) networks [52] and passive optical networks (PON) [53], where broadband switching behaviour is desirable.

Therefore, for broadband switching, instead of uniform metallic nanodiscs, we start by analyzing a periodic array of two sets of nanodisc-dimers (NDD-D_1-D_2) with progressively increasing diameters D_1 (nm) and D_2 (nm) placed on the VO_2-thin film on top of the gold coated SiO_2 substrate. Each unit cell of this periodic structure comprises of two sets of nanodisc-dimers where the diameters of the first and second set of nanodisc-dimers are 60 nm and 80 nm, respectively, as shown in the inset of Fig. 3(a). The proposed switch based on the periodic array formed by these unit cells is called NDD-60-80 in the rest of the paper for convenience. Fig. 3(a) shows the reflectance spectra for this proposed switch with the VO_2 layer in its semiconducting and metallic state. It can be observed from Fig. 3(a) that for this case — NDD-60-80 — when the VO_2 layer is in its semiconducting phase, three dips are present in the reflectance spectra at wavelength values of 720 nm (M1), 1250 nm (M2) and 1500 nm (M3). The dips at 1250 nm and 1500 nm are due to the wavelengths in the incident radiation that get coupled into the plasmonic modes arising out of the geometrical resonances of this composite nanostructure, NDD-60-80, with one metallic nanodisc dimer of a specific diameter corresponding to a one specific plasmon resonance wavelength. In this case, the wavelength of 1250 nm corresponds to the plasmon resonance wavelength of the case when uniform nanodiscs with a diameter of 60 nm are present alone (See Fig. 2(a)). Similarly, the wavelength of 1500 nm can be mapped to the plasmon resonance wavelength of the case when only uniform nanodiscs with a diameter of 80 nm are present alone. The dip in the reflectance spectra at 720 nm occurs due to the coupling of incident radiation into the plasmonic waveguide mode of this composite system [54]. A detailed discussion on the behaviour of the plasmonic modes at 1250 nm and 1500 nm and that of the plasmonic waveguide mode at 720 nm can be found in Supplementary Note 3 (See supplementary material). In addition, the proposed NDD-60-80 configuration has also been simulated without the VO_2 layer (See Supplementary Material, Note 4) to show that no switching can be obtained without VO_2.

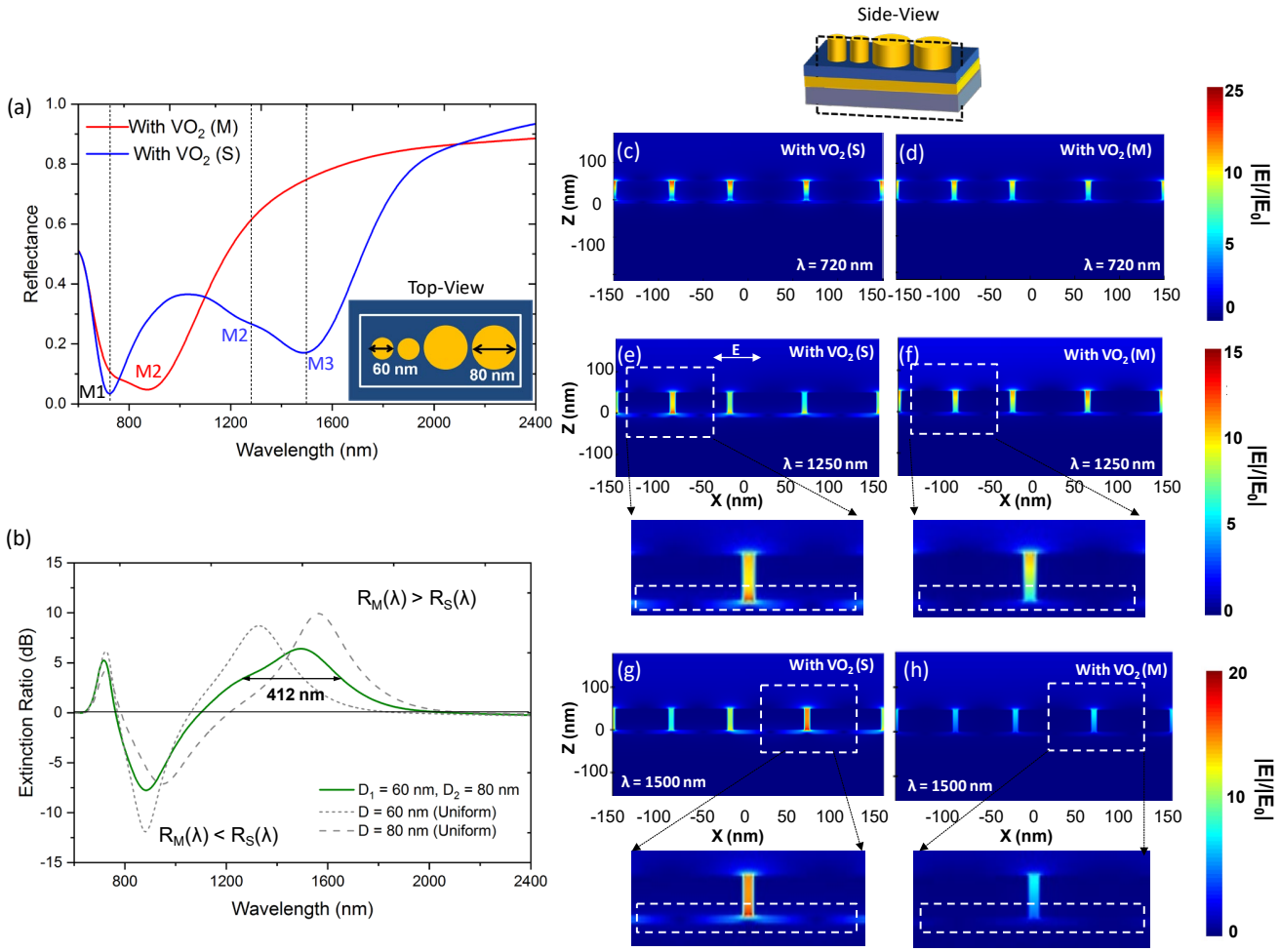


Fig. 3. (a) Reflectance spectra of the periodic array of two sets of nanodisc-dimers with the diameters being D_1 (60 nm) and D_2 (80 nm) with VO_2 is in its semiconductor phase and metallic phase. The inset shows one unit cell of the proposed periodic array. (b) Extinction ratio (ER) spectra for the proposed switch in comparison to the cases where periodic arrays of uniform nanodiscs with diameters 60 nm and 80 nm are present. E-field profile for the proposed switch at a wavelength of 720 nm with VO_2 in its (c) semiconductor phase, and (d) metallic phase. E-field profile for the proposed switch at a wavelength of 1250 nm with VO_2 in its (e) semiconductor phase, and (f) metallic phase. E-field profile for the proposed switch at a wavelength of 1500 nm with VO_2 in its (g) semiconductor phase, and (h) metallic phase. The thickness of VO_2, 't', the height of the nanodiscs, 'h', and the gap between the nanodiscs, 'g', were taken to be 10 nm, 50 nm, and 5 nm, respectively.

Fig. 3(a) also shows the reflectance spectra of the proposed NDD-60-80 switch with VO_2 in its metallic state. On being switched from semiconducting state to its metallic state, the real part of the refractive index of VO_2 decreases leading to a blue shift of the plasmonic dip in the reflectance spectra as shown in Fig. 3(a). The dip in the reflectance spectra due to the coupling of light into the plasmonic mode of the nanostructure occurs at 870 nm (M2). Further, the dip in the reflectance spectra due to the coupling of light into the plasmonic waveguide mode of the nanostructure can be seen as a shoulder around ~720 nm (M1).

Fig. 3(b) shows the extinction ratio (ER) spectrum for the switch based on the periodic array of two sets of nanodisc-dimers (NDD-60-80) on the VO_2 coated plasmonic substrate. Further, for comparison, the figure also shows the ER spectra for the periodic array of uniform nanodiscs with disc diameters of 60 nm

and 80 nm alone (discussed earlier in Fig. 2) on the VO_2 coated metallic substrate. It can be observed that the ER spectrum of the NDD-60-80 switch has peaks at the wavelengths of maximum ER for the switches with uniform nanodiscs of diameters 60 nm and 80 nm respectively, thus exhibiting an overall broadband behaviour of the ER spectrum. It can be seen from Fig. 3(b) that the proposed switch with two sets of nanodisc-dimers is capable of providing a broadband ER of 3.2 dB over an operating wavelength range of 412 nm spanning from 1250 nm to 1662 nm. This is significantly higher than the operating wavelength range of 280 nm which was achieved using periodic arrays of uniform nanodiscs of 60 nm or 80 nm alone. Moreover, the operating wavelengths for the proposed switch span the C, L, and U bands of optical communication and thus, it can be employed at any wavelength of choice over these bands.

Figs. 3(c)-(h) show the electric-field enhancement profiles for the proposed switch with periodic arrays of two sets of nanodisc dimers, with the semiconducting and metallic states of VO_2 at wavelengths of 720 nm, 1250 nm and 1500 nm. These wavelengths correspond to the dips in the reflectance spectra of the proposed switch with VO_2 in its semiconducting state. The E-field profile at 720 nm for the proposed switch with VO_2 in its semiconducting and metallic state is shown in Figs. 3(c)-(d). It can clearly be seen that at 720 nm, the field is confined equally in all the gaps between the adjacent nanodiscs as the plasmonic waveguide mode between adjacent nanodiscs depends only on the gap between the nanodiscs (i.e., the width of the insulator region of the resultant metal-insulator-metal waveguide). As the inter-nanodisc gap is uniformly 5 nm between all nanodiscs, the field is equally confined in all waveguides formed between adjacent nanodiscs, and is not dependent on the geometry of the nanodiscs. The electric-field profile of the composite structure in Fig. 3(e) (with VO_2 in the semiconducting state) shows that, at a wavelength of 1250 nm (M2), the incident energy is localized in the gap between the two nanodiscs with a diameter of 60 nm. It must be noted that the wavelength of 1250 nm corresponds to the plasmon resonance wavelength for the case when an array of uniform nanodiscs of diameter 60 nm were present alone on the VO_2-coated plasmonic substrate. For the case when VO_2 is in its metallic state (see Fig. 3(f)), the composite structure shows minimum localization of energy in the nanostructure at 1250 nm due to less plasmonic coupling of the incident radiation as compared to the semiconductor state. The transmission and reflectance spectra for the case of NDD-60-80 with VO_2 (M) is shown in Fig. S5 (Supplementary Material, Fig. S5). Additionally, the enlarged view of the E-field distribution at the boundary between VO_2 and the underlying gold substrate in Figs. 3(e)-(f) shows that while the field is present at the boundary between VO_2 and the underlying gold substrate when VO_2 is in its semiconducting state, the field vanishes as VO_2 transitions to its metallic phase. Further, at the resonance wavelength of the second plasmonic mode, i.e., at 1500 nm, the E-field profile of the composite structure in Fig. 3(g) (with VO_2 in the semiconducting state) shows the electromagnetic energy being localized between the nanodiscs with a diameter of 80 nm. As discussed earlier, Fig. 2(a) showed that a wavelength of 1500 nm is also the plasmon resonance wavelength for the case when an array of uniform nanodiscs of diameter 80 nm were present alone on the VO_2-coated plasmonic substrate. Fig. 3(h) shows that with VO_2 in its metallic state, only a minimal amount of incident radiation is coupled into the plasmonic modes of the switch - and hence incident radiation is mostly reflected from the nanostructure. Thus, it can be inferred that this composite periodic structure with two sets of nanodisc dimers (NDD-60-80) exhibits the combined plasmonic behaviour of two structures — a periodic array of uniform nanodiscs of diameter 60 nm, and another periodic array of uniform nanodiscs of diameter 80 nm on the VO_2-coated plasmonic substrate. A detailed discussion of the charge density profiles for the case of NDD-60-80 with VO_2 (S) and VO_2 (M) can be found in the supplementary section (See Supplementary Material, Note 5).

To increase the operating wavelength range of the proposed switches, a periodic array of three sets of nanodisc-dimers (NDD-60-80-100) with progressively increasing diameters (D_1 being 60 nm, D_2 being 80 nm and D_3 being 100 nm) placed on the VO_2-thin film on top of the gold coated SiO_2 substrate is analyzed in Fig. 4. It can be observed from the reflectance spectra shown in Fig. 4(a) that for this case with three sets of nanodisc dimers, when the VO_2 layer is in its semiconducting phase, four dips are present in the reflectance spectra at wavelength values of 720 nm (M1), 1250 nm (M2), 1500 nm (M3) and 1750 nm (M4), where the dip at 720 nm corresponds to the coupling of incident radiation into a plasmonic waveguide mode of the nanostructure. The plasmonic dips at 1250 nm, 1500 nm and 1750 nm closely correspond to the plasmon resonance dips of periodic arrays of uniform nanodiscs (UNDs) with diameters 60 nm, 80 nm and 100 nm on the VO_2-coated plasmonic substrate, respectively. It can be observed from the extinction ratio (ER) spectra shown in Fig. 4(b) that the proposed switch with three sets of nanodisc-dimers has peaks at three wavelengths, each corresponding to one of the peaks of the ER spectra for the three cases when uniform nanodiscs with diameter 60 nm, 80 nm and 100 nm are present on the VO_2-coated substrate (See Fig. 2). The overall ER of the NDD-60-80-100 switch shows a broadband behaviour and is capable of providing a broadband ER of 2.3 dB over an operating wavelength range of 604 nm spanning from 1304 nm to 1908 nm. This is significantly higher than the operating wavelength range of 285 nm which could be achieved using periodic arrays of uniform nanodiscs of 60 nm, 80 nm or 100 nm alone.

Figs. 4(c)-(j) show the electric-field enhancement profiles for the proposed switch with periodic arrays of three sets of nanodisc dimers, with the semiconducting and metallic states of VO_2 at wavelengths of 720 nm, 1250 nm, 1500 nm and 1750 nm. The E-field profiles corresponding to the plasmonic waveguide mode at the wavelength of 720 nm for the cases when VO_2 is present in its semiconductor and metallic state are shown in Figs. 4(c)-(d). It can be observed from Fig. 4(e), Fig. 4(g) and Fig. 4(i) that with VO_2 in the semiconducting state, the incident energy is localized in the gap between the two nanodiscs with a diameter of 60 nm, 80 nm, and 100 nm at a wavelength of 1250 nm, 1500 nm and 1750 nm, respectively. It must be noted that the wavelengths of 1250 nm, 1500 nm and 1750 nm correspond to the plasmon resonance wavelengths for the cases when an array of uniform nanodiscs of diameter 60 nm, 80 nm and 100 nm, respectively, were present alone on the VO_2-coated plasmonic substrate. Further, for the case when VO_2 is in its metallic state, Fig. 4(f), Fig. 4(h) and Fig. 4(j) show minimum localization of energy in the nanostructure at wavelengths of 1250 nm, 1500 nm and 1750 nm, respectively due to less plasmonic coupling of the incident radiation as compared to the semiconductor state. Thus, it can be inferred that this composite periodic structure with three sets of nanodisc dimers (NDD-60-80-100) exhibits the combined plasmonic behaviour of three structures — a periodic array of uniform nanodiscs of diameter 60 nm, periodic array of uniform nanodiscs of diameter 80 nm and another periodic array of uniform nanodiscs of diameter 100 nm on the VO_2-coated plasmonic substrate.

Fig. 5 shows the reflectance spectra and the ER spectra for a periodic array of a set of four nanodisc-dimers (NDD-60-80-100-120) on a VO_2 coated substrate such that the values of the diameters for the sets of nanodisc dimers are 60 nm, 80 nm, 100 nm and 120 nm, respectively (See schematic in Fig. 5(a)). Figs. 5(b)-(c) show the effect of varying the thickness of VO_2 layer on the reflectance spectra when VO_2 is in its semiconducting and metallic state, respectively. It can be noticed from Fig. 5(b) that for the case when VO_2 is in its semiconductor state and when the thickness of the VO_2 layer is less than 25 nm, there is a blue shift in the dips of the reflectance spectra when the VO_2 thickness is increased from 10 nm to 20 nm. This is due to the coupling of the incident radiation into the plasmonic waveguide mode in the VO_2 spacer layer

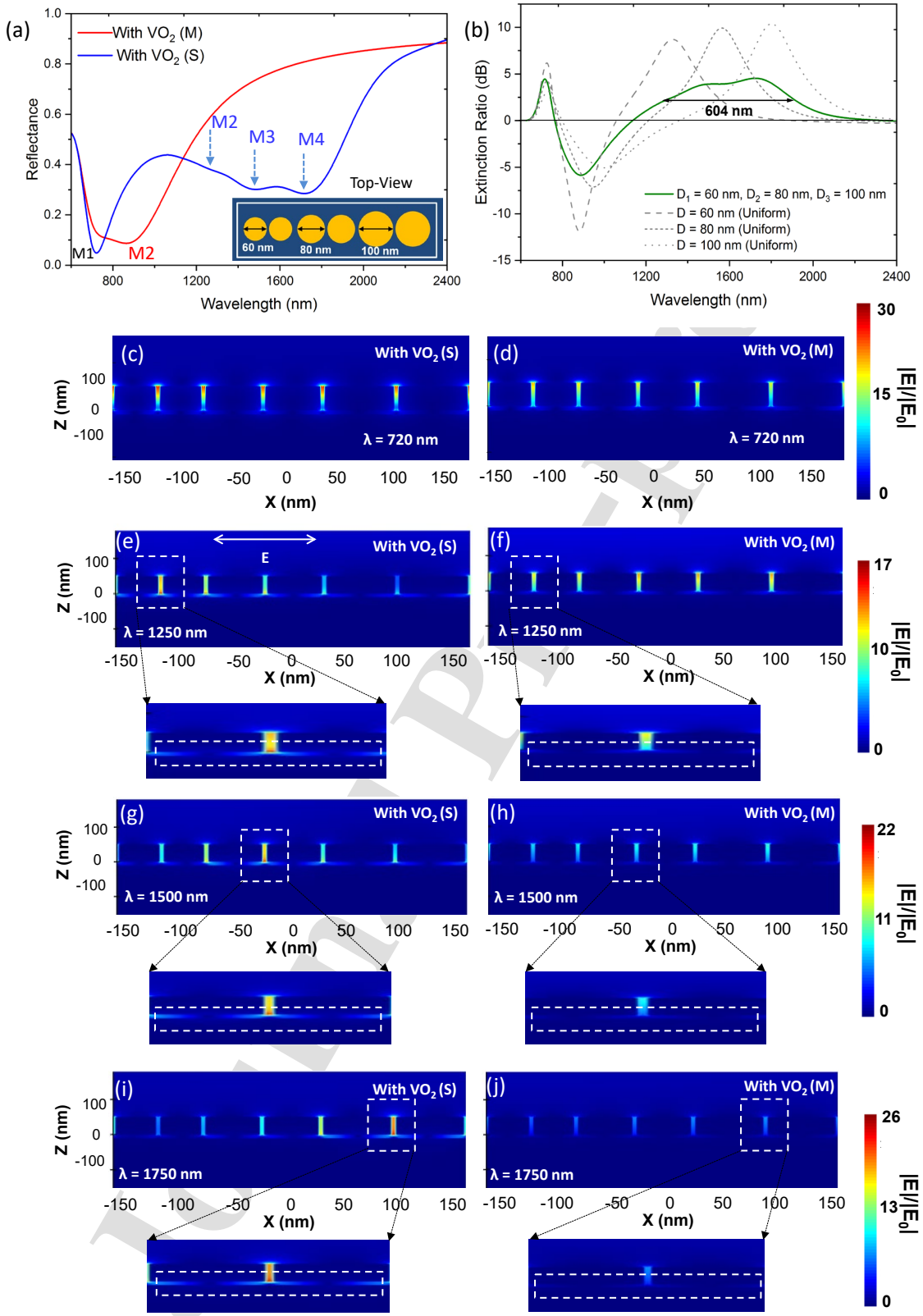


Fig. 4. (a) Reflectance spectra for the periodic array of three sets of nanodisc-dimers with the diameters being D_1 (60 nm), D_2 (80 nm) and D_3 (100 nm) when VO_2 is present in its semiconductor phase and metallic phase. The inset shows one unit cell of the proposed periodic array. (b) Extinction ratio (ER) spectra for the proposed switch in comparison to the cases when periodic arrays

of uniform nanodiscs with diameters 60 nm, 80 nm and 100 nm are present. E-field profile for the proposed switch at a wavelength of 720 nm with VO_2 in its (c) semiconductor phase, and (d) metallic phase. E-field profile for the proposed switch at a wavelength of 1250 nm with VO_2 in its (e) semiconductor phase, and (f) metallic phase. E-field profile for the proposed switch at a wavelength of 1500 nm with VO_2 in its (g) semiconductor phase, and (h) metallic phase. E-field profile for the proposed switch at a wavelength of 1750 nm with VO_2 in its (i) semiconductor phase, and (j) metallic phase. The thickness of VO_2, 't', the height of the nanodiscs, 'h', and the gap between the nanodiscs, 'g', were taken to be 10 nm, 50 nm, and 5 nm, respectively.

(between the gold nanodiscs and the underlying gold film) for sub-20 nm values of spacer thickness. Thus, the plasmon resonance wavelength blue-shifts with increasing VO_2 spacer thickness as expected [55]. However, as the thickness of VO_2 layer is further increased, this plasmonic waveguide mode becomes weak and the plasmonic characteristics of the nanostructure can be attributed to the geometrical resonance arising out of the nanodiscs on the VO_2-coated plasmonic film. When the thickness of VO_2 layer is increased beyond 20 nm, the plasmonic dips of the nanostructure exhibit a red shift with increasing thickness. This can be understood by applying the effective medium theory for the underlying substrate, i.e., VO_2 (S) on top of gold, whose effective permittivity is expected to increase with an increase in the thickness of VO_2 layer (See Supplementary Material, Note 2 for dielectric constants of VO_2 and gold). This leads to a red-shift in the plasmon resonance wavelength with an increase in the thickness of VO_2 layer. For the case when VO_2 is in its metallic state, the reflectance spectra as a function of the thickness of the VO_2 layer are shown in Fig. 5(c). It can be observed that as the thickness of the VO_2 layer is increased, there is a red-shift in the plasmon resonance wavelength. This can also be understood by applying effective medium theory to the underlying substrate, i.e., VO_2 (M) on top of gold. It is known that, for values of wavelength below 1130 nm, the permittivity of VO_2 (M) is positive. This results in an increase in the permittivity of the effective substrate as the thickness of VO_2 layer is increased, thus leading to a red-shift in the plasmon resonance wavelength. The electric-field enhancement profiles for the proposed switch with periodic arrays of four sets of nanodisc-dimers (NDD-60-80-100-120) on 10 nm thick VO_2 layer, with the semiconducting and metallic states of VO_2 at wavelengths of 1250 nm, 1500 nm, 1750 nm and 2050 nm (corresponding to the dips in the reflectance spectra of the proposed switch with VO_2 in its semiconducting state) are shown in Fig. S7 (Supplementary Material, Fig. S7).

Figs. 5(d)-(i) show the effect of varying the VO_2 thickness from 10 nm to 80 nm on the extinction ratio spectra for a periodic array of four sets of nanodisc dimers on a VO_2-coated substrate. The operating wavelength range for the switch is shown at the full width at half maxima (FWHM) of the ER for each value of thickness (shown in red colour). The range of operating wavelengths over which switching can be done is also shown with the values of the ER being 3 dB (blue), 4 dB (green) and 5 dB (orange) for comparison with other configurations of the switch and switches proposed in the past. The maximum ER, the operating wavelength range at the FWHM of the ER, and the operating wavelength ranges for the extinction ratio of 3 dB, 4 dB and 5 dB are tabulated in Table 1. The table shows that the values of ER — and the range of wavelengths over which the switch can be efficiently employed — can be optimized by varying the thickness of VO_2 layer. These switches can be employed to obtain a maximum ER of 25 dB for a 50 nm thick layer of VO_2. The table shows that an operating wavelength range of ~730 nm with an ER of 3 dB and an operating wavelength range of ~ 640 nm with an ER of 4 dB can be achieved by employing an 80 nm thick layer of VO_2. Further, an operating wavelength range of ~ 574 nm can be achieved with an ER of 5 dB by employing a 50 nm thick VO_2 layer under the periodic array of four sets of nanodisc-dimers (NDD-60-80-100-120). It can be seen from Table 1 that the proposed switch with four sets of nanodisc-dimers can be

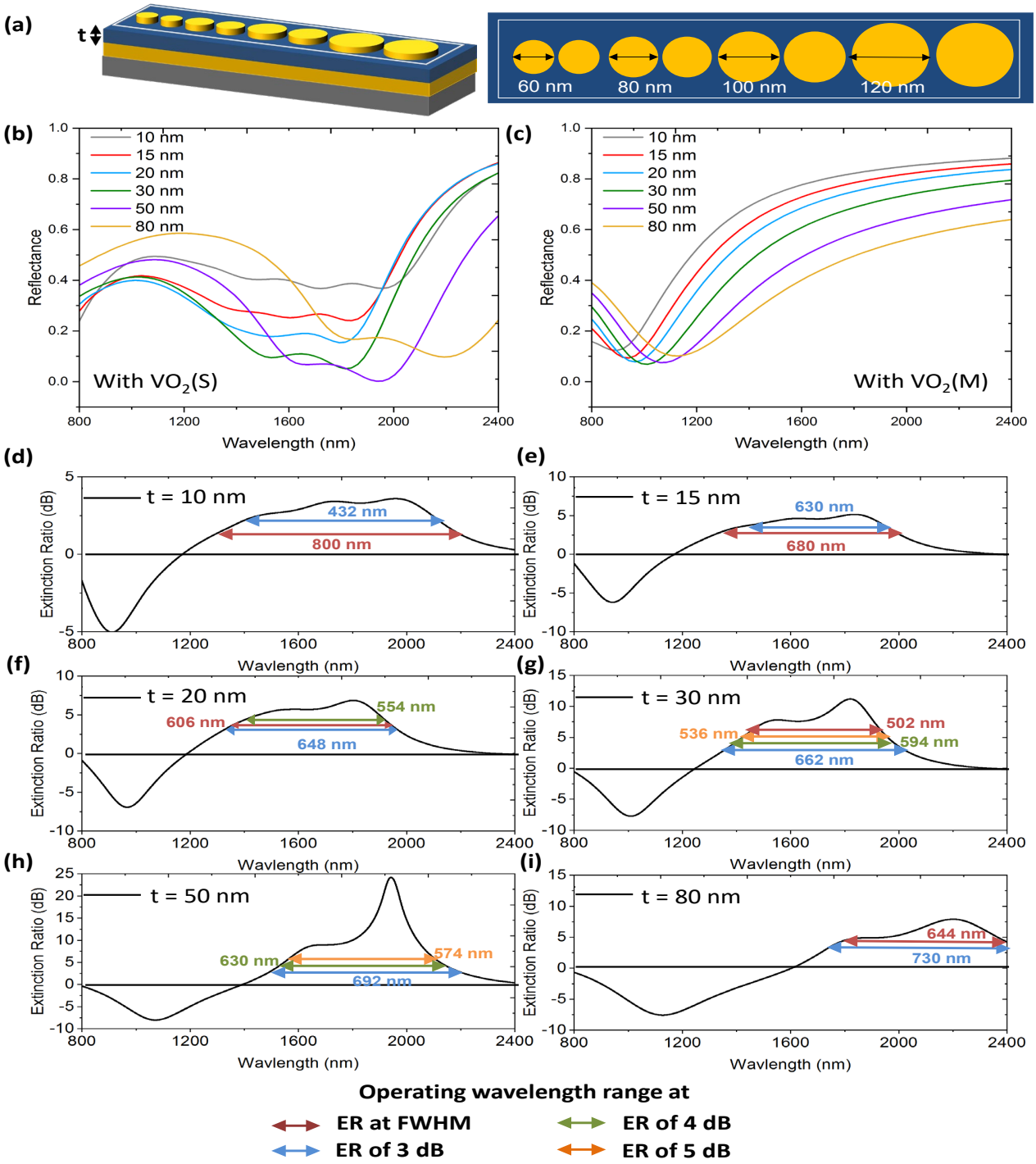


Fig. 5. (a) Schematic showing the periodic array of four sets of nanodisc-dimers on a VO_2 coated plasmonic film with the diameters of the nanodisc-dimers being D_1 (60 nm), D_2 (80 nm), D_3 (100 nm) and D_4 (120 nm). Reflectance spectra of the periodic array of four sets of nanodisc-dimers showing the effect of increasing the thickness of VO_2, 't', from 10 nm to 80 nm when VO_2 is in its (b) semiconductor phase and (c) metallic phase. Extinction ratio (ER) spectra showing the operating wavelength range for an ER of 3 dB, 4 dB, 5 dB and for the FWHM of the maximum ER when the thickness of VO_2, 't', is (d) 10 nm (e) 15 nm (f) 20 nm (g) 30 nm (h) 50 nm and (i) 80 nm. The height of the nanodiscs, 'h', and the gap between the nanodiscs, 'g', were kept constant at 50 nm, and 5 nm, respectively, for all the above cases.

employed for broadband switching in the C, L and U bands of optical communication. It should also be noted that, in this work, periodic arrays of nanodisc dimers with different diameters were chosen over arrays of single nanodiscs of variable diameters as their ER spectra exhibited a wider operating wavelength range, making them more suitable for broadband switching. The ER spectra for arrays of single nanodiscs and nanodisc-dimers of varying diameters are shown for comparison in Fig. S8 (Supplementary Material, Fig. S8). Further, a detailed analysis for the effect of reduction of diameters of the nanodisc dimers on the reflectance spectra and the extinction ratio spectra can be found in the supplementary section (See Supplementary Material, Note 6).

Table 1. Operating wavelength range and the extinction ratio for a periodic array of four sets of nanodisc-dimers with progressively increasing diameters (NDD-60-80-100-120) on a gold coated SiO_2 substrate with a VO_2 spacer when the thickness of the VO_2 spacer is varied from 10 nm to 80 nm.

Thickness of VO_2 Layer	**10 nm**	**15 nm**	**20 nm**	**30 nm**	**50 nm**	**80 nm**
Maximum ER	3.6 dB	5.2 dB	6.9 dB	11.3 dB	25 dB	7.94 dB
FWHM	800 nm (1350-2150 nm)	680 nm (1320-2000 nm)	606 nm (1345-1951 nm)	502 nm (1442-1944 nm)	-	644 nm (1768-2412 nm)
Operating wavelength range at **ER = 3 dB**	432 nm (1622-2054 nm)	630 nm (1350-1980 nm)	648 nm (1320-1968 nm)	662 nm (1356-2018 nm)	692 nm (1498-2190 nm)	730 nm (1730-2460 nm)
Operating wavelength range at **ER =4 dB**	-	442 nm (1490-1932 nm)	554 nm (1376-1930 nm)	594 nm (1390-1984 nm)	630 nm (1522-2152 nm)	640 nm (1770-2410 nm)
Operating wavelength range at **ER = 5 dB**	-	-	448 nm (1450-1898 nm)	536 nm (1422-1958 nm)	574 nm (1546-2120 nm)	404 nm (1960-2364 nm)

Fig. 6(a) shows the schematic for another configuration of the proposed switch where a periodic array of five sets of nanodisc dimers (NDD-40-60-80-100-120) are placed on the VO_2 coated substrate to allow additional wavelengths to be coupled into the plasmonic modes of the structure, thereby allowing the operating wavelength range to increase further. Figs. 6(b)-(g) show the effect of varying the thickness of the underlying VO_2 layer on the ER spectra for this configuration of the switch. The maximum ER, the operating wavelength range for the FWHM of the ER, and the operating wavelength range at the extinction ratio of 3 dB, 4 dB and 5 dB for this case are tabulated in Table 2. When a 50 nm thick VO_2 layer is employed, the proposed configuration of the switch is capable of offering an extinction ratio of 5 dB with a broadband operating wavelength range of ~ 650 nm over a wavelength range of 1460 nm to 2110 nm spanning the C, L and U bands of optical communication. The effect of varying the thickness of VO_2 layer

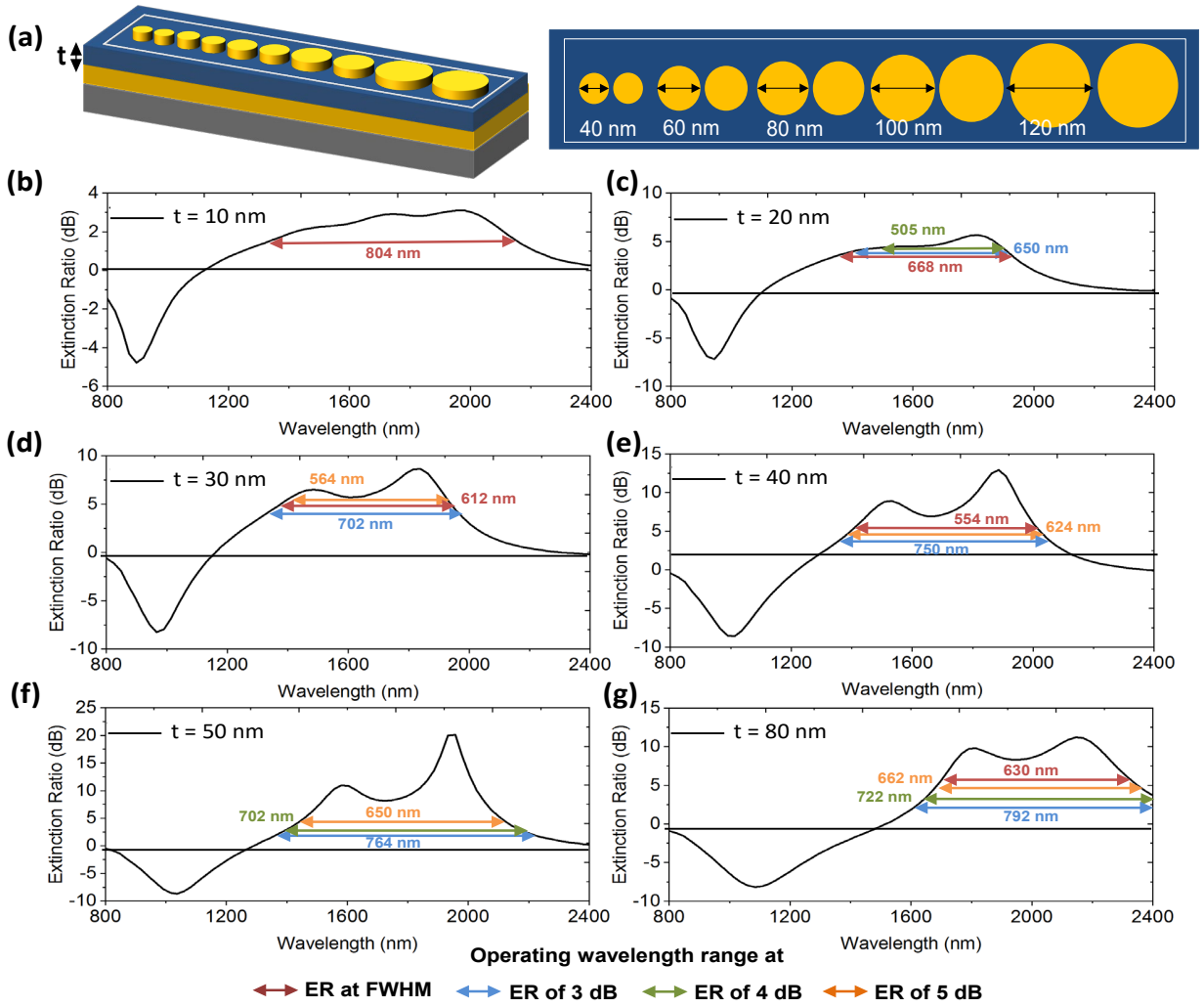


Fig. 6. (a) Schematic showing the periodic array of five sets of nanodisc-dimers on a VO_2-coated plasmonic film with the diameters of the nanodisc-dimers being D_1 (40 nm), D_2 (60 nm), D_3 (80 nm), D_4 (100 nm) and D_5 (120 nm). Extinction ratio (ER) spectra showing the operating wavelength ranges for an ER of 3 dB, 4 dB, 5 dB and at the FWHM of the ER when the thickness of VO_2, 't', is (b) 10 nm (c) 20 nm (d) 30 nm (e) 40 nm (f) 50 nm and (g) 80 nm. The height of the nanodiscs, 'h', and the gap between the nanodiscs, 'g', were kept constant at 50 nm, and 5 nm, respectively, for all the above cases.

on the reflectance spectra of the NDD-40-60-80-100-120 switch is shown in Supplementary Figs. S11(a)-(b) for the semiconducting and metallic state of VO_2, respectively (See Supplementary Material, Fig. S11).

As discussed earlier, the proposed switches can be fabricated with the state-of-the-art fabrication processes like E-beam lithography or Focused Ion beam milling (See Supplementary Material, Note 1 for proposed fabrication methodology). However, the fabrication of sub-10 nm gaps between the gold nanodisc-dimers can be challenging even by using these sophisticated fabrication methodologies. Therefore, an analysis of the effect of the inter nanodisc gap on the extinction ratio spectra is also carried out (See schematic in Fig. 7(a)). Fig. 7(b) and Fig. 7(c) show the effect of varying the gap between the nanodiscs (g) for a periodic array of a set of four nanodisc-dimers (NDD-60-80-100-120) with 10 nm and 50 nm thick layers of VO_2, respectively.

Table 2. Operating wavelength range and the extinction ratio for a periodic array of five sets of nanodisc-dimers with progressively increasing diameters (NDD-40-60-80-100-120) on a gold coated SiO_2 substrate with a VO_2 spacer when the thickness of the VO_2 spacer is varied from 10 nm to 80 nm.

Thickness of VO_2 Layer	**10 nm**	**20 nm**	**30 nm**	**40 nm**	**50 nm**	**80 nm**
Maximum ER	3.15 dB	5.68 dB	8.66 dB	12.92 dB	20.2 dB	11.2 dB
FWHM	804 nm (1346-2150 nm)	668 nm (1288-1956 nm)	612 nm (1344-1956 nm)	554 nm (1438-1992 nm)	-	630 nm (1700-2330 nm)
Operating wavelength range at **ER = 3 dB**	-	650 nm (1300-1950 nm)	720 nm (1280-2000 nm)	750 nm (1330-2080 nm)	764 nm (1402-2166 nm)	792 nm (1638-2430 nm)
Operating wavelength range at **ER =4 dB**	-	505 nm (1404-1909 nm)	638 nm (1328-1966 nm)	680 nm (1366-2046 nm)	702 nm (1432-2134 nm)	722 nm (1664-2386 nm)
Operating wavelength range at **ER = 5 dB**	-	-	564 nm (1374-1938 nm)	624 nm (1397-2021 nm)	650 nm (1460-2110 nm)	662 nm (1686-2348 nm)

Fig. 7(b) shows the extinction ratio (ER) spectra for a periodic array of a set of four nanodisc-dimers (NDD-60-80-100-120) with the thickness of VO_2 layer being 10 nm. It can be observed that for a gap of 5 nm between the nanodiscs, and the operating wavelength range at the FWHM of the ER is ~ 800 nm over a wavelength span of 1350 nm to 2150 nm. As the gap between the nanodiscs is increased from 5 nm to 20 nm, a slight drop in the extinction ratio can be observed from Fig. 7(b) along with a blue-shift in the plasmon resonance wavelengths as expected [55]. However, the operating wavelength range remains sufficiently broad band (~ 828 nm) even for higher inter-nanodisc gaps such as 20 nm which can be easily fabricated with the state-of-the-art fabrication processes. The reflectance spectra for this switch for the various values of inter-nanodisc gaps, with VO_2 in its semiconducting and metallic state are given in Supplementary Material, Fig. S12.

Fig. 7(c) shows the extinction ratio (ER) spectra for the case when a periodic array of a set of four nanodisc-dimers (NDD-60-80-100-120) are placed on a 50 nm thick VO_2 layer on top of a SiO_2 substrate. It can be seen from Fig. 7(c) that an extinction ratio of 4 dB can be achieved over an operating wavelength range of ~ 630 nm for the gap being 5 nm. As the gap is increased to 10 nm, 15 nm and 20 nm, the operating wavelength ranges for an extinction ratio of 4 dB become ~ 680 nm, ~ 684 nm and ~ 620 nm, respectively. This shows that the proposed switch can be employed even when the inter-nanodisc gaps are higher, which makes these switches more amenable to fabrication. Moreover, the wavelength range over which the switching can be obtained exhibits a blue-shift as the gap between the nanodiscs is increased [55]. Hence, the gap between the nanodiscs can also be employed to tune the broad switching band over the wavelength of choice. Moreover, the geometrical parameters of these switches can be further optimized to enable the usage of these switches for the visible range of the spectrum.

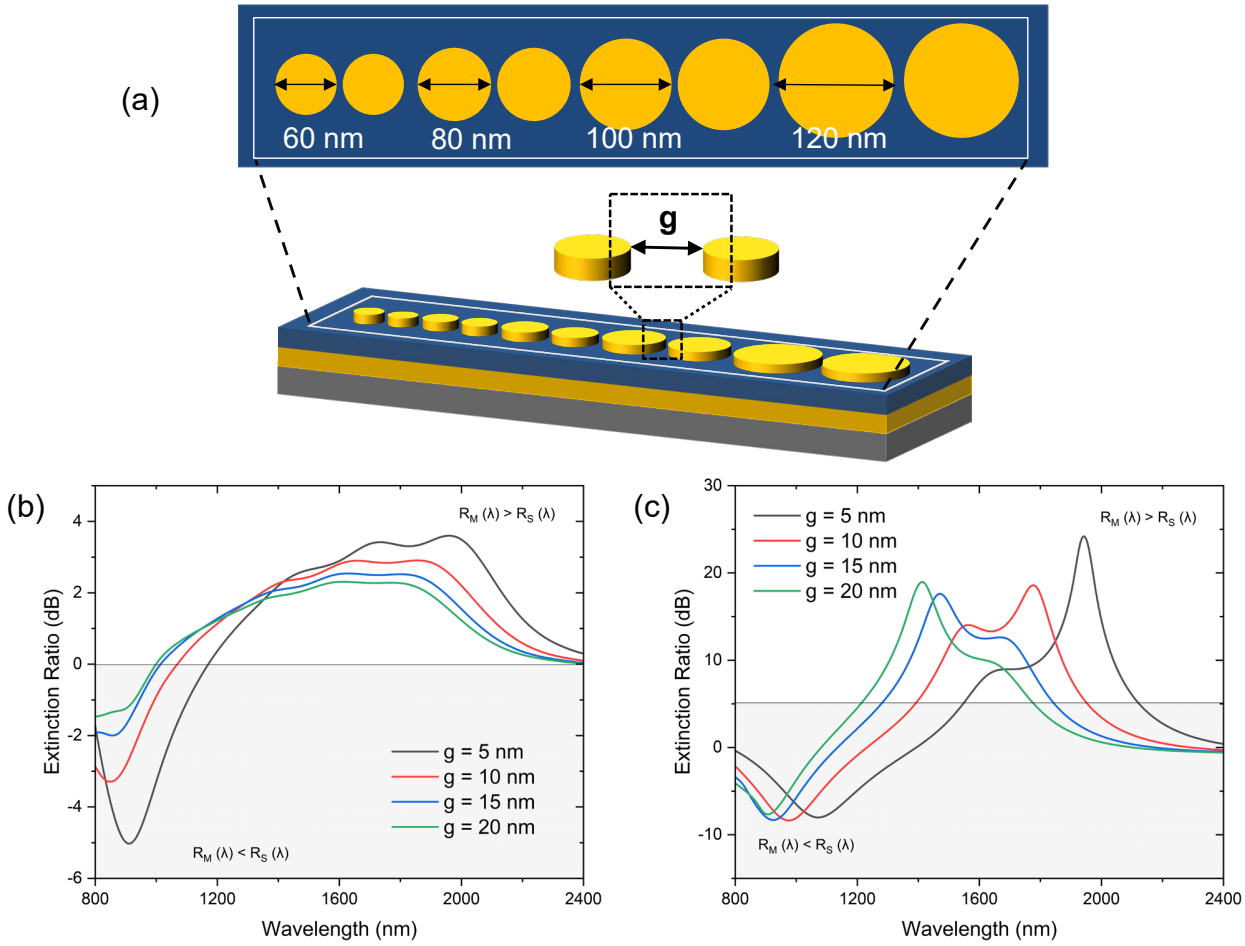


Fig. 7 (a) Schematic showing the periodic array of four sets of nanodisc-dimers on a VO_2-coated plasmonic film with the diameters of the nanodisc-dimers being D_1 (60 nm), D_2 (80 nm), D_3 (100 nm) and D_4 (120 nm). Extinction ratio (ER) spectra for the proposed switch showing the effect of varying the gap, 'g', between the nanodiscs from 5 nm to 20 nm, when the thickness of VO_2, 't', is (b) 10 nm, and (c) 50 nm. The height of the nanodiscs, 'h', is kept constant at 50 nm for both the cases.

4. CONCLUSIONS

We have proposed active broadband plasmonic switches based on a periodic array of dimers of metallic nanodiscs with progressively increasing diameters on a gold coated SiO_2 substrate with a VO_2 thin film as a spacer between the nanodiscs and the underlying plasmonic substrate. Broadband switching is demonstrated over the C, L, and U bands of optical communication with an operating wavelength range of ~ 650 nm for an extinction ratio of 5 dB, and an operating wavelength range of ~ 700 nm for an extinction ratio of 4 dB. These switches exhibit a trade-off between the extinction ratio and the operating wavelength range making them amenable to be used as narrowband switches with high extinction ratio or broadband switches with a moderate extinction ratio. Moreover, optimization of the geometrical parameters of these switches can lead to tunability of the spectral range over which broadband operation can be obtained. As these switches are based on VO_2 which can potentially exhibit ultra-fast switching, these switches are

Journal

Journal Pre-proof

capable of being employed for high-speed optical communication networks or as switching components in a plasmonic circuits.

ACKNOWLEDGEMENT

The authors wish to express their gratitude towards the Delhi Technological University (DTU) for their support. Further, the authors would also like to thank the Nanophotonics and Plasmonics Laboratory (NPPL) of the Indian Institute of Technology Delhi (IITD) for their support and resources.

DECLARATION OF COMPETING INTEREST

The authors declare no competing interests.

REFERENCES

[1] Mehrdad Djavid, Mohammad Hadi Tavakoli Dastjerdi, Moab Rajan Philip, Dipayan Datta Choudhary, Thi Tan Pham, Abdallah Khreishah, and Hieu Pham Trung Nguyen. "Photonic crystal-based permutation switch for optical networks", *Photonic Network Communications*, volume 35, pages 90-96 (2018). DOI: https://doi.org/10.1007/s11107-017-0719-7

[2] Dalai Gowri Sankar Rao, Sandip Swarnakar, Venkatrao Palacharla, Karyabhattu Seeta Rama Raju, and Santosh Kumar. "Design of all optical AND, OR, and XOR logic gates using photonic crystals for switching applications", *Photonic Network Communications*, volume 41, pages 109-118 (2021). DOI: https://doi.org/10.1007/s11107-020-00916-6

[3] Soham Saha, Benjamin T. Diroll, Joshua Shank, Zhaxylyk Kudyshev, Aveek Dutta, Sarah Nahar Chowdhury, Ting Shan Luk, Salvatore Campione, Richard D. Schaller, Vladimir M. Shalaev, Alexandra Boltasseva, and Michael G. Wood. "Broadband, High-Speed, and Large-Amplitude Dynamic Optical Switching with Yttrium-Doped Cadmium Oxide", *Advanced Functional Materials*, 30, 1908377 (2020). DOI: https://doi.org/10.1002/adfm.201908377

[4] Burak Gerislioglu, and Arash Ahmadivand. "Advances in Plasmonics and Nanophotonics", *Nanomaterials*, 11(11), 3159 (2021). DOI: https://doi.org/10.3390/nano11113159

[5] Yousef Karimi, Hassan Kaatuzian, Alireza Tooghi, and Mohammad Danaie. "All-optical plasmonic switches based on Fano resonance in an X-shaped resonator coupled to parallel stubs for telecommunication applications", *Optik*, volume 243, 167424 (2021). DOI: https://doi.org/10.1016/j.ijleo.2021.167424

[6] Stefan A. Maier. *Plasmonics: Fundamentals and Applications*. Springer, New York, USA (2007). ISBN: 9780387331508. DOI: https://doi.org/10.1007/0-387-37825-1

[7] Eleonora Petryayeva, and Ulrich J. Krull. "Localized surface plasmon resonance: Nanostructures, bioassays and biosensing-A review", *Analytica Chimica Acta*, volume 706, issue 1, pages 8-24 (2011). DOI: https://doi.org/10.1016/j.aca.2011.08.020

[8] Xinxiang Niu, Xiaoyong Hu, Qiuchen Yan, Jiankun Zhu, Haotian Cheng, Yifan Huang, Cuicui Lu, Yulan Fu, and Qihuang Gong. "Plasmon-induced transparency effect for ultracompact on-chip devices", *Nanophotonics*, vol. 8, no. 7, pp. 1125-1149 (2019). DOI: https://doi.org/10.1515/nanoph-2019-0093

[9] Nina Jiang, Xiaolu Zhuo, and Jianfang Wang. "Active Plasmonics: Principles, Structures, and Applications", *Chemical Reviews*, 118 (6), 3054-3099 (2018). DOI: https://doi.org/10.1021/acs.chemrev.7b00252

[10] Yadollah Shahamat, and Mohammad Vahedi. "Designing ultra-compact high efficiency electro-optical plasmonic switches by using of nanocavity reflectors", *Optics Communications*, volume 410, pages 25-29 (2018). DOI: https://doi.org/10.1016/j.optcom.2017.09.083

[11] Mohammad Taghinejad, and Wenshan Cai. "All-Optical Control of Light in Micro- and Nanophotonics", *ACS Photonics*, 6, 5, 1082–1093 (2019). DOI: https://doi.org/10.1021/acsphotonics.9b00013

[12] Ragip A. Pala, Ken T. Shimizu, Nicholas A. Melosh, and Mark L. Brongersma. "A Nonvolatile Plasmonic Switch Employing Photochromic Molecules", *Nano Letters*, 8, 5, 1506–1510 (2008). DOI: https://doi.org/10.1021/nl0808839

[13] Jacek Gosciniak, and Sergey I. Bozhevolnyi. "Performance of thermo-optic components based on dielectric-loaded surface plasmon polariton waveguides", *Scientific Reports*, volume 3, Article number: 1803 (2013). DOI: https://doi.org/10.1038/srep01803

[14] Wayne Dickson, Gregory A. Wurtz, Paul R. Evans, Robert J. Pollard, and Anatoly V. Zayats. "Electronically Controlled Surface Plasmon Dispersion and Optical Transmission through Metallic Hole Arrays Using Liquid Crystal", *Nano Letters*, 8, 1, 281–286 (2008). DOI: https://doi.org/10.1021/nl072613g

[15] Yu Yao, Mikhail A. Kats, Patrice Genevet, Nanfang Yu, Yi Song, Jing Kong, and Federico Capasso. "Broad Electrical Tuning of Graphene-Loaded Plasmonic Antennas", *Nano Letters*, 13, 3, 1257–1264 (2013). DOI: https://doi.org/10.1021/nl3047943

[16] Yann R. Leroux, Jean Christophe Lacroix, Kathleen I. Chane-Ching, Claire Fave, Nordin Félidj, Georges Lévi, Jean Aubard, Joachim R. Krenn, and Andreas Hohenau. "Conducting Polymer Electrochemical Switching as an Easy Means for Designing Active Plasmonic Devices", *Journal of the American Chemical Society*, 127, 46, 16022–16023 (2005). DOI: https://doi.org/10.1021/ja054915v

[17] Praveen Thoniyot, Mein Jin Tan, Anis Abdul Karim, David James Young, and Xian Jun Loh. "Nanoparticle–Hydrogel Composites: Concept, Design, and Applications of These Promising, Multi-Functional Materials", *Advanced Science*, 2: 1400010 (2015). DOI: https://doi.org/10.1002/advs.201400010

[18] G. A. Wurtz, R. Pollard, W. Hendren, G. P. Wiederrecht, D. J. Gosztola, V. A. Podolskiy, and A. V. Zayats. "Designed ultrafast optical nonlinearity in a plasmonic nanorod metamaterial enhanced by nonlocality", *Nature Nanotechnology*, volume 6, pages107–111 (2011). DOI: https://doi.org/10.1038/nnano.2010.278

[19] Simone Raoux, and Matthias Wuttig. *Phase Change Materials*. Springer, USA (2009). ISBN: 9780387848730. DOI: https://doi.org/10.1007/978-0-387-84874-7

[20] T. Ben-Messaoud, G. Landry, J. P. Gariépy, B. Ramamoorthy, P. V. Ashrit, and A. Haché. "High contrast optical switching in vanadium dioxide thin films", *Optics Communications*, volume 281, issue 24, pages 6024-6027 (2008). DOI: https://doi.org/10.1016/j.optcom.2008.09.027

[21] Seung-Yeol Lee. "Design of a plasmonic switch using ultrathin chalcogenide phase-change material", *Current Optics and Photonics,* vol. 1, issue 3, pp. 239-246 (2017). DOI: https://doi.org/10.3807/COPP.2017.1.3.239

[22] Weifeng Jiang. "Nonvolatile and ultra-low-loss reconfigurable mode (De) multiplexer/switch using triple-waveguide coupler with $Ge_2Sb_2Se_4Te_1$ phase change material", *Scientific Reports*, volume 8, Article number: 15946 (2018). DOI: https://doi.org/10.1038/s41598-018-34419-2

[23] Zhai FengXiao, Yang Kun, Wang Donglin, Liu Sujuan, Liu Nanan, Hao Yunqi, and Ren Yufen. "Bipolar resistive switching effect in laser-induced crystallization AgInSbTe films", *Optik*, volume 180, pages 271-275 (2019). DOI: https://doi.org/10.1016/j.ijleo.2018.11.100

[24] Brett A. Kruger, Arash Joushaghani, and Joyce K. S. Poon. "Design of electrically driven hybrid vanadium dioxide (VO_2) plasmonic switches", *Optics Express*, vol. 20, issue 21, pp. 23598-23609 (2012). DOI: https://doi.org/10.1364/OE.20.023598

[25] Nardeep Kumar, Armando Rúa, Jennifer Aldama, Karla Echeverría, Félix E. Fernández, and Sergiy Lysenko. "Photoinduced surface plasmon switching at VO_2/Au interface", *Optics Express*, vol. 26, issue 11, pp. 13773-13782 (2018). DOI: https://doi.org/10.1364/OE.26.013773

[26] Saranya Bhupathi, Shancheng Wang, Mohammad Abutoama, Igal Balin, Lei Wang, Peter G. Kazansky, Yi Long, and Ibrahim Abdulhalim. "Femtosecond laser-induced vanadium oxide metamaterial nanostructures and the study of optical response by experiments and numerical simulations", *ACS Applied Materials & Interfaces*, 12, 37, 41905-41918 (2020). DOI: https://doi.org/10.1021/acsami.0c03844

[27] Elizabeth E. Chain. "Optical properties of vanadium dioxide and vanadium pentoxide thin films", *Applied Optics*, vol. 30, issue 19, pp. 2782-2787 (1991). DOI: https://doi.org/10.1364/AO.30.002782

[28] G. V. Jorgenson, and J. C. Lee. "Doped vanadium oxide for optical switching films", *Solar Energy Materials*, volume 14, issues 3-5, pages 205-214 (1986). DOI: https://doi.org/10.1016/0165-1633(86)90047-X

[29] Kai Liu, Sangwook Lee, Shan Yang, Olivier Delaire, and Junqiao Wu. "Recent progresses on physics and applications of vanadium dioxide", *Materials Today*, volume 21, issue 8, pp. 875-896 (2018). DOI: https://doi.org/10.1016/j.mattod.2018.03.029

[30] Zewei Shao, Xun Cao, Hongjie Luo, and Ping Jin. "Recent progress in the phase-transition mechanism and modulation of vanadium dioxide materials", *NPG Asia Materials*, volume 10, pages 581-605 (2018). DOI: https://doi.org/10.1038/s41427-018-0061-2

[31] Yashna Sharma, Veeranjaneya A. Tiruveedhula, John F. Muth, and Anuj Dhawan. "VO_2 based waveguide-mode plasmonic nano-gratings for optical switching", *Optics Express*, vol. 23, issue 5, pp. 5822-5849 (2015). DOI: https://doi.org/10.1364/OE.23.005822

[32] Arash Joushaghani, Brett A. Kruger, Suzanne Paradis, David Alain, J. Stewart Aitchison, and Joyce K. S. Poon. (2013). "Sub-volt broadband hybrid plasmonic-vanadium dioxide switches", *Applied Physics Letters*, 102, 061101. DOI: https://doi.org/10.1063/1.4790834

[33] Miao Sun, Mohammad Taha, Sumeet Walia, Madhu Bhaskaran, Sharath Sriram, William Shieh, and Ranjith Rajasekharan Unnithan. (2018). "A photonic switch based on a hybrid combination of metallic nanoholes and phase-change vanadium dioxide", *Scientific Reports*, volume 8, Article number: 11106. DOI: https://doi.org/10.1038/s41598-018-29476-6

[34] P. Mandal, S. Mohan, S. Sharma, and M. S. Goyat. (2019). "Broadband multi-resonant circular dichroism in metal-VO_2 hybrid dagger-like plasmonic structure for switching application", *Photonics and Nanostructures - Fundamentals and Applications*, Volume 37, 100735. DOI: https://doi.org/10.1016/j.photonics.2019.100735

[35] Peijun Guo, Matthew S. Weimer, Jonathan D. Emery, Benjamin T. Diroll, Xinqi Chen, Adam S. Hock, Robert P. H. Chang, Alex B. F. Martinson, and Richard D. Schaller. "Conformal coating of a phase change material on ordered plasmonic nanorod arrays for broadband all-optical switching", *ACS Nano*, 11, 1, 693-701 (2016). DOI: https://doi.org/10.1021/acsnano.6b07042

[36] Jacob S. Schalch, Yaojia Chi, Yulian He, Yahua Tang, Xiaoguang Zhao, Xin Zhang, Qiye Wen, Richard D. Averitt. "Broadband electrically tunable VO_2-Metamaterial terahertz switch with suppressed reflection", *Microwave and Optical Technology Letters*, 62: 2782-2790 (2020). DOI: https://doi.org/10.1002/mop.32354

[37] Chunrui Han, Edward P. J. Parrott, Georges Humbert, Aurelian Crunteanu, and Emma Pickwell-MacPherson. "Broadband modulation of terahertz waves through electrically driven hybrid bowtie antenna-VO_2 devices", *Scientific Reports*, volume 7, Article number: 12725 (2017). DOI: https://doi.org/10.1038/s41598-017-13085-w

[38] Zhen Chai, Xiaoyong Hu, Feifan Wang, Xinxiang Niu, Jingya Xie, and Qihuang Gong. "Ultrafast All-Optical Switching", *Advanced Optical Materials*, 5, 1600665 (2017). DOI: https://doi.org/10.1002/adom.201600665

[39] Timothy J. Davis, Daniel E. Gómez, and Ann Roberts. "Plasmonic circuits for manipulating optical information", *Nanophotonics*, vol. 6, no. 3, pp. 543-559 (2016). DOI: https://doi.org/10.1515/nanoph-2016-0131

[40] K. Imura, K. Ueno, H. Misawa, H. Okamoto, D. McArthur, B. Hourahine, and F. Papoff. "Plasmon modes in single gold nanodiscs", *Optics Express*, vol. 22, issue 10, pp. 12189-12199 (2014). DOI: https://doi.org/10.1364/OE.22.012189

[41] Nor Ahya Hassan, and Abdul Manaf Hashim. "The effect of size of Au nanodisc on the localized surface plasmon resonance simulated using finite-difference time-domain (FDTD) method", *Materials Today: Proceedings*, volume 7, part 2, pages 607-611 (2019). DOI: https://doi.org/10.1016/j.matpr.2018.12.014

[42] Sébastien Cueff, Jimmy John, Zhen Zhang, Jorge Parra, Jianing Sun, Régis Orobtchouk, Shriram Ramanathan, and Pablo Sanchis. "VO_2 nanophotonics", *APL Photonics*, volume 5, issue 11, 110901 (2020). DOI: https://doi.org/10.1063/5.0028093

[43] Monan Liu, Xianglin Li, Siva Krishna Karuturi, Alfred Iing Yoong Tok, and Hong Jin Fan. "Atomic layer deposition for nanofabrication and interface engineering", *Nanoscale*, 4, 1522-1528 (2012). DOI: https://doi.org/10.1039/C2NR11875K

[44] Sujan Kasani, Peng Zheng, Joeseph Bright, and Nianqiang Wu. "Tunable Visible-Light Surface Plasmon Resonance of Molybdenum Oxide Thin Films Fabricated by E-beam Evaporation", *ACS Applied Electronic Materials*, 1, 11, 2389-2395 (2019). DOI: https://doi.org/10.1021/acsaelm.9b00555

[45] D. H. Kim, and H. S. Kwok. "Pulsed laser deposition of VO_2 thin film", *Applied Physics Letters*, volume 65, issue 25, 3188 (1994). DOI: https://doi.org/10.1063/1.112476

[46] Lucia Petti, Rossella Capasso, Massimo Rippa, Marianna Pannico, Pietro La Manna, Gianfranco Peluso, Anna Calarco, Eugenia Bobeico, and Pellegrino Musto. "A plasmonic nanostructure fabricated by electron beam lithography as a sensitive and highly homogeneous SERS substrate for bio-sensing applications", *Vibrational Spectroscopy*, volume 82, pages 22-30 (2016). DOI: https://doi.org/10.1016/j.vibspec.2015.11.007

[47] Shixuan He, Rong Tian, Wei Wu, Wen-Di Li, and Deqiang Wang. "Helium-ion-beam nanofabrication: extreme processes and applications", *International Journal of Extreme Manufacturing*, volume 3, number 1, 012001 (2020). DOI: https://doi.org/10.1088/2631-7990/abc673

[48] Yashna Sharma, and Anuj Dhawan. "Plasmonic "nano-fingers on nanowires" as SERS substrates", *Optics Letters*, vol. 41, issue 9, pp. 2085-2088 (2016). DOI: https://doi.org/10.1364/OL.41.002085

[49] Allen Taflove, and Susan C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Artech House, Boston, MA (2005). ISBN: 9781580538329.

[50] Hans W. Verleur, A. S. Barker, Jr., and C. N. Berglund. "Optical Properties of VO_2 between 0.25 and 5 eV", *Reviews of Modern Physics*, volume 40, issue 4, 737 (1968). DOI: https://doi.org/10.1103/RevModPhys.40.737

[51] Sotirios Papaioannou, Dimitrios Kalavrouziotis, Konstantinos Vyrsokinos, Jean-Claude Weeber, Karim Hassan, Laurent Markey, Alain Dereux, Ashwani Kumar, Sergey I. Bozhevolnyi, Matthias Baus, Tolga Tekin, Dimitrios Apostolopoulos, Hercules Avramopoulos, and Nikos Pleros. "Active plasmonics in WDM traffic switching applications", *Scientific Reports*, volume 2, Article number: 652 (2012). DOI: https://doi.org/10.1038/srep00652

[52] Stamatios V. Kartalopoulos. *Free Space Optical Networks for Ultra-Broad Band Services*. Wiley-IEEE Press (2011). ISBN: 978-0-470-64775-2.

[53] Yang Lu, Yizhen Wei, Miao Hu, Xuefang Zhou, Zhengfeng Qian, and Qiliang Li. "Triple functional shared channel in WDM PON by orthogonal modulation and network coding", *Optics Communications*, volume 336, pages 103-105 (2015). DOI: https://doi.org/10.1016/j.optcom.2014.09.070

[54] J. Britt Lassiter, Felicia McGuire, Jack J. Mock, Cristian Ciracì, Ryan T. Hill, Benjamin J. Wiley, Ashutosh Chilkoti, and David R. Smith. "Plasmonic Waveguide Modes of Film-Coupled Metallic Nanocubes", *Nano Letters*, 13, 12, 5866–5872 (2013). DOI: https://doi.org/10.1021/nl402660s

[55] Yu Huang, Lingwei Ma, Mengjing Hou, Jianghao Li, Zheng Xie, and Zhengjun Zhang. "Hybridized plasmon modes and near-field enhancement of metallic nanoparticle-dimer on a mirror", *Scientific Reports*, volume 6, Article number: 30011 (2016). DOI: https://doi.org/10.1038/srep30011

Journal Pre-proof

Journal Pre-proof

Broadband Plasmonic Switches based on Nanodisc-Dimers with Progressively Increasing Diameters on a Plasmonic Film with a VO_2 spacer

Kirti Dalal and Yashna Sharma*
Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India
*yashnasharma@dtu.ac.in

HIGHLIGHTS

- Plasmonic nanodisc dimers with progressively increasing diameters on VO_2 for active broadband plasmonic switching
- Broadband extinction ratio of 5 dB over an operating wavelength range of 650 nm
- Trade-off between the extinction ratio and the broadband operating wavelength range
- Tunability of the broadband spectral range by optimization of the geometrical parameters of the switch

Journal Pre-proof

Journal Pre-proof

Declaration of 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Alexandria Engineering Journal (2022) xxx, xxx–xxx

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

ORIGINAL ARTICLE

Compact circularly polarized 2 and 4 port multiple input multiple output antennas with bandstop filter isolation technique

Hashinur Islam [a], Saumya Das [b], Tanweer Ali [c,*], Tanushree Bose [a], Sourav Dhar [a], Bijay Rai [d], Pradeep Kumar [e]

[a] *Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim 737136, India*
[b] *Department of Information Technology, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim 737136, India*
[c] *Department of Electronics and Communication, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India*
[d] *Department of Electrical & Electronics Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim 737136, India*
[e] *Discipline of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, Durban 4041, South Africa*

Received 2 July 2022; revised 21 October 2022; accepted 18 November 2022

KEYWORDS

2 port MIMO antenna;
4 port MIMO antenna;
Bandstop filter (BSF) based decoupling network;
Circular polarization

Abstract Reduction of mutual coupling in the radiating elements of multiple input multiple output (MIMO) antenna is always an issue and a big challenge for antenna researchers. To address this issue of mutual coupling reduction, an effective method of decoupling has been proposed in this paper. Here, a bandstop filter based isolation technique has been introduced to limit the mutual coupling in each element of the MIMO antenna. A 2 port and a 4 port MIMO antenna have been designed with the proposed decoupling technique, and it is found that the decoupling technique successfully achieves isolation of more than 30 dB for both antennas. In addition to achieving high isolation, the designed MIMO antennas are also made circularly polarized to overcome poor mismatch, lower directive gain, and signal to noise problems in wireless communication. The antenna performance is observed in simulation and verified with measurements. Simulated and measured results show that the 2 port antenna exhibits a 10 dB impedance bandwidth of 5.26 % (2.78–2.93 GHz) and a 3 dB axial ratio bandwidth of 13.08 % (2.65–3.02 GHz), while those for the 4 port antenna are 2.46 % (2.41–2.47 GHz) and 13.10 % (2.36–2.69 GHz). The 2 port antenna supports a peak gain of 3.27 dBi with a maximum efficiency of 85.11 %, while the 4 port antenna successfully increases the gain to 6.49 dBi but decreases the maximum efficiency to 80.54 %. Both

* Corresponding author.
E-mail addresses: hashinur0001@gmail.com (H. Islam), saumya.das@smit.smu.edu.in (S. Das), tanweer.ali@manipal.edu (T. Ali), tanushree.b@smit.smu.edu.in (T. Bose), sourav.d@smit.smu.edu.in (S. Dhar), bijay.r@smit.smu.edu.in (B. Rai), kumarp@ukzn.ac.za (P. Kumar).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.11.029
1110-0168 © 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: H. Islam et al., Compact circularly polarized 2 and 4 port multiple input multiple output antennas with bandstop filter isolation technique, Alexandria Eng. J. (2022), https://doi.org/10.1016/j.aej.2022.11.029

the MIMO antennas show excellent diversity performance in terms of envelope correlation coefficient (ECC), Directive Gain (DG), Total active reflection coefficient (TARC), channel capacity loss (CCL), and mean effective gain (MEG) ratio. Both the structures achieve ECC of less than 0.006, DG of more than 9.9 dB, TARC of less than −19.84071 dB, CCL of less than 0.00179 bits/s/Hz, and MEG of less than −0.00577 dB in measurements. The overall dimensions of 2 port and 4 port antennas are limited to $44 \times 22 \times 1.6mm^3$ and $44 \times 52 \times 1.6mm^3$, respectively. These characteristics of the proposed antennas make them suitable for use in wireless high-speed communications.

© 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Because of their high capacity, improved spectral efficiency, and reliability, Multiple-Input Multiple-Output (MIMO) systems have been universally accepted in almost all types of wireless communications. In order to increase network efficiency, MIMO technology is being widely employed in LTE and LTE Advanced radio networks [1]. Transmission turbulence brought on by multipath has been greatly diminished with the development of MIMO technology. Another common use of MIMO technology today is in wireless local area networks (WLAN) [2]. Wireless routers with multiple antennas have become common in WLAN. With the proper implementation of MIMO technology in wireless routers and mobile devices, the data rate can be improved by a factor of two or many times. Since the advent of 5G and the internet of things (IoT), MIMO technology has become increasingly popular as it is a crucial part of any 5G network [3]. Recent innovations in the fields of radio telecommunications, robotics, augmented reality/virtual environments, and surveillance [4] depend on 5G to fulfill the requirements of their applications with high data rates. Therefore, MIMO technology is progressively becoming essential for all forms of communication in the coming generation.

One of the most important components for the proper operation of MIMO systems is the MIMO antenna, which accommodates multiple transmissions and receptions for supporting multipath propagation. A MIMO antenna requires multiple transmitters and receivers to work concurrently to improve the channel capacity. The channel capacity (bits/Hz) of MIMO antenna increases with the increase in number of transmitter and receiver antennas, as mentioned in equation (1) [5].

$$C_{MIMO} = log\left[det\left(I_{M_R} + \frac{SINR}{M_T} HH^H\right)\right] \quad (1)$$

where, M_T and M_R are the numbers of transmitters and receivers, respectively. While I_{M_R} is the identity matrix of dimension $M_R \times M_R$, and H is the channel coefficient matrix and H^H is the Hermitian transpose matrix. $SINR$ is the signal-to-interference-plus-noise ratio.

But accommodating more antennas within a confined area causes destructive interference or mutual coupling to increase, which eventually causes degradation in antenna efficiency and MIMO performance. Therefore, designing a MIMO antenna with reduced mutual coupling or increased isolation is a big challenge. Researchers have presented many techniques to address this issue. Isolation of MIMO elements has been increased by introducing a decoupling network [6–8], parasitic elements [9,10], and neutralization lines [11,12] in the structure. The defected ground plane structure was created to increase the isolation [13,14] of MIMO antenna elements. The use of metamaterials and metasurfaces has also been explored for reducing mutual coupling in many MIMO antennas [15–17]. Bandstop filters have also been placed between antenna elements to weaken the coupling current, thereby enhancing the isolation. In the research work [18], a split ring resonator based bandstop filter was used to get isolation as high as 39.25 dB for a two port MIMO antenna of dimension $44 \times 22mm^2$. The use of a bandstop filter to achieve high isolation has also been implemented in the design of a PIN diode controlled reconfigurable MIMO antenna [19]. So, the use of a bandstop filter has emerged as a good option for the design of compact MIMO antenna with high isolation.

In addition, to increase isolation, incorporation of circular polarization in the MIMO antenna becomes significant, as circularly polarized MIMO antenna can overcome poor mismatch, lower directive gain and SNR problem in wireless communication. Circularly polarized MIMO can also support wide wireless coverage, equal division of power at different feed point and superior penetration of obstacles, such as trees, walls and buildings. To accommodate circular polarization in MIMO antenna, different approaches were considered in recent articles. In the article [20], four radiating elements are placed orthogonally to the adjacent element to form a wideband circularly polarized MIMO antenna structure. The truncated corner technique was explored to get circular polarization. For increasing the circularly polarized bandwidth, an additional open circuited stub was created on the ground plane. The truncated corner patch, along with the hook-shaped open circuited stub, successfully brought fractional bandwidth of 92.08 % (1.73–3.94 GHz) and a 3-dB axial ratio bandwidth of 54.16 % (1.9–3.2 GHz) to this wearable MIMO antenna. A combination of truncated corner patches and metasurface structures can also produce wideband circular polarization in the MIMO antenna. In the paper [21], truncated corner patches and metasurface structure of periodic square metallic plates were used to achieve a 23.4 % (24.5–31 GHz) 10 dB impedance bandwidth and an overlapping 16.8 % (25–29.6 GHz), 3-dB axial ratio bandwidth for a four port MIMO antenna. In another work [22], an offset feeding technique for microstrip patches was explored to achieve circular polarization in a two port MIMO antenna. The offset feeding created two orthogonal electric fields of equal magnitudes with a 90-degree phase difference, resulting in circular polarization for MIMO antenna. In addition to offset feeding, three

grounded stubs and a defective ground structure are used to achieve high isolation and good impedance matching. This MIMO antenna structure yields an AR bandwidth of 2.2 % (2.50–2.55 GHz) and an impedance bandwidth of 3.2 % (2.47–2.55 GHz). The antenna is suitable for 2.5 GHz band Wi-Fi applications. The slot cutting technique has also been utilized to achieve circular polarization in MIMO antenna [23]. In [24], authors presented a three-port circularly polarized MIMO antenna for WLAN application. Circular polarization is achieved by forming two orthogonal slots in the ground plane. In the work [25], to produce circular polarization in MIMO antennas, a square shaped ground plane is etched with a circular slot, and two arc-shape slotted stubs along the two opposite edges are made along with that circular slot.

Based on the preceding discussions, it is expected that a bandstop filter as a decoupling network can support high isolation in a MIMO antenna structure, while orthogonal current distribution in the antenna structure can bring circular polarization. These two concepts have been primarily explored in this research work. Initially, a 2 port MIMO antenna was developed, and later, a 4 port MIMO antenna was developed to achieve higher data rates as the data rate of the MIMO antenna increases with the increase in the number of ports of the MIMO antenna. The bandstop filter isolation technique is used to obtain higher isolation on both MIMO antennas. Circular polarization has been achieved for both the MIMO antennas by generating orthogonal currents in the radiating structure. The dimensions of the MIMO antennas are restricted to $44 \times 22 \times 1.6mm^3$ and $44 \times 52 \times 1.6mm^3$ for 2 port and 4 port, respectively, for giving a compact but simple geometry of the complete structures. Different antenna performance parameters such as scattering coefficients, radiation patterns, realized gain, and radiation efficiency have been measured to ensure the reliability of antenna operation. MIMO performance indicators such as Envelope Correlation Coefficient (ECC), Diversity Gain (DG), Mean Effective Gain (MEG), Channel Capacity Loss (CCL), and Total Active Reflection Coefficient (TARC) of the proposed 2 port and 4 port antennas are also evaluated for observing the antenna performance in the MIMO environment. The proposed MIMO antennas are compared with other relevant work to find their significance. It is found that both antennas have many advantages over other types of work of the same attributes and dimensions. To the best of our knowledge, no previous work has combined bandstop filter decoupling and orthogonal current distribution for circular polarization in the same MIMO antenna design. However, this work demonstrated that these two techniques could be combined in MIMO antenna design to achieve significant improvements in antenna performance. The designed MIMO antennas could be used in LTE, LTE advanced communications, WLAN communication, and 5G IoT applications.

2. Antenna design methodology

2.1. Objectives

1. To construct a circularly polarized 2 port MIMO antenna and reduce its mutual coupling with a bandstop filter based isolation technique.
2. To extend the 2 port MIMO antenna designing technique to construct a 4 port circularly polarized MIMO antenna.

2.2. Research approach

The flow chart in Fig. 1 depicts the construction of two MIMO antennas. It includes four steps.

A step-by-step progress description of each step is mentioned below.

Step 1: The creation of a single circularly polarized monopole antenna.

In the initial phase of designing, a single circularly polarized antenna is developed on FR4 substrate ($\varepsilon_r = 4.4$, loss tangent

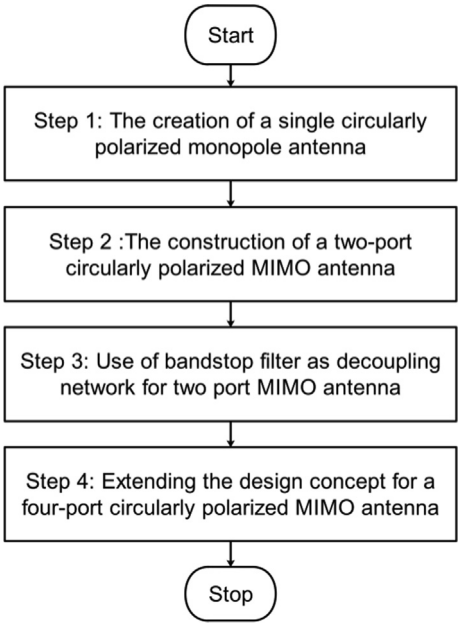


Fig. 1 Flowchart.

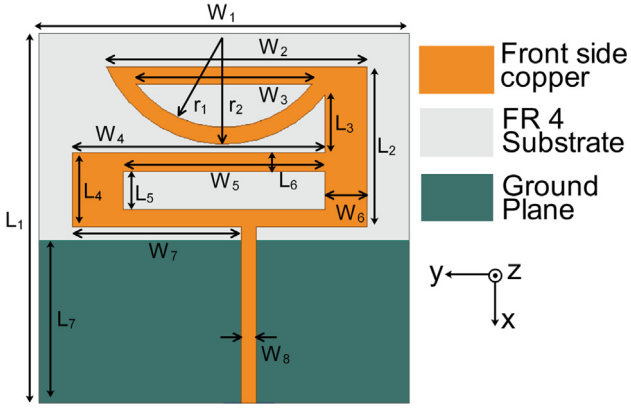


Fig. 2 Single antenna architecture.

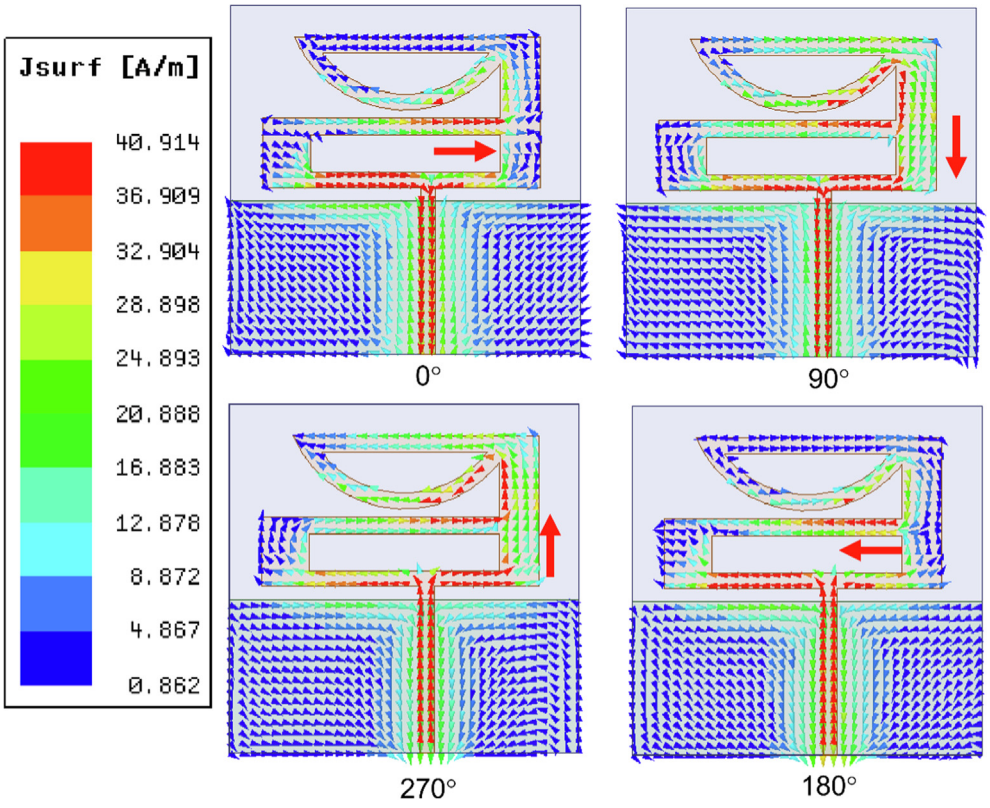


Fig. 3 Current distribution on single antenna structure.

$tan\delta = 0.02$) of thickness 1.6 mm, as presented in Fig. 2. An extended form of an 'a' shaped antenna structure with a partial ground plane is considered to bring circular polarization at 2.88 GHz in the LTE frequency band. The square shaped slot in the structure promotes orthogonal current distribution and thereby brings circular polarization at a resonance frequency of 2.88 GHz. The orthogonal current distribution can be seen in Fig. 3. To achieve the desired current path for accommodating center frequencies of 2.88 GHz with this monopole antenna

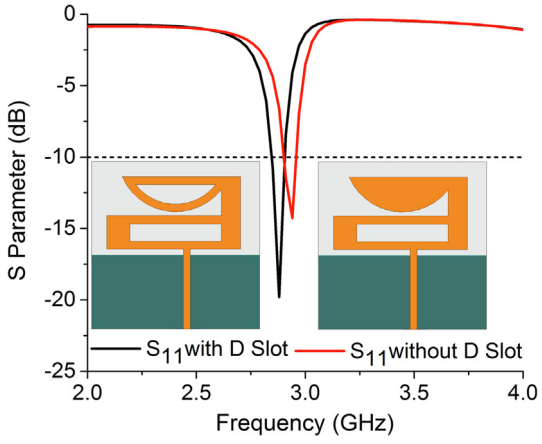


Fig. 4 The effect of D slot on reflection coefficient for single antenna.

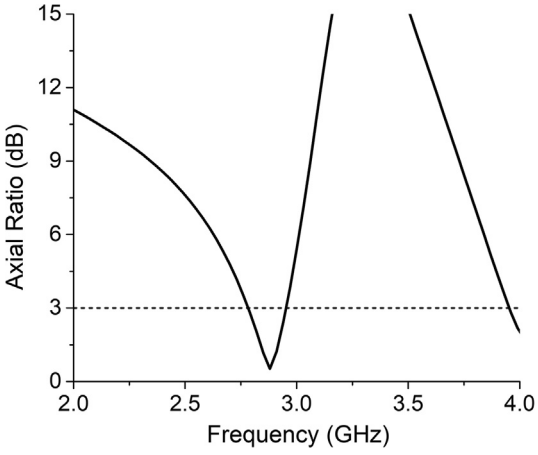


Fig. 5 Axial Ratio for single antenna structure.

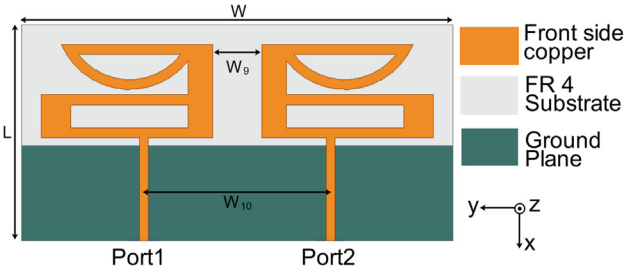


Fig. 6 2 port MIMO antenna without decoupling network.

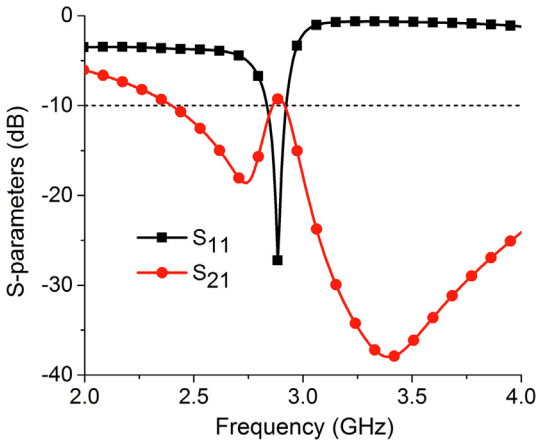


Fig. 7 S parameters for 2 port MIMO antenna without decoupling network.

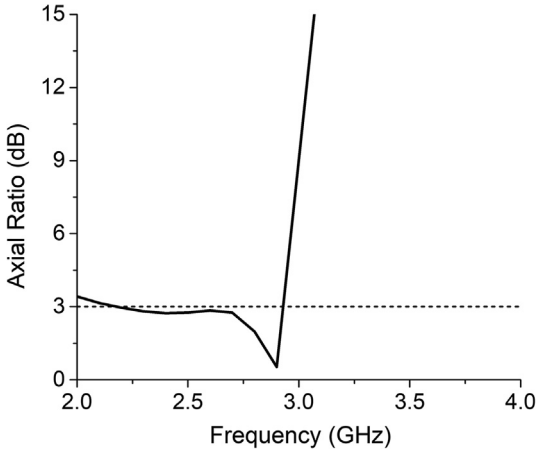


Fig. 8 Axial Ratio for 2 port antenna without decoupling network.

structure, the microstrip feed line is shifted from the middle of the antenna structure. The current distribution shown in Fig. 3 confirms the $\frac{\lambda}{4}$ current path length of 26 mm for a resonating frequency of 2.88 GHz. The partial ground plane aids in improving the impedance matching of the monopole antenna. The 'D' shaped slot in the antenna structure improves the impedance matching at the required frequency band, as illustrated with reflection coefficient in Fig. 4. The HFSS platform is used for simulation to optimize the dimensions and design parameters of a single antenna. The dimensions of various design parameters are listed as: $W_1 = 22$, $W_2 = 15.5$, $W_3 = 10.5$, $W_4 = 15$, $W_5 = 12$, $W_6 = 2.5$, $W_7 = 10$, $W_8 = 0.9$, $L_1 = 22$, $L_2 = 9.5$, $L_3 = 3.44$, $L_4 = 4.4$, $L_5 = 2.3$, $L_6 = 1.1$, $L_7 = 9.7$, $r_1 = 6.6$ and $r_2 = 7.6$ (unit: mm). The axial ratio (AR) confirms the circular polarization in the range of 2.78–2.95 GHz, as shown in Fig. 5. The single antenna structure is constructed on a $22 \times 22mm^2$ substrate.

Step 2:The construction of a two port circularly polarized MIMO antenna.

The 2 port antenna development phase begins with the placement of two single antenna elements on a 44 mm × 22 mm (W × L) substrate, as shown in Fig. 6. The distance (W_{10}) between two antenna elements is kept constant at 19 mm. It resonates with poor isolation at a center frequency of 2.88 GHz, as shown in Fig. 7. The circular polarization of the 2 port MIMO antenna is maintained in the frequency range of 2.83–2.93 GHz, as shown in the axial ratio plot in Fig. 8. The antenna is small in size, but it is unable to deliver the best performance, due to high mutual coupling between antenna elements and the ground plane. When port 1 is activated and port 2 is terminated with a 50-ohm load, the surface current distribution at 2.88 GHz confirms the existence of a significant mutual interaction between the ground plane and the remaining antenna element, as shown in Fig. 9.

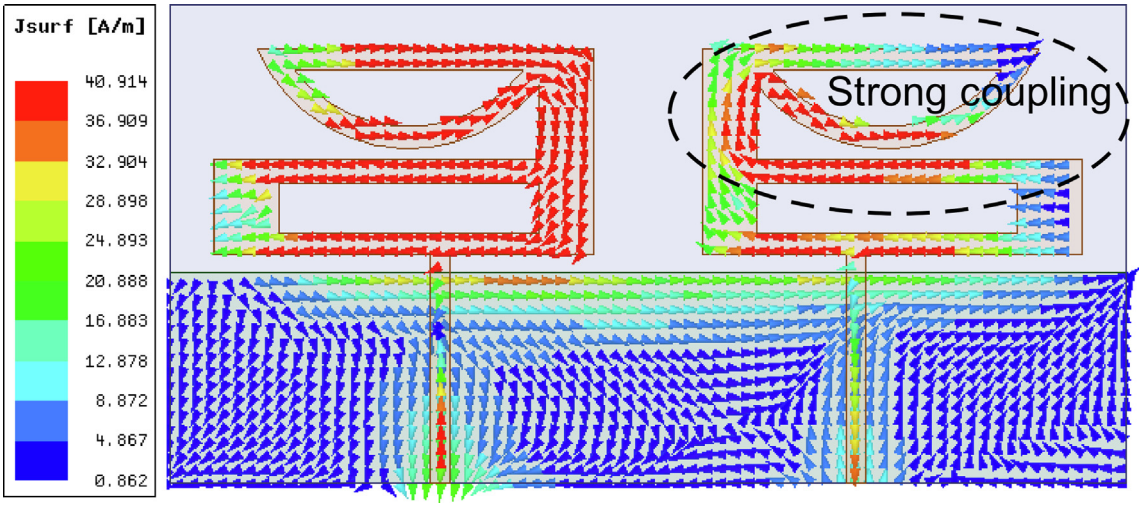


Fig. 9 Current Distribution on 2 port MIMO antenna without decoupling network.

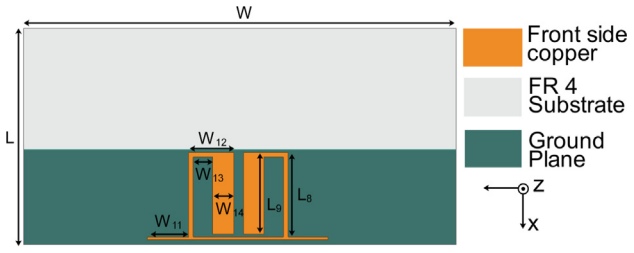


Fig. 10 Bandstop Filter Structure.

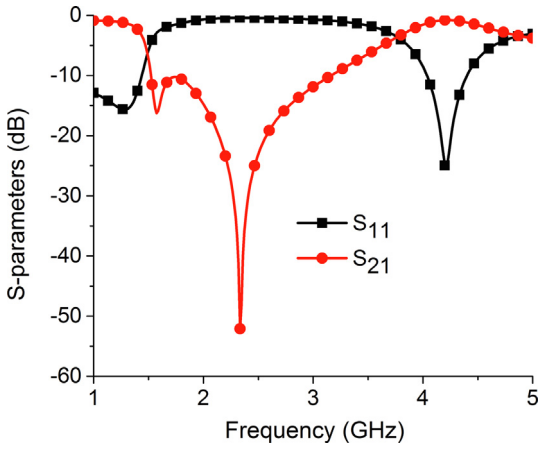


Fig. 11 S-Parameters for bandstop filter.

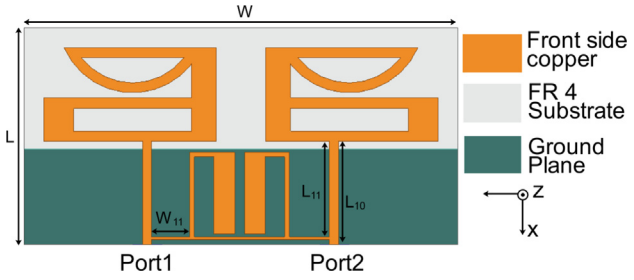


Fig. 12 Two port MIMO antenna with decoupling network (Final 2 port MIMO antenna).

Step 3: Use of bandstop filter as decoupling network for two port MIMO antenna.

In this step, an extended 'U' shape bandstop filter is created between the two antenna elements to reduce mutual coupling, as displayed in Fig. 10. The dimensions of the filter parameters (W = 44, W_{11} = 3.9, W_{12} = 4.5, W_{13} = 2, W_{14} = 2.1, L = 22, L_8 = 8.55, and L_9 = 8.25 (unit: mm)) determine the bandstop characteristics, as shown in Fig. 11. It is possible to achieve deep resonance and high isolation for the 2 port MIMO antenna by placing the bandstop filter between the two antenna elements, as shown in Fig. 12. A current distribution plot, shown in Fig. 13, confirms high isolation when port 1 is active and port 2 is terminated by a 50-ohm load. With a bandstop filter between the antenna elements, a scattering parameter plot (Fig. 14) shows resonance and isolation of more than 20 dB and 35 dB, respectively. In addition to high isolation, the MIMO antenna with bandstop filter can maintain circular polarization in the frequency band of 2.65–3.02 GHz, as shown in the AR plot in Fig. 15. The following design parameters are set for the final stage 2 port circularly polarized MIMO antenna: W = 44, L = 22, L_{10} = 10.5, L_{11} = 9.7 (unit: mm). A prototype model of this antenna structure is presented in Fig. 16.

Step 4: Extending the design concept for a four port circularly polarized MIMO antenna.

In the final step, a 4 port circularly polarized MIMO antenna (Fig. 17) is developed by extending the design concept of 2 port MIMO antenna. Four bandstop filters are used to minimize the mutual coupling between four radiating elements. A current distribution plot, shown in Fig. 18, confirms minimum mutual coupling when port 1 is active and other 3 ports are terminated by 50-ohm load. The complete structure is restricted by the $44 \times 52 mm^2 (W_A \times L_A)$ area. It is found that the 4 port MIMO antenna exhibits a return loss of 20.45 dB and isolation as high as 34.53 dB at 2.44 GHz (Fig. 19). Circular polarization is maintained with this 4 port MIMO structure for a frequency range of 2.36–2.69 GHz (Fig. 20). The final structure of the 4 port MIMO antenna has additional design parameters W_{15}, W_{16}, L_{12}, L_{13}, L_{14}, L_{15}, and L_{16} equal to

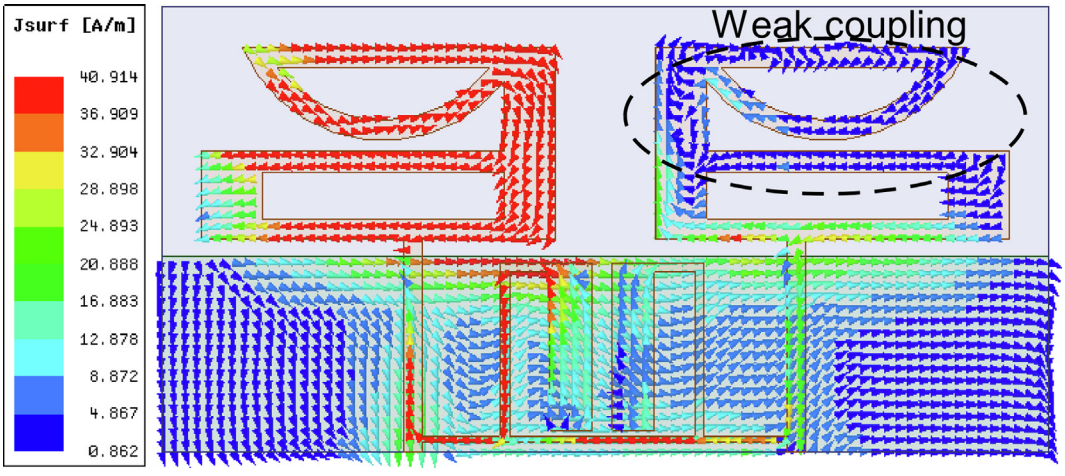


Fig. 13 Current Distribution on 2 port MIMO antenna with decoupling structure.

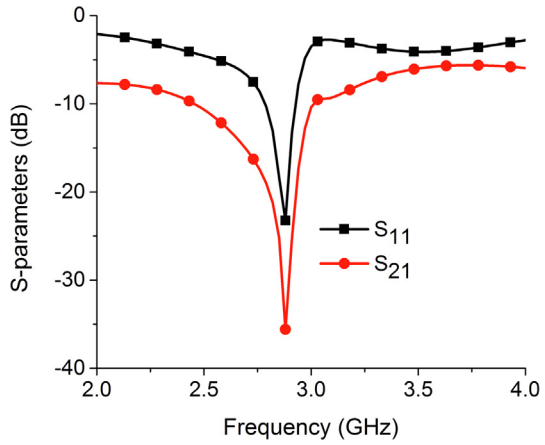


Fig. 14 Simulated S parameter for 2 port MIMO antenna with decoupling network.

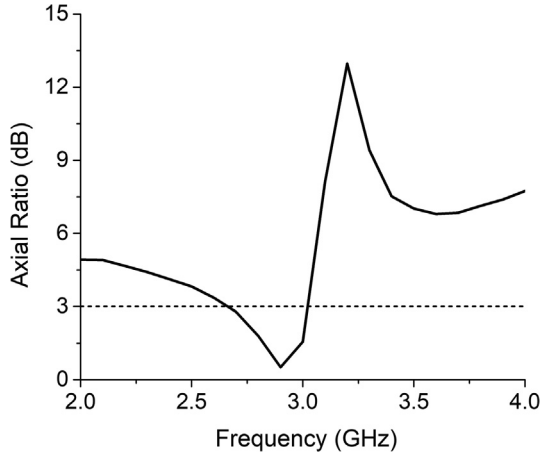


Fig. 15 Axial ratio for 2 port MIMO antenna with decoupling network.

15, 13.5, 9, 7, 12, 15.15 and 22 mm, respectively. Fig. 21 shows a fabricated model of a 4 port MIMO antenna.

3. Parametric analysis

To estimate the sizes of several structural elements in the antenna construction, a thorough parametric study is performed. L_6, L_7, W_6, W_8, W_5, W_{10} and L_9 played a significant role in improving impedance matching, achieving circular polarization, and increasing isolation for 2 port antenna structures.

L_7 represents the length of the partial ground in this antenna structure. The partial ground plane in the antenna includes additional reactance, which may provide better impedance matching. As shown in Fig. 22, the length of L_7 is optimized to be 9.7 mm to reach good impedance matching at 2.88 GHz. A change in partial ground length led to a change in resonance frequency. A change in current path length (L_6 and W_6) also caused a shift in resonance frequency, affecting isolation. To optimize the resonance and isolation at 2.88 GHz, the values of L_6 and W_6 are determined to be 1.1 mm and 2.5 mm, respectively, as shown in Figs. 23 and 24. The resonating frequency and isolation were also affected by the width of the antenna feed line (W_8). The variation in W_8 length, depicted in Fig. 25, yielded the optimum result at 0.9 mm.

The parameter W_5 affects most in achieving circular polarization. As shown in Fig. 26, a slight variation in W_5 causes significant changes in the AR value. The AR value is less than 3 dB at operating frequency (2.83–2.93 GHz) with a length of W_5 of 12 mm. Even with a change of 0.5 mm in the length of W_5, the antenna structure loses circular polarization. The AR value is also affected by the variation in the W_6 value, as seen in Fig. 27. At $W_6 = 2.5$ mm, the AR becomes less than 3 dB at the operating frequency. With the decrease in W_6 value, the antenna loses circular polarization.

The isolation value starts improving with the increase in separation distance W_{10}, as shown in Fig. 28. With $W_{10} = 19$ mm, it is possible to achieve as high isolation as 9.24 dB. But the inclusion of a bandstop filter significantly

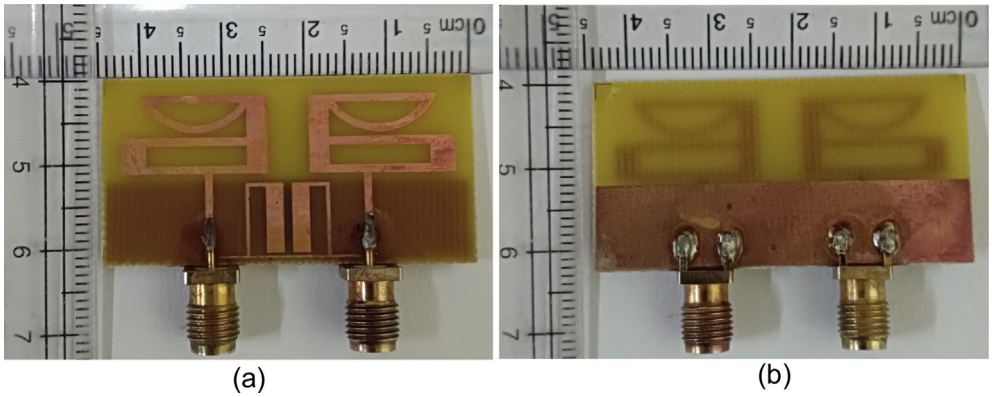


Fig. 16 Prototype Model for final 2 port MIMO antenna (a) front side (b) back side.

improves the isolation up to 35.59 dB. The length of the band-stop filter in part L_9 also has a significant impact on achieving isolation at 2.88 GHz, as shown in Fig. 29. The optimal isolation value is achieved with $L_9 = 8.25$ mm.

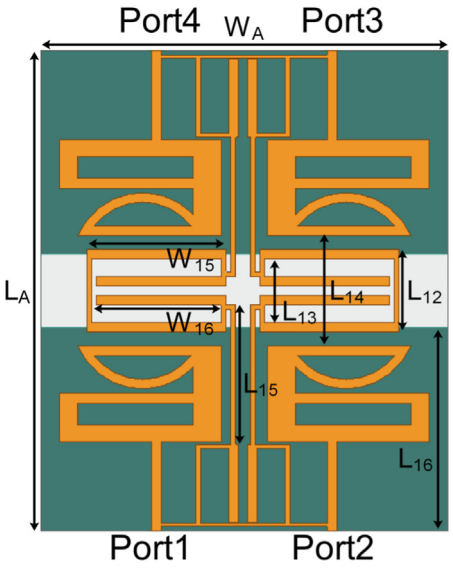


Fig. 17 Four port MIMO antenna structure.

For 4 port MIMO, the separation (L_{14}) between the upper and lower parts plays a significant role in isolation. A slight variation in the L_{14} value causes considerable differences in isolation value in the 4 port MIMO structure. Fig. 30 shows that the maximum isolation in the 4 port MIMO antenna is obtained with $L_{14} = 12$ mm.

4. Results and discussion

Results are presented for 2 port as well as 4 port MIMO antennas in three following aspects:

)A. Scattering parameters and Axial ratio for visualizing resonance, isolation, and circular polarization
)B. Gain, Efficiency and Radiation pattern for explaining the far field characteristics
)C. Diversity characteristics for describing effective MIMO operations

A. Scattering Parameters and Axial Ratio.

The simulated scattering parameters and axial ratio for 2 port and 4 port antennas are verified with measured results, as illustrated in Fig. 31. The measurement setup for carrying out all kinds of experiments is presented in Fig. 32. For 2 port circularly polarized MIMO antenna, maximum isolation of 35.59 dB and return loss of 23.24 dB are found at 2.88 GHz. The antenna yields 150 MHz (2.78–2.93 GHz), 10 dB impe-

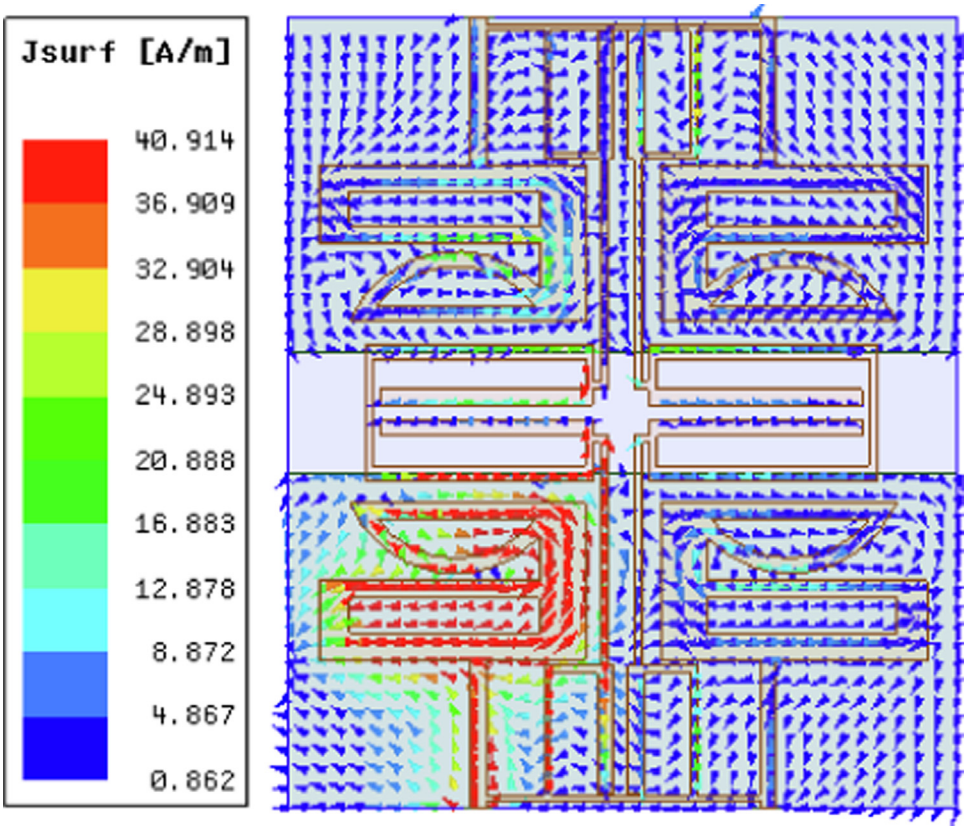


Fig. 18 Current Distribution on 2 port MIMO antenna.

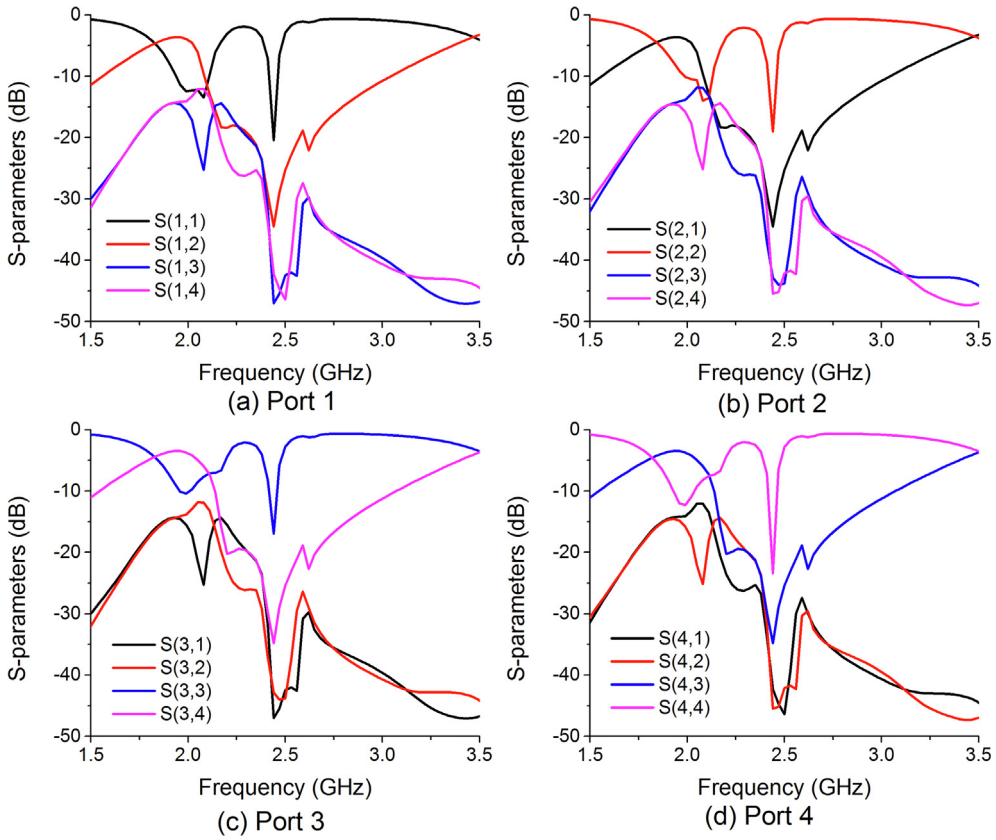


Fig. 19 Simulated s parameters for four port MIMO antenna (a) Port 1 (b) Port 2 (c) Port 3 (d) Port 4.

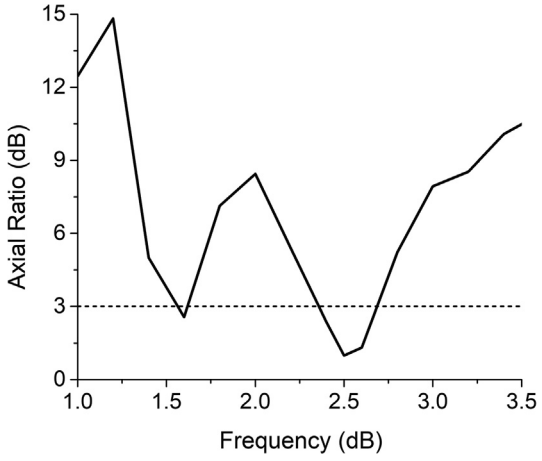


Fig. 20 Axial Ratio for 4 port MIMO antenna.

dance bandwidth, and 370 MHz (2.65–3.02 GHz), 3 dB AR bandwidth.

The 4 port circularly polarized MIMO antenna performs best at 2.44 GHz, with more than 34.53 dB isolation and 20.45 dB return loss, as shown in Fig. 31. The 10 dB impedance bandwidth and 3 dB axial ratio for the 4 port MIMO antenna are 60 MHz (2.41–2.47 GHz) and 330 MHz (2.36–2.69 GHz), respectively.

Thus, both antennas can support LTE communication in their operating frequency band. A slight difference in results is due to an error present in the fabrication process.

B. Gain, Efficiency and Radiation Pattern.

Gain, efficiency and radiation pattern were observed for understanding far field characteristics of the 2 port and 4 port antennas. It is found that, at 2.88 GHz, the 2 port antenna exhibits a peak gain of 3.27dBi with a maximum efficiency of 85.11 %. This can be seen in Fig. 33. Left Hand Circular Polarization (LHCP) and Right Hand Circular Polarization (RHCP) are illustrated in Fig. 34, and it shows a clear dominance of LHCP over RHCP for the 2 port MIMO antenna. On the other hand, the 4 port MIMO antenna yields a maximum gain of 6.49 dBi and an efficiency of 80.54 % at a frequency of 2.44 GHz, as shown in Fig. 35. Based on the radiation plot (Fig. 36), it is evident that LHCP dominates over RHCP also for the 4 port antenna structure.

So, it can be said that both antennas can support LTE communication efficiently in its surrounding region. The 4 port MIMO antenna outperforms its 2 port counterpart in terms of gain and efficiency, making it a better choice for advanced wireless communication if its larger size is not a concern.

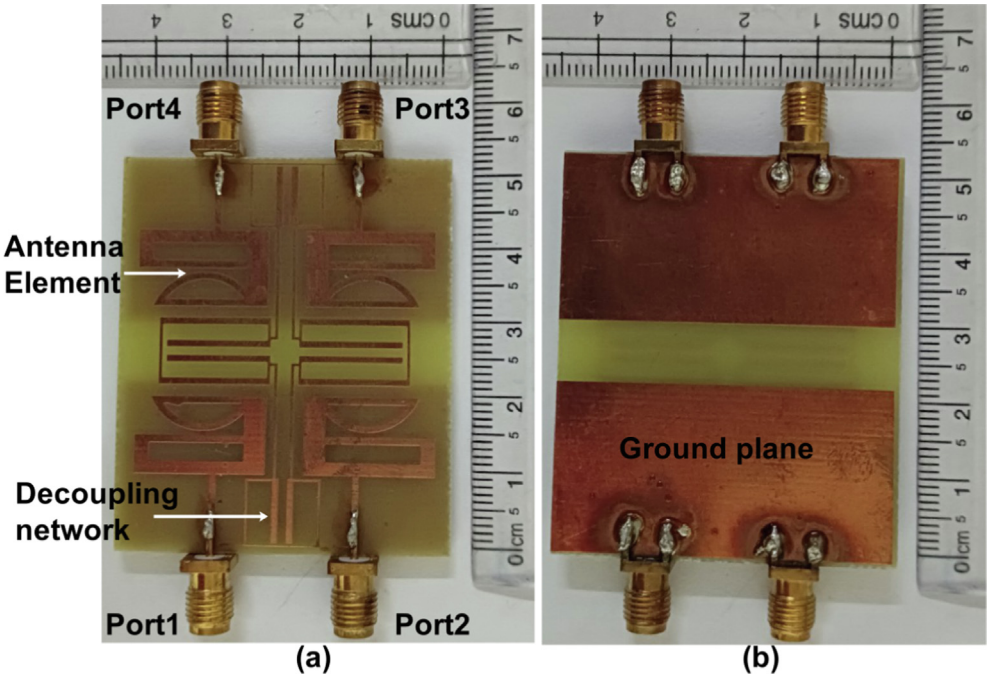


Fig. 21 Fabricated model of 4 port MIMO antenna: (a) front side (b) back side.

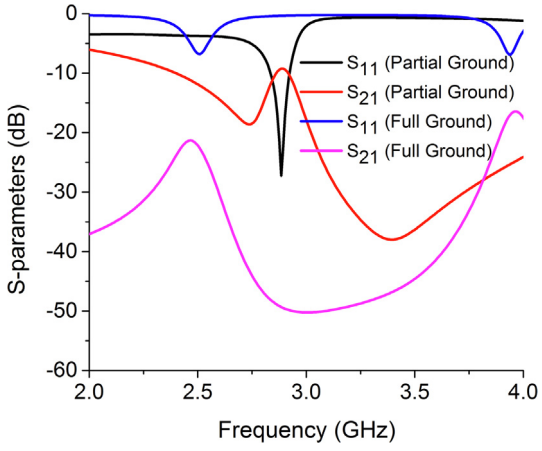


Fig. 22 Influence of ground plane on S parameters of two port antenna.

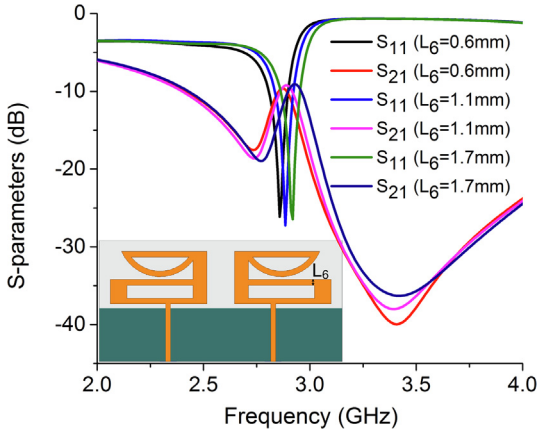


Fig. 23 Influence of L_6 on impedance matching for two port MIMO antenna.

C. Diversity Characteristics.

In this section of result analysis, MIMO performance parameters such as Envelope Correlation Coefficient (ECC), Diversity Gain (DG), Mean Effective Gain (MEG), Channel Capacity Loss (CCL), and Total Active Reflection Coefficient (TARC) of the proposed 2 port and 4 port antennas are observed to ensure satisfactory MIMO performance.

ECC is a metric used to describe the isolation or correlation of antenna elements. ECC could be calculated using Equation (2) [26] for a 2 port MIMO antenna and Equation (3), (4), (5) [27] for a 4 port MIMO antenna. ECC should ideally be zero, but in practice, anything less than 0.5 is acceptable.

$$ECC = \frac{\left|S_{11}^{*}S_{12} + S_{21}^{*}S_{22}\right|^{2}}{\left(1-\left(|S_{11}|^{2}+|S_{21}|^{2}\right)\right)\left(1-\left(|S_{22}|^{2}+|S_{12}|^{2}\right)\right)} \tag{2}$$

$$ECC_{12} = \frac{|S_{12}S_{11}{}^{*} + S_{22}S_{12}{}^{*} + S_{32}S_{13}{}^{*} + S_{42}S_{14}{}^{*}|^{2}}{\left[1-\left(|S_{11}|^{2}+|S_{12}|^{2}\right)+\left(|S_{14}|^{2}+|S_{13}|^{2}\right)\right]^{2}} \tag{3}$$

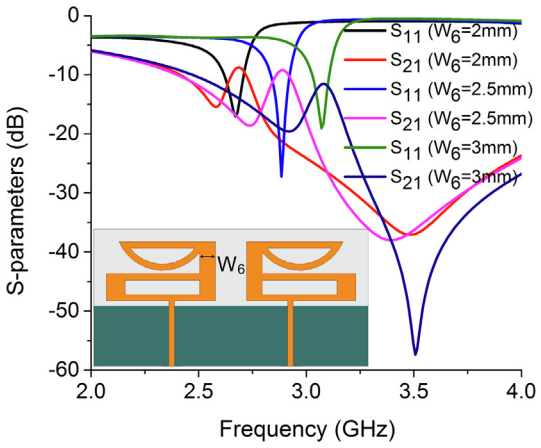


Fig. 24 Influence of W_6 on impedance matching for two port MIMO antenna.

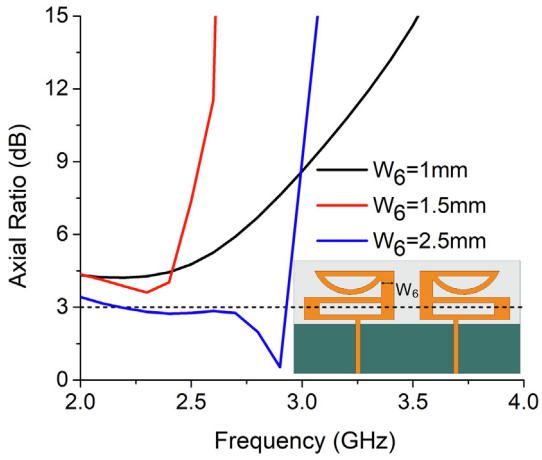


Fig. 27 Influence of W_6 on circular polarization for two port MIMO antenna.

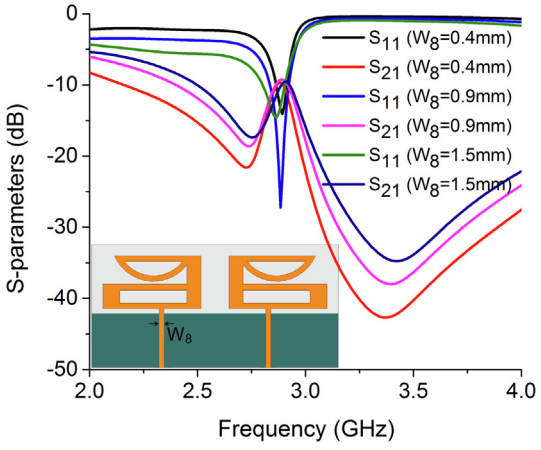


Fig. 25 Influence of W_8 on impedance matching for two port MIMO antenna.

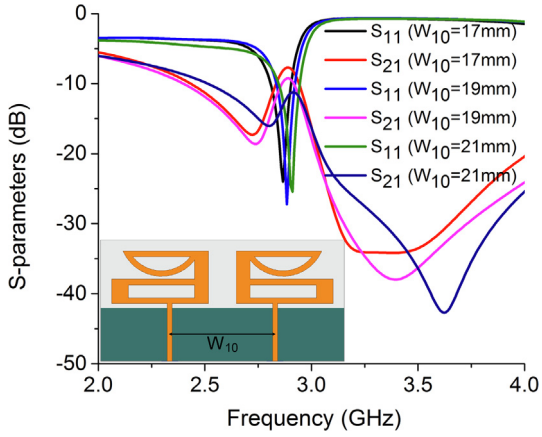


Fig. 28 Influence of W_{10} on isolation for two port MIMO antenna.

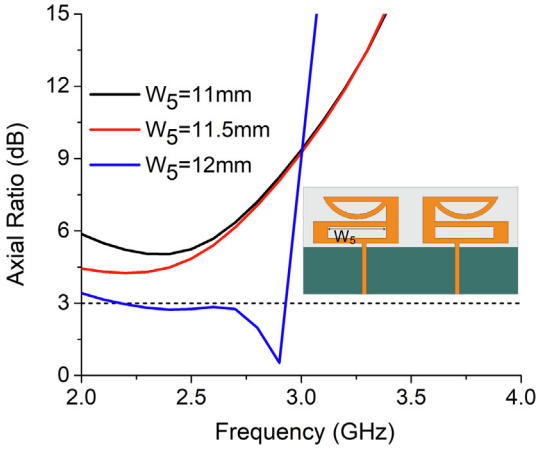


Fig. 26 Influence of W_5 on circular polarization for two port MIMO antenna.

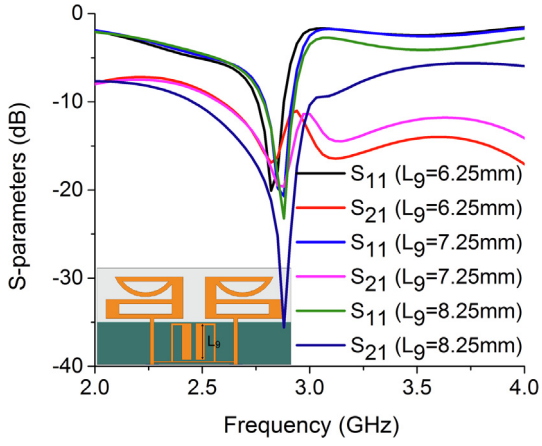


Fig. 29 Influence of L_9 on isolation for two port MIMO antenna.

$$ECC_{13} = \frac{|S_{13}S_{11}^{*} + S_{23}S_{12}^{*} + S_{33}S_{13}^{*} + S_{43}S_{14}^{*}|^2}{\left[1 - \left(|S_{11}|^2 + |S_{12}|^2\right) + \left(|S_{14}|^2 + |S_{13}|^2\right)\right]^2} \quad (4)$$

$$ECC_{14} = \frac{|S_{14}S_{11}^{*} + S_{24}S_{12}^{*} + S_{34}S_{13}^{*} + S_{44}S_{14}^{*}|^2}{\left[1 - \left(|S_{11}|^2 + |S_{12}|^2\right) + \left(|S_{14}|^2 + |S_{13}|^2\right)\right]^2} \quad (5)$$

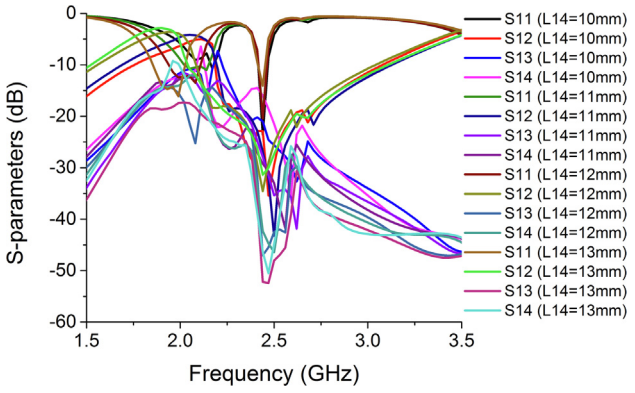


Fig. 30 Influence of L14 on S parameters of 4 port MIMO antenna.

The DG value must be high enough to ensure the quality and dependability of a wireless MIMO system. At the operating frequency, it should be close to 10 dB. The MIMO antenna ECC value can be used to calculate DG by Equation (6) [28]:

$$DG = 10 \times \sqrt{1 - |ECC|} \quad (6)$$

From Figs. 37 and 38, it can be seen that the 2 port and 4 port antennas are holding the threshold requirement for both ECC and DG values at their operational frequencies. The ECC and DG are obtained as 7.94152×10^{-5} and 10 dB, respectively, for the 2 port antenna, while the values for the 4 port antenna are less than 0.006 and 9.99 dB, respectively.

The S parameters of the MIMO antennas can be used to calculate the TARC value. TARC values range from 0 to 1. TARC value '0′ indicates that the antenna successfully radiates all available input power [29]. So, at the operating frequency, the TARC value of the MIMO antenna should be close to '0′. Equations (7) and (8) [29,30] describe the expression of TARC in terms of S parameters for two and 4 port MIMO antennas, respectively:

$$\Gamma_a^t = \frac{\sqrt{\sum_{i=1}^{N} |b_i|^2}}{\sqrt{\sum_{i=1}^{N} |a_i|^2}} \quad (7)$$

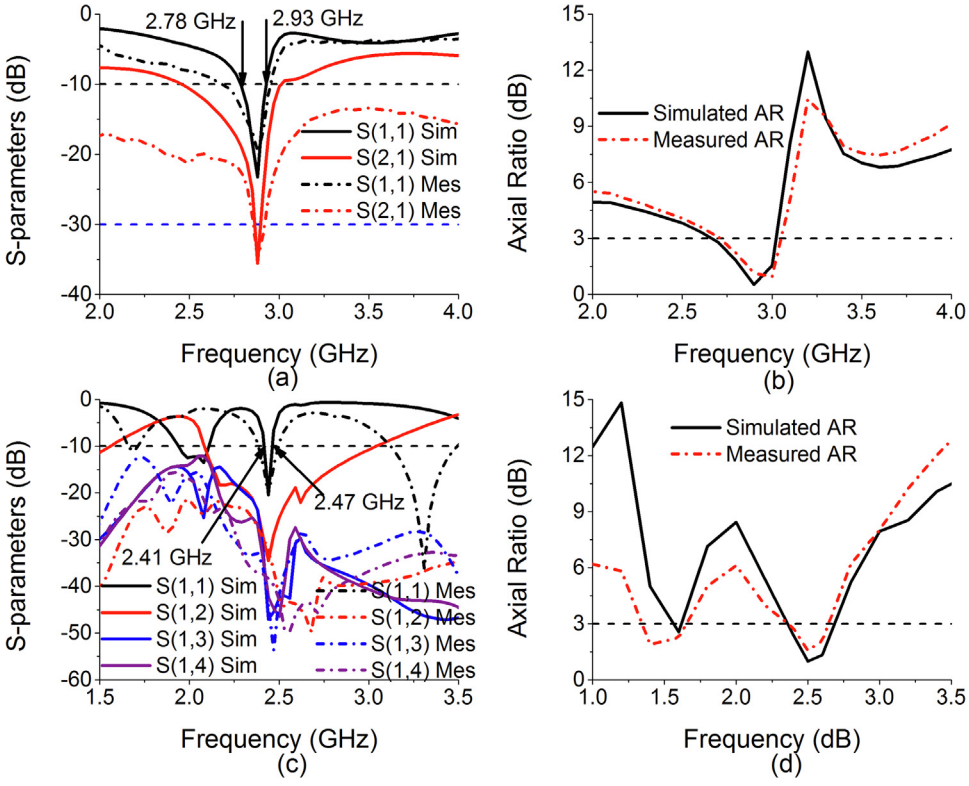


Fig. 31 S parameters and AR for 2 port and 4 port MIMO antenna (Final structure): (a) S parameters for 2 port MIMO antenna with decoupling network, (b) Axial ratio for 2 port MIMO antenna with decoupling network, (c) S parameters for 4 port MIMO antenna and (d) Axial ratio for the 4 port MIMO antenna.

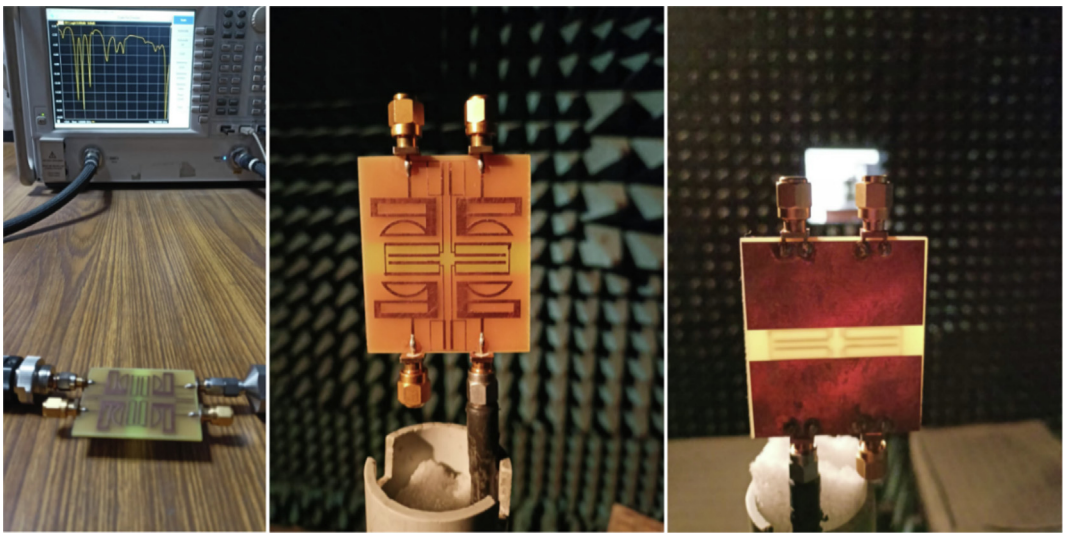


Fig. 32 Measurement set up.

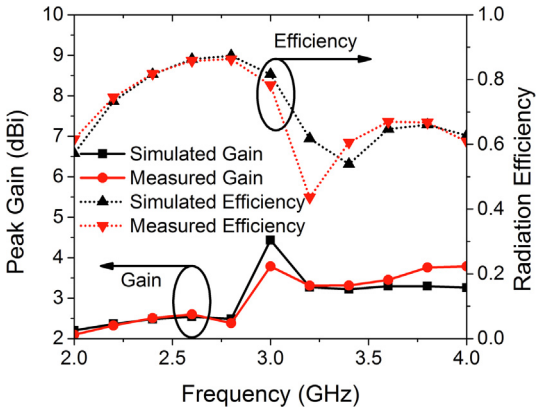


Fig. 33 Gain and efficiency of two port antenna with decoupling network.

$$\Gamma_a^t = \sqrt{\frac{\left(\left|S_{ii} + S_{ij}e^{j\theta}\right|^2 + \left|S_{ji} + S_{jj}e^{j\theta}\right|^2\right)}{2}} \tag{8}$$

Where $b_i = [s].a_i.[s]$ is the scattering matrix of the MIMO antenna, $[b] \rightarrow$ is the scattering vector and $[a] \rightarrow$ is the excitation vector. For N port antenna, TARC is related total incident power to total outgoing power.

Fig. 39 depicts the evaluated TARC value on the dB scale for the proposed MIMO antennas. It can be observed that both antennas have a satisfactory TARC value at the operating frequency.

MEG represents the gain performance of a MIMO antenna while accounting for environmental effects. Fig. 40 (a) depicts the MEG analysis performed at ports 1 and 2 of a 2 port antenna. MEG between different ports of 4 port antenna can be observed in Fig. 40 (b), 40 (c) and 40 (d). Equations (9) and (10) [31] were used for evaluating MEG from the S parameter values.

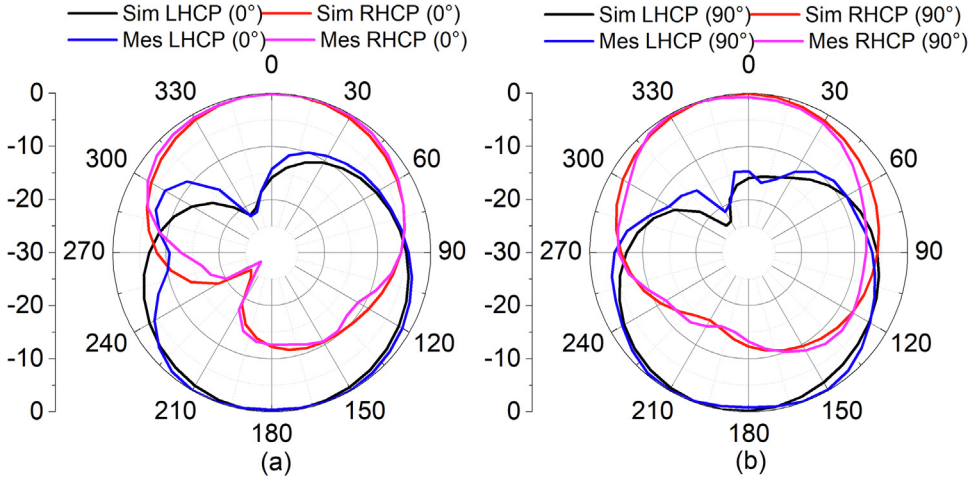


Fig. 34 LHCP and RHCP for two port MIMO antenna with decoupling network (a) at 0°, (b) at 90°.

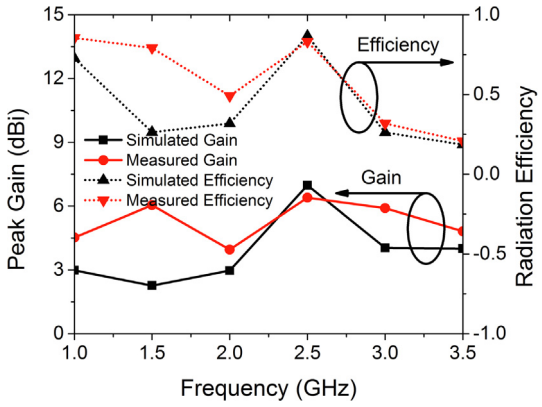


Fig. 35 Gain and efficiency for four port MIMO antenna.

$$MEG_i = 0.5\left[1 - |S_{ii}|^2 - |S_{ij}|^2\right] \quad (9)$$

$$MEG_j = 0.5\left[1 - |S_{ij}|^2 - |S_{jj}|^2\right] \quad (10)$$

To ensure a good 2 port MIMO antenna design with the same input power level at ports, the MEG-1/MEG-2 ratio must be less than 3 dB. As shown in Fig. 39 (a), the ratio is less than −0.05486 dB at the operating frequency for the proposed 2 port MIMO antenna. Similarly, the proposed 4 port MIMO antenna has MEG-1/MEG-2, MEG-1/MEG-3, and MEG-1/MEG-4 ratios lower than 0.15445 dB at 2.44 GHz.

The CCL is the maximum acceptable value of channel loss that can be allowed for successful transmission over a communication channel. For reliable communication, the CCL should be less than 0.4 bits/s/Hz. CCL can be calculated using the

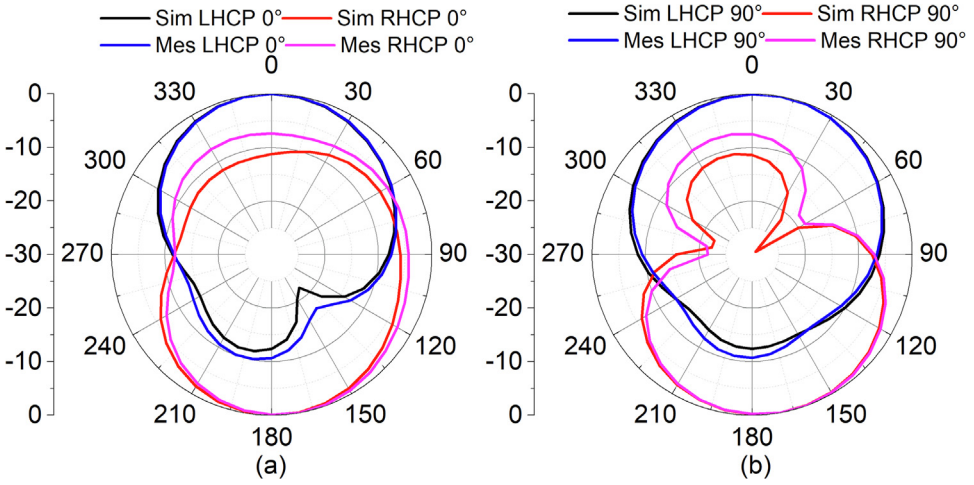


Fig. 36 LHCP and RHCP for 4 port antenna (a) at 0°, (b) at 90°.

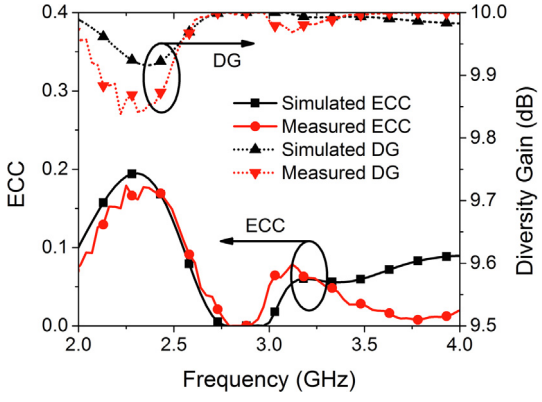


Fig. 37 ECC and DG for two port MIMO antenna.

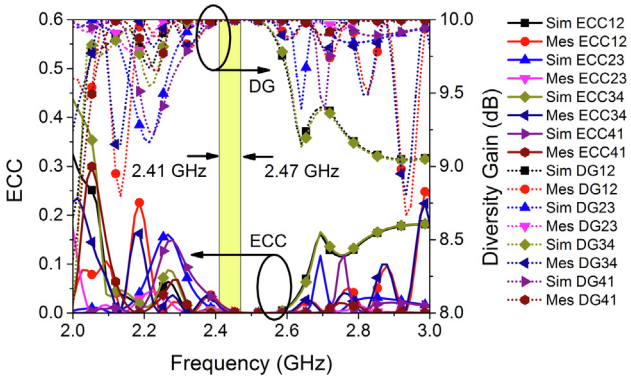


Fig. 38 ECC and DG for four port MIMO antenna.

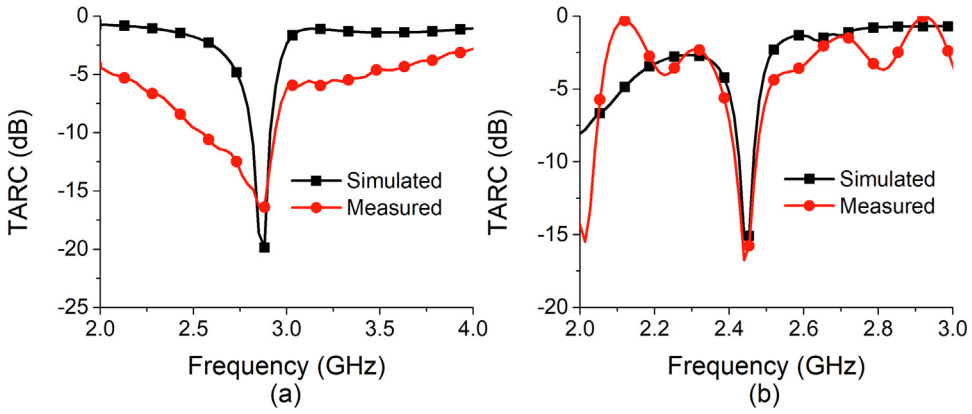


Fig. 39 TARC of (a) two port and (b) four port MIMO antennas.

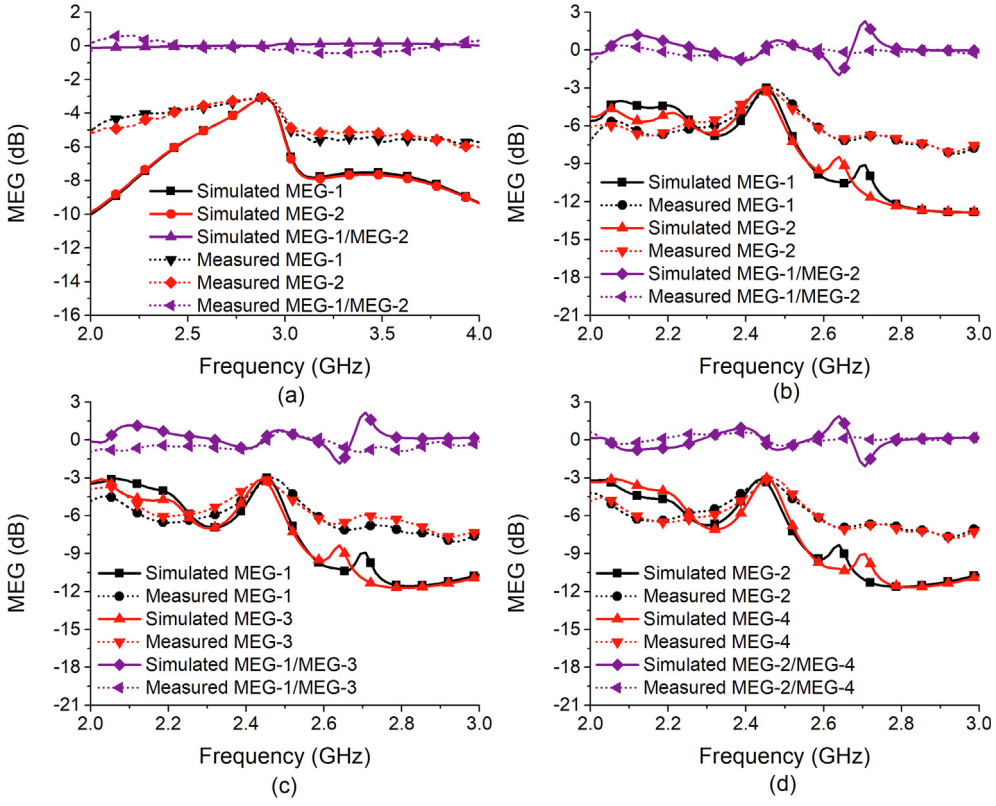


Fig. 40 MEG for MIMO antenna (a) port 1 and port 2 of two port (b) port 1 and port 2 of four port, (c) port 1 and port 3 of four port, and (d) port 2 and port 4 of four port.

Equations (11) and (12) [32,33] for two and 4 port MIMO antennas, respectively. Fig. 41 shows the CCL value, which was significantly less than 0.00179 bits/s/Hz and 0.14845 bits/s/Hz for the proposed 2 port and 4 port antennas at their operating frequencies.

$$C_{loss} = -\log_2\left|\varphi^R\right| \tag{11}$$

$$\varphi^R = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} & \varphi_{14} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{24} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} & \varphi_{34} \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & \varphi_{44} \end{bmatrix} \tag{12}$$

Where,

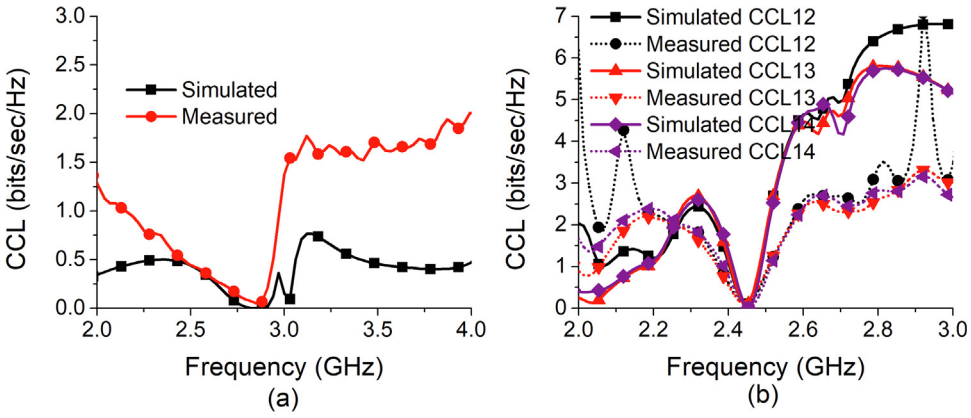


Fig. 41 CCL for MIMO antenna (a) two port (b) four port.

$$\varphi_{ii} = 1 - \left|\sum_{n=1}^{N} S_{in}{}^{*} S_{ni}\right|, for i, j = 1, 2, 3 and 4 \tag{13}$$

$$\varphi_{ij} = -\left|\sum_{n=1}^{N} S_{in}{}^{*} S_{nj}\right|, for i, j = 1, 2, 3 and 4 \tag{14}$$

4.1. Comparison analysis

A. Two Port Circularly Polarized MIMO Antenna

The proposed 2 port circularly polarized MIMO antenna is compared with a few recently reported MIMO antennas of similar attributes, as listed in Table 1. It is found that the bandstop filtering technique causes the proposed antenna to bring better isolation than its counterparts. In addition to high isolation, the proposed antenna has a smaller size than the antennas listed in [22,34–37] and a larger size than the antenna reported in [38]. At the same time, the antenna possesses comparatively lower impedance bandwidth, axial ratio bandwidth, and peak gain than the antennas mentioned in [36,37]. The proposed antenna has a lower ECC value than the structures presented in [22,35–38]. Overall, the proposed antenna has several advantages over existing MIMO antennas and may outperform them in real-world scenarios.

B. Four port circularly polarized MIMO antenna

The 4 port circularly polarized MIMO antenna is also compared with other circularly polarized 4 port MIMO antennas of similar attributes, as presented in Table 2. The main found advantages are its dimensions and isolation. The size has not been significantly increased with the increase in the number of ports. In fact, the dimensions of the proposed antenna are smaller than those of all the antennas present in Table 2. At the same time, the proposed antenna can bring better isolation than other antennas mentioned in the list. It exhibits moderate gain, lower impedance, and axial ratio bandwidth in comparison to other antennas. The ECC values are sufficiently smaller for the proposed antenna. So, the proposed 4 port circularly polarized MIMO antenna could be a good option for many MIMO communication systems.

5. Conclusion

In this work, a 2 port and a 4 port circularly polarized antennas were developed for MIMO operation. Circular polarization is brought into the antenna structure by creating orthogonal current flow, while high isolation is obtained by introducing a bandstop filter between radiating elements. This concept of circularly polarized MIMO antenna design has achieved 150 MHz (2.78–2.93 GHz) impedance bandwidth and 370 MHz (2.65–3.02 GHz) AR bandwidth with two radiator structures, while 60 MHz (2.41–2.47 GHz) and 330 MHz (2.36–2.69 GHz) AR bandwidth with four radiator structures. The gain and efficiency obtained from 2 and 4 port antennas are 3.27dBi and 85.11 %, 6.49dBi, and 80.54 %, respectively. Both the MIMO antennas show excellent diversity performance in terms of envelope correlation coefficient (ECC), Directive Gain (DG), Total active reflection coefficient (TARC), channel capacity loss (CCL) and mean effective gain (MEG) ratio. Both the structures achieve ECC of less than 0.006, DG of more than 9.9 dB, TARC of less than −19.84071 dB, CCL of less than 0.00179 bits/s/Hz, and MEG of less than −0.00577 dB in measurements. The overall dimensions of 2 port and 4 port antennas are limited to $44 \times 22 \times 1.6mm^3$ and $44 \times 52 \times 1.6mm^3$, respectively. The 2 port and 4 port antennas are compared with a few previously reported antennas of the same kind. It is found that the proposed circularly polarized antennas, with their compact dimensions and high isolation, can exhibit excellent MIMO operation and can be suitably used in LTE, LTE advanced, WLAN, 5G, and IoT communications. The concept of bandstop filter isolation proposed in this paper, in conjunction with any reconfigurable mechanism, could be very useful in the design of reconfigurable MIMO antennas for the purpose of cognitive communication.

ARTICLE IN PRESS

Table 1 Comparison list for two port antenna.

Ref.	Dimension	% BW and Freq band	3 dB AR Band	Isolation (dB)	Peak Gain (dBi)	ECC
[22]	1.25 $\lambda_0\times$ 0.83 λ_0 $\times$ 0.007 λ_0 (f = 2.5 GHz)	3.2 % (2.47–2.55 GHz)	2.2 % (2.5–2.55 GHz)	25	6.1	0.003
[34]	0.54 $\lambda_0\times$ 1.87 $\lambda_0\times$ 0.03 λ_0 (f = 5.8 GHz)	9.2 % (5.49–6.024 GHz)	1.59 % (5.772–5.864 GHz)	42.82	5.34	< 0.15
[35]	0.46 $\lambda_0\times$ 0.37 λ_0 $\times$ 0.01 λ_0 (f = 3.7 GHz)	24.32 % (3.3–4.2 GHz)	28.38 % (3.2–4.25 GHz)	> 15	2.5	< 0.10
[36]	1.03 $\lambda_0\times$ 0.59 $\lambda_0\times$ 0.05 λ_0 (f = 5.2 GHz)	13.7 % (5.10–5.85 GHz)	13.7 % (5.10–5.85 GHz)	> 20	5.8	< 0.5
[37]	0.24 $\lambda_0\times$ 0.64 λ_0 $\times$ 0.014 λ_0 (f = 5.2 GHz)	18.3 % (6.3–5.2 GHz)	18.3 % (6.3–5.2 GHz)	22	5.8	< 0.04
[38]	0.24 $\lambda_0\times$ 0.24 λ_0 $\times$ 0.016 λ_0 (f = 3 GHz)	90.94 % (3.04–8.11 GHz)	32.10 % (4.42–6.11 GHz)	> 16	2.68	< 0.004
Prop.	0.42 $\lambda_0\times$ 0.21 λ_0 $\times$ 0.015 λ_0 (f = 2.88 GHz)	5.26 % (2.78–2.93 GHz)	13.08 % (2.65–3.02 GHz)	35.59	3.27	$< 4.53698 \times 10^{-4}$

Table 2 Comparison list for four port antenna.

Ref.	Dimension	% BW and Freq band	3 dB AR Band	Isolation (dB)	Peak Gain (dBi)	ECC
[5]	1.52 $\lambda_0\times$ 1.54 $\lambda_0\times$ 0.013 λ_0 (f = 2.45 GHz)	6.94 % (2.36 – 2.53 GHz)	6.53 % (2.36 – 2.52 GHz)	> 30	7.95	$< 1.21 \times 10^{-4}$
[21]	1.0 $\lambda_0\times$ 1.0 $\lambda_0\times$ 0.04 λ_0 (f = 27.5 GHz)	23.4 % (24.5 – 31 GHz)	16.8 % (25 – 29.6 GHz)	> 30	11	less than 0.015
[39]	0.72 $\lambda_0\times$ 0.72 $\lambda_0\times$ 0.02 λ_0 (f = 3.6 GHz)	11.11 % (3.4–3.8 GHz)	6.67 % (3.46–3.7 GHz)	> 19	4.5	< 0.12
[40]	1.73 $\lambda_0\times$ 1.73 $\lambda_0\times$ 0.05 λ_0 (f = 9.62 GHz)	3.5 % (9.83–10.17 GHz)	0.83 % (9.58 – 9.66 GHz)	> 25	5.21	less than 0.0665
[41]	1.4 $\lambda_0\times$ 1.36 $\lambda_0\times$ 0.032 λ_0 (f = 6 GHz)	105.9 % (4–13 GHz)	67.7 % (4.2–8.5 GHz)	> 18	6.4	< 0.25
Prop.	0.36 $\lambda_0\times$ 0.42 $\lambda_0\times$ 0.013 λ_0 (f = 2.44 GHz)	2.46 % (2.41–2.47 GHz)	13.10 % (2.36–2.69 GHz)	> 34.53	6.49	< 0.006

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.

Acknowledgment

The authors would like to thank Sikkim Manipal University, Sikkim, India for providing TMA Pai Research Grant (Grant/Award Number: 118/SMU/REG/UOO/104/2019 and 176/SMU/REG/TMAPURF/26/2019) for this work.

ORCID

Hashinur Islam https://orcid.org/0000-0001-9239-3611.
Saumya Das https://orcid.org/0000-0001-5169-5568.
Tanushree Bose http://orcid.org/0000-0001-6945-3452.
Tanweer Ali http://orcid.org/0000-0002-1959-0480

References

[1] Y. Kim, H. Ji, J. Lee, Y.-H. Nam, B.L. Ng, I. Tzanidis, Y. Li, J. Zhang, Full dimension MIMO (FD-MIMO): The next evolution of MIMO in LTE systems, IEEE Wirel. Commun. 21 (2) (2014) 26–33.
[2] J. Yao, J. Xu, S. Luo, L. Wang, C. Yang, K. Wu, W. Lou, Comprehensive study on MIMO-related interference management in WLANs, IEEE Commun. Surv. Tutorials 21 (3) (2019) 2087–2110.
[3] K. Shafique, B.A. Khawaja, F. Sabir, S. Qazi, M. Mustaqim, Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios, IEEE Access 8 (2020) 23022–23040.
[4] A. Patel, A. Desai, I. Elfergani, A. Vala, H. Mewada, K. Mahant, S. Patel, C. Zebiri, J. Rodriguez, E. Ali, UWB CPW fed 4-port connected ground MIMO antenna for sub-millimeter-wave 5G applications, Alex. Eng. J. 61 (9) (2022) 6645–6658.
[5] R. Subhanrao Bhadade, S. Padmakar Mahajan, Circularly polarized 4× 4 MIMO antenna for WLAN applications, Electromagnetics 39 (5) (2019) 325–342.
[6] J. Deng, J. Li, L. Zhao, L. Guo, A dual-band inverted-F MIMO antenna with enhanced isolation for WLAN applications, IEEE Antennas Wirel. Propag. Lett. 16 (2017) 2270–2273.
[7] C.-D. Xue, X.Y. Zhang, Y.F. Cao, Z. Hou, C.F. Ding, MIMO antenna using hybrid electric and magnetic coupling for isolation enhancement, IEEE Trans. Antennas Propag. 65 (10) (2017) 5162–5170.
[8] S.-C. Chen, Y.-S. Wang, S.-J. Chung, A decoupling technique for increasing the port isolation between two strongly coupled antennas, IEEE Trans. Antennas Propag. 56 (12) (2008) 3650–3658.
[9] M.S. Khan, A.-D. Capobianco, M.F. Shafique, B. Ijaz, A. Naqvi, B.D. Braaten, Isolation enhancement of a wideband MIMO antenna using floating parasitic elements, Microw. Opt. Technol. Lett. 57 (7) (2015) 1677–1682.
[10] H. Yon, N. H. A. Rahman, M. A. Aris, M. H. Jamaluddin, I. Kong Cheh Lin, H. Jumaat, F. N. Mohd Redzwan, and Y. Yamada, "Development of C-Shaped Parasitic MIMO Antennas for Mutual Coupling Reduction," Electronics, vol. 10, no. 19, p. 2431, 2021.
[11] S. Zhang, G.F. Pedersen, Mutual coupling reduction for UWB MIMO antennas with a wideband neutralization line, IEEE Antennas Wirel. Propag. Lett. 15 (2015) 166–169.
[12] A. Kayabasi, A. Toktas, E. Yigit, K. Sabanci, Triangular quadport multi polarized UWB MIMO antenna with enhanced isolation using neutralization ring, AEU-Int. J. Electron. Commun. 85 (2018) 47–53.
[13] J. Banerjee, A. Karmakar, R. Ghatak, D.R. Poddar, Compact CPW-fed UWB MIMO antenna with a novel modified Minkowski fractal defected ground structure (DGS) for high isolation and triple bandnotch characteristic, J. Electromagn. Waves Appl. 31 (15) (2017) 1550–1565.
[14] Z. Niu, H. Zhang, Q. Chen, T. Zhong, Isolation Enhancement for 1 × 3 Closely Spaced E-Plane Patch Antenna Array Using Defect Ground Structure and Metal-Vias, IEEE Access 7 (2019) 119375–119383.
[15] G. Zhai, Z.N. Chen, X. Qing, Enhanced isolation of a closely spaced four-element MIMO antenna system using metamaterial mushroom, IEEE Trans. Antennas Propag. 63 (8) (2015) 3362–3370.
[16] B. Yin, X. Feng, J. Gu, A Metasurface Wall for Isolation Enhancement: Minimizing mutual coupling between MIMO antenna elements, IEEE Antennas Propag. Mag. 62 (1) (2019) 14–22.
[17] Z. Wang, C. Li, Q. Wu, Y. Yin, A metasurface-based low-profile array decoupling technology to enhance isolation in MIMO antenna systems, IEEE Access 8 (2020) 125565–125575.
[18] H. Islam, S. Das, T. Ali, P. Kumar, S. Dhar, T. Bose, Split ring resonator-based bandstop filter for improving isolation in compact mimo antenna, Sensors 21 (7) (2021) 2256.
[19] H. Islam, S. Das, T. Ali, T. Bose, S. Kumari, O. Prakash, P. Kumar, Bandstop Filter Decoupling Technique for Miniaturized Reconfigurable MIMO Antenna, IEEE Access 10 (2022) 19060–19071.
[20] A. Iqbal, A. Smida, A.J. Alazemi, M.I. Waly, N.K. Mallat, S. Kim, Wideband circularly polarized MIMO antenna for high data wearable biotelemetric devices, IEEE Access 8 (2020) 17935–17944.
[21] N. Hussain, M.-J. Jeong, A. Abbas, N. Kim, Metasurfacebased single-layer wideband circularly polarized MIMO antenna for 5G millimeter-wave systems, IEEE Access 8 (2020) 130293–130304.
[22] M.Y. Jamal, M. Li, K.L. Yeung, Isolation enhancement of closely packed dual circularly polarized MIMO antenna using hybrid technique, IEEE Access 8 (2020) 11241–11247.
[23] N. Hussain, M.-J. Jeong, J. Park, N. Kim, A broadband circularly polarized fabry-perot resonant antenna using a single-layered PRS for 5G MIMO applications, IEEE Access 7 (2019) 42897–42907.
[24] R.K. Jaiswal, K. Kumari, C.-Y.-D. Sim, K.V. Srivastava, Three-port circularly polarized MIMO antenna for WLAN application with pattern and polarization diversity, Microw. Opt. Technol. Lett. 63 (7) (2021) 1927–1934.
[25] M. Ghahremani, C. Ghobadi, J. Nourinia, M. Karamirad, B. Mohammadi, A novel circularly polarized modified slot MIMO antenna for WLAN applications, Microw. Opt. Technol. Lett. 59 (4) (2017) 876–880.
[26] S. Blanch, J. Romeu, I. Corbella, Exact representation of antenna system diversity performance from input parameter description, Electron. Lett 39 (9) (2003) 705–707.
[27] L. Malviya, R.K. Panigrahi, M.V. Kartikeyan, Four element planar MIMO antenna design for long-term evolution operation, IETE J. Res. 64 (3) (2018) 367–373.
[28] A.M. Saadh, K. Ashwath, P. Ramaswamy, T. Ali, J. Anguera, A uniquely shaped MIMO antenna on FR4 material to enhance isolation and bandwidth for wireless applications, AEU-Int. J. Electron. Commun. 123 (2020) 153316.
[29] M.S. Sharawi, Printed MIMO antenna systems: performance metrics, implementations and challenges, Forum Electromagn. Res. Methods Appl. Technol. (FERMAT) 1 (2014) 1–11.

[30] T. Kumari, G. Das, A. Sharma, and R. K. Gangwar, "Design approach for dual element hybrid MIMO antenna arrangement for wideband applications," Int. J. RF Microwave Comput. Aided Eng., vol. 29, no. 1, p. e21486, 2019.
[31] G. Das, A. Sharma, R.K. Gangwar, Dielectric resonator based circularly polarized MIMO antenna with polarization diversity, Microw. Opt. Technol. Lett. 60 (3) (2018) 685–693.
[32] G.A. Sarkar, S. Ballav, A. Chatterjee, S. Ranjit, S.K. Parui, Four element MIMO DRA with high isolation for WLAN applications, Prog. Electromagn. Res. Lett. 84 (2019) 99–106.
[33] I. Elfergani, A. Iqbal, C. Zebiri, A. Basir, J. Rodriguez, M. Sajedin, A. d. O. Pereira, W. Mshwat, R. Abd-Alhameed, and S. Ullah, "Lowprofile and closely spaced four-element MIMO antenna for wireless body area networks," Electronics, vol. 9, no. 2, p. 258, 2020.
[34] L. Malviya, R.K. Panigrahi, M. Kartikeyan, Circularly polarized 2× 2 MIMO antenna for WLAN applications, Prog. Electromagn. Res. C 66 (2016) 97–107.
[35] A.K. Dwivedi, A. Sharma, A.K. Pandey, V. Singh, Two Port Circularly Polarized MIMO Antenna Design and Investigation for 5G Communication Systems, Wirel. Pers. Commun. 120 (3) (2021) 2085–2099.
[36] H.H. Tran, N. Hussain, T.T. Le, Low-profile wideband circularly polarized MIMO antenna with polarization diversity for WLAN applications, AEU Int. J. Electron. Commun. 108 (2019) 172–180.
[37] U. Ullah, I.B. Mabrouk, S. Koziel, Enhanced-performance circularly polarized MIMO antenna with polarization/pattern diversity, IEEE Access 8 (2020) 11887–11895.
[38] R.N. Tiwari, P. Singh, B.K. Kanaujia, P. Kumar, Compact circularly polarized MIMO printed antenna with novel ground structure for wideband applications, Int. J. RF Microwave Comput. Aided Eng. 31 (8) (2021) e22737.
[39] S. Saxena, B.K. Kanaujia, S. Dwari, S. Kumar, H.C. Choi, K. W. Kim, Planar four-port dual circularly-polarized MIMO antenna for sub-6 GHz band, IEEE Access 8 (2020) 90779–90791.
[40] 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-Int. J. Electron. Commun. 121 (2020) 153241.
[41] S. Kumar, G.H. Lee, D.H. Kim, H.C. Choi, K.W. Kim, Dual circularly polarized planar four-port MIMO antenna with wide axial-ratio bandwidth, Sensors 20 (19) (2020) 5610.

Hashinur Islam received the B.Tech. degree in Electronics and Telecommunication Engineering from the Institute of Electronics and Telecommunication Engineers (IETE), India, in 2014, and the M.Tech. degree in Digital Electronics and Communication Engineering from the Sikkim Manipal Institute of Technology, Sikkim Manipal University, in 2018. He is currently pursuing the Ph.D. degree with Sikkim Manipal University.

He has published over 18 refereed papers in journals and conference proceedings. His research interests include multiband antennas, fractal antenna, circular polarization on antennas, feeding mechanism for antenna, PIFA antenna, wearable antenna, antenna for embedded systems, microstrip filters, reconfigurable filter, MIMO antenna, and so on. He is a Graduate Student Member of IEEE and Associate Member of IETE India. He is the Reviewer of IEEE Access, and International Journal of Microwave and Wireless Technologies, Cambridge University Press.

Saumya Das received the B.Tech. degree in Electronics and Telecommunication Engineering from the Institute of Electronics and Telecommunication Engineers (IETE), India, and the M.E. degree in Electronics and Communication Engineering from the Delhi College of Engineering, University of Delhi. He earned his Doctor of Philosophy in Electronics and Communication Engineering from Sikkim Manipal University.

He has more than 15 years of teaching experience with expertise in various subjects, such as electromagnetic field theory, antenna theory, microwave devices and circuits, signals and system, digital signal processing, and adaptive signal processing. He is currently an Assistant Professor with the Department of Information Technology, Sikkim Manipal Institute of Technology, India. He has published several articles in reputed peer-reviewed international journals and conferences. His research interests include different feeding techniques for microstrip and dielectric resonator antennas, flexible and wearable antenna for tracking and medical application, computational mathematics for antenna designing, RF exposure testing on human-bodies, electromagnetic compatibility, and ultra wideband antenna designing. He is an Associate Member of IETE, India. He serves as a Reviewer for Microwave and Optical Technology Letters (MOTL) (Wiley), and IETE Journal of Research (Taylor & Francis).

Dr. Tanweer Ali is currently working as an Associate Professor with the Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. He is an active researcher in the field of microstrip antennas, wireless communication, and microwave imaging. He has been listed in top 2% scientists across the world for the year 2022 and 2021 by the prestigious list published by Stanford University, USA indexed by Scopus. He has published more than 135 papers in reputed web of science(SCI) and scopus indexed journal and conferences and has filed seven Indian patents, of which three have been published. He has more than 1500 citations with an h-index of 23 and i-10 index of 39. Presently he is heading the antenna research group and Advanced Antenna Engineering lab at MIT, Manipal. He is on the board of reviewers of journals, like the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, IEEE ACCESS, IET Microwaves, Antennas and Propagation, Electronics Letter (IET), Wireless Personal Communications (WPCs) (Springer), AEU-International Journal of Electronics and Communications, Microwave and Optical Technology Letters (MOTL) (Wiley), International Journal of Antennas and Propagation (Hindawi), Advanced Electromagnetics, Progress in Electromagnetics Research (PIER), KSII Transactions on Engineering Science, International Journal of Microwave and Wireless Technologies, Frequenz, Radioengineering, IEEE open Journal of Antennas and Propagation and many more.

Tanushree Bose received the Ph.D. degree in Design of Printed Antennas using Neural Network from the Department of Electronics and Communication Engineering, BIT, Mesra, India, in 2011.
She worked in RF and microwave antennas with soft computation. She is currently an Associate Professor with the Sikkim Manipal Institute of Technology, India. She has published over 30 refereed papers in journals and conference proceedings. She has completed one Government (AICTE) Sponsored Research Project on Patch Antenna Designing Using Artificial Neural Network and successfully guided her research scholars. Her current research interests include multiband antennas for handsets and networks, miniature antennas, reconfigurable antennas, EBG application on conformal antennas, flexible and wearable antennas, smart antennas, and so on. She served on many national and international technical committees.

Sourav Dhar was born in Raiganj, West Bengal, India, in 1980. He received B.E degree from Bangalore Institute of Technology, VTU in 2002. He received M.Tech degree in Digital Electronics and Advanced Communication from Sikkim Manipal University in 2005 and PhD in 2013. Since 2003, he has been associated with the Sikkim Manipal Institute of Technology, India, where he is currently working as a Professor in ECE Department, SMIT. His current research interest includes antennas and filter designing for IoT, WSN, and remote sensing technologies. He is currently working on a project on Microwave device design funded by TMA Pai University Research Seed Grant-Major, Sikkim Manipal University, India. He is a Member of IEEE, IEEE-GRSS society, and IEI, India. He has published more than 30 SCI/Scopus indexed international journal and conference papers and guided several PG students. He also serves as the reviewer of Wireless Personal Communication, IEEE Transactions on Vehicular Technology and several other journals and conferences.

Bijay Rai has completed M. Tech in power electronics from Sikkim Manipal Institute of Technology (SMIT). He is currently a faculty in Electrical and electronics engineering department, SMIT Sikkim. His research interests include microwave device development, power electronics, and robotics.

Pradeep Kumar received his Bachelor's degree in Electronics and Communication Engineering, Master of Engineering in Electronics and Communication Engineering and Doctor of Philosophy in Electronics and Communication Engineering in 2003, 2005 and 2009, respectively. He completed his postdoctoral studies from Autonoma University of Madrid, Spain. He has over 15 years of experience in academic and research. He has held various positions such as Lecturer, Senior Lecturer, Assistant Professor and Associate Professor. He received various awards/fellowship such as MHRD fellowship, A4U fellowship, research excellence award, JW Nelson fund research award, CAES research award etc. He is registered as Professional Engineer with Engineering Council of South Africa. He is the author of many research papers published in various peer reviewed journals/conferences. He is serving as reviewer/member TPC for many journals/conferences such as IEEE Transactions on Antennas and Propagation, IEEE Access, Progress in Electromagnetic Research Journals, ACES Journal, IEEE Systems Journal, SAIEEE Research Journal, International Journal of Electronics, SATNAC, IEEE Africon etc. At present, he is working with the University of KwaZulu-Natal, South Africa. His current research areas include Antennas, Antenna Arrays, Wireless Communications etc.

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT).
India Habitat Centre, Lodhi Road, New Delhi, India. Sep 23-25, 2022

Computational Analysis of Natural Compounds as Cyclin-Dependent Kinase-5 Inhibitors for Alzheimer's and Parkinson's Disease

Dia Advani
Molecular Neurosciences and Functional Genomics Laboratory, Dept. of Biotechnology
Delhi Technological University
New Delhi – 110042, India
advanidia@gmail.com

Pravir Kumar
Molecular Neurosciences and Functional Genomics Laboratory, Dept. of Biotechnology
Delhi Technological University
New Delhi – 110042, India
pravirkumar@dtu.ac.in

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT) | 978-1-6654-9365-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBCONPT57482.2022.9938169

***Abstract*- Deregulation of cyclin-dependent kinase-5 is a hallmark of different neurodegenerative disorders, including Alzheimer's and Parkinson's disease. Herein, we have screened different natural compounds with anticancer properties against cyclin-dependent kinase-5 enzyme. The blood-brain barrier permeability and pharmacokinetic analysis, including various descriptors, screened 27 natural compounds. The docking of candidate compounds identified 3 compounds with good docking energies and stable interactions when compared to the reference drug roscovitine. Further, molecular dynamics simulation analysis including root mean square deviation, root mean square fluctuation, the radius of gyration and binding energies for 50ns revealed two compounds. Overall, we recommend two natural products apigenin and salvianolic acid as candidate inhibitors to target cyclin-dependent kinase-5 for treating Alzheimer's and Parkinson's disease.**

Keywords- Cyclin-dependent kinase-5, Alzheimer's disease, Parkinson's disease, Molecular docking, Molecular dynamics simulations

I. INTRODUCTION

Cyclin-dependent kinases (CDKs), the serine/threonine kinases, are the regulators of the cell cycle, preferentially expressed during mitosis. Unlike other Cdks, Cdk5 remains inactive during cell cycle progression and is activated in post-mitotic neurons [1]. CDK5 is considered as a crucial regulator of neuronal differentiation, survival and synaptic functions and is found to be associated with etiopathology of various neurodegenerative disorders (NDDs)[2]. In the case of Alzheimer's disease (AD), CDK5 contributes to the development of amyloid plaques and hyperphosphorylated neurofibrillary tangles, as well to other pathological events like mitochondrial dysfunction [3]. Likewise, in Parkinson's disease (PD), CDK5-dependent activation is responsible for dopaminergic neuronal cell death and activation of inflammosomes in substantia nigra [4].

CDK5 is considered an attractive therapeutic target for different NDDs and recently, many studies have been conducted to identify novel Cdk5 inhibitors with neuroprotective properties. For instance, indolinone-A has been known to protect neurons against various insults and promotes neuronal survival and synaptic plasticity. Similarly, Cdk5 inhibitor roscovitine has been reported to cross blood-brain barrier and provides neuroprotective effects against the microglial activity, cytokine production and inflammation-induced degenerative phenotype. A recent study has established the neurorestorative properties of Truncated Peptide 5 (TP5) as an inhibitor of CDK5/p25 in dopaminergic neurons [5]. Similarly, several other studies have also highlighted the potential of CDK5 inhibitors, however, none of these compounds have entered clinical trials. The most prominent one is roscovitine which has shown protective effects in various cancers but not in NDDs.

Nonetheless to potentiate the neuroprotective effects of CDK5, novel drugs need to be designed. To this end, we opted for a computational approach to identify natural compounds as potent CDK5 inhibitors.

II. METHODOLOGY

A. Data collection

The information of natural compounds was collected from the recently available literature. The natural compounds with reported anticancer activity were identified and selected for the study [6] [7]. The structural information of all the natural compounds in sdf (structure-data file) format was retrieved from PubChem database.

B. BBB permeability analysis

All the identified natural compounds were analyzed for their BBB permeability by using the online BBB prediction server- Cbligand. It is a free web tool to predict the BBB permeation of small molecules.

C. Molecular docking studies

For AD and PD, the structure of CDK5 (PDB ID 1UNL, at a resolution of 2.20 Å), was obtained from the protein data bank (PDB). CDK5 contained roscovitine as a bounded ligand which served as a control for comparing docking energies and ligand interactions. The ligand files (SDFs) were extracted from PubChem database and converted to PDBQT files by using OpenBabelGUI. Study of receptor-ligand docking was conducted with Autodock vina with 40 x 40 x 40 box size. The grid points are centered on the ATP binding site of CDK5, with a 0.375 Å spacing value, to cover most of the protein and the active sites. In order to obtain a clean protein structure, functional ligands, water molecules were removed from the structure of human MAO B. In this study, the active sites were predicted by redocking the crystallized ligand to the ATP binding site of CDK5. The analysis of docked complexes was performed by using Biovia Drug discovery studio visualizer 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 23,2022 at 05:11:57 UTC from IEEE Xplore. Restrictions apply.

D. Molecular dynamics simulation

To understand the stability of protein-ligand complexes, molecular dynamic studies were conducted on complex files with the best binding energy. A molecular dynamics simulation of a selected protein-ligand complex was performed with Gromacs-2019. The forcefield coordinates were obtained by downloading the ligand topology from the Prodrug server. A steepest descent algorithm was used for pre-processing the system using 1500 steps of vacuum minimization. The simulation system was prepared in octahedral box, hydrated by TIP3P water model and the neutralization of counterions was done by 0.1M NaCl solution. The final production run of each structure from the equilibration phase was conducted using an ensemble of NPT (Number of atoms in the system, pressure of the system and temperature of the system) simulations for 50ns. Furthermore, the trajectories in Gromacs were utilized to conduct various structural analysis like root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg) and Coulombic energy (binding energy) for the period of 50ns.

E. Pharmacokinetic analysis and toxicity profile prediction

The bioavailability of a drug is influenced by different physicochemical properties like molecular size, lipophilicity, water solubility and permeability as determined by Lipinski's rule of five. Similarly, the prediction of ADME (absorption, distribution, metabolism, excretion) properties is necessary for drug development and prioritization. The prediction of various physicochemical parameters and ADME (absorption, distribution, metabolism, excretion) properties was done by the online tool SwissADME, and toxicity profiling was performed by the Pro-Tox server.

III. RESULTS

A. Identification of natural compounds and BBB permeability analysis

We retrieved 77 natural compounds with reported anticancer activity from the literature. BBB permeability analysis resulted in 27 compounds with BBB penetrant activity and can be used for further analysis. We selected 0.02 value as threshold and all the compounds with higher scores were considered as BBB penetrant drugs. The BBB scores of various compounds varies from 0.02 to 0.18 with allicin showing highest score and ursolic acid showing lowest score.

B. Binding mechanism of CDK5 inhibitor compounds

Understanding of physical interactions of protein and ligands is an essential step in computational drug designing. To understand the binding affinities of query ligands to the protein, we first docked the reference drug roscovitine to CDK5 and identified the interacting residues. The substrate binding site of CDK5 was analyzed and presented in **Figure 1(a).** We found that roscovitine interacted with CDK5 with a docking score -8.7 Kcal/mol. The drug positioned itself in the hydrophobic pocket of CDK5 consisting of Ile10, Gly11, Glu12, Gly13, Val18, Ala31, Lys33, Val64, Phe80, Phe82, Asp84, Gln85, Asp86, Lys89, Gln130, Leu133, Ala143, Asn144 residues where there is one H-bond formed with Cys83 residue with a distance of 2.81Å. The docking analysis of all the 27 natural compounds with CDK5 with docking scores and interacting residues are represented in **Table 2**.

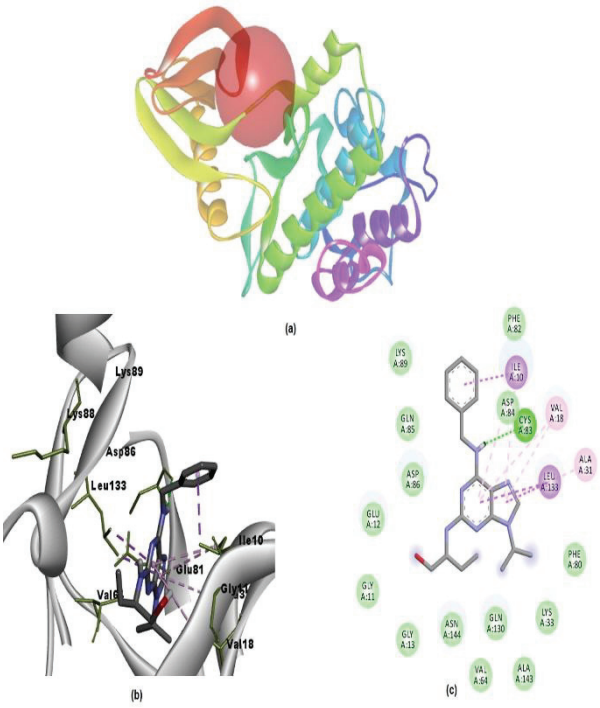


Fig. 1. (a) Structure of CDK5 chain A is shown where the red circle represents the substrate binding site. (b) 3D structure of the binding mode of roscovitine with CDK5 (c) 2D structure of the binding pattern of the reference drug roscovitine with CDK5. Dark green lines interactions represent H-bonds, light green represents Vander Waals, pink represents alkyl and pi-alkyl, and blue represent pi-sigma interactions.

We observed that three compounds apigenin, salvianolic acid and silymarin had good docking scores and also these compounds interacted with CDK5 through important catalytic residues. We found two compounds crytotanshinone and tanshinone had more docking scores compared to reference drug but had no significant interactions. Thus we omitted these two compounds from further analysis. H-bond interaction analysis had shown that apigenin formed two H-bonds with CDK5 while salvianolic acid and silymarin formed four and two H-bonds, respectively. We have seen that Ile10, Val18, Ala31, Val64, Phe80, Glu81, Phe82, Cys83, Lys89, and Asn144 were the common interacting residues among all the three compounds and are also present in the experimental structure in CDK5-roscovitine interaction. The 2D and 3D interactions of roscovitine and the best three compounds with CDK5 are presented in **Figure 1 (b),(c) and figure 2.**

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 23,2022 at 05:11:57 UTC from IEEE Xplore. Restrictions apply.

TABLE I. DOCKING DETAILS OF THE COMPOUNDS ALONG WITH THE REFERENCE DRUG INCLUDING DOCKING ENERGIES, H-BONDS AND OTHER. THE

Compound	Docking energies	H-bond with distance	Non polar interactions
Roscovitine (Reference)	-8.7	**Cys83** (2.81Å)	Val18, Ala31, Ile10, Leu133 , Gly11, Glu12, Gly13, Lys33, Val64, Phe80, Phe82, Asp84, Gln85, Asp86, Lys89, Gln130, Ala143, Asn144
Apigenin	-8.5	Glu81 (2.61 Å) **Cys83** (2.91 Å)	Val18, Ala31, Ile10, Leu133 , Lys33, Val64, Phe80, Phe82, Asp84, Gln85, Asp86, Lys89, Asn144
Allicin	-6.6	-	Ile10, Ala31, Lys33, Val64, Phe82, Cys83, Leu133 , Gly11, Phe80 , Glu81, Ala143, Asn144
Borbonol	-6.6	Asp86 (2.10 Å)	Ile10, Val18, Ala31, Val64, Phe80, Leu133, Ala143, Gly13 , Phe80 , Lys33, Glu81, Phe82, Cys83, Gln85, Gln130, Asn144
Beta-elemene	-6.9	-	Ile10, Ala31, Lys33, Val64, Cys83, Val18, Leu133, Gly11, Glu81, Asp86, Asn144
Chamomillol	-7.7	Asn144	Val18, Ala31, Leu133, Ile10, Glu12, Gly13, Val64, Phe80, Cys83, Asp86, Ala143
Cryptotanshinone	-10.3	-	Ala31, Leu133, Gly13, Val18, Phe80 , Ile10, Glu12, Lys33, Val64, Glu81, Phe82, Cys83, Asp86, Gln130, Ala143, Asn144
Cepharanthine	-8.4	-	Lys89, Asp86, Ile10, Gly11, Glu12, Gly13, Thr14, Val18, Gln85, Lys88, Gln130, Leu133, Asn144
Capsaicin	-6.8	**Cys83** (2.37Å)	Val18, Lys33, Lys89, Gln85, Ile10, Asp86, Phe80, Glu12, Gly13, Ala31, Val64, Phe82, Asp84, Gln130, Leu133, Ala143, Asn144
Gingerol	-7.1	Asn144 (3.08Å)	Val18, Ala31, Lys33,Lys89, Leu133, Val64, Phe82, Cys83, Asp84, Gln85, Asp86, Ala143
Genipin	-6.9	Glu81 (2.04Å) Asn144 (3.09Å)	Ile10, Val18, Ala31, Leu133, Gly13, Lys33, Val64, Phe80, Phe82, Cys83
Kaempferol	-7.8	Lys33 (3.35Å)	Ile10, Ala31, Asp86, Val18, Leu133, Gly11, Glu12, Gly13, Thr14, Val64, Glu81, Phe80, Phe82,Gln130, Asn144
Melatonin	-7.2	-	Val64, Ala143, Ile10, Ala31, Val64, Val18, Phe80, Leu133, Cys83, Asn144
Noscapine	-7.4	Asp86 (3.24 Å)	Ala31, Phe82, Cys83, Val18, Ile10, Leu133, Gly11, Glu12, Gly13, Asp84, Gln85, Gln130, Asn144
Paradol	-7.1	**Cys83** (2.00 Å) Asp86 (3.03 Å) Asn144 (2.13 Å)	Val18, Ala31, Lys33, Val64, Lys89, Ala143, Phe80, Ile10, Leu133, Glu81, Phe82, Asp84, Gln85
Pterostilbene	-6.8	-	Ile10, Lys33, Val64, Phe82, Ala143, Asn144, Val18, Ala31, Phe80, Gly11, Glu12, Glu81, Asp86, Gln130
Resveratrol	-7.5	Glu81 (2.13 Å) Asp84 (3.08 Å)	Val18, Ala31, Ile10, Leu133, Phe80 , Val64, Cys83, Gln85, Asp86, Lys89, Asn144
Sulforaphane	-4.5	Asn144 (3.25 Å)	Ile10, Leu133, Val18, Val64, Phe80, Cys83, Ala143
Safranal	-6.8	Asn144 (2.23 Å)	Val18, Ala31, Val64, Leu133, Ile10, Lys33, Phe80, Glu81, Phe82, Cys83, Ala143
Salvianolic acid	-9.3	Glu12 (2.06 Å) Glu81 (2.36 Å) Asp86 (3.32 Å) Asn144 (2.94 Å)	Val18, Ala31, Phe80, Ile10, Gly13, Thr14, Lys33, Val64, Phe82, Cys83, Asp84, Gln85, Lys89, Lys128, Gln130, Asn131
Silymarin	-8.7	Glu81 (2.46 Å)	Lys88 , Gly11, Val18, Ala31, Val64, Lys88, Lys89 , Asp86 , Ile10, Leu133, Glu12, Glu51, Phe82, Cys83, Asp92, Ala143, Asn144
Shogaol	-4.5	**Cys83** (2.42 Å) Asn144 (3.11 Å)	Val18, Val64, Lys33, Ala143, Leu133, Ile10, Phe80 , Gly13, Ala31, Phe82, Asp84, Gln85, Asp86
Salvicin	-7.5	Asp86 (2.72 Å)	Val18 , Ile10, Gly11, Glu12, Glu13, Ala31, Lys33, Glu51, Val64, Phe80, Cys83, Gln130, Leu133, Ala143
Thymol	-6.9	-	Val18, Ala31, Cys83, Leu133, Lys33, Val64, Phe80, Glu81, Phe82, Ala143
Thymoquinone	-7.1	**Cys83** (3.01 Å)	Ile10, Ala31, Val18, Phe80, Leu133, Lys33, Val64, Phe82, Ala143, Asn144
Triptolide	-8.1	-	Gly13, Ile10, Gly11, Gly12, Val18, Cys83, Gln85, Asp86, Lys128, Gln130, Asn131, Leu133, Asn144
Tanshinone	-9.7	-	Asp84, Lys89, Ile10, Ala31, Lys33, Val18, Leu133, Val64, Phe80, Glu81, Cys83, Gln85, Asp86, Asn144
Ursolic acid	-8.0	Gln282 (3.11 Å) Phe289 (3.12 Å)	Phe108, Lys112, Leu281, Pro292, Glu278, Phe286, Ser287, Asp288

SIGNIFICANT H-BONDS ARE HIGHLIGHTED IN BOLD.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 23,2022 at 05:11:57 UTC from IEEE Xplore. Restrictions apply.

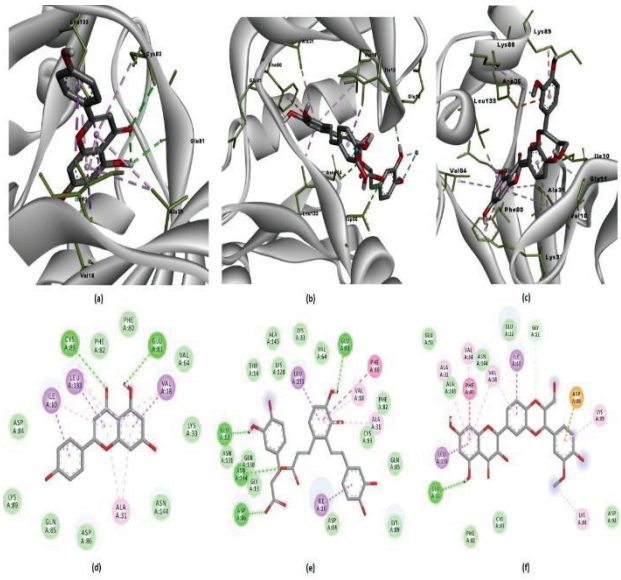


Fig. 2. (a) and (d) 3D and 2D structures of the binding pattern of apigenin with CDK5 (b) and (e) 3D and 2D structures of the binding pattern of salvianolic acid with CDK5 (c) and (f)) 3D and 2D structures of the binding pattern of silymarin with CDK5. Dark green lines interactions represent H-bonds, light green represents Vander Waals, pink represents alkyl and pi-alkyl, and blue represent pi-sigma interactions. Dark green lines interactions represent H-bonds, light green represents Vander Waals, pink represents alkyl and pi-alkyl, dark pink represents pi-pi interactions, blue represent pi-sigma interactions and orange represents anionic interactions.

C. *Structural deviation, flexibility, compactness and binding free energy analysis*

To understand the binding pattern and molecular interactions of the selected natural compounds with CDK5, molecular dynamics simulation was performed for 50ns time period. We found that among the 27 docked complexes, three compounds apigenin, salvianolic acid and silymarin were the best candidates after simulation analysis. As shown in figure 3 (a), the RMSD analysis of protein backbone atoms, roscovitine (reference drug) and the three natural compounds are compared. We found that all the systems were remain in steady state and did not deviate more than 0.3 nm. Further, all the three compounds had less deviation as compared to the reference drug and there were comparable fluctuations for the three candidate inhibitor compounds. To understand the residue flexibility of all the systems, RMSF analysis was carried out for the complete simulation period. From Fig 3 (b), it was clear that all the systems were fluctuating differently, with apigenin had the lowest fluctuation. We found that for roscovitine residues 220-250, for salvianolic acid residues 150-158, and for silymarin residues 285-290 were showing variations. For compactness analysis, we performed Rg analysis of all the systems. As shown in **figure 3 (c),** roscovitine and apigenin have shown less fluctuation while salvianolic acid and silymarin have higher values of Rg indicative of slightly labile nature of the systems. The binding affinity of various inhibitors to CDK5 protein was done by comparing the Coulomb's energy of various inhibitors with roscovitine as presented in **figure 4 (a).** We found that the average Coulomb's energy of both apigenin and salvianolic acid were higher than roscovitine while for silymarin the energy value was lower.

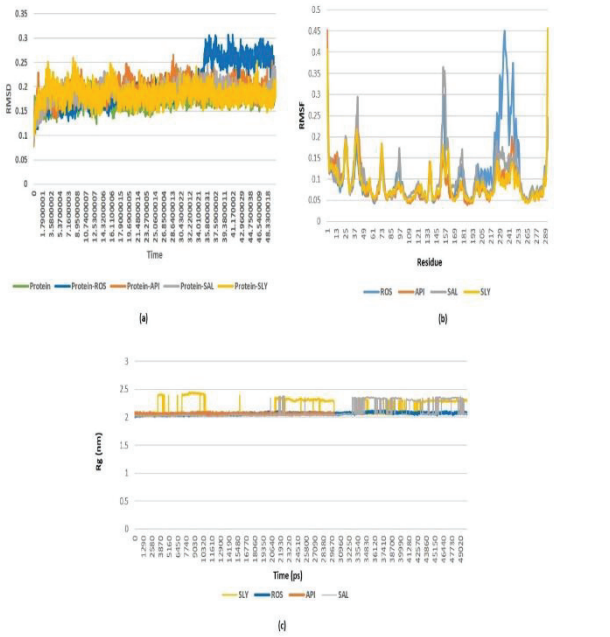


Fig. 3. (a) Root mean square deviation (RMSF) analysis of the protein, protein-reference drug and protein-natural compounds for 50ns time period. (b) Root mean square fluctuation (RMSF) analysis of all the three protein-inhibitor complexes versus protein-reference drug complex. (c) Radius of gyration (Rg) analysis showing compactness of protein-inhibitor complexes compared to protein-reference drug complex. The color bars are represented in the figures to denote various complexes.

D. *Assessment of pharmacokinetic properties and toxicity profile*

The prediction of physiochemical parameters and determination of ADMET properties are considered as crucial factors in drug designing process.The important physicochemical features including molecular weight (MW), number of hydrogen bond donors (HBD) and acceptors (HBA), number of rotatable bonds, and topological polar surface area (TPSA), and various ADMET descriptors were estimated for all the 27 compounds along with the reference drug. Furthermore, the physicochemical parameters of the selected compounds apigenin, silymarin and salvianolic acid were compared to roscovitine as presented in **Figure 4 (b)**. We found that the values of physicochemical properties were higher than threshold for salvianolic acid and silymarin but for apigenin the scores were acceptable within the threshold range**.** In keeping with the rule of five all the three compounds were in accordance with the set parameters and no violation was found. Likewise, the ADME parameters including water solubility, octanol/water partition coefficient (log P), gastrointestinal absorption, and others were found to be suitable. The toxicity (hepatotoxicity, immunogenicity and mutagenicity) analysis has shown that both apigenin and salvianolic acid were non-toxic while apigenin was moderately toxic.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 23,2022 at 05:11:57 UTC from IEEE Xplore. Restrictions apply.

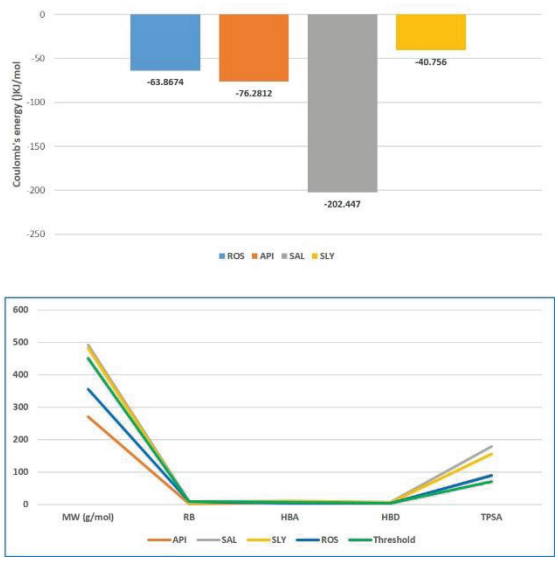


Fig. 4. (a) Contribution of Coulombs' energy to binding free energy of various inhibitors with the protein along with the reference drug. (b) Distribution of various physicochemical parameters such as molecular weight (MW), number of rotatable bonds (RB), number of H-bond donors (HBD), number of H-bond acceptors (HBA) and total polar surface area (TPSA)

IV. DISCUSSION

CDKs are Ser/Thr kinases which are highly expressed in proliferating cells. Unlike different classic CDKs, CDK5 does not participate in cell cycle progression and plays vital role in post-mitotic neurons. A plethora of studies have considered CDK5 as an attractive target for various neurological disorders that involve defective learning and memory functions [8]. A study has highlighted the pharmacological importance of CDK5 for the treatment of brain inflammation evoked by Aβ toxicity. Likewise, a study has reported that CDK5 inhibitory peptide (CIP) protected dopaminergic neurons from MPTP induced toxicity and also alleviated behavioral deficits. Although a number of known CDK5 inhibitors are available and are tested in AD and PD but none of them have shown clinical success due to poor availability and low specificity. The current study aimed to identify some natural compounds as CDK5 inhibitors for the treatment of AD and PD. We developed a computational pipeline to screen natural compounds by molecular docking and simulation studies. We used roscovitine as the reference drug as it has been known to cross BBB and has shown neuroprotective functions and alleviation of neuronal death in various brain diseases.

The substrate binding site of CDK5 comprised of a Gly-rich loop, an activation loop and the hinge region. A deeper analysis suggested that Lys33, Phe80, Glu81, Cys83, and Asn144 are the key residues regulating ATP binding and all CDK5 inhibitors occupy the ATP-binding site of the protein. It has been reported that roscovitine strongly binds to CDK5 through intermolecular H-bonds involving residues Cys83, Asp86 and Gln130. The interactions of all the identified natural compounds with CDK5 was analyzed by applying molecular docking approach. We identified apigenin, salvianolic acid and silymarin as potent compounds with good docking energies and significant interactions when compared to the reference drug. We observed that apigenin interacted with CDK5 by forming two H-bonds and four hydrophobic interactions. A detailed mechanistic analysis revealed that both carbonyl and hydroxyl oxygens were involved in H-bonding while benzene and pyran rings were majorly involved in hydrophobic interactions such as pi-alkyl and pi-sigma interactions. Likewise, in salvianolic acid, three hydroxyl oxygens and one carbonyl oxygen were involved in H-bonding. The compound had five hydrophobic interactions where phenol ring had the major contribution. For silymarin, one H-bond and ten different hydrophobic interactions were obtained. We found that one hydroxyl oxygen of was interacting with Glu81.

To further validate our discovery, 50ns simulations were carried out on the selected docked compounds along with protein and reference drug. The RMSD analysis of all the systems revealed that the reference drug has shown higher amplitude than candidate inhibitor compounds. We presume that the candidate compounds may slightly may be more stable than the reference drug. The RMSF analysis indicated that all the residues were fluctuating less over the simulation time. The compactness analysis revealed that all the three compounds remained stable over the simulation time when compared to the reference drug. Further, apigenin was found to be more stable with comparable values of gyration with the reference drug. The binding free energy analysis confirmed the stable interactions of candidate inhibitor drugs with CDK5 where silymarin was having lower energy values.

V. CONCLUSION

Pharmacological inhibition of CDK5 is an attractive strategy to combat AD and PD. Our study has employed a computational approach including molecular docking and molecular dynamics simulations to identify potent natural compounds as CDK5 inhibitors. The information of molecular interactions of roscovitine with CDK5 was used to screen natural compounds. The information of natural compounds was extracted from the literature then the compounds were screened based on BBB permeability and ADMET analysis. The interaction analysis revealed that three compounds apigenin, salvianolic acid and silymarin formed stable H-bonds and further confirmed by molecular dynamics simulation analysis. The binding energy analysis confirmed the affinities of apigenin and salvianolic acid with CDK5. Altogether, from docking, molecular dynamics simulations and energy calculations, we proposed apigenin and salvianolic acid as potent CDK5 inhibitor compounds that can be verified experimentally for the management of AD and PD.

ACKNOWLEDGMENT

We would like to thank the senior management of Delhi Technological University (DTU) and the Department of Biotechnology (DBT), India's Government, for their constant support and financial assistance

REFERENCES

[1] L. H. Tsai, T. Takahashi, V. S. Caviness, and E. Harlow, "Activity and expression pattern of cyclin-dependent kinase 5 in the embryonic mouse nervous system," *Development*, vol. 119, no. 4, pp. 1029–1040, Dec. 1993.

[2] Z. Cheung, N. I.-N. letters, and undefined 2004, "Cdk5: mediator of neuronal death and survival," *Elsevier*.

[3] "Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles - Google Search." [Online]. Available: https://www.google.com/search?q=Aberrant+Cdk5+activation+by+p25+triggers+pathological+events+leading+to+neurodegeneration+and+neurofibrillary+tangles&oq=Aberrant+Cdk5+activation+by+p25+triggers+pathological+events+leading+to+neurodegeneration+and+ne

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 23,2022 at 05:11:57 UTC from IEEE Xplore. Restrictions apply.

urofibrillary+tangles&aqs=chrome..69i57.354j0j7&sourceid=chrome&ie=UTF-8. [Accessed: 18-May-2022].

[4] P. Zhang *et al.*, "Cdk5-Dependent Activation of Neuronal Inflammasomes in Parkinson's Disease," *Mov. Disord.*, vol. 31, no. 3, pp. 366–376, Mar. 2016.

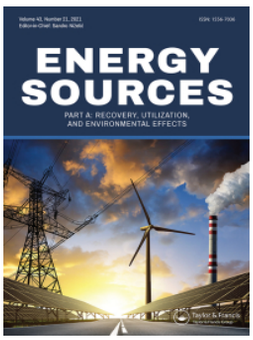
[5] J. Tran *et al.*, "Therapeutic effects of TP5, a Cdk5/p25 inhibitor, in in vitro and in vivo models of Parkinson's disease," *Curr. Res. Neurobiol.*, vol. 2, p. 100006, Jan. 2021.

[6] P. Garcia-Oliveira *et al.*, "Status and challenges of plant-anticancer compounds in cancer treatment," *Pharmaceuticals*, vol. 14, no. 2, pp. 1–28, 2021.

[7] C. A. Dehelean *et al.*, "Plant-Derived Anticancer Compounds as New Perspectives in Drug Discovery and Alternative Therapy," *Molecules*, vol. 1096, pp. 1–29, 2021.

[8] K. Shah and D. K. Lahiri, "Cdk5 activity in the brain - multiple paths of regulation," *J. Cell Sci.*, 2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 23,2022 at 05:11:57 UTC from IEEE Xplore. Restrictions apply.



Energy Sources, Part A: Recovery, Utilization, and Environmental Effects

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ueso20

Control algorithm for coordinated operation of wind-solar microgrid standalone generation system

Sombir Kundu, Madhusudan Singh & Ashutosh K. Giri

To cite this article: Sombir Kundu, Madhusudan Singh & Ashutosh K. Giri (2022) Control algorithm for coordinated operation of wind-solar microgrid standalone generation system, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 44:4, 10024-10044, DOI: 10.1080/15567036.2022.2143943

To link to this article: https://doi.org/10.1080/15567036.2022.2143943

Published online: 11 Nov 2022.

Submit your article to this journal

Article views: 42

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
2022, VOL. 44, NO. 4, 10024–10044
https://doi.org/10.1080/15567036.2022.2143943

Taylor & Francis
Taylor & Francis Group

Check for updates

Control algorithm for coordinated operation of wind-solar microgrid standalone generation system

Sombir Kundu[a], Madhusudan Singh[a], and Ashutosh K. Giri[b]

[a]Department of Electrical Engineering, Delhi Technological University, Delhi, India; [b]Government Engineering College, Bharuch, India

ABSTRACT

In this paper, a wind-driven self-excited induction generator (SEIG) combining solar photovoltaic (SPV) and battery storage has been successfully demonstrated. A second-order sequence filter-based frequency locked loop (SOSF-FLL) control is implemented for operating the voltage source converter (VSC). The SOSF-FLL method is designed by replacing two second-order low pass filter (LPF) pure integrators with two reduced order generalized integrators (ROGIs). The SOSF-FLL has a nonlinear control feedback system. The major contributions in this study include: 1. Implementation of SOSF-FLL control to generate reference source currents for VSC and evaluate the amplitude of load current component under intermittent condition of renewable energy sources (RESs) and nonlinear load, 2) The control algorithm with VSC performs numerous functions such as voltage/frequency regulation, load leveling, harmonic reduction, active/reactive power compensation and reduces the overall power quality problems, 3) Through DC-link voltage management, a bidirectional converter (BDC) with battery storage optimizes the best power flow to provide power balancing between the sources and the unbalanced load, 4) Under diverse operating situations, an incremental conductance (INC)-based maximum power point tracking approach with BDC control provides quick system convergence and support frequency. Moreover, a comparison is carried out between SOSF-FLL control and conventional control algorithms. The settling time of the proposed control is very less than one cycle (20 ms) within 2% described limit, whereas the conventional control algorithms show oscillations of more than 5% during the load dynamics. Despite the load current THD 24.41%, the proposed control is able to obtain supply current THD at 4.23%, which is less than 5% as per IEEE-519 standard. The proposed microgrid (MG) is constructed in MATLAB/Simulink and experimental prototype environment, and its dynamic response is investigated under dynamic conditions of RESs and nonlinear load conditions.

ARTICLE HISTORY
Received 10 December 2021
Revised 12 October 2022
Accepted 13 October 2022

KEYWORDS
Battery system; bi-directional converter; harmonic analysis; microgrid; Second-order sequence filter-based frequency locked loop

Introduction

Renewable Energy Sources (RESs) are paving the way to a greener world that can make us to rely less on fossil fuels and reduce greenhouse gas emissions. Wind and solar energy have witnessed the fastest growth among RESs in the past four decades. They are becoming less expensive than fossil fuels due to advancements in manufacturing technology. The inclusion of RESs has been described as a pivotal role to diminishing pollutant emissions and improving living environment quality (Hoang, Viet Pham, and Nguyen 2021). Wind energy is clean, abundant and available around the world. Wind driven electricity generation can supplement an existing utility supply as well as meet the demand of electricity in isolated applications. The solar photovoltaic (SPV) system is one of the faster growing commercial RES's and researcher estimates that with advancement in solar technologies world can achieve 100% energy

CONTACT Sombir Kundu sskundu46@gmail.com Department of Electrical Engineering, Delhi Technological University, Delhi, India

© 2022 Taylor & Francis Group, LLC

production by 2050 (Kumar et al. 2020; Roy et al. 2022). A new VSC-based controller is developed to maintain the voltage profile and system frequency of islanded SEIG during changes in wind velocity and load demand, in conjunction with a battery system (Kasal and Singh 2011). A SEIG integrated with a PV system is also developed to support system frequency through real power compensation under varying load demands (Sombir and Singh 2020). The gating pulses for VSC generated using 1-phase synchronous d-q frame theory are described. The compensator reduces the harmonics, produced by consumer loads and provides reactive power support (Singh and Chilipi 2014). An FPGA-based closed-loop technique is used for peak point power operation of the SEIG. The 3-phase generator terminals and 1-phase grid are interfaced with a 2-stage power converter (Arthishri, Kumaresan, and AmmasainGounden 2019). For standalone power generation, various VSC schemes based on voltage regulators are discussed (Kasal and Singh 2009). A modified least mean square control based on adaptive theory is proposed to maintain VSC for power quality enhancement. This VSC-based compensation provides load leveling, compensation of neutral current, and load current distortion in wind driven SEIG system at standalone application with single phase loads (Giri, Raj Arya, and Maurya 2019). The battery system is used for dual purpose; it regulates voltage profile and frequency at the common point of interface and ensures wind generator synchronization and disconnection during wind variation by compensating for real and reactive power. In different operating modes, the interfacing converter and voltage feedback loop provide a continuous power supply to critical loads (Chauhan et al. 2019). The effect of sudden load changes on voltage and frequency are controlled by the battery system and the frequency and voltage profile of the microgrid (MG) are regulated quickly. Together with additional services such as support for reactive power, mitigation of generator current harmonics and decrease of voltage harmonics at PCI, the voltage profile and frequency control makes the real power balance (Nižetić et al. 2021). A photovoltaic thermal-based phase change materials approach for thermal management is described to enhance the efficiency of solar collectors (Kewat, Singh, and Hussain 2018). An INC MPPT control has a fast convergence speed and good tracking accuracy in different operating conditions (Sharma, Kewat, and Singh 2020). A BDC control is utilized to maintain DC-link voltage and provide active power balance under load perturbations (Chandran, Kewat, and Singh 2020; Kewat and Singh 2020).

In wind-based distributed generation systems, the LAF algorithm neutralizes the effect of wind variations on the generator's voltage profile and frequency. The LAF-based VSC improves the system's power quality; controls the voltage profile and frequency in the permissible range and provides load balance, compensation of neutral current, harmonic reduction, and support for reactive power (Giri et al. 2018). Different types of FLL techniques (DROGI-FLL, DSOGI-FLL) are implemented and their LTP modeling for different time domain signal decomposition methods is developed (Saeed Golestan et al. 2021). A DSP-based generalized impedance controller for synchronization operation of variable speed standalone SEIG is discussed. The frequency is followed in the system using digital PLL implemented on dSPACE (Chauhan et al. 2015). For the estimation of the phase, frequency and peak value of the 1-phase grid voltage, a FLL-based nonlinear technique is utilized (Ahmed, Amamra, and Bierhoff 2008). An LMMN control method is developed for harmonic eradication and evaluates amplitude of load current component (Chishti, Murshid, and Singh 2019). A harmonic decoupling network based on MSOGI-FLL is used to regulate the frequency of the system with MSOGI-FLL (Rodriguez et al. 2011). Using 3-phase alpha-beta transformation, the fundamental component extraction of detected load current is presented using a VCO-less PLL control. This method significantly enhances the power quality of the distributed generation system (Giri et al. 2019). An isolated MG system is developed for regulating voltage profile and frequency using a generalized filter-based control approach and enhancing the quality of power (Seema and Singh 2017). A CDSC+DSOGI-FLL method gives a fast response with small overshoots, cancels undesired negative and harmonic components, and removes unreal disturbance in case jump of a phase angle in the estimated frequency (Simek, Skramlik, and Valouch 2019). An ESOGI-FLL is utilized for an accurate evaluation of grid parameters like phase angle, amplitude, and operating frequency. ESOGI-FLL provides DC-offset rejection in a 1-phase VSC for both isolated and grid-connected mode (Kherbachi, Chouder, and Bendib 2019). FLL estimates the operating frequency and reduces the high-frequency noises produced by differential operators (Fang et al. 2019). To improve the power quality, an ASMC method is used. An ASMC evaluates the reference active power of the

designed system to regulate the frequency and balance the energy in the battery system for an isolated system (Kumar Kalla et al. 2018). The fundamental load current components are extracted using CC-ROGI-based FLL to generate the reference grid current for the VSC. A VSC-based CC-ROGI-FLL capable of harmonic elimination, enhancement of power factor and grid current leveling (Kumar et al. 2020) and coordination control of robust normalized mixed norm control algorithm-based wind-solar-BES MG system is designed for standalone applications (Chishti, Murshid, and Singh 2020).

As also available in the literature with the increase in the number of power quality problems in recent years, custom power devices for monitoring and compensation have received much interest. Furthermore, advanced power converter-based RESs are expected to supply auxiliary services to assist reduce power quality problems. In such conditions, it is preferable to use time-domain signal decomposition and SOSF-based FLL algorithms. Nonlinear feedback control schemes with two or more frequency-adaptive filters that are dynamically interactive are used in this control algorithm. A second-order sequence filter is built by restoring two integrators of a second-order LPF with two ROGIs (Rajput et al. 2019; Saeed Golestan et al. 2019).

In this article, the author's main contributions are as follows:

The solar-wind hybrid energy systems (HES's) are becoming popular as standalone power generation for providing electricity in isolated places and hilly areas, especially where grid supply is unavailable. The hybrid energy systems provide improved system availability and greater balance in energy supply and demand and hence voltage and frequency stability in isolated applications. Many control strategies exist for voltage and frequency regulations in standalone energy systems (Chauhan et al. 2019; Giri et al. 2019; Kewat, Singh, and Hussain 2018; Kumar Kalla et al. 2018).

(1) In the proposed wind-solar MG, a second-order sequence filter-based frequency locked loop (SOSF-FLL) is developed for the estimate of phase, frequency, and evaluates the amplitude of fundamental load current component.
(2) A VSC with SOSF-FLL provides harmonic eradication, load leveling, real/reactive power support, regulate voltage/frequency of wind-driven SEIG and enhances the power quality of MG system in isolated applications under different operating conditions.
(3) An INC MPPT control draws peak power from a SPV under the intermittent condition of solar insolation. By giving compensation for real/reactive power through VSC, the BDC control technique maintains the charging or discharging mode of the battery system, as well as the system frequency and voltage profile of the standalone MG system.
(4) A lithium ion battery bank is used as battery storage to consume extra power at the DC interface and balance energy exchange during critical load and scarce generation of RESs.
(5) A comparative assessment is carried out to demonstrate that the proposed control shows less steady-state oscillations and better harmonics abatement capability as compared to the existing control approach.

System configuration

The main components of the standalone 3-phase 3-wire wind-solar MG system have been depicted in Figure 1. The proposed wind-solar MG consists of a three-phase wind turbine driven SEIG, 3-phase capacitor bank, a PV array with boost converter INC MPPT control, a VSC, a three-phase nonlinear load and battery storage. A SOSF-FLL estimates the peak value of active load current component and improves the performance of MG in distorted conditions. To interface the point of common coupling (PCC) with the AC side of the VSC, filter inductors are used. An appropriate value of three-phase capacitor enables the voltage buildup at no load across the SEIG terminal. At the point of common coupling, a 3-leg VSC is interfaced in parallel to load for providing reactive power compensation, harmonic mitigation, load leveling and improved power quality with varying three-phase nonlinear load. The VSC is made of six isolated bipolar transistors that are connected in the three legs to form a bridge configuration. The SPV system

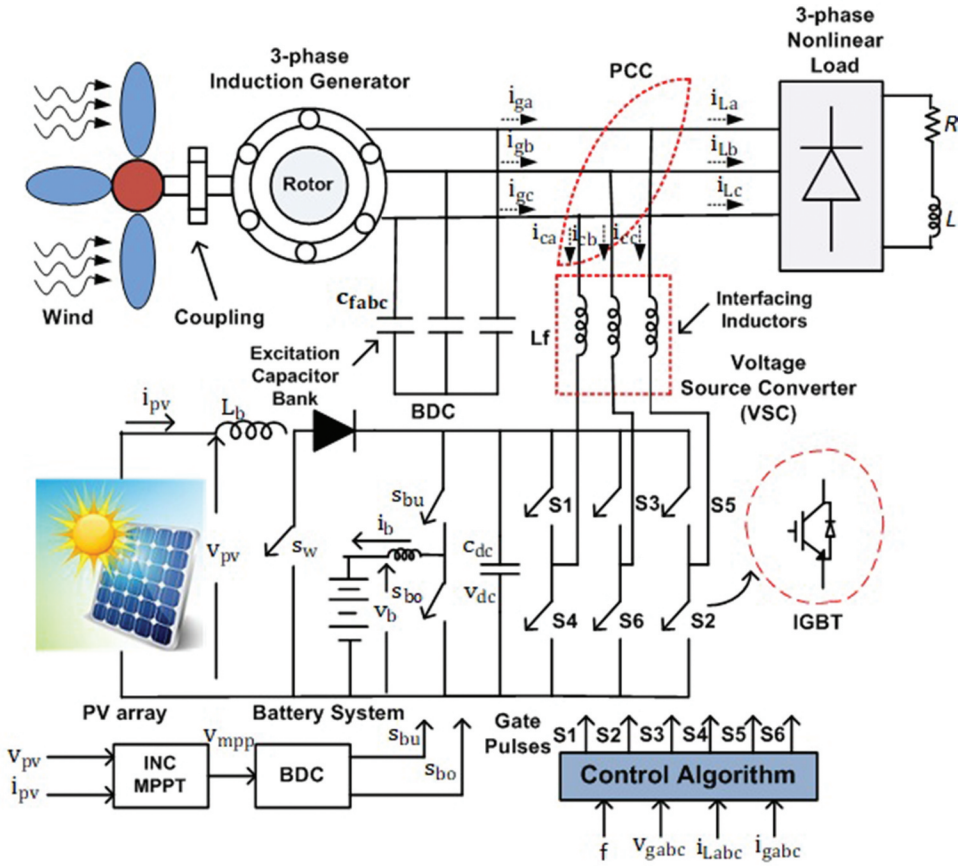


Figure 1. Standalone wind-solar MG configuration.

consists of a SPV and a BDC converter that has cleared the way for maximum PV power to be extracted for a shared DC link.

SOSF-FLL control algorithm

The effectiveness of the proposed wind-solar MG (MG) depends on the SOSF-FLL control algorithm that is utilized to regulate frequency, voltage and enhance the overall quality of power under varying wind velocity, solar insolation, and nonlinear load. However, the performance of SOSF-FLL is superior during frequency changes, as it provides accurate phase angle estimation. The overview of the control technique is given as below:

SOSF-FLL for fundamental extraction of load current and phase angle

The proposed SOSF-FLL control scheme is depicted in Figure 2. The SOSF-FLL was built to extract and evaluate the amplitude of fundamental load current component. Complex gain constants are applied between the alpha-beta axes to improve the performance of proposed control. The gain parameters (K and λ) are used to control the SOSF-FLL algorithm. The transfer function of the proposed control algorithm is illustrated as

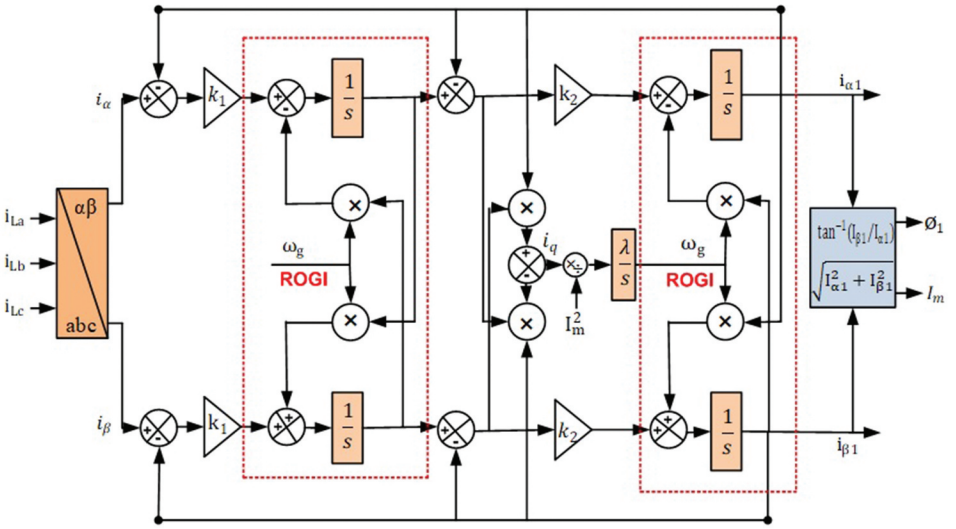


Figure 2. SOSF-FLL control algorithm.

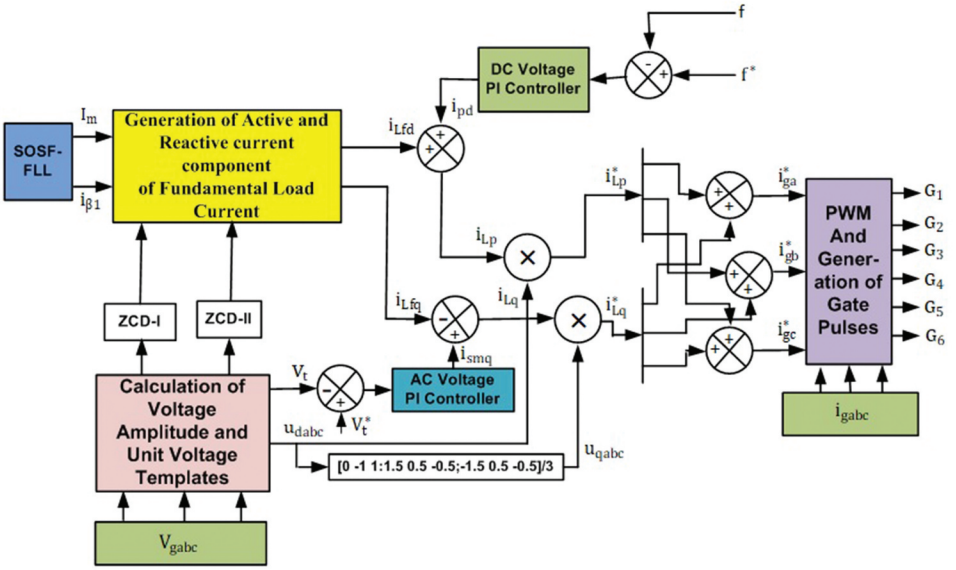


Figure 3. Control scheme of wind-solar microgrid.

$$M(s) = \frac{w_n^2}{s^2 + 2\xi w_n s + w_n^2} \tag{1}$$

where ξ and w_n are represents the damping ratio and undamped natural frequency. The open loop and closed loop transfer functions of proposed control algorithm are given as

$$T(jw) = \frac{I_m}{i_L} = \frac{k_1 k_2}{(s - jw + k_2)(s - jw)} \tag{2}$$

$$T_1(s) = \frac{I_m}{i_L} = \frac{k_1 k_2}{(s - jw' + k_2)(s - jw' + k_1 k_2)} \tag{3}$$

The peak value of the proposed control for closed transfer function is evaluated as

$$|T_1(jw)| = \frac{I_m}{i_L} = \frac{k_1 k_2}{\sqrt{(k_1 k_2 - (w - w')^2 + k_2^2 (w - w')^2}} \tag{4}$$

where k_1 andk_2 are the constant terms of the proposed SOSF control algorithm given in Figure 2. The SOSF control yields rapid response during dynamic conditions of MG system. $_1, \omega_g$ and I_m represent the evaluated values of phase, frequency, and amplitude of fundamental load current component. The SOSF-based FLL is developed using two ROGIs (Rajput et al. 2019). The three load currents component are converted into two-phase orthogonal alpha-beta coordinates (i_α, i_β), which are evaluated as

$$i_\alpha = I_m \cos(\emptyset_1), \text{ and } i_\beta = I_m \sin(\emptyset_1) \tag{5}$$

The fundamental estimated in-phase ($I_{L\alpha 1}$) and quadrature current ($I_{L\beta 1}$) quantities are evaluated as

$$I_{L\alpha 1} = I_m \cos(\emptyset_1), \text{and} I_{L\beta 1} = I_m \sin(\emptyset_1) \tag{6}$$

where the amplitude and phase angle of fundamental load current component are expressed as

$$I_m = \sqrt{I_{L\alpha 1} + I_{L\beta 1}}, \text{and} \emptyset_1 = \tan^{-1}(I_{L\beta 1}/I_{L\alpha 1}) \tag{7}$$

Evaluation of voltage amplitude and unit template generation

The sensed line voltages at PCC are converted into phase voltage that is formed as

$$v_{ga} = \frac{1}{3}(2v_{ab} + v_{bc}), v_{gb} = \frac{1}{3}(-v_{ab} + v_{bc}), \text{and } v_{gc} = \frac{1}{3}(-v_{ab} - 2v_{bc}) \tag{8}$$

The peak value of 3-phase SEIG terminal voltage (V_t) can be evaluated as

$$V_t = \left[2\left(v_{ga}^2 + v_{gb}^2 + v_{gc}^2\right)/3\right]^{\frac{1}{2}} \tag{9}$$

where v_{ga}, v_{gb}andv_{gc} are 3-phase instantaneous voltage values of the proposed MG system. The unit vectors can be assessed by dividing the instantaneous voltages by the voltage amplitude (V_t).

$$u_{da} = v_{ga}/V_t, u_{db} = v_{gb}/V_t, u_{dc} = v_{gc}/V_t \tag{10}$$

The voltage templates of the quadrature unit are calculated using the in-phase vectors and evaluated as

$$u_{qa} = -\frac{u_{da}}{\sqrt{3}} + \frac{u_{dc}}{\sqrt{3}}, u_{qb} = \frac{\sqrt{3}u_{da}}{2} + \frac{(u_{db} - u_{dc})}{2\sqrt{3}}, u_{qc} = -\frac{\sqrt{3}u_{da}}{2} + \frac{(u_{db} - u_{dc})}{2\sqrt{3}} \tag{11}$$

where (u_{da}, u_{db}, u_{dc}) are in-phase unit voltage vectors of all three phases and (u_{qa}, u_{qb}, u_{qc}) are in quadrature phase computing of unit voltage templates (Nižetić et al. 2021).

Gate pulse generation for coordinated control of proposed wind-solar microgrid

The SOSF-FLL algorithm is used to calculate the amplitude of the load current component (I_m), and the instantaneous value is determined by multiplying the amplitude of load current component (I_m) and the in-phase unit vector. Using two distinct zero-crossing detectors (ZCD) and sample and hold (S/H) circuits, the peak value of the real and reactive components of the fundamental load current can be determined. The instantaneous fundamental in-phase load current and the quadrature-phase unit vector give input to ZCD-I; the ZCD-I produces the edge triggering signals for the S/H-I as shown in Figure 3.

The actual output value generated by the S/H-I circuit followed by LPF gives the value of active load current component (i_{Lfd}). Similarly, a parallel computation path is used to produce the maximum value of the reactive load current component (i_{Lfq}), in which in-phase unit vector is used for the edge triggering signal of the S/H-II circuit at the input of the ZCD-II.

The PCC voltage is fed to the 3-phase PLL and the actual value of frequency (f) is evaluated. The difference in reference frequency (f^*) and actual frequency (f) is fed to PI controller. The addition of PV current (i_{pv}) and the output of the PI controller (i_{pf}) is subtracted from active component of current (i_{Lfd}) of fundamental load current, which gives resultant active component of reference source current. The frequency at the s^{th} sampling instant is sampled and the difference is obtained as follows:

$$f_e(s) = f^*(s) - f(s) \tag{12}$$

After that, the error from (12) is given to the PI controller to get current (i_{pf}).

The extracted active component of the reference source current is evaluated as

$$i_{Lp} = i_{pf} + i_{pv} - i_{Lfd} \tag{13}$$

$$i^*_{Lp} = i_{Lp} u_{dabc} \tag{14}$$

The voltage error (v_{terr}) of a PCC voltage at s^{th} sampling instant is evaluated as

$$V_{terr}(s) = V^*_t(s) - V_t(s) \tag{15}$$

To get the current (i_{smq}), the output obtained from (15) is given to the PI controller.

$$i_{smq} = k_p v_{terr} + k_i \int_0^{i_d} v_{terr} dt \tag{16}$$

The peak value of the reference source current (i_{Lq}) reactive component is evaluated as

$i_{Lq} = i_{smq} - i_{Lfq}$ and the reference current of reactive component is computed as below

$$i^*_{Lq} = i_{Lq} u_{qabc} \tag{17}$$

The following relationship calculates the reference source currents

$$i^*_{gabc} = i^*_{Lp} + i^*_{Lq} \tag{18}$$

The process of gate pulses generation is depictedin Figure 3, the sensed source currents (i_{gabc}) and extracted reference source current (i^*_{gabc}) are given to the hysteresis controller (Giri et al. 2019).

MPPT control method for SPV generation

The PV array is developed for a power generation of 3 kW, with 6 series and 2 parallel modules. An INC MPPT algorithm is used to operate the boost converter under varying solar insolation and the VSC maintains the voltage of the DC link at 400 V. The duty cycle (D) for the boost converter is evaluated as

$$\text{Duty cycle}(D) = 1 - \frac{v_{pv}}{v_{dc}} \tag{19}$$

where v_{pv} and v_{dc} are the output voltage of the INC MPPT control and DC link voltage. To generate the duty cycle of the converter, the incremental conductance (INC) approach is applied. By employing the relationship between dP/dV and (-I/V), the MPP can be calculated.

The INC MPPT technique equation is evaluated as

$$\frac{dP}{dV} = \frac{d(V.I)}{dV} = I\frac{dV}{dV} + V\frac{dI}{dV} = I + V\frac{dI}{dV} \tag{20}$$

Using the equation Equation 20, the value of MPP when $dP/dV = 0$ and $\frac{dI}{dV} = -\frac{I}{V}$

$$\frac{dP}{dV} > 0 \text{ then } V_p < V_{mpp} \tag{21}$$

$$\frac{dP}{dV} = 0 \text{ then } V_p = V_{mpp} \tag{22}$$

$$\frac{dP}{dV} < 0 \text{ then } V_p > V_{mpp} \tag{23}$$

The key benefits of the INC algorithm are that it can accurately evaluate the path to the MPP, reduce steady-state oscillations, improve PV system response under fast varying conditions, enhance SPV efficiency, and reduce system cost and power loss (Kewat, Singh, and Hussain 2018; Nižetić et al. 2021).

Bi-directional control strategy

The variation in solar insolation level and load perturbation is the causes of imbalance active power. The problem of active power imbalance arises due to improper regulation of DC-link voltage. A bi-directional controller is used at DC side of the VSC to regulate the DC-link voltage and performing the buck/boost operation to manage charging or discharging of the battery. The proposed BDC control provides synchronization between power generation and the load demand by consuming the extra power in the battery system. In the same way, if the power generation is less than the load demand, the battery supplies the deficit power. In this work, a BDC control with two PI controllers is utilized to control the DC-link voltage and under varying conditions.

The equation for DC-link voltage regulation is managed by a proportional integral (PI) controller, which is described as

$$i_b^*(s) = i_b^*(s-1) + k_{dp}\{v_{dce}(s) - v_{dce}(s-1)\} + k_{di}v_{dce}(s) \tag{24}$$

where k_{dp} and k_{di} are the gains of the PI controller, respectively.

Where,

$$v_{dce}(s) = v_{mpp}(s-1) - v_{dc}(s) \tag{25}$$

The duty cycle of the BDC control is governed by a battery current PI controller and it is computed as

$$D_{dc}(s) = D_{dc}(s-1) + k_{bp}\{i_{ber}(s) - i_{ber}(s-1)\} + k_{bi}i_{ber}(s) \tag{26}$$

where k_{bp} and k_{bi} are the gains of battery current PI controller.

The battery current error (i_{ber}) is evaluated using sensed battery current and reference battery current (Chandran, Kewat, and Singh 2020; Sharma, Kewat, and Singh 2020).

$$i_{ber}(s) = i_b^*(s-1) - i_b(s) \tag{27}$$

The converter's pulse width modulation (PWM) pulses are generated by comparing the duty cycle to a saw-tooth signal.

Results and discussion

The performance of the wind-solar MG system is analyzed under different operating conditions through a dynamic model designed in the MATLAB/Simulink and experimental prototype. The simulated and test result represents the dynamic response of the controller. A SOSF-FLL-based VSC is used to eliminate harmonics, maintains voltage and frequency control, provides reactive power support, and enhances the power quality of standalone MG.

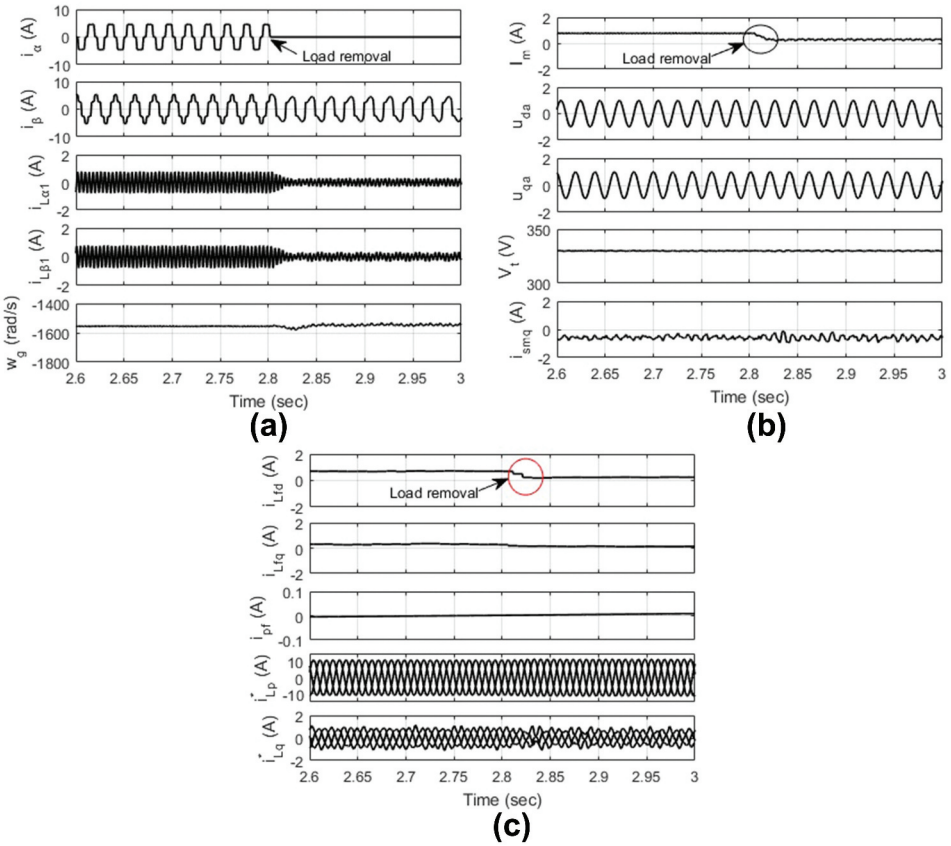


Figure 4. Estimation of control signal under unbalanced load (a) i_{α}, i_{β}, $i_{L\alpha1}$, $i_{L\beta1}$ and ω_g (b) I_m, u_{da}, u_{qa}, V_t and i_{smq} (c) i_{Lfd}, i_{Lfq}, i_{pf}, i^*_{Lp} and i^*_{Lq}.

Control signals of SOSF-FLL algorithm

The proposed wind-solar MG system's dynamic response has been built under varying nonlinear load conditions, as shown in Figure 4(a–c), which exhibit all internal control signals of the SOSF-FLL algorithm under nonlinear load conditions. The phase "a" of nonlinear load is eliminated at t = 2.8s. Figure 4(a) shows the internal signal of control algorithm i_{α}, i_{β}, $i_{L\alpha1}$, $i_{L\beta1}$ and ω_g. The variables i_{α}, i_{β} and ω_g are the extracted component from load current by abc to alpha-beta clarke transform and ω_g is the center frequency of SOSF-FLL control algorithm. Figure 4(b) shows the internal control signals I_m(extracted fundamental load current component), u_{da} (in-phase unit templates), u_{qa} (quadrature unit templates), i_{smq} (output of AC PI controller), and V_t (voltage amplitude). Figure 4(c) shows the internal control signals i_{Lfd}, i_{Lfq}, i_{pf}, i^*_{Lp}, and i^*_{Lq}, where, i^*_{Lp} and i^*_{Lq} are the extracted active and reactive component of reference source current. The summation of the extracted active and reactive component is used to generate sinusoidal reference generator current (i^*_{gabc}). These control signals yield the rapid response of the control algorithm and enhance the power quality under different operating condition.

Dynamic response of wind-solar microgrid with fixed wind velocity and varying solar insolation feeding nonlinear load

Figure 5(a,b) demonstrate the dynamic response of the wind-solar MG feeding nonlinear load with constant wind velocity and varying solar insolation. Figure 5(a) shows variation in solar irradiance 1000w/m^2 to 600w/m^2 and the temperature 35°C to 25°C and the wind velocity (V_w) of the wind

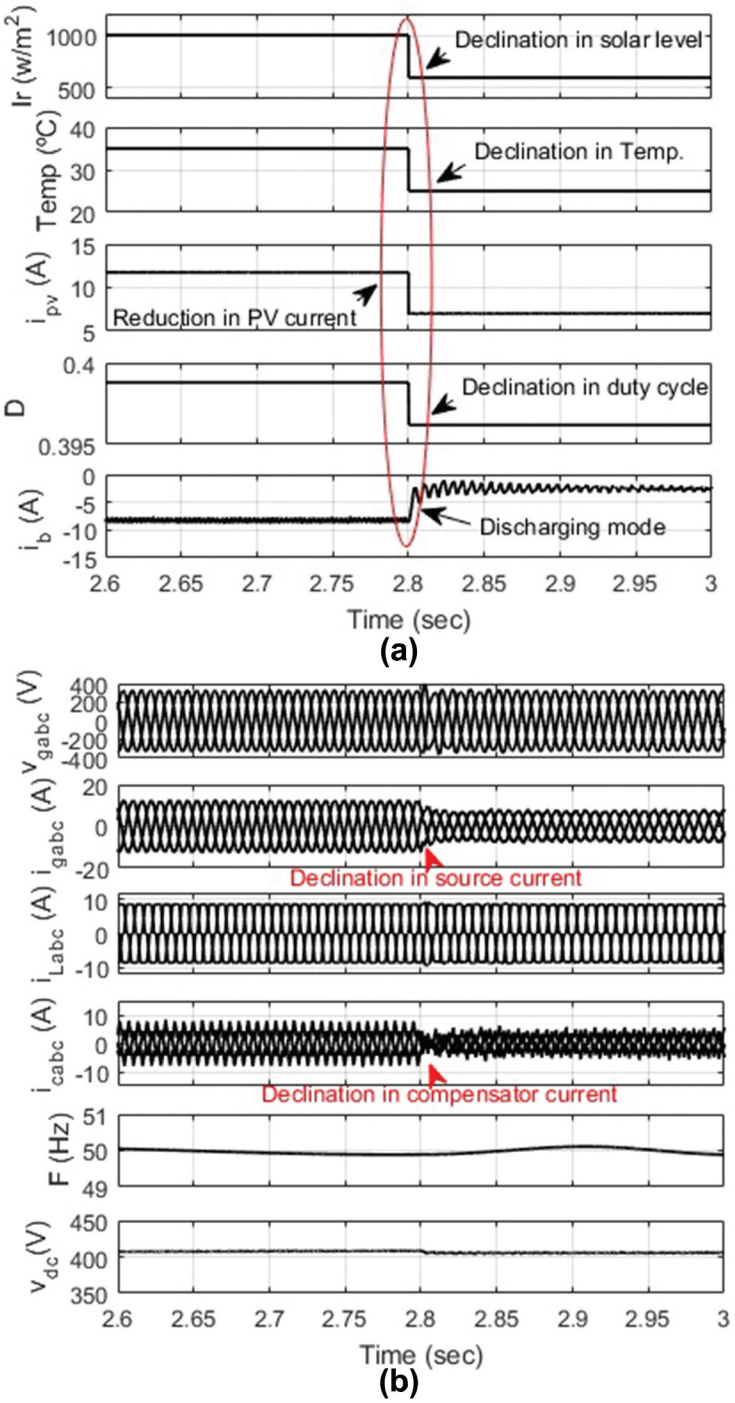


Figure 5. (a) and (b) Dynamic response of wind-solar MG with fixed wind velocity and varying solar insolation feeding nonlinear load.

turbine is fixed at 13 m/s. As a result of the decrease in solar insolation, the PV current (i_{pv}), PV power (p_{pv}), and duty cycle are reduced. Due to variation in solar insolation, the generated power, load power, and PV power of MG system are 4.34 kW, 3.02 kW, and 1.65 kW, respectively. The battery system is shifted charging mode to discharging mode at t = 2.8s and provides power support at PCC through VSC to maintain the power balance during this intermittent condition of solar power. Figure 5(b) shows that during the reduction in solar power at t = 2.8s a small distortion of 0.02s in source voltage, after this the controller maintains the voltage profile of MG system. The source current and compensator current are diminished due to reduction in solar power. As the solar irradiance changes, the battery storage reacts quickly. The battery storage is used to regulate the active power balance in an islanded mode of wind-solar MG under varying solar insolation. The controller maintains PCC voltage (v_{gabc}) and frequency (f) of the MG and other parameters are kept constant.

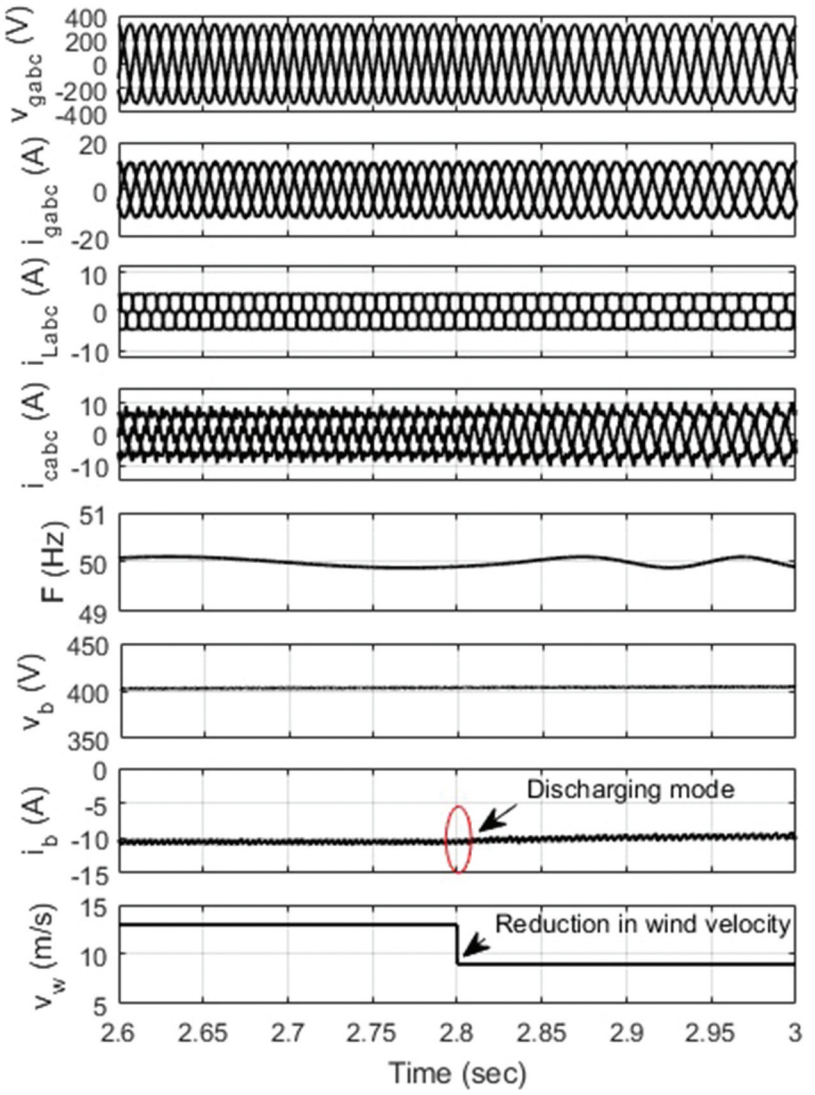


Figure 6. Dynamic response of wind-solar microgrid system with varying wind velocity and fixed solar insolation feeding nonlinear load.

Dynamic response of wind-solar microgrid with varying wind velocity and fixed solar insolation feeding nonlinear load

Figure 6 demonstrates the dynamic response of the MG feeding nonlinear load, with varying wind velocity and fixed solar insolation. The solar irradiance is fixed for 1000W/m^2 at temperature 35°C and the wind velocity (v_w) of the wind turbine is reduced from 13 m/s to 9 m/s at t = 2.8s. During the reduction in wind speed the generated power is diminished to 5.5 kW. With the decrease in wind velocity, at t = 2.8s, the battery charging current (i_b) decreases. Hence, the battery is shifted charging mode to discharging mode and BDC controller with the help of VSC provides active power support at PCC and maintains the voltage profile and current profile of isolated MG system. Moreover, the compensator current is slightly increasing and the other parameters such as frequency and DC-link voltage are kept constant. The result shows that voltage profile and frequency of the MG are controlled by the SOSF-FLL-based controller under differing wind velocity.

Dynamic response of wind-solar microgrid with fixed wind velocity/solar insolation feeding varying nonlinear load

Figure 7 demonstrates the dynamic response of wind-solar MG with fixed wind velocity/solar insolation feeding varying nonlinear load. At t = 2.7s from to t = 2.9s, phase "a" of nonlinear load is kept open. During the load removal, the generated power is increased to 5.89 kW and load power is reduced to 1.1 kW. The

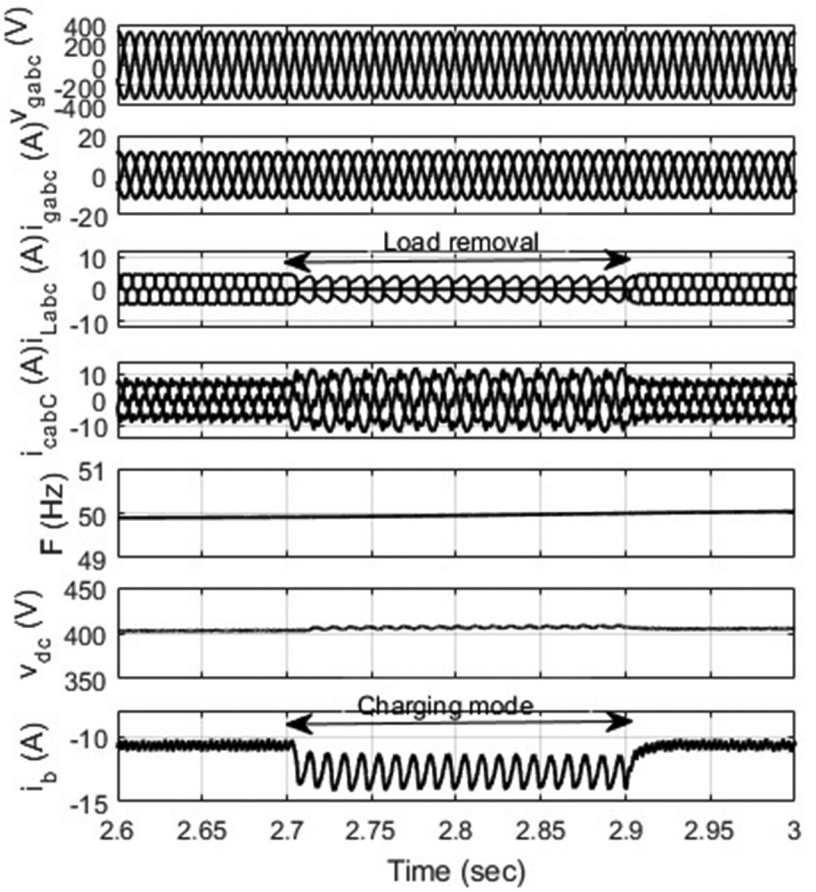


Figure 7. Dynamic response of wind-solar microgrid with fixed wind velocity/solar insolation feeding varying nonlinear load.

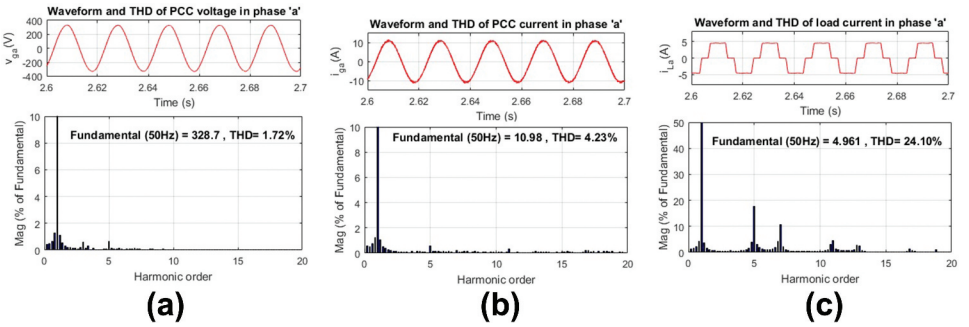


Figure 8. (a)-(c) Total Harmonic Distortions (Thd's) of PCC voltage, PCC current and load current.

results represent the voltage (v_b) and current (i_b) of the battery is increasing and load current (i_{Labc}) is decreasing due to the removal of load and other parameters are maintained during this period. The SEIG works in constant power mode as even the load is increased, at t = 2.9s the PCC voltage (v_{gabc}) and PCC current (i_{gabc}) are constant. Based on the accessibility of load demand and renewable energy, this system can alter the dynamical power sharing among the various RESs. During one phase of load removal, the battery charging current is increased. In the same way, when the PCC is connected to a load, the battery charging current is reduced while the generated voltage remains sinusoidal. The SOSF-FLL-based controller is effectively regulating the voltage and frequency of MG under varying nonlinear load.

Harmonic distortion in wind-solar microgrid system

Figure 8(a–c) represents the steady-state harmonic analysis of wind-solar MG. The PCC voltage (v_{gabc}), PCC current (i_{gabc}), and load current (i_{Labc}) harmonics are 1.72%, 4.23%, and 24.10%, respectively. The MG voltage and current harmonics are under acceptable limit under worst case of nonlinear load, and the generated source currents in the proposed MG are almost sinusoidal even though the load current of the MG is non-sinusoidal.

Comparative performance of SOSF-FLL control with conventional control algorithms

The performance of the proposed SOSF-FLL control versus conventional control algorithms under imbalanced load is depicted in Figure 9(a,b). Figure 9(a) shows the load of phase "a" is removed at t = 2.7s. The proposed control in comparison of MLMS and SOGI-FLL control yields fast dynamic response and less dynamic oscillations with reduced error of fundamental load current component (I_m) under load removal. Figure 9(b) shows during the insertion of load at t = 2.9s, the proposed control yields fast dynamic response and less steady oscillations under load perturbation.

Experimental results

A prototype of 3P3W wind-battery-based system is designed in the laboratory, as depicted in Figure 10. The Hall-effect sensors circuit is utilized to measure the voltages and currents of the standalone system. For a balanced three-phase three-wire standalone system, two Hall-effect sensors (LEM LA-25) are utilized for sensing PCC current of phases "a" and "b", respectively. Similarly, three current sensors (LEM LA-25) are utilized for sensing the load current side. These current sensors operate at $\pm$ 15 V DC supply. The output of the sensors is fed to the ADC channel of DSP-dSPACE 1104 through appropriate buffer circuitry. Moreover, the sensed voltage signal from the voltage sensor is fed to the buffer circuitry. The buffer circuitry is designed using an operational amplifier (OP-07) with appropriate resistance values. A buffer has a unity gain that has high input resistances and extremely low output resistances. The system's dynamics performance results were recorded using the four channels of the

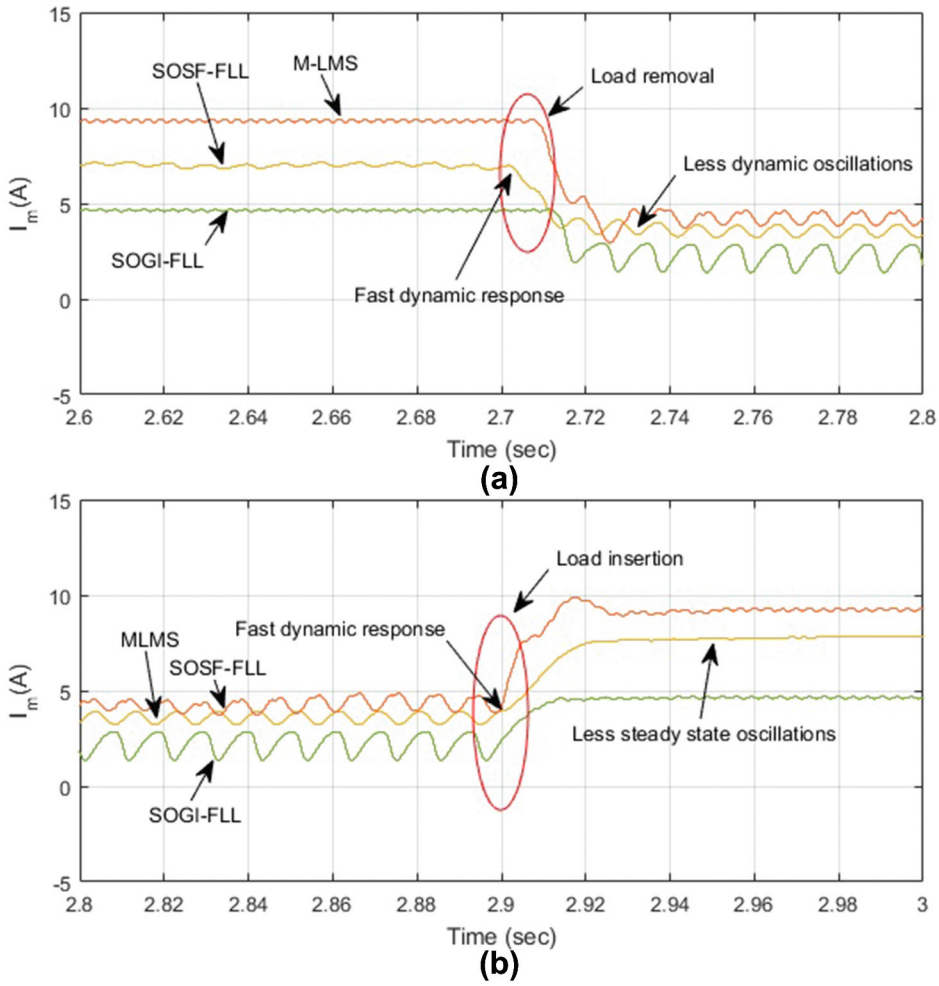


Figure 9. (a) and (b) Comparative performance of SOSF-FLL control with conventional control algorithms under load perturbation.

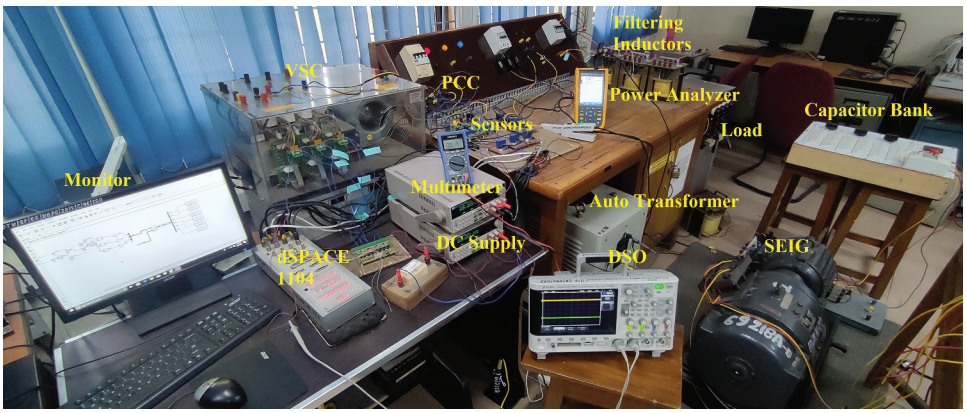


Figure 10. Experimental prototype of standalone system.

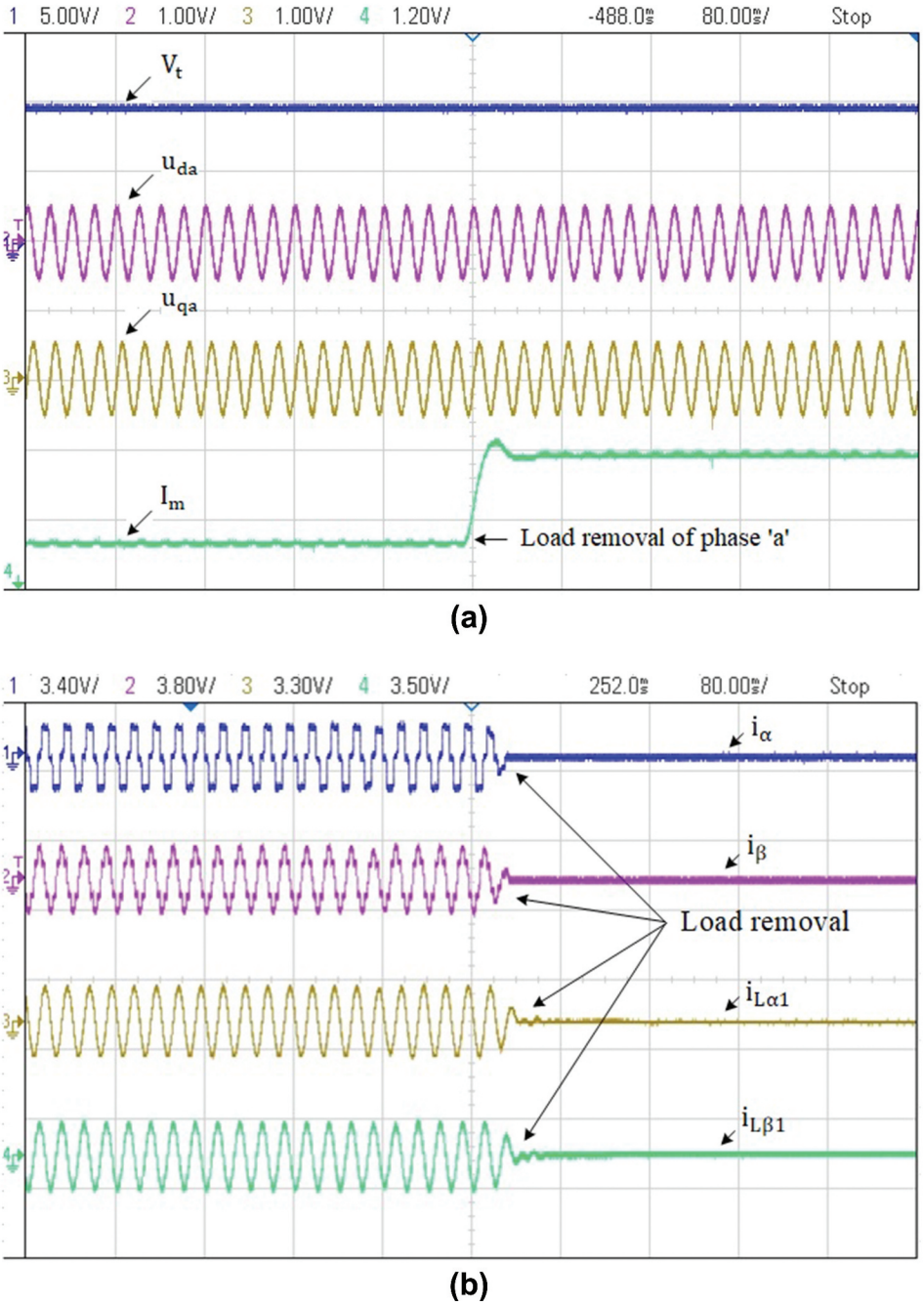


Figure 11. Internal signals of SOSF-FLL control algorithm (a) V_t, u_{da}, u_{qa} and I_m and (b) i_α, i_β, $i_{L\alpha1}$ and $i_{L\beta1}$.

KEYSIGHT INFINII VISION DSO-X 2024A. The results were recorded in the steady state using the HIOKI PQ3100 power quality analyzer. The observed experimental results and harmonic indices of the standalone system have been shown in Figures 11(a,b), Figure 12(a,b) and Figure 13(a,b). The dynamic response of SOSF-FLL control approach and standalone system is expressed as follows.

Internal signals of SOSF-FLL control algorithm

Figure 11(a,b) represents the internal signals of SOSF-FLL control algorithm under unbalanced load conditions. Figure 11(a) shows the internal signals of voltage amplitude (V_t), in-phase unit template

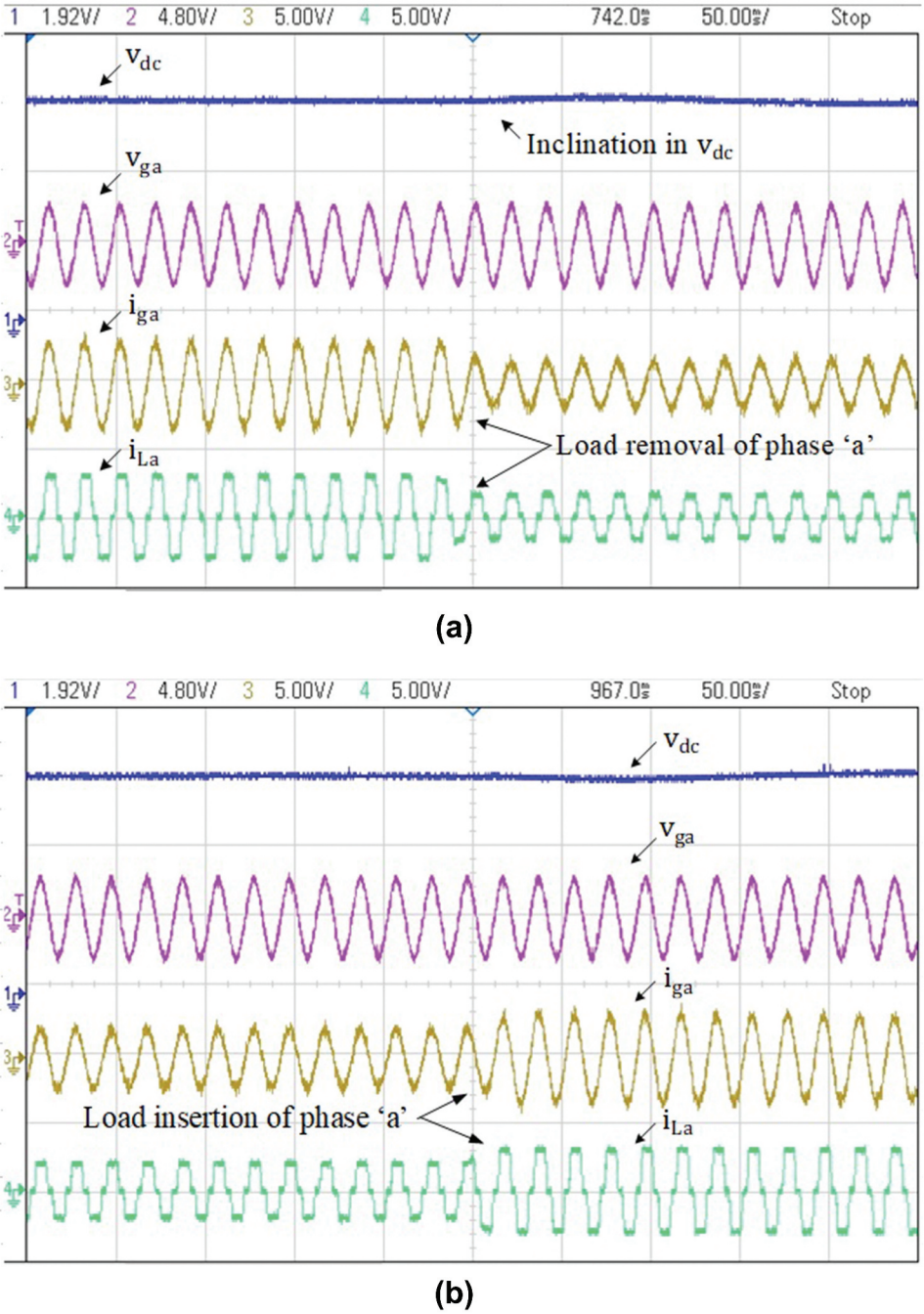
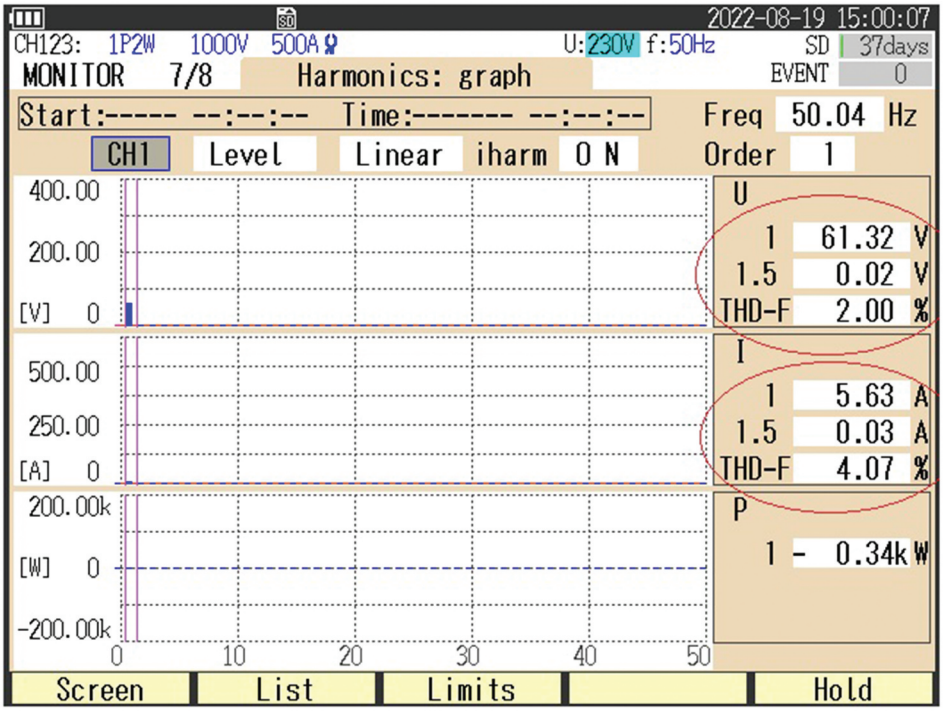
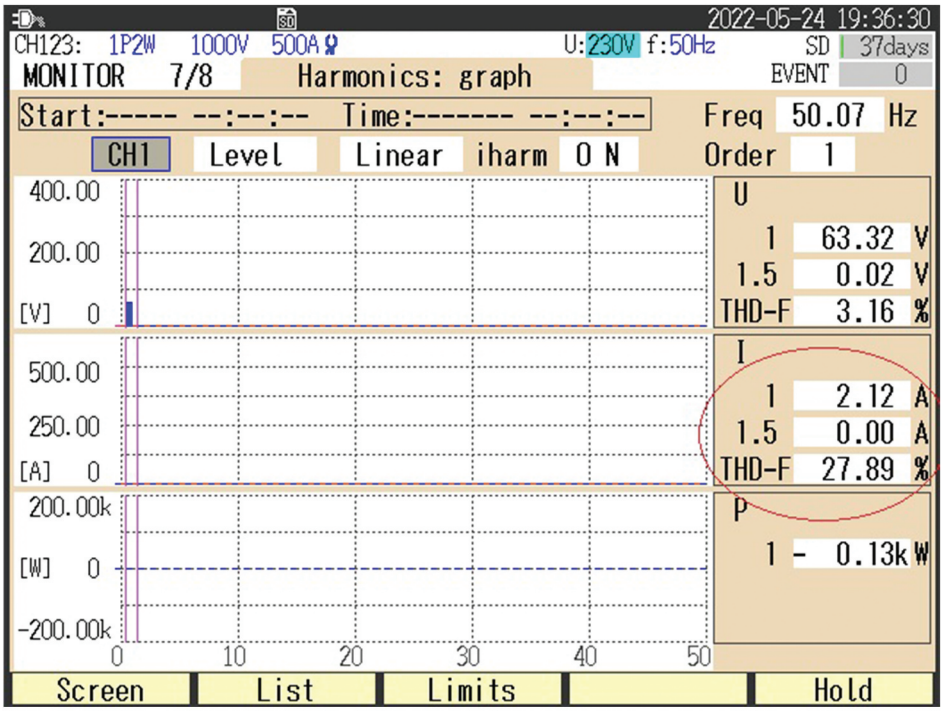


Figure 12. (a) and (b) Dynamic response of standalone system under unbalanced load.



(a)



(b)

Figure 13. (a) THDs of PCC voltage, PCC current and (b) load current.

(u_{da}), quadrature unit template (u_{qa}), and fundamental load current component (I_m). The voltage amplitude is maintained constant and the amplitude of the fundamental load current component is increased during the removal of load. On the other side in Figure 11(b), the i_α and i_β are computed using abc to αβ transformation from the nonlinear load current, and $i_{L\alpha 1}$ and $i_{L\beta 1}$ are the extracted component of nonlinear load current. During the load removal duration, the extracted component i_α, i_β, $i_{L\alpha 1}$ and $i_{L\beta 1}$ are attained the zero value.

Dynamic response of standalone system

Figure 12(a,b) represents the dynamic response of 3P3W standalone system under unbalanced load condition. Figure 12(a) shows the dynamic response of DC-link voltage, PCC voltage, PCC current and load current under removal of load of phase "a." During the removal of load, the dc link voltage is slightly increased and the PCC current (i_{gabc}) and load current (i_{Labc}) is reduced. During this operation, extra power is consumed by the battery system and the battery is shifted to charging mode. On the other side in Figure 12(b), during the insertion of load of phase "a", the PCC current (i_{gabc}) and load current (i_{Labc}) are increased. The charging or discharging modes of battery system are controlled by BDC. The battery system is utilized to control the dynamics of standalone system under transient conditions. The SOSF-FLL-based controller successfully maintains the parameters and stables the current profile and DC-link voltage of the standalone system.

Harmonic analysis of standalone system

The THD of a standalone system PCC voltage (v_{gabc}) and PCC current (i_{gabc}) is always well within the permissible limit 5%, while a highly nonlinear load is given to the generator. In SOSF-FLL control approach, the THDs of standalone system PCC voltage and PCC current as displayed in Figure 13(a) are 2.0% and 4.07%, respectively. These PCC voltage and PCC current THDs at PCC are within the IEEE 519 standard limit. As shown in Figure 13(b), the non-linear load current has a THD of 27.89%. For reactive power support during transient conditions, the required reactive power is fed by the converter.

Conclusion

A proposed wind-solar MG based on SOSF-FLL control algorithm for a three-phase three-wire feeding nonlinear load has been developed and analyzed under different operating conditions. The steady state and dynamic response of the MG system are demonstrated under dynamic condition of RESs and unbalanced load, and the behavior of the SOSF-based FLL current control technique has been described in detail.

- The proposed control efficiently evaluates the amplitude of fundamental load current component and functions effectively even in the worst load conditions of load current distortion. Moreover, The SOSF-FLL-based control with VSC successfully provides compensation for reactive power, harmonic suppression, load leveling, reduces power quality problems, and improves the dynamic performance the MG power.
- To draw maximum power from an SPV, an INC-based MPPT method was utilized under different irradiation levels. The battery system interfaced at DC link provides real/reactive power support at PCC through VSC under intermittent condition of RES's and unbalanced load. Further, A BDC controller has been effectively regulates DC-link voltage and managing the charging/discharging of the battery current.
- The comparison of dynamic performance of the standalone system with different control algorithms reveals that the proposed SOSF-FLL control outperforms conventional controls such as MLMS and SOGI-FLL.

- The dynamic results demonstrate the performance of MG under different operating conditions and the controller effectively maintains PCC voltage, PCC current, load current, DC link voltage, and frequency of the standalone MG system.
- The THD of PCC voltages and PCC currents is less than 5%, which is acceptable according to IEEE-519. In future, the proposed work with suitable controller can be used for electrical vehicle battery charging and supply to single-phase loads in remote areas with reduced cost, minimum inconvenience, and easier installation.

Nomenclature

SOSF, VSC, PCC	Second-order sequence filter, voltage source converter, and point of common coupling
SPV, INC, MPPT, LAF	Solar photovoltaic, incremental conductance, and maximum power point tracking, and Lorentzian norm-based adaptive filter
$v_{ga}, v_{gb}, v_{gc}; v_{ab}, v_{bc}, v_{ca}$	Phase and line voltage of all the phases
i_{Lfd} and i_{Lfq}	Active and reactive fundamental component
v_{pv}, i_{pv} and p_{pv}	Voltage, current, and power of solar photovoltaic
V_t^*, V_t and V_{terr}	Reference and sensed voltage amplitude with their error
v_{dc}, v_{mpp}, and v_{dce}	DC-link voltage, maximum power point voltage with their error
i_b^*, i_b and i_{ber}	Reference and sensed battery current with error

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Sombir Kundu http://orcid.org/0000-0001-6848-3721

References

Ahmed, H., S.-A. Amamra, and M. Bierhoff. 2008. Frequency locked loop based estimation of single-phase grid voltage parameters. *IEEE Transactions on Industrial Electronics* 66 (11):8856–59. doi:10.1109/tie.2018.2873527.

Arthishri, K., N. Kumaresan, and N. AmmasainGounden. 2019. Analysis and application of three-phase SEIG with power converters for supplying single-phase grid from wind energy. *IEEE System Journal* 13 (2):1813–22. doi:10.1109/jsyst.2018.2875761.

Chandran, V., S. Kewat, and B. Singh. 2020. Multi-objective control and operation of grid connected small hydro-solarPV-battery energy storage based distributed generation. *IET Renewable Power Generation* 14: 3259–72. doi:10.1049/iet-rpg.2020.0212

Chauhan, P. J., J. K. Chatterjee, H. Bhere, B. V. Perumal, and D. Sarkar. 2015. Synchronized operation of DSP based generalized impedence controller with variable speed isolated SEIG for noval voltage and frequency control. *IEEE Transactions on Industry Applications* 51: 0093–9994. doi:10.1109/tia.2014.2356642

Chauhan, P. J., B. D. Reddy, S. Bhandari, and S. K. Panda. 2019. Battery energy storage for seamless transitions of wind generator in standalone microgrid. *IEEE Transactions on Industry Applications* 55 (1). doi: 10.1109/tia.2018.2863662.

Chishti, F., S. Murshid, and B. Singh. 2019. LMMN based adaptive control for power quality improvement of grid intertie wind-PV system. *IEEE Transactions on Industrial Informatics* 15 (9):1551–3203. doi:10.1109/tii.2019.2897165.

Chishti, F., S. Murshid, and B. Singh. 2020. Robust normalized mixed-norm adaptive control scheme for PQ improvement at PCC of a remotely located wind–solarpv-BES microgrid. *IEEE Transactions on Industrial Informatics* 16 (3):1708–21. doi:10.1109/tii.2019.2923641.

Fang, J., R. Zhang, H. Li, and Y. Tang. 2019. Frequency derivative based inertia enhancement by grid connected power converters with a frequency locked loop. *IEEE Transactions on Smart Grid* 10 (5):1949–3053. doi:10.1109/TSG.2018.2871085.

Giri, A. K., S. Raj Arya, B. Chiiti Babu, and B. C. Babu. 2018. Power quality improvement in stand-alone SEIG-based distributed generation system using lorentzian norm adaptive filter. *IEEE Transactions on Industry Applications* 54 (5):5256–66. doi:10.1109/tia.2018.2812867.

Giri, A. K., S. Raj Arya, and R. Maurya. 2019. Compensation in power quality problems in wind based renewable energy system for small consumer as isolated loads. *IEEE Transactions on Industrial Electronics* 66 (11):9023–31. doi:10.1109/tie.2018.2873515.

Giri, A. K., S. Raj Arya, R. Maurya, and B. Chiiti Babu. 2019. VCO-less PLL control-based voltage source converter for power quality improvement in distributed generation system. *IET Electric Power Applications* 13 (8):1114–24. doi:10.1049/iet-epa.2018.5827.

Hoang, A. T., V. Viet Pham, and X. P. Nguyen. 2021. Integrating renewable sources into energy system for smart city as a sagacious strategy towards clean and sustainable process. *Journal of Cleaner Production* 305:127161. ISSN 0959-6526. doi:10.1016/j.jclepro.2021.127161.

Kasal, G., and B. Singh. 2009. Design and control of voltage regulators for a standalone power generation. *IETE Technical Review* 26 (2). doi: 10.4103/0256-4602.49099.

Kasal, G. K., and B. Singh. 2011. Voltage and frequency controllers for an asynchronous generator-based isolated wind energy conversion system. *IEEE Transactions Energy Conversion* 26 (2):402–16. doi:10.1109/tec.2010.2102029.

Kewat, S., and B. Singh. 2020. Improved reweighted zero-attracting quaternion-valued LMS algorithm for islanded distributed generation system at IM load. *IEEE Transactions on Industrial Electronics* 67 (5):3705–16. doi:10.1109/tie.2019.2916301.

Kewat, S., B. Singh, and I. Hussain. 2018. Power management in PV-battery-hydro based standalone microgrid. *IET Renewable Power Generation* 12 (4):391–98. doi:10.1049/iet-rpg.2017.0566.

Kherbachi, A., A. Chouder, and A. Bendib. 2019. Enhanced structure of second order generlized integrator frequency locked loop suitable for DC-offset rejection in single-phase systems. *Electric Power Systems Research* 170:0378–7796. doi:10.1016/j.epsr.2019.01.029.

Kumar, P. S., R. P. S. Chandrasena, V. Ramu, G. N. Srinivas, and K. V. S. M. Babu. 2020. Energy management system for small scale hybrid wind solar battery based microgrid. *IEEE Access* 8:8336–45. doi:10.1109/access.2020.2964052.

Kumar Kalla, U., S. S. Bhim Singh, C. J. Murthy, and K. Kant. 2018. Adaptive sliding mode control of standalone single phase microgrid using hydro, wind and solar PV array based generation. *IEEE Transactions on Smart Grid* 9 (6):1949–3053. doi:10.1109/TSG.2017.2723845.

Kumar, A., S. Kewat, B. Singh, and R. Jain. 2020. CC-ROGI-FLL based control for grid tied photovoltaic system at abnormal grid conditions. *IET Generation Transmission Distribution* 14 (17):3400–11. doi:10.1049/iet-gtd.2019.0765.

Nižetić, S., M. Jurčević, D. Čoko, M. Arıcı, and A. Tuan Hoang. 2021. Implementation of phase change materials for thermal regulation of photovoltaic thermal systems:comprehensive analysis of design approaches. *Energy* 228:120546. ISSN 0360-5442. doi:10.1016/j.energy.2021.120546.

Rajput, A., N. Beniwal, B. Singh, and S. Mishra. 2019. Second-order sequence filter based control algorithm for single phase grid interfaced solar PV system. *IET Generation, Transmission & Distribution* 13 (12):2382–90. doi:10.1049/iet-gtd.2018.6560.

Rodriguez, P., A. Luna, I. Candela, R. Mujal, R. Teodorescu, and F. Bllabjerg. 2011. Multiresonant frequency locked loop for grid synchronization of power converters under distorted grid conditions. *IEEE Transactions on Industrial Electronics* 58 (1):127–38. doi:10.1109/tie.2010.2042420.

Roy, P., J. He, T. Zhao, and Y. V. Singh. 2022. Recent advances of wind-solar hybrid renewable energy systems for power generation: A review. *IEEE Open Journal of the Industrial Electronics Society* 3:81–104. doi:10.1109/ojies.2022.3144093.

Saeed Golestan, J. M. G., J. C. Vasquez, A. M. Abusorahh, and Y. A. Turki. 2019. A study on three-phase FLLs. *IEEE Transactions on Power Electronics* 34 (1):0885–8993. doi:10.1109/tpel.2018.2826068.

Saeed Golestan, J. M. G., J. C. Vasquez, A. M. Abusorrah, and Y. A. Turki. 2021. Harmonic linearization and investigation of three-phase parallel structured signal decomposition algorithms in grid-connected applications. *IEEE Transactions on Power Electronics* 36: 0885–8993. doi:10.1109/tpel.2020.3021723

Seema, and B. Singh. 2017. Analysis and control of a standalone PV-BESS-DG based microgrid. *7th India International Conference on Power Electronics (IICPE)*, 978-1-5090-4530-3. doi: 10.1109/iicpe.2016.8079483.

Sharma, R., S. Kewat, and B. Singh. 2020. SyRG-PV-BES-based Standalone microgrid using approximate multipliers based adaptive control algorithm. *IEEE Transactions on Industry Applications* 56 (3):2913–24. doi:10.1109/tia.2020.2969393.

Simek, P., J. Skramlik, and V. Valouch. 2019. A frequency locked loop strategy for synchronization of inverters used in distributed energy sources. *Electrical Power and Energy Systems* 107:0142–0615. doi:10.1016/j.ijepes.2018.11.004.

Singh, B., and R. S. R. Chilipi. 2014. STATCOM-based controller for a three-phase SEIG feeding single-phase loads. *IEEE Transactions Energy Conversion* 29 (2). doi: 10.1109/tec.2014.2299574.

Sombir, S., and M. Singh. 2020. Voltage and frequency control of self excited induction generator integrated with PV system. *IECON*, the 46th annual conference of the IEEE industrial electronics society, 978-1-7281-5414-5/20. doi: 10.1109/iecon43393.2020.9254608.

Appendix

Appendix: Rating of Machines and Parameters of solar-wind proposed microgrid system

S.No.	Quantity	Parameters
1	Wind driven SEIG and filtering inductors	3-Phase, 415 V, 3.7 kW, 4-Poles, 50 Hz and star connected. stator resistance ($\mathbf{R_s}$)=1.405 Ω, stator inductance ($\mathbf{L_s}$)=0.0078, rotor resistance ($\mathbf{R_r}$) =0.39, rotor inductance $(L_r) =$ 0.0078 H, Inertia = 0.138, Friction factor = 0, Pole Pairs = 2 and L_f = 5mH.
2	Solar photovoltaic (SPV) array	$v_{pv} = 252.32V, i_{pv} = 11.69A$
3	Battery system	400 V, 7.5Ah
4	Nonlinear load	R = 17 Ω and L = 200μH
5	Gain parameters	SOSF-FLL: $k_1 = k_2 = k_3 = k_4 = k = 140$, λ = 8000, PI controller- in a terminal voltage control loop, $k_p = 0.7$ and $k_i = 0.07$, in frequency control loop, $k_p = 0.09$ and $k_i = 0.01$.

Earth Science Informatics
https://doi.org/10.1007/s12145-022-00892-7

CORRECTION

Correction to: Integrating the Particle Swarm Optimization (PSO) with machine learning methods for improving the accuracy of the landslide susceptibility model

Sunil Saha[1] · Anik Saha[1] · Bishnu Roy[1] · Raju Sarkar[2] · Dhruv Bhardwaj[2] · Barnali Kundu[1]

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Correction to: Earth Science Informatics
https://doi.org/10.1007/s12145-022-00878-5

The original published version of the above article contained errors in the Abstract and the Study Area section, along with the Fig. 1 image. The proper Abstract, Study Area section and Fig. 1 are as follows [Bold text used to highlight corrected area]:

Abstract

Landslide is one of the serious concerns due to which, the safety and sustainability of hilly areas across the globe, become vulnerable. Therefore, preparing of landslide susceptibility maps (LSMs) with the sound methods is a preliminary task for the safe and sustainable land use planning and design. In this research, Particle Swarm Optimization (PSO) was integrated with the pre-existing machine learning techniques such as Artificial Neural Network (ANN), Radial Basis Functional Neural Network (RBF-net), Reduced Error Pruning Tree (REPTree), Random Tree, Multivariate Adaptive Regression Splines (MARS) and M5tree to increase their efficiency and accuracy of prediction of landslide susceptibility in **upper catchment of Meenachil river of Kerala**. For the advancement of the ongoing research, a total of 189 landslide locations were analysed and datasets were obtained. To prepare LSMs, 70% of the total datasets was utilized for training and the rest 30% was used for the validation purposes. In this research, methods like: ROC, Precision, Proportion Incorrectly Classify, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Taylor diagram were applied for the validation of the models. Based on the prior pieces of literature, a total of twelve landslide conditioning factors (LCFs) were chosen. However, none of them found to be possessing multi-collinearity. It is challenging to select features from a dataset through optimization. In this regards PSO is effective because, it is straightforward with efficient universal optimization techniques. The PSO algorithm has updated and optimized the weights of models, and as results, the efficiency of the used models in predicting landslide susceptibility has increased. Nearly, 13% of the study area is very highly susceptible to landslide. The area under the curve (AUC) value of the Random Tree-PSO integrated model is the highest, 86.38% for the training dataset and 88.05% for the validation dataset. According to the sensitivity analysis elevation is most sensitive factor (0.285) and curvature is very less sensitive factor (0.115). As a result, it can be concluded that, of all the models evaluated, it is the most suitable for predicting a landslide tragedy.

Study area

In this research study, we considered **upper catchment of Meenachil river of Kerala (Fig. 1). It lies in the Western Ghats. The total area of upper catchment of Meenachil**

The original article can be found online at https://doi.org/10.1007/s12145-022-00878-5.

✉ Raju Sarkar
rajusarkar@dce.ac.in

Sunil Saha
sunilgeo.88@gmail.com

Anik Saha
aniksahamalda@gmail.com

Bishnu Roy
bishnuroyy8348@gmail.com

Dhruv Bhardwaj
dhruv97bhardwaj@gmail.com

Barnali Kundu
barnalimalda29@gmail.com

[1] Department of geography, University of Gour Banga, Malda, West Bengal 732103, India

[2] Department of Civil Engineering, Delhi Technological University, Delhi 110042, India

Published online: 04 November 2022

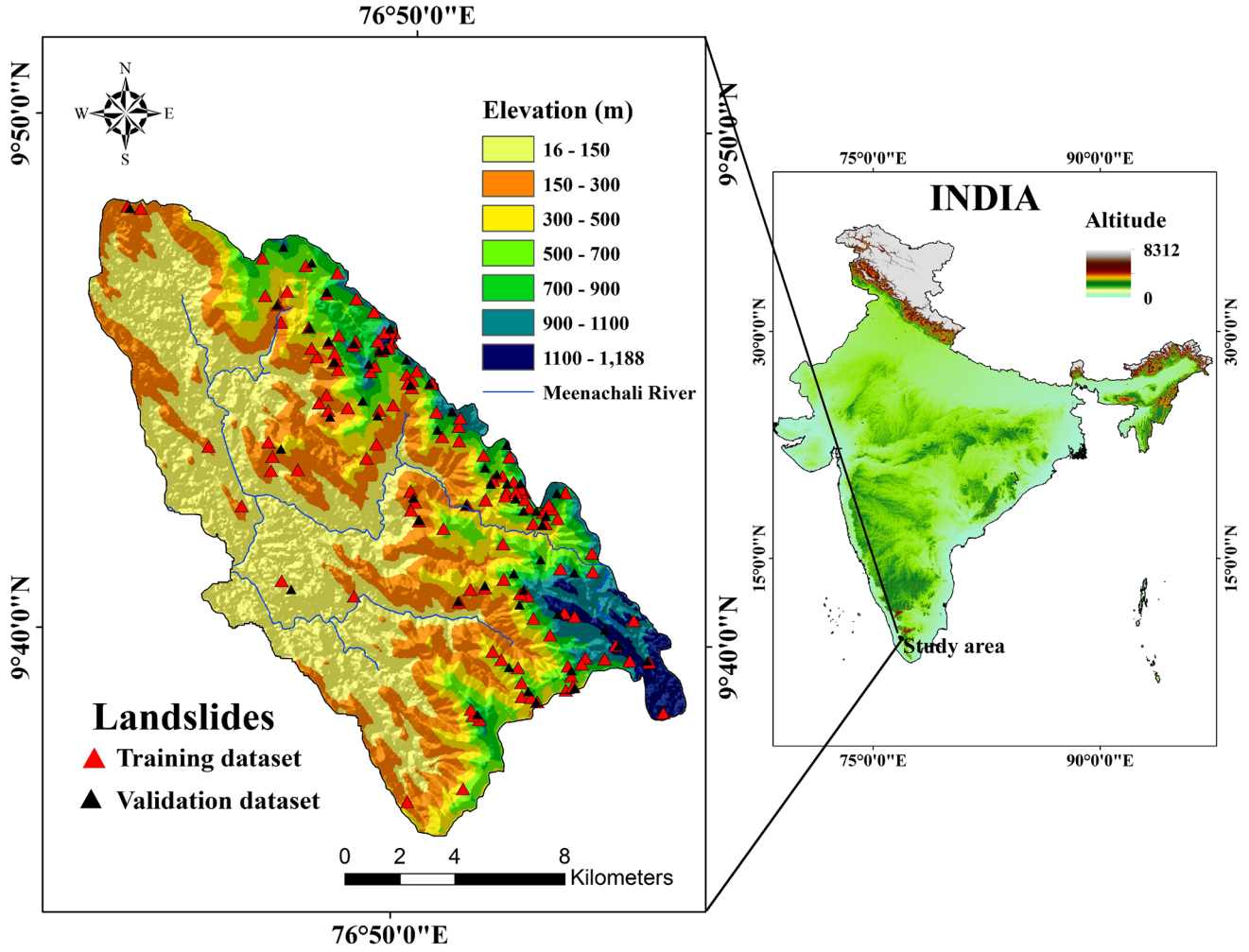


Fig. 1 Location of the study area

River is 216.46 sq. km. The study area is located between the geographical coordinates: $9^0$36′00” N to $9^0$49′00” N and $76^0$45′0″ E to $76^0$55′0″ E. The highest elevation in the study area is 1188 m. Mainly, the study area consists of four soil textural classes: sandy loam, foothills gravelly lateritic, foothills gravelly and mountain forest gravelly. Mountain forest gravelly is the most commonly found soil texture in the area and it occurs in the midland and hilly areas. Five lithological classes have been found in the study region i.e. Dharwar group, unclassified granite, quilon & warkalli bed, alluvium and charnockites respectively. Land use/land cover classes in present study area consist of five types which are built up area, cropped land, fellow land, vegetation cover and water bodies. Most of the part of the study area is covered by vegetation. The western portion of the study area is frequently affected by landslides. For managing the loss substantiated by the landslides, it becomes extremely important to implement appropriate strategies. A comprehensive assessment of landslide susceptibility is among the most effective and reliable methods for the management of landslide induced losses by making aware the people who are living in the areas of high landslide susceptibility.

The original article has been corrected.

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

Neural Processing Letters
https://doi.org/10.1007/s11063-022-11060-9

COVID-19 Detection from Chest X-rays Using Trained Output Based Transfer Learning Approach

Sanjay Kumar[1] · **Abhishek Mallik**[1]

Accepted: 16 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
The recent Coronavirus disease (COVID-19), which started in 2019, has spread across the globe and become a global pandemic. The efficient and effective COVID-19 detection using chest X-rays helps in early detection and curtailing the spread of the disease. In this paper, we propose a novel Trained Output-based Transfer Learning (TOTL) approach for COVID-19 detection from chest X-rays. We start by preprocessing the Chest X-rays of the patients with techniques like denoising, contrasting, segmentation. These processed images are then fed to several pre-trained transfer learning models like InceptionV3, InceptionResNetV2, Xception, MobileNet, ResNet50, ResNet50V2, VGG16, and VGG19. We fine-tune these models on the processed chest X-rays. Then we further train the outputs of these models using a deep neural network architecture to achieve enhanced performance and aggregate the capabilities of each of them. The proposed model has been tested on four recent COVID-19 chest X-rays datasets by computing several popular evaluation metrics. The performance of our model has also been compared with various deep transfer learning models and several contemporary COVID-19 detection methods. The obtained results demonstrate the efficiency and efficacy of our proposed model.

Keywords COVID-19 · Chest X-ray · Deep transfer learning models · Ensemble learning · Image classification · Medical diagnosis

1 Introduction

The COVID-19 outbreak is very severe and fatal and has resulted in a global pandemic and caused millions of deaths across the globe. It originated in late December 2019 in the Wuhan province of China. Since then, it has become a global health hazard spreading across the globe through personal and community based transmission [1]. It has symptoms similar to

✉ Sanjay Kumar
sanjay.kumar@dtu.ac.in

Abhishek Mallik
abhishekmallik265@gmail.com

[1] Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110042, India

Published online: 28 October 2022

Pneumonia [2] but is more fatal than that. It is difficult to obtain a standard medication for the COVID-19 disease as it mutates quickly, forming different variants. In March 2020, the World Health Organization (WHO) declared the COVID-19 as a pandemic. The rampant spread of the disease and the unavailability of a standard medication has caused there to be 3.7 million deaths across the globe by May 2021 [3]. Thus, it is very essential to diagnose the disease in the patients at an early stage so that the transmission and spread of the disease can be curbed and the pandemic can be combated. RT-PCR (Reverse Transcriptase Polymerase Chain Reaction) is the most popularly used molecular type test for COVID-19 detection in human beings. An RT-PCR test conducted within three weeks after the occurrence of symptoms has a 90% accuracy in detecting Covid-19 cases positively [4]. But there are several limitations related to the RT-PCR test like it is not so widely available, and it takes more than 6 hours to get the test results. Lateral Flow Test (LFT) is another test that can be used to detect Covid-19, but it is not as reliable as the RT-PCR test. In situations where molecular tests like RT-PCR are not available and the time is constrained, the analysis of the individual chest radiography images can be used to detect Covid-19 efficiently and immediately. In addition to detection, the chest radiography analysis also reveals the percentage of the infection present, which can aid in prescribing a more suitable treatment for the patient. Some of the well-known chest radio-graphical techniques are Chest X-rays and CT Scan. There are various advantages and disadvantages associated with both CT Scan and chest Xray. The instrument required for a CT Scan are more expensive as compared to that required for chest Xray. Moreover, the disinfection time for CT scan is approximately 15 minutes which is far greater than chest Xray as there is a portable version of it which exposes the patient to a lesser amount of bacteria. The ionizing radiation to which the patient is exposed is also greater in CT Scan as compared to chest Xray. Since chest X-rays are readily available in the hospitals and they have low ionizing radiation exposure to the patients, so it is more preferable as compared to the CT scans [5].

Various medical professionals and radiologists have acknowledged the efficacy of the use of chest radiology to detect COVID-19 which gives a comparable performance with the RT-PCR test having an accuracy of 90% [6]. The manual diagnosis of COVID-19 from chest X-rays on a large scale is impractical and cumbersome. Hence, the Computer-Aided Diagnosis (CAD) tools are often used by radiologists to diagnose COVID-19 from chest X-rays and other chest radiographs. Various researchers have worked extensively on developing efficient and reliable CAD tools for detecting COVID-19 from chest X-rays. Initial research was done on using machine learning-based approaches for detecting pathogenic lungs infected with COVID-19 [7]. Methods using Fuzzy color-based preprocessing techniques have also been presented on chest X-ray images [8]. The features are then extracted from the processed images using the pre-trained models, namely MobileNetV2 and SqueezeNet. The discriminant features are then extracted by stacking the extracted features and using Social Mimic optimization. Then a Support Vector Machine was used to exploit the discriminant features and make the final COVID-19 detection. A Generative Adversarial Network (GAN) based approach was proposed by Rasheed et al. [9]. They used Generative Adversarial Network (GAN) for data augmentation and then employed a Convolutional Neural Network for representation learning. They augmented the dataset using GAN and then used a Convolutional Neural Network for representation learning. The higher dimensional features are reduced to lower-dimensional features using Principal Component Analysis (PCA), and then Logistic Regression was used to make the final COVID-19 classification. The data augmentation using GAN and the dimensionality reduction helped the authors achieve improved performance. Recently, transfer learning models have emerged to achieve regularization and generalization of deep learning models [10]. Another work was proposed, which performed feature extrac-

tion as well as classification using a single model instead of using different models [11]. Several pre-trained CNN models such as Xception, VGG16, DenseNet121, EfficientNet, and NASNet were employed and fine-tuned for COVID-19 detection from chest X-rays [12].

The detection of COVID-19 using a radiological expert is time-consuming, and an error-prone task [13]. Hence, an automated framework is required for COVID-19 detection using chest X-rays. Most of the existing works in the field of COVID-19 detection uses several deep neural network architectures, but they only use a single model. These methods are only dependent on unilateral decisions and rely on findings from a single model. These methods do not use any technique to combine the classification power of other methods. This motivates us to present this work. In this work, we present a novel Trained Output-based Transfer deep Learning (TOTL) approach for COVID-19 detection from chest X-rays. In this paper, we introduced an improved framework for COVID-19 Detection from Chest X-rays using the Trained Output based Transfer Learning Approach. We combine the results of various transfer learning models by training a deep neural network instead of conventional ensemble techniques while most of the existing works use conventional techniques like majority voting. Overall, the proposed work provides better results for COVID-19 detection by combining the strengths of various deep transfer learning models in terms of the extracted features. Moreover, the conformity of the deep learning models towards general problem statements also helps us to use medical imagery-based COVID-19 detection using fine-tuning. Most of the existing works done in this field have used classical performance metrics like Accuracy, Precision, Recall, and F1 Score. However, we have used classical as well as some other important performance metrics. Moreover, we also present the simulations on chest X-ray instead of CT scans which has been extensively used in recent works. The reason for using chest X-rays is that it is readily available in hospitals and they, have low ionizing radiation exposure to the patients, hence it is more preferable as compared to the CT scans. We performed technical analysis like loss convergence of our model, the effect of data augmentation, hyperparameter analysis relating to epochs, optimizer, and confidence intervals. In general, the performed experimental analysis reveals the better performance of our method as compared to the existing baselines and recently proposed techniques. We start by processing chest X-rays images using various techniques like denoising with Gaussian filter, contrasting using CLAHE, segmentation using U-Net, and then cropping around the lungs region to capture only the relevant details from the images. These processed images are then reshaped and passed through various deep transfer learning-based pre-trained models, namely InceptionV3, InceptionResNetV2, Xception, MobileNet, ResNet50, ResNet50V2, VGG16, and VGG19. The models are fine-tuned on the processed chest X-ray images, and the classification outputs are generated. The classification outputs are then fed to a deep neural network to train the outputs of the various pre-trained models and achieve a more balanced output combining the multiple capabilities of all the models. This model is then finally used to detect COVID-19 from the chest X-ray images. We performed extensive experiments on four benchmarked datasets and evaluated several performance metrics. The performance of our model was also compared with several classical deep transfer learning models and multiple contemporary COVID-19 detection frameworks. The experimental results reveal the utility of our proposed framework as an efficient COVID-19 detection model. The major contributions of this work are as follows:

(i) We proposed a novel Trained Output-based Transfer Learning (TOTL) approach for COVID-19 detection from chest X-rays. It trains the combined outputs of several fine-tuned deep transfer learning-based pre-trained models to capture the performance capabilities of every model and achieve improved performance.

(ii) We process the chest X-ray images various preprocessing techniques like denoising using Gaussian filter, contrasting using CLAHE, segmentation using U-Net.
(iii) We fine-tune several deep transfer learning-based pre-trained models like InceptionV3, InceptionResNetV2, Xception, MobileNet, ResNet50, ResNet50V2, VGG16, and VGG19 using preprocessed chest X-ray images.
(iv) An automated framework for COVID-19 detection using chest X-ray has been proposed having real-world prevalence as is demonstrated by extensive experimentations done by us.

The rest of the paper is organised as follows. Section 2 discusses some of the existing work in the field of COVID-19 detection using chest X-rays. The preliminary concepts required for a better understanding of this manuscript are mentioned in Sect. 3. In Sect. 4 we discuss the various datasets and the evaluation metrics used by us for this study. Section 5 illustrates our proposed Trained Output based Transfer Learning approach for COVID-19 detection from chest X-rays. The experimental results and analysis are discussed in Sect. 6. Finally, the concluding remarks and scope for future extensions are given in Sect. 7.

2 Related Work

Due to the recent COVID-19 pandemic, a lot of work is being done in detecting and tackling it. Numerous research works have been published in this field. Recent advancements in the field of computer vision have paved the way for the application of deep learning and convolutional neural networks in the field of medical image diagnosis like retinopathy identification [14–16], tumor detection [17–19], and also for COVID-19 detection [20]. In this section, we discuss several recent research works done in the field of COVID-19 detection using chest X-rays and CT scans. The various works done in the field of COVID-19 detection involves the usage of deep neural networks, convolutional neural networks, transfer learning, ensemble learning, etc.

Luz et al. [21] presented a deep learning based approach to COVID-19 detection. They used the Covidx dataset for their study and to evaluate the efficiency of their model. Their proposed model achieved a 93.9% accuracy and a 96.8% sensitivity. Hemdan et al. [22] proposed an automatic COVID-19 detection framework, COVIDX-Net to detect COVID-19 from the chest X-rays. They used seven different deep learning architectures. They used a dataset containing 50 chest X-ray images. VGG16 and DenseNet201 both provided a 90% accuracy on the dataset. Basu et al. [23] presented an extension to the detection framework, which used a transfer learning based method. As part of their extension, they presented an algorithm for various abnormality detection caused by the COVID-19 disease in the patients as present in the chest X-rays. They used a Gradient Class Activation Map for feature extraction from the chest X-ray images. For detecting features from X-ray images, they used Gradient Class Activation Map. They evaluated their performance on the NIH chest X-ray dataset and achieved an accuracy of 95.3%. Ozturk et al. [24] proposed an automated COVID-19 detection framework, DarkCovidNet, which uses chest X-ray images. They used a dataset containing 125 images. They achieved a binary class accuracy of 98.08% and a multi-class accuracy of 87.02%. One of the drawbacks of their approach is the small dataset size.

Das et al. [25] proposed a COVID-19 detection framework using deep learning techniques for COVID-19 detection using chest X-ray images. They attained an accuracy of 97.4% through their approach. Tuncer et al. [26] presented an automated COVID-19 detection method for chest X-ray images. They performed feature extraction using a residual

exemplar local binary pattern. Then they selected important features through iterative Relief. The chosen features are then fed into several well-known classifiers. They evaluated their approach on a small dataset having just 321 chest X-ray images and obtained a classification accuracy of 99% using the SVM classifier. Methods using Fuzzy color-based preprocessing techniques have also been presented on chest X-ray images [8]. The features are then extracted from the processed images using the pre-trained models, namely MobileNetV2 and SqueezeNet. The discriminant features are then extracted by stacking the extracted features and using Social Mimic optimization. Then a Support Vector Machine was used to exploit the discriminant features and make the final COVID-19 detection. They achieved a classification accuracy of 98.25% and 97.81% with MobileNetV2 and SqueezeNet, respectively. A COVID-19 detection framework called Covid-Net was proposed by Wang et al. [27]. Covid-Net presented a multi-class classification study for detecting Normal, Pneumonia, and Covid-19 infected chest X-ray images. Covid-Net gave a better performance as compared to the VGG19 and ResNet50 frameworks.

Data augmentation techniques have been extensively used for regularization and generalization of deep learning models [28–31]. Zheng et al. [32], worked on improving the performance of deep convolutional neural networks by proposing a full-stage simultaneous data augmentation for the training and testing phase. The consistent data augmentation technique helps in achieving the appropriate transfer of the domain specific knowledge to the model thereby improving the performance. Shen et al. [33], proposed a metaheuristic based approach for Covid-19 detection from chest Xrays. They performed feature extraction using a pre-trained ResNet18 architecture. The obtained features further undergo feature selection using a discrete social learning particle swarm optimization algorithm (DSLPSO). The selected features are then passed through an SVM classifier for final classification. Lung segmentation is often used in the detection of lung diseases to achieve an improved performance [34, 35]. A lung segmentation approach using CT scans to extract the lung region infected by CT scans was proposed by Oulefki et al. [36]. They used an efficient Kapur entropy-based multilevel thresholding unsupervised procedure. Further work on lung segmentation was used by Liu et al. [37] using a weakly supervised scribble annotation algorithm. Several researchers have also used respiratory sounds based on coughing, breathing, and voice parameters [38]. Lella et al. [39] proposed a 1D convolutional network framework for detecting Covid-19 from the respiratory parameters. They later improved this work by using deep convolutional network [40] and a lightweight convolutional neural network with modified MFCC and enhanced GFCC [41].

A COVID-19 study based on deep transfer learning frameworks was performed by Narin et al. [42]. They used ResNet50, InceptionV3, InceptionResNetV2. They obtained a classification accuracy of 98% for COVID-19 detection based on the chest X-ray images. An ensemble learning based approach was proposed by Chouhan et al. [43]. They presented an ensemble model comprising of deep transfer learning based frameworks, namely GoogleNet, AlexNet, ResNet18, InceptionV3, and DenseNet121. Their approach achieved an accuracy of 96.4% Shelke et al. [44] used DenseNet-161 for the analysis and diagnosis of the COVID-19 from the chest X-ray images. They performed experiments on a very small dataset containing only twenty-two X-ray images and achieved an accuracy of 98.9%. A hybrid approach combining ResNet50V2 and Xception was proposed by Rahimzadeh et al. [45] for discerning the identifying structural patterns in the chest X-rays for COVID-19 detection. Another framework named DeTraC was proposed by Abbas et al. [46] for COVID-19 detection. Pre-trained CNN architecture was used for feature extraction from the images. Thereby, local features were extracted from the images using a class decomposition method. For making the final classification, they used a gradient descent based method. The obtained classification results

are further enhanced using a class composition layer. They used a dataset of 105 chest X-ray images for their experiments and attained an accuracy of 95.12%.

We can see from the discussion above that most of the works done in the field of COVID-19 detection do not have considerable detection accuracy to be of any real-world applications. Moreover, most of the existing techniques use pre-trained transfer learning models for either feature extraction or for fine-tuning. But only a few provide an ensemble based approach to combining the predictive powers of various models. The few ensemble approaches that are even presented only create a strong classifier from weak classifiers. Also, the datasets used in the existing works only use datasets having very few data samples. Through our work, we present a novel approach of combining several pre-trained deep transfer learning models by training their outputs using a deep neural network to achieve improved performance. The datasets used by us contain more data samples belonging to each category, namely Normal, COVID-19, and Pneumonia this makes our results more reliable.

3 Preliminaries

In this section, we explain several basic concepts like a deep neural networks, convolutional neural networks, and transfer learning. Knowledge and understanding of these concepts would help the readers to understand and implement our Trained Output based Transfer Learning (TOTL) approach for COVID-19 detection from chest X-rays.

3.1 Deep Neural Networks

Deep neural networks are inspired by the structure of the brain's cerebral cortex. At the basic level, a perceptron is the scientific portrayal of a natural neuron existing in the brain of humans. It combines the inputs to the neuron into a single output using a pre-defined function. An activation function is used by the neuron to convert the outputs generated to a usable form. A deep neural network consists of an input layer, a network of hidden layers, and an output layer. The input layer takes the input as a feature set for the data points. The hidden layers process the input, where each layer contains multiple stacked neurons. The output layer finally makes the prediction for the data points. The prediction can be regression, classification, or clustering-based. The hidden layers iteratively adjust the weights associated with each connection between the neurons. The ideal weights are determined using back-propagation techniques, which adjust the weights using an optimizer like gradient descent. This is accompanied by a loss function like Mean Squared Error or Categorical Cross-Entropy. The quality of the prediction made is determined by the loss function, and the deep neural network aims to minimize this loss and achieve the most optimal results.

3.2 Convolutional Neural Networks

Convolutional Neural Networks are variations of the deep neural networks used for various computer vision tasks like image classification, object detection, image segmentation, image recognition, etc. It exploits template matching techniques to complete a vision-associated task. A stack of convolution layers is used for extracting essential features from the input images. It contains various pooling layers, channels (Kernels), and finally fully connected layers (FC). The last layer uses an activation function to make the final prediction for the input sample in a probabilistic range of 0 to 1 [47]. A convolutional neural network usually

consists of several convolutional layers. The convolutional layers are made up of filters of varying sizes which extract different features and use an activation mapping to map the inputs to their corresponding outputs. The activation maps are combined by stacking them to obtain a substantial output volume and confidence. Various CNN's take input of a pre-defined fixed size and produce outputs of a different shape. A convolutional layer takes as input a lot of parameters like the Number of filters, stride, and activation, which describe the convolutional layer. A pooling layer always follows a convolutional layer to combine the outputs generated by various filters. The pooling layers can be average pooling, max-pooling, etc., making it more efficient. Convolutional layers reduce overfitting and make neural networks concentrate on the more essential details [48].

3.3 Transfer Learning Based Deep Convolutional Networks

Transfer learning refers to using a large and complex pre-trained model, which is trained on a large and diverse dataset to achieve a generalized framework. The trained model is used on a similar problem statement as the original model was trained but with relatively smaller datasets and allows only a few layers to be fine-tuned while all other layers are frozen. The obtained generic framework of a complex architecture allows giving an impressive performance on a variety of problem statements. Transfer learning can be used for feature extraction and fine-tuning based classification. For feature extraction, the classification part of the pre-trained transfer learning model is removed, and the remaining model can be treated as a feature extractor which can be used with another classification algorithm [49]. While for classification, we freeze some of the layers of the pre-trained model, which are not trained again on the new dataset, the rest all the layers are fine-tuned by the model, and the classification is made thus [50]. Transfer learning is beneficial when the data available is scarce. Also, having a generalized framework helps with overfitting and randomization effects on the weights and the training and fine-tuning process.

For our study, we use several pre-trained deep convolutional neural networks like VGG16, VGG19, ResNet50, ResNet50V2, MobileNet, InceptionV3, Xception, and InceptionResNetV2., which have been pre-trained on the Imagenet dataset [51]. VGG16 is a 16 layers deep model which improves the performance of AlexNet, which was amongst the first winners of the Imagenet competition, by increasing the depth of the model and reducing the size of the kernel filters from 11 and 5 to 3. VGG19 is similar to VGG16 with 19 layers. While ResNet50 is a deep residual network having residual blocks to avoid the vanishing gradient problem. It uses skip connections to connect the output of one layer to some other layer by skipping several layers in between. ResNet50V2 is a modified version of the ResNet50 architecture by improving the propagation formulation of the connections between blocks. As compared to ResNet50, MobileNet is a lightweight mobile application-friendly deep neural network having fewer trainable parameters by using depthwise separable convolutions. InceptionV3 is a 42 layers deep model. It improves the performance of its predecessors by incorporating Factorization into Smaller Convolutions, Spatial Factorization into Asymmetric Convolutions, Utility of Auxiliary Classifiers, and Efficient Grid Size Reduction. Xception model, on the other hand, is 78 layers deep and uses modified depthwise separable convolution to perform better than InceptionV3. InceptionResNetV2 is an amalgamation of Inception and ResNet architectures. It is 164 layers deep. It combines the multiple-sized convolutional filters with residual connections to avoid the degradation problem and reduce the training time.

4 Datasets and Evaluation Metrics

In this section, we describe the various datasets and evaluation metrics used by us. We used four real-life COVID-19 X-ray image based datasets collected from various online data sources. We also used several benchmarked evaluation metrics to obtain a measure of the performance of our proposed algorithm. The description of the datasets and the performance metrics are as follows:

4.1 Datasets

We used four different datasets for the purpose of our study. Figure 1 shows the sample images of various datasets belonging to the COVID-19, Normal, and Pneumonia categories. The used datasets are described below:

1. Dataset 1: This dataset was proposed by Asraf. It is obtained from Kaggle and contains 1525 images of COVID-19, Normal and Pneumonia each. The dataset can be accessed from https://www.kaggle.com/amanullahasraf/covid19-pneumonia-normal-chest-xraypa-dataset
2. Dataset 2: This dataset was proposed by Chowdhury et al. at the University of Dhaka and Qatar University along with several medical practitioners and collaborators. It contains 3616 COVID-19 samples along with 10192 Normal and 1345 Viral Pneumonia chest X-rays. The dataset is available at https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
3. Dataset 3: This dataset is also a chest X-rays based COVID-19 detection image dataset. It contains 137 COVID-19 samples, 90 Normal samples, and 90 Pneumonia samples. The dataset can be accessed from https://www.kaggle.com/pranavraikokte/covid19-image-dataset
4. Dataset 4: This dataset contains 576 COVID-19 image samples, 1583 normal image samples, and 4273 pneumonia images. Since the COVID-19 samples are relatively less than the normal and pneumonia samples, we exclude the Pneumonia samples from our experimentations and use the COVID-19 and the Normal image samples. The dataset can be obtained from https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia.

4.2 Evaluation Metrics

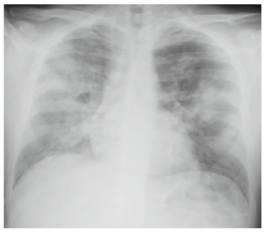
We test the performance of our proposed model on several benchmarked evaluation metrics for the classification tasks, namely, Accuracy, Precision, Recall, F1 Score. The description of these performance metrics is given below:

1. Accuracy: Accuracy is defined as the measure of the model to correctly classify the samples to their corresponding labels. Mathematically, it can be explained as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

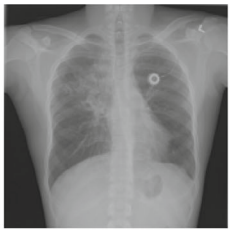
2. Precision: Precision is used as a measure to estimate the exactness of the algorithm. It gives a measure of the agreement of the exact labels with respect to the positive labels predicted by the model. Mathematically it can be expressed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$



(a) Covid

(b) Normal



(c) Pneumonia

(a) Dataset 1

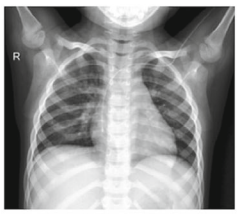
(a) Covid

(b) Normal

(c) Pneumonia

(b) Dataset 2

(a) Covid

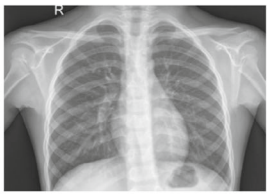


(b) Normal

(c) Pneumonia

(c) Dataset 3

(a) Covid



(b) Normal

(c) Pneumonia

(d) Dataset 4

Fig. 1 Examples of Chest X-rays in all considered datasets having COVID-19, normal, and pneumonia samples

3. Recall: Recall is also known as sensitivity of a model and it signifies the completeness of the model. Higher value of Recall suggests lower false positives being predicted. Mathematically it can be expressed as follows:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4. F1 Score: F1 score shows the balance between Precision and Recall. In other words, it is a measure of the data distribution. Mathematically it can be expressed as follows:

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

Here, TP (True Positive) suggests that a patient is COVID-19 positive, and the model has correctly identified him as positive. In the same way, FP (False Positive) represents that a patient doesn't have COVID-19 but the model has detected him as COVID-19 positive. While TN (True Negative) shows that a patient doesn't have COVID-19, and the model has predicted him not to have COVID-19. Similarly, FN (False Negative) implies that a patient has COVID-19, but the model has detected him not to have COVID-19.

5 Proposed Work

In this section, we describe our proposed Trained Output based Transfer Learning approach for more accurate COVID-19 detection based on chest X-ray images. Our approach consists of three phases: (i) Preprocessing phase, (ii) Training Deep Transfer Learning models, and finally, (iii) Training outputs. The details about these phases is given below.

5.1 Preprocessing

The chest X-ray available in the datasets used by us contains noise, varied brightness, and irregular shapes. To counter such effects, we preprocess our dataset. We start by first removing the noise from the images using a Gaussian filter, which stabilizes the variance in the images [52]. The contrast in these images is then improved using the Contrast Limited Adaptive Histogram Equalization (CLAHE) approach [53]. It improves the visibility of all the images across the dataset. Then we perform lung segmentation to extract only the lung region using U-Net [54]. The parameters used for the U-Net are as follows: ResNet34 is used as the backbone, imagenet weights form the encoder weights, while we use transpose as the decoer block types. We train the U-Net on the Chest Xray Masks and Labels dataset. We achieve a training loss and Intersection Over Union (IOU) of 0.0421 and 0.9886, respectively. The entire image is then cropped around the lungs to only capture details essential to the problem statement of COVID-19 detection. The entire process is shown in Fig. 2. Moreover, we also augment data to obtain a more generalized framework. Since neural networks are complex architectures having lots of neurons and numerous parameters, hence to obtain an appropriate learning curve, there should be a sufficient amount of data. Hence, to compensate for the lack of available data, we augment the training and validation using rotation by 15 degree, vertical flipping, horizontal flipping, and slant angle (0.2) for shear transformation. The processed images are then re-sized as per the pre-trained transfer learning models we used. We use $224 \times 224 \times 3$ shaped images for the VGG16, VGG19, ResNet50, ResNet50V2, MobileNet architectures and $299 \times 299 \times 3$ shaped images for the InceptionV3, Xception, InceptionResNetV2 models. Then we split the entire dataset into training, validation, and testing. We first split the entire dataset in an 80:20

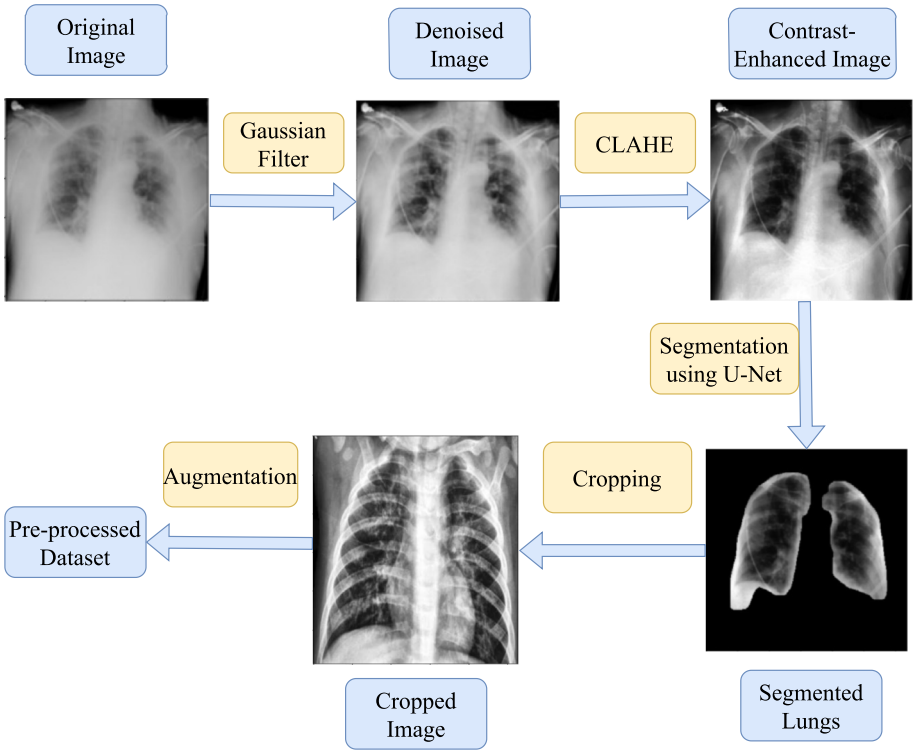


Fig. 2 Various preprocessing techniques used by us in our approach

ratio. The 20% dataset is kept for testing while the 80% dataset is split again by keeping 70% for training while 10% for validating. We use a larger share of the dataset for the training purpose as the model trains and learns parameters from it. Moreover, increasing the training dataset further might result in the model being over-fitted to the training dataset and losing its robustness for the testing dataset.

5.2 Training Deep Transfer Learning models

In this phase, we extract several semantic spatial representations and other crucial features from the COVID-19 X-rays. It is not very efficient and effective to train a deep Convolutional Neural Network architecture from scratch as there is a lack of sufficient quality and quantity of data related to COVID-19. To overcome this, we use several state-of-the-art pre-trained Transfer Learning models, VGG16, VGG19 [55], ResNet50, ResNet50V2 [56], InceptionV3 [57], Xception [58], MobileNet [59] and InceptionResNetV2 [60]. These models are selected after an extensive experimental study which suggests their use due to the contribution made by them in the improvement of the classification performance based on the essential features extracted by them. The models are used with the pre-trained imagenet weights along with fine-tuning on our dataset. The various statistical details consisting of the shape of the input, number of parameters, layers, layers that are frozen, and the size of the spatial features obtained is mentioned in Tab. 1. Figure 3 shows the implementation of transfer learning for our proposed approach. We use the pre-trained complex Convolutional neural networks which

Table 1 Statistical details about the various transfer learning models used by us

Model	Layers	Parameters	Frozen layers	Input size	Features
InceptionV3	48	23851784	46	299×299×3	2048
Xception	36	22910480	34	299×299×3	2048
InceptionResNetV2	164	55873736	162	299×299×3	1536
VGG16	16	138357544	13	224×224×3	4096
VGG19	19	143667240	16	224×224×3	4096
ResNet50	50	25636712	48	224×224×3	2048
ResNet50V2	50	25613800	48	224×224×3	2048
MobileNet	28	4253684	26	224×224×3	1024

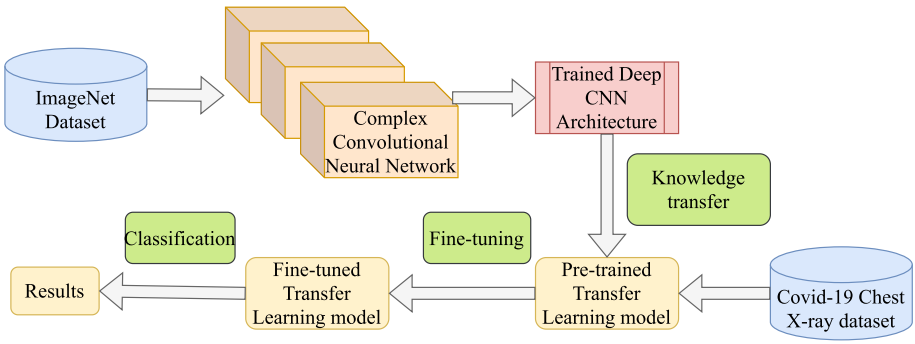


Fig. 3 The implementation of transfer learning for the task of COVID-19 detection

are pre-trained on the ImageNet dataset. The pre-trained models are then fine-tuned on the COVID-19 chest X-ray dataset. The fine-tuned models are then used to make the COVID-19 predictions. Following along these lines, we use several deep transfer learning based models like InceptionV3, InceptionResNetV2, Xception, MobileNet, ResNet50, ResNet50V2, VGG16, and VGG19. All the chosen models are thereby fine-tuned on the training datasets as mentioned in Sect. 4.1. These fine-tuned models are then used in the next section to make the final COVID-19 classifications.

5.3 Training Outputs

The models trained in the previous phase give varied outputs for the same input image. Some perform well on some images while others perform well on other images. This shows that we need a framework to optimally combine the outputs of these models and achieve a consensus which produces better results. To do this we feed the outputs of all the pre-trained models trained in the previous phase to a deep neural network efficiently combines and assigns proper weightage and importance to the outputs of these models. The outputs generated by the fine-tuned transfer learning model act as the features for the deep neural network which is trained against the ground-truth labels of the chest Xrays. This combines the features extracted by the various transfer learning models to obtain an improved performance. We use a six-layered deep neural network with the first five layers having 64, 128, 256, 128, 64, neurons with ReLU activation function [61] and 0.2 dropout rate. The final layer is a classification layer having

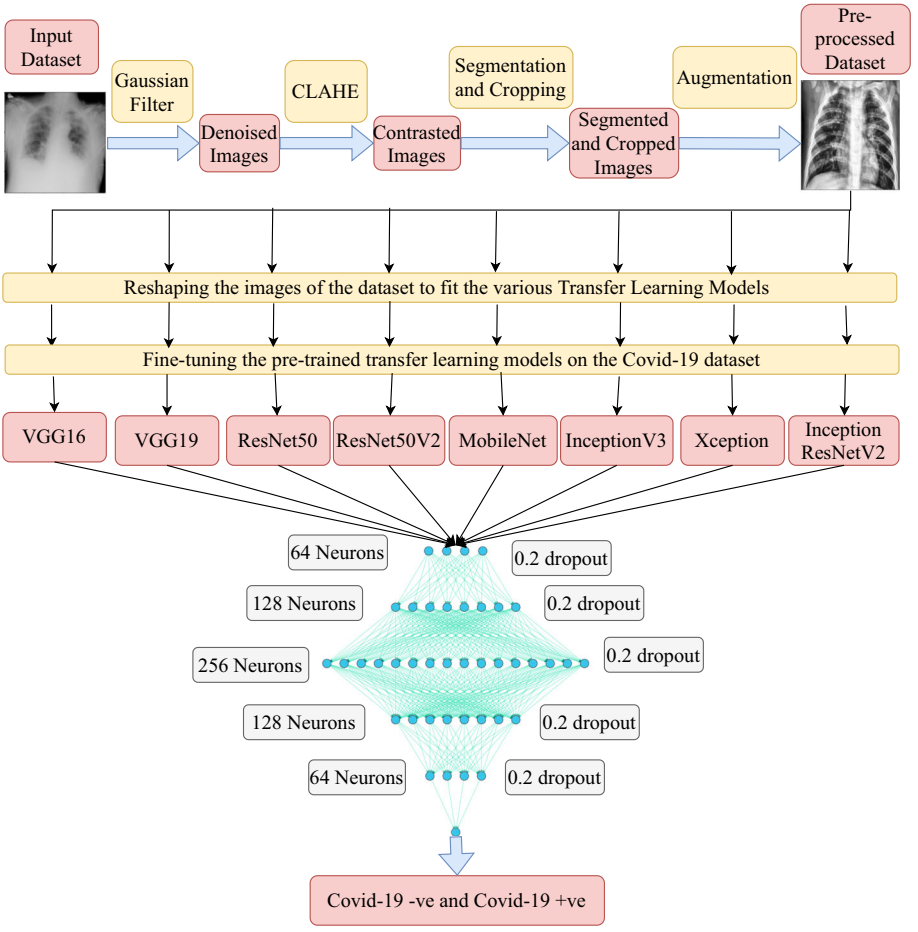


Fig. 4 The Algorithmic flowchart of our proposed trained output based transfer learning (TOTL) approach

softmax activation. The entire framework including the image preprocessor mentioned in Sect. 5.1, transfer deep learning models discussed in Sect. 5.2 and the output combining deep neural network described in this section are used to make the final COVID-19 classification based on the input chest X-ray images.

Figure 4 shows our proposed Trained Output based Deep Transfer Learning approach for COVID-19 detection using chest X-rays. We start by preprocessing the raw images as described in Sect. 5.1. Then we resize the images to fit the input requirements of various transfer learning based models. The images are then fed to these transfer learning models to fine tune the model for the COVID-19 detection task. Then we use the outputs of these models to train our deep neural network architecture for the final COVID-19 detection based on the input chest X-ray images.

5.4 Hyperparameters

Hyperparameters refer to the technical design which describes the architectural details of a model. The various hyperparameters of our model are given in Tab. 2. The hyperparameters

Table 2 The values of the various parameter used in the study

Parameters	Values
Activation function	ReLu and Softmax
Loss function	Categorical-crossentropy
Optimizer	SGD (lr = 1e-4, momentum = 0.9)
Parameters for InceptionV3	23851784
Parameters for Xception	22910480
Parameters for InceptionResNetV2	55873736
Parameters for VGG16	138357544
Parameters for VGG19	143667240
Parameters for ResNet50	25636712
Parameters for ResNet50V2	25613800
Parameters for MobileNet	4253684
Dropout rate	0.25
Decay Rate	0.9
Number of Epochs	25
Callbacks	ReduceLROnPlateau

are chosen with the help of the Random Search technique. It works by selecting and evaluating random combinations of hyperparameters from a search space to choose the most optimal parameters for the model. The activation functions chosen are ReLu and Softmax, the loss function is Categorical-crossentropy, and the optimizer is chosen to be SGD (lr = 1e-4, momentum = 0.9). The decay rate for the optimizer is 0.9. The dropout rate is set to be 0.25, the callback is chosen to be ReduceLROnPlateau while the entire framework is run for 25 epochs or iterations.

6 Experimental Analysis

In this section, we discuss the performance of our proposed Trained Output based Transfer Learning (TOTL) approach for COVID-19 detection from chest X-rays. We performed experiments on all the datasets mentioned in Sect. 4.1 based on the performance metrics mentioned in Sect. 4.2. We compared the performance of our proposed approach with several recent Chest X-ray based COVID-19 detection approaches, namely ResNet50 + SVM [62], CovidGAN [63], COVIDX-Net [22], CAAD [64], Dark COVID-Net [24], CORO-Net [65], Covid-Net [27], CNN-LSTM [66], ResNet+PSO [33], and DNN+XGBoost [67]. Since, we proposed an integrational technique combining the outputs of several state-of-the-art pre-trained Transfer Learning models, VGG16, VGG19 [55], ResNet50, ResNet50V2 [56], InceptionV3 [57], Xception [58], MobileNet [59] and InceptionResNetV2 [60], so we also compared our obtained results with the results obtained by these models, individually. The entire code is developed in the python programming language. To obtain results on existing techniques, we used several python libraries like sklearn, NumPy, pandas, etc., and some publicly available GitHub repositories. We performed simulations on a personal computer with Intel i7 11th generation processor, 16 GB RAM, and RTX 3070 graphics card. The training loss of the converged model for the various datasets is shown in Fig. 5. From the

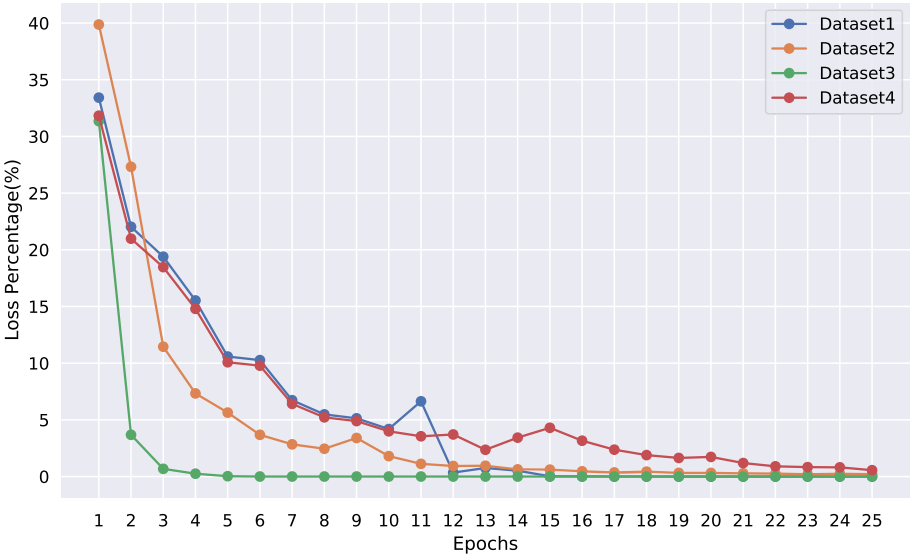


Fig. 5 Training loss of the converged model for various datasets

figure, we can clearly see that our proposed model converges around 25 epochs for all the datasets. This justifies the choice of the number of epochs for our framework. The further experimental results are discussed below.

6.1 Dataset 1

Tab. 3 shows the results obtained by various algorithms for all the evaluation metrics for Dataset 1. We see that our proposed model performs exceptionally well in terms of Accuracy with a percentage of 96.82%. ResNet50 + SVM is the best performer in terms of contemporary models. Also, InceptionV3 performs the best in terms of the classical transfer learning-based algorithms with an accuracy of 95.01%. InceptionV3 is the best performer in terms of Precision with a score of 98.54%, while the contemporary COVIDX-Net obtains the best results for Recall with a score of 98.42%. Our proposed TOTL approach performs reasonably well for these metrics as well. For the F1 Score, our proposed TOTL and the ResNet50 + SVM perform the best with an almost equal score. The obtained results also show that only our proposed algorithm gives stable results throughout the performance metrics. This can be attributed to the proper image denoising, contrasting, segmenting-based pre-processing techniques used by us. Moreover, the ability of our model to optimally combine the results of all the chosen pre-trained models also helps it capture better details and generate improved performance compared to other models.

6.2 Dataset 2

Tab. 4 shows the results obtained by the various algorithms on dataset 2 to evaluate different performance metrics. We see that in terms of all the performance metrics, our proposed Trained Output-based Transfer Learning (TOTL) approach performs the best compared to

Table 3 Results obtained for all evaluation metrics for various transfer learning algorithms for dataset 1

Methods	Accuracy	Precision	Recall	F1 Score
VGG16	93.12	97.36	88.65	92.8
VGG19	93.93	97.78	89.91	93.68
ResNet50	90.76	98.16	83.08	89.99
ResNet50V2	90.44	97.1	83.37	89.71
MobileNet	89.14	97.03	80.76	88.15
InceptionV3	95.01	98.54	91.36	94.82
Xception	89.06	93.91	83.54	88.42
InceptionResNetV2	88.67	93.98	82.64	87.95
ResNet50 + SVM	95.38	95.96	97.54	96.75
CovidGAN	95.29	93.28	97.65	95.41
COVIDX-Net	90.58	85.01	98.42	91.24
CAAD	72.77	95.94	60.66	74.33
Dark COVID-Net	85.88	84.66	87.5	86.05
CORO-Net	88.92	90.15	87.72	88.92
Covid-Net	90.59	89.41	91.11	90.25
CNN + LSTM	91.65	90.02	93.59	91.77
ResNet+PSO	93.95	80.91	90.27	84.58
DNN+XGBoost	90.15	86.29	93.96	89.96
TOTL	96.82	95.97	97.54	96.74

Table 4 Results obtained for all evaluation metrics for various transfer learning algorithms for dataset 2

Methods	Accuracy	Precision	Recall	F1 Score
VGG16	83.47	78.83	90.56	84.29
VGG19	85.43	81.95	92.15	86.75
ResNet50	89.29	86.93	93.2	89.95
ResNet50V2	89.01	85.91	92.49	89.07
MobileNet	88.94	85.97	93.35	89.51
InceptionV3	91.11	88.58	94.78	91.57
Xception	91.53	89.22	94.76	91.91
InceptionResNetV2	88.06	94.09	81.23	87.19
ResNet50 + SVM	83.94	87.41	79.85	83.45
CovidGAN	89.68	95.36	83.27	88.91
COVIDX-Net	89.87	92.85	86.18	89.39
CAAD	91.78	96.28	87.41	91.63
Dark COVID-Net	87.02	88.2	85.95	87.06
CORO-Net	95.63	99.03	92.73	95.77
Covid-Net	93.06	94.41	92.18	93.28
CNN + LSTM	97.04	98.92	93.88	96.34
ResNet+PSO	95.46	83.67	87.34	85.45
DNN+XGBoost	88.62	87.95	89.29	88.61
TOTL	98.43	99.21	97.69	98.44

Table 5 Results obtained for all evaluation metrics for various transfer learning algorithms for dataset 3

Methods	Accuracy	Precision	Recall	F1 Score
VGG16	70.87	72.28	69.52	70.87
VGG19	85.92	90.1	82.73	86.26
ResNet50	87.86	86.79	89.32	88.04
ResNet50V2	84.47	85.19	85.19	85.19
MobileNet	85.44	91.92	80.53	85.85
InceptionV3	86.41	85.19	88.46	86.79
Xception	87.38	81.25	94.79	87.5
InceptionResNetV2	70.39	67.19	81.9	73.82
ResNet50 + SVM	66.5	59.85	85.42	70.39
CovidGAN	75.24	71.56	79.59	75.36
COVIDX-Net	82.04	72.73	98.97	83.84
CAAD	86.89	84.35	91.51	87.78
Dark COVID-Net	95.63	99.03	92.73	95.77
CORO-Net	82.04	76.64	82.04	85.02
Covid-Net	82.52	80.51	87.96	84.07
CNN + LSTM	84.07	82.25	88.9	86.52
ResNet+PSO	94.97	87.18	82.78	84.93
DNN+XGBoost	89.23	91.13	87.09	89.02
TOTL	96.6	98.92	93.88	96.34

all the contemporary algorithms and all the transfer learning algorithms. Amongst the various transfer learning algorithms, Xception performs the best for accuracy, recall, and F1 Score, while InceptionResNetV2 performs the best for precision. CNN + LSTM performs the best throughout the various evaluation metrics amongst all the chosen contemporary algorithms. The better performance of our approach as compared to other algorithms can be attributed to the use of appropriate pre-processing techniques to make the input images more clearer and less noisy. The segmentation and cropping of images across the lung regions also help our model to focus on only the relevant details. The proper feature extraction done by various transfer learning models also help our model to capture various crucial details regarding COVID-19 infection from the chest X-rays. We see that our proposed algorithm outperforms all the other algorithms by a considerable margin, thus establishing its efficacy as a COVID-19 detection framework.

6.3 Dataset 3

In Tab. 5, we present the experimental results obtained for Dataset 3 for all algorithms across all the performance metrics. We infer from the results that our proposed Trained Output-based Transfer Learning (TOTL) approach outperforms all the other algorithms in terms of Accuracy and F1 Score. For Precision, our algorithm falls short only to Dark COVID-Net with a minute 0.11%, while for Recall, it lags behind COVIDX-Net and Xception with a very tiny fraction. But by obtaining an improved result for the F1 Score, our proposed algorithm proves its utility in getting balanced results for Precision and Recall as compared to other contemporary algorithms. Moreover, a high accuracy signifies the high classification

Table 6 Results obtained for all evaluation metrics for various transfer learning algorithms for dataset 4

Methods	Accuracy	Precision	Recall	F1 Score
VGG16	90.2	85.23	97.69	91.04
VGG19	90.98	85.71	95.58	90.38
ResNet50	95.29	93.28	97.66	95.42
ResNet50V2	88.63	86.89	89.08	87.97
MobileNet	89.8	90.7	89.31	90.0
InceptionV3	87.45	85.92	89.92	87.88
Xception	90.98	88.72	93.65	91.12
InceptionResNetV2	71.76	82.28	52.84	64.36
ResNet50 + SVM	67.06	78.87	44.8	57.14
CovidGAN	70.98	78.41	55.65	65.09
COVIDX-Net	76.86	81.82	66.39	73.3
CAAD	79.22	89.0	67.94	77.06
Dark COVID-Net	82.35	86.41	74.17	79.82
CORO-Net	83.92	85.59	80.8	83.13
Covid-Net	90.59	85.03	98.42	91.24
CNN + LSTM	83.92	77.27	95.2	85.3
ResNet+PSO	94.21	84.33	85.49	84.98
DNN+XGBoost	95.49	86.99	91.76	89.31
TOTL	96.47	97.81	95.71	96.75

prowess of our algorithm. The better performance of our approach can be attributed to the proper training of the outputs of various transfer learning approaches with a deep neural network instead of using trivial ensemble techniques like majority voting. Moreover, the use of several transfer learning methods exploits the dataset to extract different features from the dataset appropriately, as per the capabilities of the candidate models. This also helps our approach to give better results.

6.4 Dataset 4

In this section, we discuss the experimental results obtained for Dataset 4. Tab. 6 presents the results obtained by various transfer learning algorithms and several contemporary COVID-19 detection algorithms across all the chosen evaluation metrics for Dataset 4. The results exhibit the superior performance of our proposed Trained Output-based Transfer Learning (TOTL) approach as compared to other chosen algorithms. Our model gives better results for all the evaluation metrics, apart from recall. For recall, our algorithm falls short only to VGG16, ResNet50, and COVID-Net. The proper pre-processing techniques can be attributed to the better results obtained by our approach as they improve the quality of the input images and highlight the lung region necessary for detecting COVID-19 from the chest X-rays. Combining various transfer learning models also helps our model to extract various salient features as per the capabilities of each transfer learning model. The above discussion demonstrates the utility of our proposed Trained Output-based Transfer Learning (TOTL) approach as an optimal framework for COVID-19 detection from the chest X-rays of the patients.

Table 7 Accuracy results obtained for all the datasets obtained by various contemporary algorithms along with our proposed TOTL framework with and without data augmentation

Methods	Dataset 1	Dataset 2	Dataset 3	Dataset 4
VGG16	93.12	83.47	70.87	90.2
VGG19	93.93	85.43	85.92	90.98
ResNet50	90.76	89.29	87.86	95.29
ResNet50V2	90.44	89.01	84.47	88.63
MobileNet	89.14	88.94	85.44	89.8
InceptionV3	95.01	91.11	86.41	87.45
Xception	89.06	91.53	87.38	90.98
InceptionResNetV2	88.67	88.06	70.39	71.76
ResNet50 + SVM	95.38	83.94	66.5	67.06
CovidGAN	95.29	89.68	75.24	70.98
COVIDX-Net	90.58	89.87	82.04	76.86
CAAD	72.77	91.78	86.89	79.22
Dark COVID-Net	85.88	87.02	95.63	82.35
CORO-Net	88.92	95.63	82.04	83.92
Covid-Net	90.59	93.06	82.52	90.59
CNN + LSTM	91.65	97.04	84.07	83.92
ResNet+PSO	93.95	95.46	94.97	94.21
DNN+XGBoost	90.15	88.62	89.23	95.49
TOTL (without Aug)	90.53	90.98	92.15	95.63
TOTL (with Aug)	96.82	98.43	96.6	96.47

6.5 Results Without Data Augmentation

This section presents the results obtained for all the datasets obtained by various contemporary algorithms along with our proposed TOTL framework with and without data augmentation. The obtained results are presented in Tab. 7. From the results, we can infer that the results obtained by our algorithm without data augmentation are inferior compared to that obtained with data augmentation. The obtained results are even inferior to several transfer learning-based approaches. This can be attributed to the class bias of our framework to the class having a larger number of samples due to the over-fitting during the training phase. However, the results obtained by our framework without data augmentation are still better than several contemporary algorithms like COVIDX-Net, CAAD, Dark COVID-Net, and CORO-Net.

6.6 Hyperparameter Analysis

This section illustrates the impact of the hyperparameters on the final results. Tab. 8, shows the accuracy results obtained by our model on various datasets as the various hyperparameters are varied. From Tab. 8, we can see that how the accuracy improves as the number of epochs are increased for all the four datasets. It also shows the impact of the choice of optimizer on the performance of our model. We see that the chosen SGD optimizer gives the best performance as compared to other optimizers. The obtained results demonstrate the efficacy and efficiency of our chosen hyperparameters.

Table 8 The accuracy results obtained by our model on various datasets as the various hyperparameters are varied

Epochs	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy over the epochs				
5	81.32	84.94	86.19	83.87
10	88.06	94.97	91.43	87.32
15	92.50	97.19	91.71	91.45
20	95.74	97.69	94.09	95.07
25	96.82	98.43	96.6	96.47
Optimizer	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy based on the optimizer				
RMSprop	95.34	98.14	95.45	95.16
Adam	96.07	98.03	95.89	95.28
SGD	96.82	98.43	96.6	96.47

Table 9 Statistical significance of the model performance based on the confidence interval on accuracy for all the datasets

Confidence	Dataset 1	Dataset 2	Dataset 3	Dataset 4
90%	[96.8243, 96.8157]	[96.8251, 96.8149]	[96.8260, 96.8140]	[96.8267, 96.8133]
95%	[98.4317, 98.4283]	[98.4320, 98.4280]	[98.4324, 98.4276]	[98.4326, 98.4274]
98%	[96.6167, 96.5833]	[96.6200, 96.5800]	[96.6237, 96.5763]	[96.6263, 96.5737]
99%	[96.4738, 96.4662]	[96.4745, 96.4655]	[96.4754, 96.4646]	[96.4759, 96.4641]

Tab. 9 presents the statistical significance of the model performance based on the confidence interval on accuracy for all the datasets. From Tab. 9, we can see that the obtained confidence results are obtained are very close to the accuracy results achieved by us. This shows that the performance of our model is mathematically justified and is superior to the other algorithms. Through the above discussion, we infer that our proposed Trained Output based Transfer Learning (TOTL) approach performs reasonably well for the task of COVID-19 detection using the chest X-rays. Our model also achieves a consistent performance throughout the datasets and performance metrics and outperforms various chosen transfer learning approaches and contemporary COVID-19 detection frameworks. The good results of our method can be attributed to the proper use of preprocessing techniques like denoising using Gaussian filter, contrasting using CLAHE, segmentation using U-Net, and cropping around the lung region to capture only the details relevant to COVID-19 detection. We also further train the outputs generated by several transfer learning models to assign appropriate importance to the outputs obtained by each of them. Thus achieving an enhanced performance as compared to their individual results. The extensive size of the datasets used by us also provide us with sufficient data to optimally train the model parameters.

7 Conclusion

COVID-19 is a global pandemic that has spread across the globe and has caused millions of deaths. The existing COVID-19 screening methods are expensive, time-consuming,

and menial. In this paper, we introduced a novel Trained Output-based Transfer Learning approach for COVID-19 detection using chest X-rays. We started by preprocessing the chest X-rays using various techniques, nemely, denoising using Gaussian filters, contrasting using CLAHE algorithm, segmentation using U-Net, and then cropping around the lungs region to only use relevant details. Then we passed these images through various deep transfer learning-based pre-trained models, namely InceptionV3, InceptionResNetV2, Xception, MobileNet, ResNet50, ResNet50V2, VGG16, and VGG19. The models are then fine-tuned on the chest X-ray images, and their outputs are fed to a deep neural network. The deep neural network assisted in allocating appropriate importance to the outputs of each model. This trained framework is then used to detect COVID-19 from chest X-rays. We performed extensive experimentations using four benchmarked datasets and evaluated several popular performance metrics. We also compared the performance of our model with several classical transfer learning models and various contemporary COVID-19 detection frameworks. The results obtained exhibit the efficiency of our proposed algorithm for COVID-19 detection. This work can be extended further by using several other symptoms or other organ infections to detect COVID-19. Moreover, we can also use other features like age, gender, medical history, genetic history, location info, etc., along with chest X-rays to detect COVID-19. Various respiratory parameters can also be considered in future research. Several other diseases can be explored using computer vision based approaches.

Declarations

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

References

1. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, Nguyen TT, Cao TM, Pham QD (2020) Importation and human-to-human transmission of a novel coronavirus in Vietnam. N Engl J Med 382(9):872–874
2. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P (2020) Covid-CT-dataset: a CT scan dataset about covid-19. arXiv preprint arXiv:2003.13865
3. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, Zimmer T, Thiel V, Janke C, Guggemos W et al (2020) Transmission of 2019-NCoV infection from an asymptomatic contact in Germany. N Engl J Med 382(10):970–971
4. Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, Dittrich S, Emperador D, Takwoingi Y, Cunningham J, et al. (2021) Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-COV-2 infection. Cochrane Database Syst Rev (3)
5. Yu X, Kang C, Guttery DS, Kadry S, Chen Y, Zhang YD (2020) Resnet-SCDA-50 for breast abnormality classification. IEEE/ACM Trans Comput Biol Bioinf 18(1):94–102
6. Dangis A, Gieraerts C, De Bruecker Y, Janssen L, Valgaeren H, Obbels D, Gillis M, Van Ranst M, Frans J, Demeyere A, et al. (2020) Accuracy and reproducibility of low-dose submillisievert chest CT for the diagnosis of covid-19. Radiol Cardiothorac Imaging 2(2):e200196
7. Mohamadou Y, Halidou A, Kapen PT (2020) A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of covid-19. Appl Intell 50(11):3913–3925
8. Toğaçar M, Ergen B, Cömert Z (2020) Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches. Comput Biol Med 121:103805
9. Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F (2021) A machine learning-based framework for diagnosis of covid-19 from chest x-ray images. Interdiscip Sci Comput Life Sci 13(1):103–117
10. Kumar S, Mallik A, Khetarpal A, Panda B (2022) Influence maximization in social networks using graph embedding and graph neural network. Inf Sci

11. Jain R, Gupta M, Taneja S, Hemanth DJ (2021) Deep learning based detection and analysis of covid-19 on chest x-ray images. Appl Intell 51(3):1690–1700
12. Nigam B, Nigam A, Jain R, Dodia S, Arora N, Annappa B (2021) Covid-19: automatic detection from x-ray images by utilizing deep learning methods. Expert Syst Appl 176:114883
13. Cellina M, Orsi M, Toluian T, Pittino CV, Oliva G (2020) False negative chest x-rays in patients affected by covid-19 pneumonia and corresponding chest CT findings. Radiography 26(3):e189–e194
14. Shaik NS, Cherukuri TK (2021) Lesion-aware attention with neural support vector machine for retinopathy diagnosis. Mach Vis Appl 32(6):1–13
15. Bodapati JD, Shaik NS, Naralasetti V (2021) Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification. J Ambient Intelli Hum Comput 1–15
16. Bodapati JD, Shaik NS, Naralasetti V (2021) Deep convolution feature aggregation: an application to diabetic retinopathy severity level prediction. Signal Image Video Process pp 1–8
17. Bodapati JD, Shaik NS, Naralasetti V, Mundukur NB (2021) Joint training of two-channel deep neural network for brain tumor classification. SIViP 15(4):753–760
18. Bodapati JD, Shareef SN, Naralasetti V, Mundukur NB (2021) Msenet: multi-modal squeeze-and-excitation network for brain tumor severity prediction. Int J Pattern Recognit Artif Intell 2157005
19. Shaik NS, Cherukuri TK (2021) Multi-level attention network: application to brain tumor classification. Signal Image Video Process 1–8
20. Mohamadou Y, Halidou A, Kapen PT (2020) A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of covid-19. Appl Intell 50(11):3913–3925
21. Luz E, Silva P, Silva R, Silva L, Guimarães J, Miozzo G, Moreira G, Menotti D (2021) Towards an effective and efficient deep learning model for covid-19 patterns detection in x-ray images. Res Biomed Eng 1–14
22. Hemdan EED, Shouman MA, Karar ME (2020) Covidx-net: a framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055
23. Basu S, Mitra S, Saha N (2020) Deep learning for screening covid-19 using chest x-ray images. In: 2020 IEEE symposium series on computational intelligence (SSCI). IEEE, pp 2521–2527
24. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of covid-19 cases using deep neural networks with x-ray images. Comput Biol Med 121:103792
25. Das NN, Kumar N, Kaur M, Kumar V, Singh D (2020) Automated deep transfer learning-based approach for detection of covid-19 infection in chest x-rays. IRBM
26. Tuncer T, Dogan S, Ozyurt F (2020) An automated residual exemplar local binary pattern and iterative relieff based covid-19 detection method using chest x-ray image. Chemom Intell Lab Syst 203:104054
27. Wang L, Lin ZQ, Wong A (2020) Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Sci Rep 10(1):1–12
28. Zheng Q, Zhao P, Li Y, Wang H, Yang Y (2021) Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification. Neural Comput Appl 33(13):7723–7745
29. Zheng Q, Zhao P, Zhang D, Wang H (2021) Mr-dcae: manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification. Int J Intell Syst 36(12):7204–7238
30. Jin B, Cruz L, Gonçalves N (2021) Face depth prediction by the scene depth. In: 2021 IEEE/ACIS 19th international conference on computer and information science (ICIS). IEEE, pp 42–48
31. Jin B, Cruz L, Gonçalves N (2020) Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis. IEEE Access 8:123649–123661
32. Zheng Q, Yang M, Tian X, Jiang N, Wang D (2020) A full stage data augmentation method in deep convolutional neural network for natural image classification. Discrete Dyn Nat Soc 2020
33. Shen C, Zhang K, Tang J (2021) A covid-19 detection algorithm using deep features and discrete social learning particle swarm optimization for edge computing devices. ACM Trans Internet Technol (TOIT) 22(3):1–17
34. He J, Zhu Q, Zhang K, Yu P, Tang J (2021) An evolvable adversarial network with gradient penalty for covid-19 infection segmentation. Appl Soft Comput 113:107947
35. Mu N, Wang H, Zhang Y, Jiang J, Tang J (2021) Progressive global perception and local polishing network for lung infection segmentation of covid-19 CT images. Pattern Recognit 120:108168
36. Oulefki A, Agaian S, Trongtirakul T, Laouar AK (2021) Automatic covid-19 lung infected region segmentation and measurement using CT-scans images. Pattern Recognit 114:107747
37. Liu X, Yuan Q, Gao Y, He K, Wang S, Tang X, Tang J, Shen D (2022) Weakly supervised segmentation of covid19 infection with scribble annotation on CT images. Pattern Recognit 122:108341
38. Lella KK, PJA A (2021) A literature review on covid-19 disease diagnosis from respiratory sound data. arXiv preprint arXiv:2112.07670

39. Lella KK, Pja A (2021) Automatic covid-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice. AIMS Public Health 8(2):240
40. Lella KK, Pja A (2022) Automatic diagnosis of covid-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath. Alex Eng J 61(2):1319–1334
41. Kranthi Kumar L, Alphonse P (2022) Covid-19 disease diagnosis with light-weight CNN using modified MFCC and enhanced GFCC from human respiratory sounds. Eur Phys J Spec Top 1–18
42. Narin A, Kaya C, Pamuk Z (2021) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Anal Appl 1–14
43. Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaševičius R, De Albuquerque VHC (2020) A novel transfer learning based approach for pneumonia detection in chest x-ray images. Appl Sci 10(2):559
44. Shelke A, Inamdar M, Shah V, Tiwari A, Hussain A, Chafekar T, Mehendale N (2021) Chest x-ray classification using deep learning for automated covid-19 screening. SN Comput Sci 2(4):1–9
45. Rahimzadeh M, Attar A (2020) A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2. Inf Med Unlocked 19:100360
46. Abbas A, Abdelsamea MM, Gaber MM (2021) Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. Appl Intell 51(2):854–864
47. Prabhu R (2018) Understanding of convolutional neural network (CNN)-deep learning. Medium Com pp 1–11
48. O'Shea K, Nash R (2015) An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458
49. Orenstein EC, Beijbom O (2017) Transfer learning and deep feature extraction for planktonic image data sets. In: 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1082–1088
50. Guo Y, Shi H, Kumar A, Grauman K, Rosing T, Feris R (2019) Spottune: transfer learning through adaptive fine-tuning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4805–4814
51. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 248–255
52. Wang M, Zheng S, Li X, Qin X (2014) A new image denoising method based on gaussian filter. In: 2014 international conference on information science, electronics and electrical engineering. IEEE, vol 1, pp 163–167
53. Yadav G, Maheshwari S, Agarwal A (2014) Contrast limited adaptive histogram equalization based enhancement for real time video system. In: 2014 international conference on advances in computing. communications and informatics (ICACCI). IEEE, pp 2392–2397
54. Yadav G, Maheshwari S, Agarwal A (2014) Contrast limited adaptive histogram equalization based enhancement for real time video system. In: 2014 international conference on advances in computing. communications and informatics (ICACCI). IEEE, pp 2392–2397
55. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
56. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
57. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
58. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
59. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861
60. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence
61. Agarap AF (2018) Deep learning using rectified linear units (RELU). arXiv preprint arXiv:1803.08375
62. Sethy PK, Behera SK (2020) Detection of coronavirus disease (covid-19) based on deep features
63. Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR (2020) Covidgan: data augmentation using auxiliary classifier GAN for improved covid-19 detection. IEEE Access 8:91916–91923

64. Zhang J, Xie Y, Pang G, Liao Z, Verjans J, Li W, Sun Z, He J, Li Y, Shen C, et al. (2020) Viral pneumonia screening on chest x-ray images using confidence-aware anomaly detection. arXiv preprint arXiv:2003.12338
65. Khan AI, Shah JL, Bhat MM (2020) Coronet: a deep neural network for detection and diagnosis of covid-19 from chest x-ray images. Comput Methods Programs Biomed 196:105581
66. Islam MZ, Islam MM, Asraf A (2020) A combined deep CNN-LSTM network for the detection of novel coronavirus (covid-19) using x-ray images. Inf Med Unlocked 20:100412
67. Nasiri H, Hasani S (2022) Automated detection of covid-19 cases from chest x-ray images using deep neural network and xgboost. Radiography

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.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0368-492X.htm

Crowdsourcing a wellspring of value co-creation: an integration of social capital and organisational learning mechanisms

Divya Mishra and Nidhi Maheshwari
University School of Management and Entrepreneurship (USME), Delhi Technological University, Delhi, India

Received 19 April 2022
Revised 2 July 2022
25 October 2022
Accepted 27 October 2022

Abstract

Purpose – With the advent of Internet technologies, shorter product life cycles and increasing competition, organisations have started looking for innovation sources outside the organisational boundaries. The external community of crowds can be used as a valuable source of co-creation in a company's innovation process to generate value. Despite its growing popularity, organisations often face difficulty capturing value from crowdsourcing due to the lack of proper mechanisms behind crowdsourcing-based value co-creation between a crowd and an organisation and their impact on organisational learning and innovation performance. The present study seeks to understand the crowdsourcing-based co-creation mechanism that influences knowledge transfer effectiveness and the organisation's absorptive capacity, resulting in improved innovation performance.

Design/methodology/approach – The model was empirically tested using online survey data received from 300 managers of IT firms. Partial least squares structural equation modelling was used to test the model.

Findings – The empirical results reveal that crowdsourcing-based value co-creation causes structural, cognitive and relational linkages between a crowd and a firm, among which crowdsourcing-based cognitive linkage contributes more to organisational value capture. Further, an organisation's effective knowledge transfer and absorptive capacity play an important role in influencing the crowdsourcing-based-co-creation organisational learning-innovation performance framework.

Originality/value – This is the first and foremost study that has developed an integrated model using social capital dimensions to understand the entire mechanism behind crowdsourcing-based value co-creation between a crowd and an organisation and their impact on organisational learning and innovation performance. The study provides organisations with theoretical and practical implications of using crowdsourcing as a value co-creation tool and its effects on enhancing organisational learning and value capture.

Keywords Crowdsourcing, Value co-creation, Social capital, Organisational learning, Organisational value capture, Innovation performance

Paper type Research paper

1. Introduction

Emerging disruptive technologies are becoming a key for companies to survive, stay competitive and cope with the flexibility and high dynamics of the innovation process. Taking the first-mover advantage in advanced technologies requires adopting responsive and agile innovation strategies. Henceforth, advances in information technologies have resulted in "crowdsourcing," which involves crowds' collective intelligence to innovate and reach a broader spectrum of people (Heeks, 2010; Bullinger *et al.*, 2010), which adds richness and depth to the innovation process. Crowdsourcing has gained management's attention as a co-creation tool for companies to engage the virtual community of the crowd to solve a problem and innovation needs of the organisation (Afuah and Tucci, 2012; Boudreau and Lakhani, 2013).

High-tech firms have started leveraging crowds collective intelligence to share knowledge, optimise cost and involve in continuous innovation and co-creation that is rare, valuable, non-substitutable and inimitable (Barney, 1991; Teece, 1993). Organisations engage with the

Kybernetes
© Emerald Publishing Limited
0368-492X
DOI 10.1108/K-04-2022-0580

"wisdom of crowds" to gather innovative ideas and then keep making changes as per their needs, thereby accelerating the rate of innovation acceptance among crowds (Marjanovic *et al.*, 2012). For instance, the "Battleofconcepts" is a crowdsourcing platform where winning ideas are rewarded and winners co-create their ideas into realisation. This results in meaningful crowd-firm engagement and value capture. Similarly, "DHL's innovation centres," a logistic company, invited experts and their hardcore fans to actively collaborate in solving complex problems of voyage sequencing and package delivery. Thus, crowdsourcing-based co-creation allows companies to work with their crowd community and fulfil their requirements per their needs. The underlying premise of such a collaborative business model is that crowd-firm engagement specific to value capture in companies can improve innovation performance. Although literature exhibits the need to understand the role of crowd-firm co-creation in crowdsourcing, studies focusing on the mechanism behind the influence of crowdsourcing-based co-creation on value capture are not yet explored.

A review of past scholarly research gives important insights into the effect of crowdsourcing-based co-creation. First, past studies highlight the benefits and risks associated with crowd-value co-creation. On the one hand, the collective intelligence of a crowd helps in the strategic renewal of an enterprise (Nevo and Kotlarsky, 2020), solving complex problems (Brabham, 2008; Vukovic, 2009), performing tedious tasks (Afuah and Tucci, 2012), gathering novel ideas (Boudreau and Lakhani, 2013) and brand awareness (Schulze *et al.,* 2011). Conversely, crowd value co-creation leads to culture, coordination, intellectual and control risk (Lehner, 2013; Wazny, 2017). The results indicate that the overall effect of crowdsourcing-based value co-creation on organisational innovation depends on moderating and mediating factors. Second, limited studies focused on crowdsourcing-based crowd-firm value co-creation from an organisational perspective. The implication of crowd-firm co-creation in crowdsourcing for organisational innovation is not well-known. As crowdsourcing facilitates organisations to harness the crowd's collective intelligence to enhance innovation performance, it becomes imperative to explore how crowdsourcing-based co-creation influences an organisation's value capture.

This gap propels the researchers to unravel how crowdsourcing-based crowd-firm co-creation influences organisations' value capture (measured by organisational innovation performance). More specifically, the important research question investigated in the present study is:

RQ1. What mechanism influences the relationship between crowdsourcing-based crowd-firm co-creation and organisational value capture?

Different works of research by Liao *et al.* (2007), Levy (2009) and Poetz and Schreier (2012) unveil that the amount and variety of information received from connections and interactions with external networks help shape creative competence and gain competitive advantage and organisational innovation. Even though past researchers have talked about its importance in management and research practices, a conceptual understanding of the role of crowd-firm co-creation in crowdsourcing is still missing. Taking Iyengar *et al.*'s (2015) model as a base, the present study takes crowdsourcing-based social linkage as a co-creation mechanism leading to organisational learning via knowledge transfer effectiveness and absorptive capacity capability. Social capital, organisational learning and innovation performance are presented in a nomological framework to examine the process of crowdsourcing-enabled value co-creation.

While previous literature has revealed the significance of information technologies in influencing value co-creation, the studies linking IT-mediated technologies to the crowd-focused value co-creation process are limited. Therefore, this study is the first step towards understanding the role of IT-mediated technologies for crowd-firm interactions and value capture in organisational innovation. With the advent of Web 2.0, there had been an

advancement in Information and Communication Technologies (ICT), Internet access, and the collaborative and networked world (Gassmann *et al.*, 2014), resulting in crowdsourcing proliferation. However, to the author's knowledge, published studies have not yet empirically investigated the role of crowdsourcing as a facilitator for crowd co-creation and improved organisational innovation performance. Despite the growing importance of crowdsourcing usage in developed countries, the role of crowdsourcing as a co-creation tool to enhance organisations' value capture is not yet studied in developing countries like India. Given the gaps above, the present study explores how crowdsourcing use complements crowd-firm social linkage, its effect on organisational learning and value capture in the form of organisational innovation. The empirical research is based on Indian employees working on different crowdsourcing projects in the information and technology (IT) industry. The self-administered survey will be directed to the managers of IT firms involved in crowdsourcing.

The paper is structured in the following manner: Section 2 reviews the advances in the literature on crowdsourcing-based co-creation and its outcomes on organisational value capture. The research framework and hypotheses are discussed in Section 3. Section 4 discusses the data collection, study context and measurement development. Section 5 explains the research methodology used to validate the established research framework. Section 6 presents a discussion of the study findings and hypothesis validations. Finally, Section 7 summarises the research's findings, implications, limitations and suggested future research directions.

2. Literature review

2.1 Crowdsourcing-based co-creation mechanism

A firm's external network of the crowd consists of information, knowledge and data that is not available internally and reduces economic and social frictions (Bapna *et al.*, 2017), value capture and co-creation of value between crowds and firms (Piyathasanan *et al.*, 2017). Thus, the voluntary engagement, contribution and interaction between crowd-firm results in social capital building. Resource-based view states that knowledge and resources are heterogeneously distributed among the crowd that is valuable, immobile, scarce and difficult to substitute (Grover and Kohli, 2012). Similarly, the relationship view states that an organisation's internal resources help gain a competitive advantage from the wisdom of crowd-firm and crowd-crowd relationships (Rai *et al.*, 2012). The views are the logic behind co-creation and value capture between the firm and its community of crowds.

Social capital emerging from crowd-firm interactions is a valuable and productive resource that diminishes information-sharing and organisational learning barriers. It enables value creation, knowledge sharing, improved performance and organisational development to gain a competitive advantage (Abili, 2011; Allameh, 2018). In particular, several organisational studies have considered social capital as a facilitator of knowledge contribution, acquisition, creation and innovation (Nahapiet and Ghoshal, 1998; Tsai and Ghoshal, 1998; Wasko and Faraj, 2005). Therefore, the study focuses on the perspective of social capital-based social linkage that acts as a catalyst for the value co-creation mechanism.

Social capital mainly consists of structural, cognitive and relational linkage (Nahapiet and Ghoshal, 1998). Hence, the study intellectualises the three dimensions emerging from the association between a crowd and a firm as a mechanism of crowd-firm value co-creation. The co-creation mechanism leads to real-time interactions between the firm and the crowd's wisdom. Therefore, the focus of the study is to understand the relationships formed, shared understanding and the mutual relationship of crowdsourcing-based crowd-firm interactions for improving innovation capability (value capture). The structural dimension describes the overall relationship pattern between the crowd and the firm during crowdsourcing. During crowdsourcing, the cognitive dimension describes how the crowd and firm understand each

other's perspectives through shared language, vision, objectives and values (Wasko and Faraj, 2005). Finally, the relational dimension refers to the level of respect and trust, the working pattern between crowd and firm during crowdsourcing (Chow and Chan, 2008).

2.2 Co-creation outcome

The crowdsourcing-based crowd-firm co-creation for value capture in innovation is based on the premise of organisational learning. Past reviews highlight that the organisational learning capability is improved with a firm's connection and the knowledge acquired from external networks (Schlagwein and Andersen, 2014; Xu *et al.*, 2015; Devece *et al.*, 2019). Schulze and Hoegl (2008) further identified the effect of crowdsourcing on learning the distinctive competence of organisations. These learning competencies result in developing a continuous learning system (Hine *et al.*, 2010) that gives an organisation a competitive edge. The collective intelligence gathered from the external networks acts as a market research tool to obtain tacit knowledge about trends, customers, products and services as per the requirements. Thus, organisations that implement crowdsourcing as a value capture tool have a greater capacity to learn, be very receptive to market tendencies and foresee changes easily compared to competitors.

Previous research (e.g. Ye and Kankanhalli, 2015; Schlagwein and Anderson, 2014; Xu *et al.*, 2015; Devece *et al.*, 2019; Lenart-Gansiniec and Sułkowski, 2020) suggested that an IT-enabled crowdsourcing platform drives organisational learning for improved innovation capability. However, from a broader perspective, crowdsourcing-enabled value capture is a process that involves interacting with external specialists having diverse knowledge, resources, abilities, mastery and heterogeneous skills (Whitla, 2009). Firms that use crowdsourcing practices as a facilitator of co-creation and value capture have excellent organisational learning capability to promote firms' adaptation and continuous improvement and integrate external resources into the innovation process. With the case study of crowdsourcing in LEGO, Schlagwein and Andersen (2014) proposed that "organisational learning with crowdsourcing is a novel, unique (e.g. external, non-professional, IT-shaped, and entrepreneurial) form of organisational learning. It is not just a cheaper, digital re-enactment of traditional employee–based organisational learning, but rather an alternative learning mode". Therefore, we intellectualise crowdsourcing-based crowd-firm co-creation in innovation as an organisational learning process.

Iyengar *et al.*'s (2015) "Learning mechanism-outcome performance" model (Figure 1) linked learning and improved organisational performance as an organisational learning mechanism. The outcome of organisational learning capability is enhanced through the firm's knowledge transfer propensity and absorptive capacity. Knowledge transfer refers to the functional competencies and proficiencies to acquire external resources to strengthen operational activities. Absorptive capacity refers to a dynamic capability of integrating external resources into a firm's innovation capacity. Therefore, we intellectualise crowdsourcing-based value capture as an organisational learning process by taking the Iyengar model as a base. The study considers knowledge transfer propensity as acquiring tacit and implicit knowledge from connections and interactions with external networks and absorptive capacity as an acclimatisation capability to identify and assimilate the knowledge acquired for shaping innovative performance. Crowdsourcing-based co-creation is a means of enhancing acquisition (knowledge transfer propensity) and acclimatisation (absorptive capacity) capabilities, thus, resulting in value capture (organisational innovation capability).

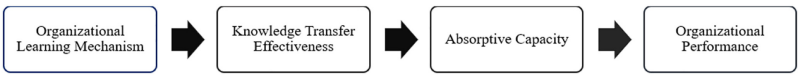


Figure 1.
Organisational learning mechanism outcome performance model purported by Iyengar *et al.* (2015)

2.3 Organisational value capture

Co-creating values with IT-enabled tools have five essential components influencing organisational value capture, namely, complementary resources, knowledge-sharing routines, effective governance, new information technologies and complementary capabilities (Grover and Kohli, 2012). Value capture is the capacity of an organisation to capture the crowd's collective intelligence to improve innovation performance (Lepak *et al.*, 2007). Organisational innovation refers to the innovative processes, procedures and practices that an organisation utilises to meet the changing needs and expectations, resulting in competitive advantage (Rezaei and Moghanlo, 2014). From the study perspective, co-creation refers to a collective process between a firm and a crowd to acquire knowledge and assimilate the available resources to enhance an organisation's research and innovation capabilities. Organisations such as IBM (Innovation Jam), Adidas, Foldit, BMW, Starbucks, Fiat Mio (Bjelland and Wood, 2008; Saldanha *et al.*, 2014), MyStarbucksIdea, Threadless and Innocentive (Acar, 2019) have embraced crowdsourcing-based co-creation to enhance their research and innovation capabilities leading to value capture in operations. Table 1 lays out the elements of co-creation and value capture used in the present study and discussed in different empirical studies at the organisational level.

3. Theoretical framework and hypothesis development

3.1 Crowdsourcing-based social capital and organisational learning

Value co-creation research often focuses on the notion that co-creation is a facilitator of enhancing an organisation's acquisition and assimilation capabilities (Grover and Kohli, 2012). For instance, Lusch and Nambiasan (2015) argued that value co-creation allows the customer and the firm to exchange knowledge and resources to enhance operational capabilities. Similarly, Johannessen and Olsen (2010) stated that co-creation between the firm and the crowd results in knowledge acquisition and integration of diverse expertise, thus resulting in enhanced capabilities of organisations. Prior research on social capital (Wasko and Faraj, 2005; Chang and Chuang, 2011) also claimed that social capital leads to knowledge contribution. For instance, Yan *et al.* (2019) proposed a relationship between social capital and knowledge contribution among online user communities. Nov *et al.* (2012) asserted that a strong connection between a firm and a crowd leads to knowledge sharing and acclimatisation of resources acquired. The social capital theory claims that social capital facilitates cooperation and coordination for mutual benefit (Putnam, 1995), enabling knowledge acquisition and exploitation to accelerate innovation in technology-based firms (Tsai and Ghosal, 1998; Nahapiet and Ghosahal, 1998; Yli-Renko *et al.*, 2001).

The social capital theory offers numerous reasons providing an impetus to propose that social capital entrenched in the crowdsourcing-based interaction between a firm and a crowd leads to organisational learning with enhanced knowledge transfer propensity and absorptive capacity. The crowdsourcing-based structural linkage between a crowd and a firm reduces transaction costs for information acquisition, assimilation, diffusion and collective understanding (Chiu *et al.*, 2006). Structural linkages formed during crowdsourcing facilitate interaction, transferring diverse knowledge, resources, abilities, mastery and heterogeneous skills (Whitla, 2009; Ansari *et al.*, 2012; Ye and Kankanhalli, 2015) between parties, namely, a crowd and a firm. Thus, crowdsourcing-based structural linkage enhances explicit and tacit knowledge broadening the base of minds resulting in more top-down and bottom-up connections, conversations, dialogue and discussions in organisations (Wagner *et al.*, 2014; Nishikawa *et al.*, 2017; Allen *et al.*, 2018) to fulfil the crowd's requirement. Prior research (Yli-Renko *et al.*, 2001; Chang *et al.*, 2012; Krupowicz *et al.*, 2020) suggests that strong ties, connections and interactions in a network provide a mechanism to enhance knowledge sharing behaviour; achieve and sustain intensity; breadth, frequency of knowledge transfer; and

Table 1.
Elements of co-creation and value capture used in our study

Empirical studies	Complementary resources	Knowledge-sharing routines	Effective governance	New information technologies	Complementary capabilities	Organisation value capture integrating innovation capability	Organisational innovation
Forbes and Schaefer (2018)	✓	✓			✓		
Chanal and Fasan (2008)	✓	✓		✓	✓		
Afuah and Tucci (2012)			✓		✓		
Marjanovic *et al.* (2012)							✓
Mehtala *et al.* (2016)	✓	✓			✓		✓
Palacious *et al.* (2016)	✓				✓		✓
Hammon and Hippner (2012)		✓		✓		✓	✓
Zhu *et al.* (2016)					✓		✓
Ye and Kankanhalli (2015)						✓	✓
Xu *et al.* (2015)							✓
Modaresnezhad *et al.* (2020)				✓		✓	
Elements Focussed in Study	✓	✓		✓	✓	✓	✓

Note(s): The Table depicts the study's novelty compared to the other studies conducted

absorptive capacity. For instance, the success of Danish LEGO Group, a toy manufacturer company, lies in the structural linkage that fosters social interaction between the company and users to develop innovative product concepts and get feedback directly from users and the company for further development (Rass *et al.*, 2013; Schlagwein and Andersen, 2014).

Further, crowdsourcing-based cognitive linkage impacts knowledge transfer by enhancing the ability of the crowd to access and contribute knowledge to a community (Nahapiet and Ghoshal, 1998; Tsai and Ghoshal, 1998). In a knowledge exchange process, the crowd discusses ideas and answers with others, receives updates on the best solutions, and reads and comments on other people's thoughts and solutions. Additionally, cognitive linkage enables mutual understanding of collective goals between a crowd and a firm, which is indispensable to augment the dynamic capability of an organisation to apply knowledge and resources (Nambisan and Baron, 2010; Lusch and Nambisan, 2015). For instance, Threadless, an apparel company, uses shared codes and languages between the crowd and a firm to create, inspect, approve, improve and select shirt designs. The cognitive linkage formed during crowd-firm interaction facilitates Threadless to understand the preferences and needs of crowds correctly and then proceed with customised shirt design (Piller, 2010). Therefore, cognitive linkages impact the absorptive capacity by enhancing three capabilities: identifying innovative ideas, acclimatising new insights and making changes accordingly in the product and services (Iyengar *et al.*, 2015).

Finally, crowdsourcing-based relational linkage in the study focuses on respect and trust built by the crowd-firm interactions that affect the quality and quantity of knowledge exchange (Gu and Jarvenpaa, 2003; Dholakia *et al.*, 2009; Wiertz and De Ruyter, 2007; Porte and Donthu, 2008; Wagner *et al.*, 2014). Respect and trust in the crowdsourcing community enhance knowledge-sharing intentions and collaboration capability by creating a positive atmosphere to exchange knowledge, assimilate, interpret and integrate within the organisation (Nevo and Wade, 2011; Prpic *et al.*, 2015). The outcome of the crowd-firm collaboration is improved organisational learning capability resulting in information acquisition, assimilation, diffusion and collective understanding to be applied for innovation purposes (Maktabi and Khazaei, 2014; Onag *et al.*, 2014). Therefore, we tested the following hypotheses:

H1. Crowdsourcing-based social capital formed with structural, cognitive and relational linkages between crowd-firm positively relates to an organisation's knowledge transfer propensity.

H1a. Crowdsourcing-based structural linkage positively affects an organisation's knowledge transfer propensity.

H1b. Crowdsourcing-based cognitive linkage positively affects an organisation's knowledge transfer propensity.

H1c. Crowdsourcing-based relational linkage positively affects an organisation's knowledge transfer propensity.

H2. Crowdsourcing-based social capital formed with structural, cognitive and relational linkages between crowd-firm positively relates to the organisation's absorptive capacity.

H2a. Crowdsourcing-based structural linkage positively affects an organisation's absorptive capacity.

H2b. Crowdsourcing-based cognitive linkage positively affects an organisation's absorptive capacity.

H2c. Crowdsourcing-based relational linkage positively affects an organisation's absorptive capacity.

3.2 Organisational learning and organisational value capture

The essence of absorptive capacity lies in the knowledge transfer effectiveness of an organisation. Organisations go beyond their boundaries to acquire and integrate the crowd's contribution to solving innovation problems. Prior research (Roberts *et al.*, 2012; Prpic *et al.*, 2015) has demonstrated that organisations' absorptive capacity depends on knowledge transfer effectiveness. For instance, in the case of "Xiaomi," a mobile Internet company used crowdsourcing to acquire knowledge to develop a mobile operating system further. The interaction between the crowd-firm led to a knowledge intersection that enhances developers' understanding to draw an association between new innovative ideas and existing ones. The association helped Xiaomi developers to understand the potential of creative ideas and absorb knowledge for further product development (Argote and Miron-Sektor, 2011; Liu *et al.*, 2018). Therefore, we hypothesise that absorptive capability is developed by effective knowledge transfer:

H3. The knowledge transfer propensity is positively related to an organisation's absorptive capacity.

Past studies have confirmed the relationship between absorptive capacity and organisational performance (Roberts *et al.*, 2012; Xie *et al.*, 2018; Muller *et al.*, 2021) in the context of new product development. A firm's absorptive capacity is a dynamic capability to acquire, integrate, enhance knowledge transfer effectiveness and use external knowledge and resources to significantly improve financial growth (Iyengar *et al.*, 2015). Organisations with a higher absorptive capacity can continually expand their skill and knowledge base, absorb new information from the external environment and develop the core competencies to apply information in enhancing innovation performance (Xun and Xuehan, 2014). Hence, organisations that use crowdsourcing as a co-creation tool have a greater absorptive capacity that makes it possible to create new external knowledge, keep updating the competencies, and understand the development process of knowledge management capabilities to achieve innovative performance difficult to imitate.

Similarly, excellent absorptive capacity can better use acquired knowledge to respond to external changes. However, suppose the organisations want to survive the cut-throat competition and build innovation strategies per demand. In that case, there is a need to continuously absorb new information and apply it internally to achieve greater competitiveness (Sancho-Zamora *et al.*, 2022). Therefore, we tested the following hypothesis:

H4. The organisation's absorptive capacity positively relates to organisational value capture from an innovation perspective.

3.3 Organisational learning as mediator

Knowledge transfer effectiveness beyond organisational boundaries facilitates organisations to reduce knowledge management costs, speed up the product development process, and respond quickly to external market changes for innovations and further improvement (Xun and Xuehan, 2014). Thus, this supports the notion that a high level of dynamic capability, namely absorptive capacity, uses past knowledge and depends more on transferring external knowledge resources (Iyengar *et al.*, 2015). Effective knowledge transfer makes it possible for organisations with social linkage to supplement, refocus and strengthen the knowledge base to grasp innovation opportunities. On the same lines, the knowledge-based view by Trantopoulos *et al.* (2017) highlighted the role of new knowledge acquired in enhancing a firm's absorptive capacity. Therefore, we propose the following hypotheses:

H5. The relationship between social capital and absorptive capacity is mediated by the efficacy of knowledge transfer during crowdsourcing.

H5a. The relationship between structural linkage and absorptive capacity is mediated by the efficacy of knowledge transfer during crowdsourcing.

H5b. The relationship between cognitive linkage and absorptive capacity is mediated by the efficacy of knowledge transfer during crowdsourcing.

H5c. The relationship between relational linkage and absorptive capacity is mediated by the efficacy of knowledge transfer during crowdsourcing.

Absorptive capacity is a facilitating factor between crowdsourcing-based social linkage and organisational innovation capability. Organisations that are continuously involved in the absorption of external knowledge have a greater ability to innovate, seize the opportunities present in the market, identify and fulfil the demands of the community of crowd and gain a competitive advantage against competitors (Nieto and Quevedo, 2005; Roberts *et al.*, 2012). Even though social linkage formed during crowd-firm interaction helps in acquiring the crowd's wisdom for further innovation and improvements, absorptive capability helps the organisations absorb innovative knowledge and apply it commercially to improve processes, products and organisational innovations (Schmidt and Rammer, 2006; Saldanha *et al.*, 2017). Thus, absorptive capacity enhances an organisation's ability to recognise market trends and innovation prospects through knowledge acquisition acclimatisation and application for further value capture.

Crowdsourcing-based social linkage formed with crowd-firm interactions improves the quantity and quality of information sharing, understanding and applying knowledge for crowd-enabled innovativeness (Roberts *et al.*, 2012). The increase in competition and sudden market changes forces organisations to interact and acquire external resources to combine them internally through absorptive capacity and respond quickly to the community demands through innovation. Hence, we stipulate that co-creation results in structural, cognitive and relational linkages that affect an organisation's absorptive capacity and innovation performance. Therefore, we tested the following hypotheses:

H6. The relationship between social capital formed and organisational innovation is mediated by an organisation's absorptive capacity.

H6a. The relationship between structural linkage formed and organisational innovation is mediated by an organisation's absorptive capacity.

H6b. The relationship between cognitive linkage formed and organisational innovation is mediated by an organisation's absorptive capacity.

H6c. The relationship between relational linkage formed and organisational innovation is mediated by an organisation's absorptive capacity.

Iyengar *et al.* (2015) proposed that organisations' absorptive capacity to assimilate, convert and apply knowledge determines the relationship between knowledge transfer effectiveness and firms' innovation performance. In the fast-changing environment, mere crowd-firm interaction and knowledge sharing are not enough. Instead, organisations should build greater absorptive capacity to recognise valuable information and its commercial application to develop innovation strategies. Hence, we tested the following hypothesis:

H7. An organisation's absorptive capacity mediates the relationship between effective knowledge transfer and organisational innovation performance.

4. Research methodology

The proposed research model is shown in Figure 2.

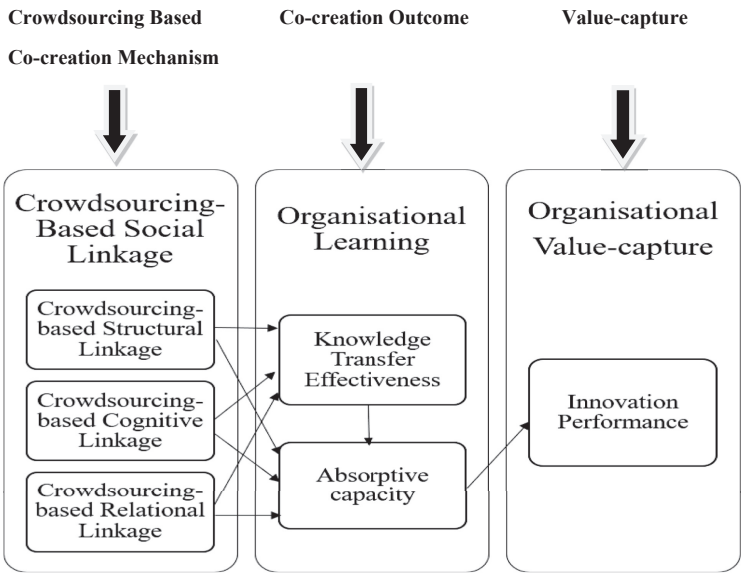


Figure 2.
Research model

4.1 Sample and data collection

The Information and Technology (IT) industries were considered the most suitable for conducting empirical study because of their highly dynamic nature of technological discontinuity, where research and development efforts are very high to achieve the ultimate aim of innovation. Such firms rely heavily on the crowd's knowledge to anticipate market changes and keep updating their products and services per the crowd's requirements, which is vital to success.

The study considered a sample of Indian IT firms that have used crowdsourcing to interact with the crowd community to enhance their innovation capability. Primarily, employees involved in different crowdsourcing projects of Indian IT firms were contacted to collect their views. Data collection for the study was carried out between August 2021 and November 2021 via mail questionnaire and by sharing Google form links on LinkedIn. The study targeted mainly middle and senior-level managers of IT firms as they are a rich source of knowledge and have reasonable perspectives about organisational performance and operations. The survey hyperlink was shared with 400 managers, out of which 325 responses (response rate of 81.25%) were received. After removing the invalid and incomplete responses, 300 responses were used to test the proposed research model. Table 2 presents the demographic characteristics of the respondents.

4.2 Measurement development

The measurement items for the construct were adapted from the existing validated scales, as presented in Table 3. A five-point Likert scale (1 = strongly disagree to 5 = strongly agree) was used to measure the items. The four-item scale of structural linkage was adapted from Wagner *et al.* (2014) and Hsu and Hung (2013). The four-item scale of cognitive linkage was adapted from Wagner *et al.* (2014) and Chiu *et al.* (2006). The relational linkage three-item scale was adapted from Wagner *et al.* (2014). The measurement items of knowledge transfer effectiveness and absorptive capacity were adapted from Iyengar *et al.* (2015). The scales for organisational innovation performance were adapted from Chen *et al.* (2009).

Measure	Category	Frequency	%
Gender	Female	115	38.2
	Male	185	61.8
Age	Under 20	20	6.8
	20–30	127	42.2
	30–40	110	36.8
	Over 40	43	14.2
Position	Middle-level managers	125	41.67
	Senior-level managers	175	58.33
IT Firms	Wipro	25	8.33
	IBM	24	8
	Philips	16	5.33
	Deloitte	18	6
	TCS	23	7.67
	Infosys	24	8
	Cognizant	16	5.33
	HCl technologies	24	8
	Genpact	16	5.33
	Tata Group	26	8.67
	Capgemini	16	5.33
	Microsoft	24	8
	Dell	19	6.34
	Google	15	5
	Apple	14	4.67

Source(s): Primary data

Table 2. Respondent's demography

Construct and source	Category/Type	Scales
Structural linkage (STL) adapted from Wagner *et al.* (2014) and Hsu and Hung (2013)	Reflective construct/ Independent variable	(1) The organisation's crowdsourcing platform enables crowds to express their innovative knowledge simultaneously (STL1) (2) The organisation regularly uses a crowdsourcing platform to share innovation-related communications (STL2) (3) A community of crowds can discuss innovation-related concerns on the organisation's crowdsourcing platform (STL3) (4) The organisation and its external crowds can communicate through crowdsourcing platforms to ensure an effective and efficient innovation process (STL4)
Cognitive Linkage (COL) adapted from Wagner *et al.* (2014) and Chiu *et al.* (2006)	Reflective construct/ Independent variable	(1) When communicating on a crowdsourcing platform, the organisation and its external crowd members utilise the same terminology (COL1) (2) The organisation and its crowds connect on a crowdsourcing platform using a comprehensible communication pattern (COL2) (3) The organisation and its external crowds can use a crowdsourcing platform to discuss innovation-related issues (COL3) (4) The organisation and its crowds can post innovation-specific messages on crowdsourcing platforms using easily accessible narrative styles (COL4)

(*continued*)

Table 3. Measurement items

Construct and source	Category/Type	Scales
Relational Linkage (REL) adapted from Wagner *et al.* (2014)	Reflective construct/ Independent variable	(1) During the innovation process in crowdsourcing platforms, crowds and the organisation have a high mutual trust and respect (REL1) (2) External crowds and organisations routinely consult one other for improved innovation performance in crowdsourcing platforms (REL2) (3) During the innovation process in crowdsourcing platforms, crowds and the organisation work closely together
Knowledge transfer effectiveness (KTE) adapted from Iyengar *et al.* (2015)	Reflective construct/ Mediator variable	(1) Crowds offer organisations critical and practical innovation-related knowledge (KTE1) (2) Crowds have provided organisations with an abundance of innovation-related information (KTE2) (3) External crowds are seen as a valuable source of innovation-specific information by the organisation (KTE3) (4) External crowds frequently provide the organisation with relevant innovation-specific information (KTE4) (5) Overall, crowd-firm interactions have improved the firm's innovation-related knowledge (KTE5)
Absorptive Capacity (ABC) adapted from Iyengar *et al.* (2015)	Reflective construct/ Mediator variable	(1) The organisation has established excellent procedures for identifying and acquiring new innovation-specific data and expertise (ABC1) (2) The organisation has suitable procedures in place to examine the innovation-specific data and knowledge gathered (ABC2) (3) The organisation has suitable procedures in place to integrate new innovation-specific knowledge and information (ABC3) (4) In innovation, the organisation may successfully integrate current knowledge with new information and expertise gained from crowds (ABC4) (5) The organisation can effectively incorporate internal and external data and knowledge into the innovation process (ABC5) (6) The organisation is good at putting innovation-related knowledge to use (ABC6)
Innovation Performance (INP) adapted from Chen *et al.* (2009)	Formative construct/ Dependent variable	(1) The organisations' merchandise quality can be improved through innovation (INP1) (2) Through innovation, the organisation can speed up the commercialisation of new products (INP2) (3) The organisation improved innovation performance have the potential to generate a significant profit (INP3)

Table 3.

4.3 Common method bias

The data collection was done only one time from a single survey. Therefore, we conducted different statistical analyses to assess the common method bias. First, Harman's one-factor test was performed using all items' in the principal component analysis (Podsakoff, 2003). The exploratory factor analysis results gave four factors with eigenvalues greater than 1.0, accounting for 66% of the total variance. The findings revealed that no single factor accounted for most of the variance, implying that common method bias was not an issue in this study. Second, the correlation matrix technique was performed (Pavlou *et al.*, 2007). Previous studies have mainly suggested that intra-construct correlations higher than 0.90 depict the issue of common method bias. The highest intra-construct correlation in our study

was 0.79, which is below the threshold limit of 0.90. Therefore, from the results of the two statistical analyses, it was inferred that common method bias is not an issue in the study.

5. Data analysis and results

A two-step approach (Anderson and Gerbing, 1988) was used to validate the instrument using a measurement model. Then the structural model was assessed to test the research hypothesis. Partial least squares (PLS), a component-based structural equation modelling, was used to validate the study model. The bootstrap operation was performed using the SmartPLS 3.2.7 software package with 5,000 resamples. We mainly used SmartPLS to analyse data and test hypotheses for the following reasons. First, PLS is advantageous when working with complicated models that include reflective and formative constructs. Second, PLS is a prediction-oriented model meant to evaluate exploratory models. Third, the sample size was small which drove the choice of PLS over other co-variance-based techniques like LISREL and AMOS. Lastly, PLS avoids factor indeterminacy and inadmissible solutions, making it a better choice for estimating complicated interaction effects.

5.1 Measurement model

Confirmatory factor analysis (CFA) was used to check the reliability and validity of the first-order reflective constructs. Three criteria for assessing the measurement model were discriminant validity, convergent validity and multicollinearity.

Convergent validity was established when item loadings, Cronbach's alpha (CA) and composite reliability (CR) exceeded the threshold limit of 0.70 (Straub *et al.*, 2004). Additionally, the average variance extracted (AVE) exceeds the threshold of 0.50 (Fornell and Larcker, 1981). Table 4 shows that all the Cronbach's values were higher than 0.70, and composite reliability values ranged between 0.831 and 0.908. The indicator loadings exceeded the threshold limit of 0.32 (Tabachnick and Fidell, 2001). All the reflected constructs' AVE values were between 0.558 and 0.733. Therefore, the results exhibited good convergent validity.

Discriminant validity was established using three criteria: Fornell–Larcker, cross loads and the HTMT (Heterotrait–Monotrait). Each construct's AVE square root was higher than the other construct's correlation values (Hair *et al.*, 2016). The Fornell and Larcker criterion results in Table 5 suggest that reflective constructs have adequate discriminant validity as each construct's AVE square root was higher than the other construct's correlation values (Hair *et al.*, 2016). Table 6 shows that all the indicator loadings were loaded better on their intended constructs than the other, reflecting the discriminant validity. Finally, the HTMT criterion was used to estimate that the correlation between all constructs was perfectly reliable and measured without errors (Hair *et al.*, 2016). Table 7 shows that all the HTMT values were less than 0.85 (Ritcher *et al.*, 2016), thus indicating discriminant validity.

Multicollinearity was validated by variance inflation factor values (VIF) of constructs with a threshold limit of 5 or below (Hair *et al.*, 2014). As shown in Table 4, all the variance inflation values lie between 1.078 and 3.590, indicating a lack of multicollinearity issues in our constructs.

Three criteria examined the measurement model of the second-order formative constructs, namely outer loading, outer weights and multicollinearity. The outer weights of all the formative indicators were statistically significant except for one (INP1). Therefore, to decide whether the insignificant indicator should be dropped or retained, we further assessed the outer loadings of all formative constructs items as suggested by Hair *et al.* (2016). As shown in Table 8, all the item loadings were above the limit of 0.50; therefore, we retained INP1 for data analysis as it forms a vital part of the innovation construct. Lastly, the variance inflation

Construct	Indicator	Item loadings >0.32	Cronbach's α (CA) >0.70	Composite reliability (CR) >0.70	Average variance extracted (AVE) >0.50	Variance inflation factor values (VIF) <5
STL	STL1	0.846	0.742	0.844	0.586	2.562
	STL2	0.879				2.747
	STL3	0.786				1.539
	STL4	0.486				1.078
COL	COL1	0.540	0.723	0.831	0.558	1.107
	COL2	0.842				1.881
	COL3	0.826				1.849
	COL4	0.742				1.443
REL	REL1	0.865	0.819	0.892	0.733	1.829
	REL2	0.830				1.796
	REL3	0.873				1.843
KTE	KTE1	0.793	0.871	0.908	0.666	1.762
	KTE2	0.882				3.590
	KTE3	0.864				3.007
	KTE4	0.865				2.935
	KTE5	0.654				1.300
ABC	ABC1	0.700	0.875	0.905	0.615	2.097
	ABC2	0.780				2.261
	ABC3	0.778				2.726
	ABC4	0.770				2.748
	ABC5	0.828				2.484
	ABC6	0.842				2.685

Source(s): Primary data

Table 4. Confirmatory factor analysis (CFA) results – reflective constructs

	STL	COL	REL	KTE	ABC
STL	*0.765*				
COL	0.739	*0.747*			
REL	0.749	0.739	*0.856*		
KTE	0.756	0.740	0.784	*0.816*	
ABC	0.425	0.511	0.442	0.525	*0.784*

Source(s): Primary data

Table 5. Discriminant validity results – reflective constructs Fornell and Larcker criterion

factor values (VIF) were below the threshold limit of 5 (Hair *et al.*, 2014), so multicollinearity was not a severe issue in our formative constructs. Therefore, the results exhibited the reliability and validity of the formative constructs used in the study.

5.2 Structural model

The structural model was evaluated by a four-step process given by Hair *et al.* (2017). First, the absence of collinearity was assessed. Second, the relevance and significance of the relationships of variables were assessed and calculated. Third, the effect size f^2 and the coefficient of determination R^2 were assessed. Lastly, the predictive capacity Q^2 of the model was evaluated.

The collinearity assessment was done using the VIF values of the model. Table 9 shows that all the variance inflation values were below the threshold limit of 3.3 (Diamantopoulos and Siguaw, 2006), indicating a lack of collinearity issues.

Following this, the relationship coefficients and the significance of the relationships between variables were assessed. Tables 10–12 show the values of relationship coefficients,

Items	ABC	COL	INP	KTE	REL	STL
ABC1	0.734	0.376	0.459	0.527	0.415	0.404
ABC2	0.802	0.475	0.636	0.591	0.508	0.456
ABC3	0.753	0.169	0.58	0.106	0.122	0.125
ABC4	0.777	0.258	0.545	0.238	0.264	0.223
ABC5	0.843	0.343	0.596	0.337	0.323	0.315
ABC6	0.847	0.235	0.675	0.294	0.202	0.194
COL1	0.361	0.894	0.346	0.382	0.189	0.224
COL2	0.365	0.847	0.272	0.598	0.652	0.615
COL3	0.269	0.857	0.187	0.702	0.672	0.686
COL4	0.376	0.745	0.273	0.47	0.595	0.468
INP1	0.521	0.365	0.679	0.478	0.421	0.35
INP2	0.683	0.166	0.888	0.243	0.217	0.179
INP3	0.579	0.324	0.754	0.416	0.277	0.278
KTE1	0.365	0.631	0.328	0.785	0.652	0.615
KTE2	0.312	0.603	0.271	0.879	0.633	0.622
KTE3	0.299	0.623	0.259	0.856	0.609	0.607
KTE4	0.334	0.652	0.24	0.877	0.643	0.691
KTE5	0.59	0.434	0.522	0.646	0.494	0.449
REL1	0.341	0.7	0.32	0.697	0.859	0.689
REL2	0.308	0.643	0.211	0.57	0.837	0.511
REL3	0.38	0.655	0.272	0.705	0.858	0.653
STL1	0.195	0.58	0.098	0.61	0.563	0.832
STL2	0.248	0.621	0.175	0.622	0.616	0.873
STL3	0.285	0.612	0.211	0.598	0.636	0.773
STL4	0.444	0.326	0.403	0.346	0.336	0.461

Note(s): STL – Crowdsourcing-based structural linkage; COL – Crowdsourcing-based Cognitive linkage; REL – Crowdsourcing-based relational linkage; KTE – Knowledge transfer effectiveness; ABC – Absorptive Capacity; INP – Innovation performance

Table 6.
Item cross-loadings

	Absorptive capacity	Crowdsourcing-based cognitive linkage	Knowledge transfer effectiveness	Crowdsourcing-based relational linkage
Crowdsourcing-based Cognitive linkage	0.611			
Knowledge transfer effectiveness	0.506	0.528		
Crowdsourcing-based relational linkage	0.447	0.662	0.640	
Crowdsourcing-based structural linkage	0.476	0.630	0.637	0.680

Table 7.
Heterotrait and Monotrait (HTMT) criteria

mean values, standard deviation, t values and significance level. The confidence interval values of 2.5 and 95% were also reported, as recommended by the authors (Hair *et al.*, 2016; Henseler *et al.*, 2016; Roldan *et al.*, 2012). The results of the structural model analysis shown in Figure 3 depict the causal relationships between constructs. The research hypothesis is tested using path coefficient values (β) and p-values.

The results of the structural equation model for the hypothesis testing showed that crowdsourcing-based structural linkage ($\beta = 0.285$, $p < 0.001$), cognitive linkage ($\beta = 0.237$, $p < 0.05$) and relational linkage ($\beta = 0.396$, $p < 0.001$) positively influenced organisational knowledge transfer effectiveness. Therefore, H1a, H1b and H1c were supported. Furthermore, crowdsourcing-based cognitive linkage ($\beta = 0.338$, $p < 0.05$) also influenced organisations absorptive capacity. Thus, supporting H2b. Whereas, crowdsourcing-based structural linkage ($\beta = -0.064$, $p > 0.05$) and relational linkage ($\beta = -0.029$, $p > 0.05$) had nonsignificant effects on the organisation's absorptive capacity. Thus, it provides evidence to reject H2a and H2c. Meanwhile, organisational knowledge transfer effectiveness positively influenced their absorptive capacity capability ($\beta = 0.285$, $p < 0.05$), and the absorptive capacity capability of an organisation positively influenced organisational value capture, i.e. innovation capability ($\beta = 0.731$, $p < 0.001$). Hence, this provides strong support to confirm

Table 8. Confirmatory factor analysis (CFA) results – formative constructs

Formative constructs	Indicators	Outer weights	Outer loadings	Variance inflation factor values (VIF)
Organisational Innovation	INP1	0.164 n/s	0.701***	1.941
	INP2	0.717***	0.913***	1.268
	INP3	0.327***	0.706***	1.811

Source(s): Primary data, ***$p < 0.001$

Table 9. Structural model variables – VIF values

	Knowledge transfer effectiveness	Absorptive capacity	Innovation performance
Knowledge transfer effectiveness		3.147	
Crowdsourcing-based structural linkage	2.750	3.022	
Crowdsourcing-based cognitive linkage	2.654	2.842	
Crowdsourcing-based relational linkage	2.761	3.185	
Absorptive capacity			1.000

Source(s): Primary data

Table 10. Structural model relationships – direct effect results

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	*T* statistics (|O/STDEV|)	2.5%	97.5%	*P* values
ABC → INP	0.731	0.735	0.035	21.18	0.667	0.80	0.001
COL → ABC	0.338	0.345	0.092	3.669	0.149	0.518	0.000
COL → KTE	0.237	0.233	0.055	4.321	0.114	0.341	0.002
REL → ABC	−0.029	−0.035	0.096	0.306	−0.215	0.151	0.76
REL → KTE	0.396	0.398	0.063	6.308	0.277	0.518	0.000
STL → ABC	−0.064	−0.068	0.097	0.662	−0.262	0.113	0.508
STL → KTE	0.285	0.288	0.052	5.479	0.184	0.388	0.001
KTE → ABC	0.285	0.286	0.084	3.411	0.459	0.459	0.001

Note(s): STL – Crowdsourcing-based structural linkage; COL – Crowdsourcing-based Cognitive linkage; REL – Crowdsourcing-based relational linkage; KTE – Knowledge transfer effectiveness; ABC – Absorptive Capacity; INP – Innovation performance

	Original sample (O)	Sample mean (M)	Standard Deviation (STDEV)	Statistics t (|O/STDEV|)	2.5%	97.5%	P values
REL → KTE → ABC → INP	0.083	0.084	0.029	2.816	0.032	0.142	0.005
COL → ABC → INP	0.248	0.254	0.071	3.486	0.102	0.396	0.001
REL → KTE → ABC	0.113	0.114	0.039	2.86	0.045	0.199	0.004
COL → KTE → ABC → INP	0.049	0.049	0.018	2.679	0.018	0.089	0.008
COL → KTE → ABC	0.068	0.066	0.025	2.728	0.024	0.119	0.007
STL → KTE → ABC	0.081	0.082	0.028	2.904	0.036	0.138	0.004
REL → ABC → INP	−0.022	−0.026	0.071	0.303	−0.157	0.108	0.762
STL → ABC → INP	−0.047	−0.05	0.072	0.652	−0.193	0.081	0.515
STL → KTE → ABC → INP	0.059	0.06	0.021	2.873	0.026	0.102	0.004
KTE → ABC → INP	0.209	0.21	0.062	3.345	0.092	0.342	0.001

Table 11. Structural model relationships – indirect effect results

	Original sample (O)	Sample mean (M)	Standard Deviation (STDEV)	Statistics t (|O/STDEV|)	2.5%	97.5%	P values
ABC → INP	0.731	0.735	0.035	21.18	0.667	0.8	0.001
COL → ABC	0.406	0.411	0.09	4.511	0.22	0.583	0.001
COL → INP	0.297	0.303	0.07	4.219	0.154	0.44	0.002
COL → KTE	0.237	0.233	0.055	4.321	0.114	0.341	0.002
REL → ABC	0.084	0.079	0.092	0.908	−0.098	0.257	0.364
REL → INP	0.061	0.058	0.068	0.9	−0.073	0.194	0.369
REL → KTE	0.396	0.398	0.063	6.308	0.277	0.518	0.000
STL → ABC	0.017	0.014	0.098	0.175	−0.186	0.202	0.861
STL → INP	0.013	0.01	0.072	0.173	−0.141	0.142	0.863
STL → KTE	0.285	0.288	0.052	5.479	0.184	0.388	0.000
KTE → ABC	0.285	0.286	0.084	3.411	0.127	0.459	0.001
KTE → INP	0.209	0.21	0.062	3.345	0.092	0.342	0.001

Table 12. Structural model relationships – total effect results

H3 and H4. The model explained 25.9 and 53.5% variance in an organisation's absorptive capacity and value capture.

Additional analyses were conducted to assess the structural model and the individual relationships. The structural model was evaluated using a coefficient of determination R^2. While the model's predictive accuracy was evaluated using predictive relevance Q^2. The values of the coefficient of determination R^2 and the adjusted R^2 are reported in Table 13.

The effect size f^2 was also assessed to find the individual effects of endogenous constructs variables (Henseler *et al.*, 2016; Roldan *et al.*, 2012). Hair *et al.* (2017) described that f^2 values of 0.01 have a negligible effect, 0.15 has a moderate effect and 0.35 has a significant impact. The f^2 values reported in Table 14 show that an organisation's absorptive capacity has a strong f^2 size effect on innovation performance. Crowdsourcing-based structural linkage has a negligible impact on the organisation's knowledge transfer effectiveness and absorptive capacity. Crowdsourcing-based cognitive linkage has a negligible effect on the organisation's knowledge transfer effectiveness and absorptive capacity. Crowdsourcing-based relational linkage has a moderate effect on knowledge transfer effectiveness.

Finally, the model's predictive capacity (Q^2) was assessed using the blindfolding process. The Q^2 values shown in Table 15 indicate that all the endogenous variables have positive values, thus, depicting the predictive relevance of the model.

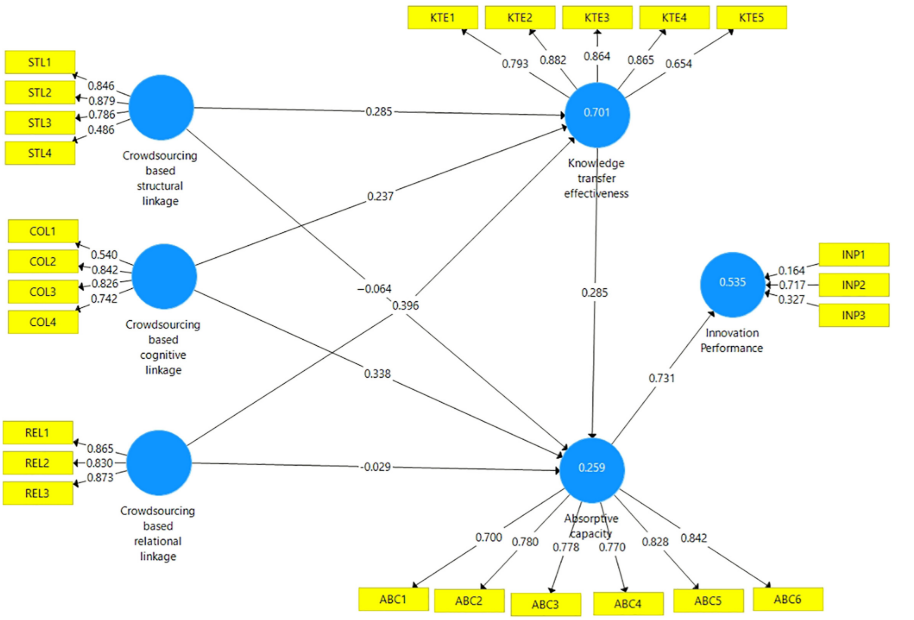


Figure 3.
Path coefficient analysis

Table 13.
Coefficient of determination – R^2 values

	R^2	Adjusted R^2
Knowledge transfer effectiveness	0.701	0.699
Absorptive Capacity	0.259	0.251
Innovation performance	0.535	0.251

Source(s): Primary data

Table 14.
Coefficient of determination effect size-f^2 values

	Knowledge transfer effectiveness	Absorptive capacity	Innovation performance
Knowledge transfer effectiveness		0.033	
Crowdsourcing-based structural linkage	0.099	0.002	
Crowdsourcing-based Cognitive linkage	0.071	0.054	
Crowdsourcing-based relational linkage	0.19		
Absorptive Capacity			1.15

Source(s): Primary data

Table 15.
Predictive capacity values – Q^2

	Q^2
Knowledge transfer effectiveness	0.457
Absorptive capacity	0.141
Innovation performance	0.319

Source(s): Primary data

5.3 Testing mediating effects

The bootstrapping method was used to check the mediating role of organisational knowledge transfer effectiveness and absorptive capacity capability. Mediation analysis results are shown in Table 11. The results indicated that organisational knowledge transfer effectiveness mediated the relationship between crowdsourcing-based structural linkage formed and absorptive capacity (β = 0.081, CI = [0.031, 0.141]), crowdsourcing-based cognitive linkage formed and absorptive capacity (β = 0.068, CI = [0.023, 0.123]), crowdsourcing-based relational linkage formed and absorptive capacity (β = 0.113, CI = [0.043, 0.194]). Hence, this provides strong support to confirm H5a, H5b and H5c. Meanwhile, the organisation's absorptive capability mediated the relationship between crowdsourcing-based cognitive linkage and organisational value capture (β = 0.248, CI = [0.103, 0.396]). At the same time, an organisation's absorptive capacity does not mediate the relationship between crowdsourcing-based structural linkage formed and organisational value capture (β = −0.047, CI = [−0.192, 0.091]), and crowdsourcing-based relational linkage formed and organisational value capture (β = −0.022, CI = [−0.153, 0.115]). Thus, H6b was supported, and H6a and H6c were not supported. Lastly, an organisation's absorptive capacity mediated the relationship between organisational knowledge transfer effectiveness and organisational value capture (β = 0.209, CI = [0.063, 3.290]). Thus, supporting H7.

6. Discussions and conclusion

The study's primary goal was to understand how crowdsourcing use complements crowd-firm social linkage, its effect on organisational learning as a co-creation outcome and value capture in the form of organisational innovation. One of the interesting findings was that the crowdsourcing-based social capital formed between crowd and firm positively influenced organisational knowledge transfer effectiveness. Along the same lines, only cognitive linkage formed between crowd-firm during crowdsourcing influenced the organisations' absorptive capacity capability. Conversely, crowdsourcing-based structural and relational capital formed between crowd-firm was not enough to develop the organisation's absorptive capacity for the following reasons. Social interactions, respect and trust between crowds and firms enhance an organisation's market knowledge and novel insights. But organisation absorptive capacity is mainly determined by individual members' shared understanding and previous knowledge base (Roberts *et al.*, 2012). Thus, organisations and crowds with a common understanding and similar backgrounds having the potential to comprehend and employ the knowledge base in a specific domain ultimately speed up the knowledge transfer well, which further improves the organisation's absorptive capacity. Further, the results indicated a positive relationship between an organisation's knowledge transfer effectiveness and absorptive capacity. Additionally, an organisation's absorptive capacity has a significant role in determining organisational value capture (improved innovation performance).

The crowdsourcing-based co-creation between crowd-firm results in organisational learning, thus enabling organisational knowledge transfer effectiveness and absorptive capacity and creating a synergetic effect on organisations' value capture. Therefore, our research findings reveal that the organisation's knowledge transfer effectiveness mediated the relationship between the crowdsourcing-based social linkage and the organisation's absorptive capacity. The results thus indicate that social capital formed between crowd and firm affects an organisation's absorptive capacity by enhancing knowledge transfer capability from crowds to the firm.

On the other hand, the organisation's absorptive capacity mediated the relationship between cognitive linkage formed during crowdsourcing and organisation value capture. At the same time, the organisation's absorptive capacity mediated the relationship between organisational knowledge transfer effectiveness and value capture (innovation performance).

In contrast, organisation's absorptive capacity does not mediate the relationship between structural and relational linkage formed during crowdsourcing and organisation value capture. The rationale behind insignificant relationships is that the development of an organisation's absorptive capacity depends on a specific path which is mainly domain and firm-specific (Roberts *et al.*, 2012) and cannot be easily incorporated into an organisation. Though structural and relational capital formed during crowdsourcing creates an information-rich learning environment, the application and acclimatisation of new knowledge acquired depend on the transfer of common understanding and collective goals (Cohen and Levinthal, 1990) between crowd and firm.

6.1 Theoretical implications

Using Iyengar *et al.*'s (2015) organisational learning model, the present study explored the impact of crowdsourcing-based crowd-firm co-creation on organisational learning. The outcome of the current study supports the Iyengar *et al.* (2015) notion that organisational learning positively impacts organisational knowledge transfer effectiveness and absorptive capacity capability. The collective intelligence gathered during crowdsourcing and crowd-firm interaction can improve an organisation's learning capability when an organisational environment facilitates knowledge transfer. The study also promotes that crowdsourcing enriches the learning capabilities of organisations, which further enhances the innovative capacity of the firms.

Further, the study used social capital theory to distinguish the crowdsourcing-based co-creation mechanism into three dimensions. The findings show that social linkages between a crowd and a firm enhance knowledge transfer effectiveness. Crowdsourcing-based structural linkage will result in an acquisition of explicit and tacit knowledge in organisations that further broadens the base of minds, resulting in more top-down and bottom-up connections, conversations, dialogue and discussions within organisations. Crowdsourcing-based cognitive linkage formed between crowd-firm will impact knowledge transfer by enhancing the ability of the crowd to access and contribute knowledge to a community. Crowdsourcing-based relational linkage formed between crowd-firm will enhance knowledge-sharing intentions and collaboration capability, thus, creating a positive atmosphere to exchange knowledge, assimilate, interpret and integrate within the organisation. The study is the pioneer in using the social capital dimensions to understand the entire mechanism behind the formation of crowdsourcing-based social linkage for effective knowledge transfer and absorptive capacity.

6.2 Practical implications

Social linkages between a crowd and a firm in crowdsourcing act as a catalyst for effective knowledge transfer and the organisation's absorptive capacity, further improving innovation performance. The findings apprise practitioners of the fact that crowdsourcing-based co-creation can be used as a panacea to improve innovation performance only when complemented with the organisational capabilities to build social linkages, effective knowledge transfer and better absorptive capacity. Therefore, organisations should work on the improvement of three social linkages formed with the crowd to enhance organisational innovation performance. First, managers should build interaction channels in crowdsourcing to enable new connections and tie-up between the crowd and a firm to interact and share their knowledge. For instance, Dell's IdeaStorm (www.ideastorm.com) and IBM's Innovation Jam (www.collaborationjam.com) enable rich interactions and discussions with external crowds to acquire experience and knowledge for improving innovation performance. Second, organisations should build standard codes and languages, namely cultural and business assumptions, descriptions of technology and task requirements that enable the formation of cognitive linkage (Yang and Li, 2016) between a crowd and an RandD team for further

innovation opportunities. Thus, the cognitive linkage formed results in shared understanding and learnings between a crowd and a firm. For instance, Threadless, an apparel company, uses shared codes and languages between the crowd and a firm to create, inspect, approve, improve and select shirt designs. The cognitive linkage formed during crowd-firm interaction facilitates Threadless to correctly understand the preferences and needs of crowds and then proceed with customised shirt designs (Piller, 2010). Third, managers should develop a co-creation process and acclimatise these processes in organisations to build a crowd-firm relationship. Additionally, certain guidelines and regulations should be formulated to reduce crowd concerns related to privacy, security and intellectual property rights in the co-creation processes.

Another interesting finding of the study is that effective knowledge transfer and an organisation's absorptive capacity create a synergetic effect on organisations' value co-creation and capture. Therefore, managers should use crowdsourcing as a co-creation platform to enrich the relationship between effective knowledge sharing and an organisation's absorptive capacity. Moreover, organisations should create different activities and processes to reach a wider crowd, resulting in the acquisition of quality information between a diverse crowd and an RandD team. Organisational innovation capability depends on the breadth of knowledge acquisition (Cohen and Levinthal, 1990). Therefore, organisations should use crowdsourcing as a co-creation platform to enable effective knowledge sharing and further enhance organisational capacity to acquire knowledge and assimilate the available resources to enhance an organisation's research and innovation capabilities. Crowdsourcing is mainly a co-creation tool to create knowledge within organisations and improve RandD's capabilities to innovate well. Our research has indicated that organisations using social linkages in crowdsourcing are more innovative and enable innovation through new ideas and knowledge transfer, enhancing employees' capability to comprehend and employ collective intelligence.

7. Limitations and future research

The study has a few limitations that need to be acknowledged in future research. First, the operationalisation of three dimensions of social capital was adapted from the study of Wagner *et al.* (2014). Such conceptualisation may limit the research findings to a single aspect. For instance, different studies suggest that structural capital is an overall interaction pattern and plays a significant role in network decentralisation (Robert *et al.*, 2008), increased communication intensity (Chiu *et al.*, 2006), etc. Our study defined structural capital as an interaction mechanism between a firm and a crowd during crowdsourcing for improved innovation performance. Second, future studies may use different qualitative techniques to understand better the three dimensions used and the co-creation mechanism from the perspective of social capital. Third, our study's outcome of crowd-firm co-creation was organisational learning; therefore, future research studies may take a different outcome perspective of co-creation. Fourth, the study used only cross-sectional data to examine the proposed research model. Thus, future research may be conducted based on longitudinal data to investigate the co-creation processes and outcomes over time. Finally, the co-creation performance between a crowd and a firm may differ at different stages of the innovation process. Hence, future researchers may explore the role of co-creation and its outcome at different stages of organisational innovation performance.

References

Abili, K. (2011), "Social capital management in Iranian knowledge-based organizations", *Electronic Journal of Knowledge Management*, Vol. 9 No. 3, pp. 204-210.

Acar, O.A. (2019), "Motivations and solution appropriateness in crowdsourcing challenges for innovation", *Research Policy*, Vol. 48 No. 8, 103716.

Afuah, A. and Tucci, C.L. (2012), "Crowdsourcing as a solution to distant search", *Academy of Management Review*, Vol. 37 No. 3, pp. 355-375.

Allameh, S.M. (2018), "Antecedents and consequences of intellectual capital: the role of social capital, knowledge sharing and innovation", *Journal of Intellectual Capital*, Vol. 19 No. 5, pp. 858-874.

Allen, B.J., Chandrasekaran, D. and Basuroy, S. (2018), "Design crowdsourcing: the impact on new product performance of sourcing design solutions from the crowd", *Journal of Marketing*, Vol. 82 No. 2, pp. 106-123.

Anderson, J.C. and Gerbing, D.W. (1988), "Structural equation modeling in practice: a review and recommended two-step approach", *Psychological Bulletin*, Vol. 103 No. 3, p. 411.

Ansari, S., Munir, K. and Gregg, T. (2012), "Impact at the 'bottom of the pyramid: the role of social capital in capability development and community empowerment", *Journal of Management Studies*, Vol. 49 No. 4, pp. 813-842.

Argote, L. and Miron-Spektor, E. (2011), "Organizational learning: from experience to knowledge", *Organization Science*, Vol. 22 No. 5, pp. 1123-1137.

Bapna, R., Gupta, A., Rice, S. and Sundararajan, A. (2017), "Trust and the strength of ties in online social networks: an exploratory field experiment", *MIS Quarterly*, Vol. 41 No. 1, pp. 115-130.

Barney, J. (1991), "Firm resources and sustained competitive advantage", *Journal of Management*, Vol. 17 No. 1, pp. 99-120.

Bjelland, O.M. and Wood, R.C. (2008), "An inside view of IBM's' Innovation Jam", *MIT Sloan Management Review*, Vol. 50 No. 1, p. 32.

Boudreau, K.J. and Lakhani, K.R. (2013), "Using the crowd as an innovation partner", *Harvard Business Review*, Vol. 91 No. 4, pp. 60-69.

Brabham, D.C. (2008), "Crowdsourcing as a model for problem-solving: an introduction and cases", *Convergence*, Vol. 14 No. 1, pp. 75-90.

Bullinger, A.C., Neyer, A.-K., Rass, M. and Moeslein, K.M. (2010), "Community-based innovation contests: where competition meets cooperation", *Creativity and Innovation Management*, Vol. 19 No. 3, pp. 290-303.

Chanal, V. and Caron-Fasan, M.L. (2008), "How to invent a new business model based on crowdsourcing: the Crowdspirit® case", *Conférence de l'Association Internationale de Management Stratégique*, pp. 1-27.

Chang, H.H. and Chuang, S.S. (2011), "Social capital and individual motivations on knowledge sharing: participant involvement as a moderator", *Information and Management*, Vol. 48 No. 1, pp. 9-18.

Chang, Y.Y., Gong, Y. and Peng, M.W. (2012), "Expatriate knowledge transfer, subsidiary absorptive capacity, and subsidiary performance", *Academy of Management Journal*, Vol. 55 No. 4, pp. 927-948.

Chen, Y.S., Lin, M.J.J. and Chang, C.H. (2009), "The positive effects of relationship learning and absorptive capacity on innovation performance and competitive advantage in industrial markets", *Industrial Marketing Management*, Vol. 38 No. 2, pp. 152-158.

Chiu, C.M., Hsu, M.H. and Wang, E.T. (2006), "Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories", *Decision Support Systems*, Vol. 42 No. 3, pp. 1872-1888.

Chow, W.S. and Chan, L.S. (2008), "Social network, social trust and shared goals in organizational knowledge sharing", *Information and Management*, Vol. 45 No. 7, pp. 458-465.

Cohen, W.M. and Levinthal, D.A. (1990), "Absorptive capacity: a new perspective on learning and innovation", *Administrative Science Quarterly*, Vol. 35 No. 1, pp. 128-152.

Devece, C., Palacios, D. and Ribeiro-Navarrete, B. (2019), "The effectiveness of crowdsourcing in knowledge-based industries: the moderating role of transformational leadership and organisational learning", *Economic Research-Ekonomska Istraživanja*, Vol. 32 No. 1, pp. 335-351.

Dholakia, U.M., Blazevic, V., Wiertz, C. and Algesheimer, R. (2009), "Communal service delivery: how customers benefit from participation in firm-hosted virtual P3 communities", *Journal of Service Research*, Vol. 12 No. 2, pp. 208-226.

Diamantopoulos, A. and Siguaw, J.A. (2006), "Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration", *British Journal of Management*, Vol. 17 No. 4, pp. 263-282.

Forbes, H. and Schaefer, D. (2018), "Crowdsourcing in product development: current state and future research directions", pp. 579-588.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Gassmann, O., Friesike, S. and Daiber, M. (2014), "Crowdsourcing as an innovation tool", *Management of the Fuzzy Front End of Innovation*, Springer, Cham, pp. 75-87.

Grover, V. and Kohli, R. (2012), "Cocreating IT value: new capabilities and metrics for multifirm environments", *MIS Quarterly*, Vol. 36 No. 1, pp. 225-232.

Gu, B. and Jarvenpaa, S. (2003), "Online discussion boards for technical support: the effect of token recognition on customer contributions, ICIS Proceedings", *24th International Conference on Information Systems*, Seattle, Washington, Paper 10.

Hair, J.F., Jr., Sarstedt, M., Hopkins, L. and Kuppelwieser, V.G. (2014), "Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research", *European Business Review*, Vol. 26 No. 2, pp. 106-121.

Hair, J.F. Jr, Hult, G.T.M., Ringle, C. and Sarstedt, M. (2016), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage Publications, Thousand Oaks, CA.

Hair, J., Hollingsworth, C.L., Randolph, A.B. and Chong, A.Y.L. (2017), "An updated and expanded assessment of PLS-SEM in information systems research", *Industrial Management and Data Systems*, Vol. 117 No. 3, pp. 442-458.

Hammon, L. and Hippner, H. (2012), "Crowdsourcing", *Business and Information Systems Engineering*, Vol. 4 No. 3, pp. 163-166.

Heeks, R. (2010), "Do information and communication technologies (ICTs) contribute to development?", *Journal of International Development*, Vol. 22 No. 5, pp. 625-640.

Henseler, J., Hubona, G. and Ray, P.A. (2016), "Using PLS path modeling in new technology research: updated guidelines", *Industrial Management and Data Systems*, Vol. 116 No. 1, pp. 2-20.

Hine, D.C., Parker, R. and Ireland, D. (2010), "The knowledge exchange intermediary as service provider: a discussion and an Australian case", *The Service Industries Journal*, Vol. 30 No. 5, pp. 713-729.

Hsu, J.S.C. and Hung, Y.W. (2013), "Exploring the interaction effects of social capital", *Information and Management*, Vol. 50 No. 7, pp. 415-430.

Iyengar, K., Sweeney, J.R. and Montealegre, R. (2015), "Information technology use as a learning mechanism", *MIS Quarterly*, Vol. 39 No. 3, pp. 615-642.

Johannessen, J.A. and Olsen, B. (2010), "The future of value creation and innovations: aspects of a theory of value creation and innovation in a global knowledge economy", *International Journal of Information Management*, Vol. 30 No. 6, pp. 502-511.

Krupowicz, W., Czarnecka, A. and Grus, M. (2020), "Implementing crowdsourcing initiatives in land consolidation procedures in Poland", *Land Use Policy*, Vol. 99, 105015.

Lehner, O.M. (2013), "Crowdfunding social ventures: a model and research agenda", *Venture Capital*, Vol. 15 No. 4, pp. 289-311.

Lenart-Gansiniec, R. and Sułkowski, Ł. (2020), "Organizational learning and value creation in local governance: the mediating role of crowdsourcing", *The Learning Organization*, Vol. 27 No. 4, pp. 321-335.

Lepak, D.P., Smith, K.G. and Taylor, M.S. (2007), "Value creation and value capture: a multilevel perspective", *Academy of Management Review*, Vol. 32 No. 1, pp. 180-194.

Levy, M. (2009), "WEB 2.0 implications on knowledge management", *Journal of Knowledge Management*, Vol. 13 No. 1, pp. 120-134.

Liao, S.H., Fei, W.C. and Chen, C.C. (2007), "Knowledge sharing, absorptive capacity, and innovation capability: an empirical study of Taiwan's knowledge-intensive industries", *Journal of Information Science*, Vol. 33 No. 3, pp. 340-359.

Liu, Q., Zhao, X. and Sun, B. (2018), "Value co-creation mechanisms of enterprises and users under crowdsource-based open innovation", *International Journal of Crowd Science*, Vol. 2 No. 1, pp. 2-17.

Lusch, R.F. and Nambisan, S. (2015), "Service innovation", *MIS Quarterly*, Vol. 39 No. 1, pp. 155-176.

Maktabi, S.H. and Khazaei, A. (2014), "The impact of organizational learning on organizational performance and organizational innovation: evidence from Bank Industry of Iran", *International Journal of Economy, Management and Social Sciences*, Vol. 3 No. 10, pp. 569-573.

Marjanovic, S., Fry, C. and Chataway, J. (2012), "Crowdsourcing based business models: in search of evidence for innovation 2.0", *Science and Public Policy*, Vol. 39 No. 3, pp. 318-332.

Mehtälä, J., Kauranen, I., Karjalainen, J. and Nyberg, T. (2016), "A crowdsourcing model for new idea development in the fashion industry", *2016 IEEE International Conference on Service Operations and Logistics, and Informatics*, IEEE, pp. 29-36.

Modaresnezhad, M., Iyer, L., Palvia, P. and Taras, V. (2020), "Information Technology (IT) enabled crowdsourcing: a conceptual framework", *Information Processing and Management*, Vol. 57 No. 2, 102135.

Müller, J.M., Buliga, O. and Voigt, K.I. (2021), "The role of absorptive capacity and innovation strategy in the design of industry 4.0 business Models-A comparison between SMEs and large enterprises", *European Management Journal*, Vol. 39 No. 3, pp. 333-343.

Nahapiet, J. and Ghoshal, S. (1998), "Social capital, intellectual capital, and the organizational advantage", *Academy of Management Review*, Vol. 23 No. 2, pp. 242-266.

Nambisan, S. and Baron, R.A. (2010), "Different roles, different strokes: organizing virtual customer environments to promote two types of customer contributions", *Organization Science*, Vol. 21 No. 2, pp. 554-572.

Nevo, D. and Kotlarsky, J. (2020), "Crowdsourcing as a strategic is sourcing phenomenon: critical review and insights for future research", *The Journal of Strategic Information Systems*, Vol. 29 No. 4, 101593.

Nevo, S. and Wade, M. (2011), "Firm-level benefits of IT-enabled resources: a conceptual extension and an empirical assessment", *The Journal of Strategic Information Systems*, Vol. 20 No. 4, pp. 403-418.

Nieto, M. and Quevedo, P. (2005), "Absorptive capacity, technological opportunity, knowledge spillovers, and innovative effort", *Technovation*, Vol. 25 No. 10, pp. 1141-1157.

Nishikawa, H., Schreier, M., Fuchs, C. and Ogawa, S. (2017), "The value of marketing crowdsourced new products as such: evidence from two randomized field experiments", *Journal of Marketing Research*, Vol. 54 No. 4, pp. 525-539.

Nov, O., Ye, C. and Kumar, N. (2012), "A social capital perspective on meta-knowledge contribution and social computing", *Decision Support Systems*, Vol. 53 No. 1, pp. 118-126.

Onağ, A.O., Tepeci, M. and Başalp, A.A. (2014), "Organizational learning capability and its impact on firm innovativeness", *Procedia-Social and Behavioral Sciences*, Vol. 150, pp. 708-717.

Palacios, M., Martinez-Corral, A., Nisar, A. and Grijalvo, M. (2016), "Crowdsourcing and organizational forms: emerging trends and research implications", *Journal of Business Research*, Vol. 69 No. 5, pp. 1834-1839.

Pavlou, P.A., Liang, H. and Xue, Y. (2007), "Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective", *MIS Quarterly*, Vol. 31 No. 1, pp. 105-136.

Piller, F.T. (2010), "Open innovation with customers: crowdsourcing and co-creation at Threadless", available at: SSRN 1688018.

Piyathasanan, B., Mathies, C., Patterson, P.G. and de Ruyter, K. (2017), "Continued value creation in crowdsourcing from creative process engagement", *Journal of Services Marketing*, Vol. 32 No. 1, pp. 19-33.

Podsakoff, N.P. (2003), "Common method biases in behavioral research: a critical review of the literature and recommended remedies", *Journal of Applied Psychology*, Vol. 885 No. 879, pp. 10-1037.

Poetz, M.K. and Schreier, M. (2012), "The value of crowdsourcing: can users really compete with professionals in generating new product ideas?", *Journal of Product Innovation Management*, Vol. 29 No. 2, pp. 245-256.

Porter, C.E. and Donthu, N. (2008), "Cultivating trust and harvesting value in virtual communities", *Management Science*, Vol. 54 No. 1, pp. 113-128.

Prpić, J., Shukla, P.P., Kietzmann, J.H. and McCarthy, I.P. (2015), "How to work a crowd: developing crowd capital through crowdsourcing", *Business Horizons*, Vol. 58 No. 1, pp. 77-85.

Putnam, R.D. (1995), "Tuning in, tuning out: the strange disappearance of social capital in America", *PS: Political Science and Politics*, Vol. 28 No. 4, pp. 664-683.

Rai, A., Pavlou, P.A., Im, G. and Du, S. (2012), "Interfirm IT capability profiles and communications for co-creating relational value: evidence from the logistics industry", *MIS Quarterly*, Vol. 36 No. 1, pp. 233-262.

Rass, M., Dumbach, M., Danzinger, F., Bullinger, A.C. and Moeslein, K.M. (2013), "Open innovation and firm performance: the mediating role of social capital", *Creativity and Innovation Management*, Vol. 22 No. 2, pp. 177-194.

Rezaei, R. and Moghanlo, F. (2014), "The effect of intellectual capital on organizational innovation in the agricultural services ventures of Zanjan province", *Journal of Entrepreneurship Development*, Vol. 7 No. 4, pp. 653-673.

Richter, N.F., Sinkovics, R.R., Ringle, C.M. and Schlägel, C. (2016), "A critical look at the use of SEM in international business research", *International Marketing Review*, Vol. 33 No. 3, pp. 376-404.

Robert, L.P., Jr., Dennis, A.R. and Ahuja, M.K. (2008), "Social capital and knowledge integration in digitally enabled teams", *Information Systems Research*, Vol. 19 No. 3, pp. 314-334.

Roberts, N., Galluch, P.S., Dinger, M. and Grover, V. (2012), "Absorptive capacity and information systems research: review, synthesis, and directions for future research", *MIS Quarterly*, Vol. 36 No. 2, pp. 625-648.

Roldán, J.L. and Sánchez-Franco, M.J. (2012), "Variance-based structural equation modeling: guidelines for using partial least squares in information systems research", *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems*, IGI global, pp. 193-221.

Sancho-Zamora, R., Hernández-Perlines, F., Peña-García, I. and Gutiérrez-Broncano, S. (2022), "The impact of absorptive capacity on innovation: the mediating role of organizational learning", *International Journal of Environmental Research and Public Health*, Vol. 19 No. 2, p. 842.

Saldanha, F.P., Cohendet, P. and Pozzebon, M. (2014), "Challenging the stage-gate model in crowdsourcing: the case of Fiat Mio in Brazil", *Technology Innovation Management Review*, Vol. 4 No. 9, pp. 28-35.

Saldanha, T.J., Mithas, S. and Krishnan, M.S. (2017), "Leveraging customer involvement for fueling innovation: the role of relational and analytical information processing capabilities", *MIS Quarterly*, Vol. 41 No. 1, pp. 267-286.

Schlagwein, D. and Bjorn-Andersen, N. (2014), "Organizational learning with crowdsourcing: the revelatory case of LEGO", *Journal of the Association for Information Systems*, Vol. 15 No. 11, p. 3.

Schmidt, T. and Rammer, C. (2006), *The Determinants and Effects of Technological and Non Technological Innovations–Evidence from the German CIS IV*, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim.

Schulze, A. and Hoegl, M. (2008), "Organizational knowledge creation and the generation of new product ideas: a behavioral approach", *Research Policy*, Vol. 37 No. 10, pp. 1742-1750.

Schulze, T., Seedorf, S., Geiger, D., Kaufmann, N. and Schader, M. (2011), "Exploring task properties in crowdsourcing–An empirical study on Mechanical Turk".

Straub, D., Boudreau, M.C. and Gefen, D. (2004), "Validation guidelines for IS positivist research", *Communications of the Association for Information Systems*, Vol. 13 No. 1, p. 24.

Tabachnick, B.G. and Fidell, L.S. (2001), *Using Multivariate Statistics*, Allyn & Bacon, Boston.

Teece, D.J. (1993), "Profiting from technological innovation: implications for integration, collaboration, licensing and public policy", *Research Policy*, Vol. 22 No. 2, pp. 112-113.

Trantopoulos, K., von Krogh, G., Wallin, M.W. and Woerter, M. (2017), "External knowledge and information technology: implications for process innovation performance", *MIS Quarterly*, Vol. 41 No. 1, pp. 287-300.

Tsai, W. and Ghoshal, S. (1998), "Social capital and value creation: the role of intrafirm networks", *Academy of Management Journal*, Vol. 41 No. 4, pp. 464-476.

Vukovic, M. (2009), "Crowdsourcing for enterprises", *2009 Congress on Services-I*, IEEE, pp. 686-692.

Wagner, H.T., Beimborn, D. and Weitzel, T. (2014), "How social capital among information technology and business units drives operational alignment and IT business value", *Journal of Management Information Systems*, Vol. 31 No. 1, pp. 241-272.

Wasko, M.M. and Faraj, S. (2005), "Why should I share? Examining social capital and knowledge contribution in electronic networks of practice", *MIS Quarterly*, Vol. 29 No. 1, pp. 35-57.

Wazny, K. (2017), "Crowdsourcing ten years: a review", *Journal of Global Health*, Vol. 7 No. 2, pp. 1-13.

Whitla, P. (2009), "Crowdsourcing and its application in marketing activities", *Contemporary Management Research*, Vol. 5 No. 1.

Wiertz, C. and de Ruyter, K. (2007), "Beyond the call of duty: why customers contribute to firm-hosted commercial online communities", *Organization Studies*, Vol. 28 No. 3, pp. 347-376.

Xie, X., Zou, H. and Qi, G. (2018), "Knowledge absorptive capacity and innovation performance in high-tech companies: a multi-mediating analysis", *Journal of Business Research*, Vol. 88, pp. 289-297.

Xu, Y., Ribeiro-Soriano, D.E. and Gonzalez-Garcia, J. (2015), "Crowdsourcing, innovation and firm performance", *Management Decision*, Vol. 53 No. 6, pp. 1158-1169.

Xun, D. and Xuehan, G. (2014), "The effect of absorptive capacity on innovation performance: empirical research on the mediating effects of knowledge transfer efficiency", *BioTechnology an Indian Journal*, Vol. 10 No. 7, pp. 1735-1740.

Yan, J., Leidner, D.E., Benbya, H. and Zou, W. (2019), "Social capital and knowledge contribution in online user communities: one-way or two-way relationship?", *Decision Support Systems*, Vol. 127, 113131.

Yang, X. and Li, G. (2016), "Factors influencing the popularity of customer-generated content in a company-hosted online co-creation community: a social capital perspective", *Computers in Human Behavior*, Vol. 64, pp. 760-768.

Ye, H.J. and Kankanhalli, A. (2015), "Investigating the antecedents of organizational task crowdsourcing", *Information and Management*, Vol. 52 No. 1, pp. 98-110.

Yli-Renko, H., Autio, E. and Sapienza, H.J. (2001), "Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms", *Strategic Management Journal*, Vol. 22 Nos 6-7, pp. 587-613.

Zhu, H., Sick, N. and Leker, J. (2016), "How to use crowdsourcing for innovation?: a comparative case study of internal and external idea sourcing in the chemical industry", *2016 Portland International Conference on Management of Engineering and Technology*, IEEE, pp. 887-901.

Further reading

Coleman, J.S. (1988), "Social capital in the creation of human capital", *American Journal of Sociology*, Vol. 94, pp. S95-S120.

Qing-liang, M. and Xin-hui, X. (2018), "A dynamic decision model of users' sustained participation in crowdsourcing innovation", *International Journal of Services Operations and Informatics*, Vol. 9 No. 3, pp. 223-241.

Zhou, K., Yang, C., Li, L., Miao, C., Song, L., Jiang, P. and Su, J. (2021), "A folksonomy-based collaborative filtering method for crowdsourcing knowledge-sharing communities", *Kybernetes*.

About the authors

Divya Mishra is a Ph.D. Research Scholar in the Department of University School of Management and Entrepreneurship (USME), Delhi Technological University, New Delhi. She holds a Bachelors's and Master's Degree in Commerce from the University of Allahabad with a specialisation in Human Resource Management and Marketing. She received three Gold Medals for outstanding performance in the academic field. Her research interests include human resource management, organisational behaviour and technological innovation. Divya Mishra is the corresponding author and can be contacted at: divyamishra919@gmail.com

Dr. Nidhi Maheshwari is an Associate Professor at Delhi Technological University, New Delhi. Experienced Associate Professor with a demonstrated history of working in the higher education industry. Strong education professional skilled in Market Research, Management, Organisational Development and Training. She has been in the research and teaching profession for over a decade. She has completed her Ph.D. and LL.B. degree from Mohan Lal Sukhadiya University, Rajasthan. She also holds an MBA degree from Banasthali Vidyapeeth. During her doctoral studies in the area of Emotional and Spiritual quotient, she had widely published papers in refereed international and national journals. Her areas of research include Spirituality, Value system and Indian ethos. She was awarded the "Best research paper" in the IIM – Shillong Sustainability conference (2015).

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-14083-5

Deep auto encoder based on a transient search capsule network for student performance prediction

Rahul [1] · Rahul Katarya [1]

Received: 13 May 2022 / Revised: 27 July 2022 / Accepted: 21 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Prediction of Student performance through a machine predicts a student's future success. It can be considered an essential procedure to determine the students' academic excellence and identify them at high risk for academic performance. Prediction of student performance also provides universities with a high reputation and ranking. The evaluation of 'What students can do with their learning' is still a tedious task. There are many challenging factors to solve this problem, mainly owing to the enormous amount of data collected from students. Most of the research works have focused on developing new methodologies for student performance prediction. But all the existing work has some performance limitations. Here, a new model called *transient search capsule network based on the deep Autoencoder (TSCNDE*) is introduced to detect student performance. The TSCNDE method is implemented with the help of the PYTHON tool. The performance prediction process has been completed with the help of the OULA dataset. The obtained results are assessed on accuracy (99.2%), precision, (99.8%), specificity (98.7%), and sensitivity (98.9%) parameters. The results obtained showed that the TSCNDE method is about 99.2% more accurate than the other related method. Also, the obtained results are compared with some existing deep learning and machine learning methods.

Keywords Deep learning · Student performance prediction · OULA dataset · Classification · Deep autoencoder · Capsule network

1 Introduction

Education is considered the most powerful tool for the development of our country. Today, educational sectors are similar to business organizations and strongly compete daily to obtain better profit. In addition, these different sectors try to provide quality education to students for a better

✉ Rahul
rahulzhere023@gmail.com

[1] Big Data Analytics and Web Intelligence Laboratory, Department of Computer Science & Engineering, Delhi Technological University, New Delhi, India

future. Now, modern universities face many problems related to student efficiency [22]. It is highly effective for every educational institution to develop high quality education to the students. A student can be exposed to better performance if the teacher provides support based on identifying strong and weak sides in academics. Therefore, it is significant to identify the informative behaviour of the students. Student performance prediction helps the learners and educators modify their teaching process and learning process [3]. Thus, an early warning can be given over weak people to improve the academic status. Thus, the students can reconstruct their mode of study, which leads to good result. But today, due to a large amount of data, universities cannot process student performance [18]. Also, the student performance prediction methods of different institutes fail to understand students' learning patterns and adversely affecting students' performance [2]. To conquer this, machine learning techniques are employed for this application.

Many researchers focused on developing new approaches for the academic evaluation of students as the students face so many problems which negatively affect the whole future. So many factors have the greatest dependency on the student curriculum [8]. Students' learning environment is a salient aspect that has the greatest dependency among students. If the learning environment is good, half of the students' problems can be solved [35]. A good environment always creates a positive attitude towards scholars; this allows them to concentrate more on their studies easily. Many researchers have done their work related to student work environment. Apart from that, teacher is another factor who positively and negatively influences the students. A learner can achieve good results only if the teacher supports the students well. Today many institutes are up to date with good relationships between teachers and students [10].

Likewise, so many factors exist for promoting good results for students. The students' family background is the most concerning factor in achieving good results [30]. Some students face problems within their families, which is a great impact in their academic area. But some families are giving greater support to their children to perform well. Moreover, the mental state of the student also depends on their performance. Therefore, it is very important to measure the student's academic performance. But it seems to be a very difficult task for many researchers. The growth of the students as well as the growth of the institution, is always measured by the successful result of the students [26].

To improve quality education, early prediction of student performance is necessary. Student performance prediction in academic acts as a major concern for every institution as it is connected to job opportunities and organization's reputation. So many methods are developed for performance analysis. The data mining technique acts as the most common method [37]. However, the analysis is done based on some attributes. The most commonly used attributes are CGPA, internal assessment, previous exam grades etc. Apart from this, some researchers use age, family income, family status, and hobbies as attributes based on academics [7]. KNN and SVM are the most commonly used methods, providing very effective results in many research works. KNN is an ML-based supervised algorithm and can be applied for regression prediction and classification [4]. It is also considered a lazy algorithm because it lacks a specialized training phase and utilizes all training information during the classification process. Another feature of KNN is non-parametric because it doesn't have any assumption regarding the underlying data [24]. SVM is also a simple algorithm used for regression and classification. It is the most abandoned method in the classification process, but it is sometimes the most effective technique in the regression process [9]. Apart from this, many ML techniques are widely applied to measure student performance. Even though ML techniques are widely adopted to predict the performance of students, there are certain drawbacks emerging like lack of investigations based on the factors that influence the performance of students, inaccurate identification of attributes, prediction of student performance at later stages and so on. Recently, DL

strategy has obtained increased attention because of automatic ability in prediction process. In particular, it tends to quite easier with the ability of DL to analyse the huge data in accurate prediction. Some of the DL methodologies like convolutional neural network (CNN), Long-short term memory (LSTM) and deep neural network (DNN). The CNN model improves the training ability, LSTM offers better prediction and better flexibility, scalability can be attained in DNN approach. DL approaches can be utilized in applications including fingerprint vein recognition [31–33], image classification [5], forgery detection [36], object recognition [6] and so on. Even through, DL prediction methods performs well than ML techniques, there exists some time complexity issues and deterioration of prediction performance. To conquer these issues, a capsule network based deep auto encoder is used to predict the student performance at early stages with improved accuracy and less complexity.

1.1 Motivation

Evaluation always gives a guide to achieving good performance. Likewise, student performance evaluation possesses a vital role in achieving good outcomes. But it is a very difficult task to achieve since the learning process takes place inside the student's brain. Therefore 'what students can do with their learning' can only be understood through their academic performance. But it is challenging to achieve better results due to the rapid growth of the student's dataset. With the advancement of technologies, many researchers are trying to develop to measure students' performance. But all developed methods fail at some point. The difficulties faced by many educational centres to identify their students' performance draw our attention. Since the students are the backbone of the future world, giving the right instruction on their path is necessary. Many students drop out of their careers due to some problems. This is the main motivation for us to analyze a better solution for the early detection of student performance. Thus, we can build each child with great potential and make them succeed in their entire lives.

1.2 Contributions

- To improvise the performance of student prediction using hybrid transient search Capsule Network-based Deep Auto Encoder to identify the strong and weak sides in academics.
- To determine the accurateness of the proposed TSCNDE method in terms of diverse evaluation metrics to predict student performance and to provide high quality education to the students.
- To enhance the accuracy of student performance prediction over the existing methods and compare it with different ML and DL methods.

Organization: A brief introduction to student education and its importance is given in Section 1. Section 2 discusses the existing works associated with our proposed TSCNDE algorithm, and Section 3 provides the TSCNDE approach. The outcome of the proposed algorithm is provided in Section 4, and the overall work is concluded in Section 5.

2 Related works

Some of works on Deep learning (DL) and Machine Learning (ML) methods are as follows:

An efficient method based on capsule network was presented in [15] for predicting the student entities in corpus. The named entity recognition was explored for leveraging big

data for security. LSTM (Long Short-Term Memory) was considered one of the best DL methods that only use the temporary data in the input data [21]. It was constructed using RNN. Here the performance was achieved using the recurring relationship between the neurons. A memory cell was considered as a special neuron structure found in an LSTM. This memory cell can retain data forever. Three gates such as entrance, forget, and exit gates, were responsible for regulating the transfer of information. The input neuron provides the same feedback as each gate in the LSTM. In addition, each gate has an activation method. If the input gate detects a large activation frequency, the memory cell accumulates the input. The output gate opens the accumulated data to the following neurons if it senses a high degree of activation. If the gate detects a high activation level, the memory cell will be vacant. RNN (Recurrent NN) was the furthermost promising DL method widely utilized to solve many problems [23]. It was a model with neural sequences which uses the state of the art performance. AE (Auto-Encoder) was a neural network version whose input and output were the same [16]. This network operate by reducing the input into a latent-space illustration and rebuilding the output from this illustration.

AE was a kind of artificial neural network adopted to learn effective coding unsupervised. In the modern period, AE has been an emerging field of research in many areas, including anomaly detection. ML was a form of data processing, and it's a branch of AI that focuses on the premise that computers can learn from data and recognize patterns. Here the decision-making process was carried out without any human input. So much research work exists in student performance prediction by using machine learning methods [17]. Here four ML techniques like KNN, DT, NB and ES were considered. KNN (K-Nearest Neighbour) was a commonly used ML technique for many regression, predictive and classification problems [29]. Its popularity was due to the low calculation time and ease of interpretation. It was also considered a supervised learning technique. The K-NN algorithm believes that the original and previous data were identical, where the new case was placed in more similar group to the existing ones. It accumulates all obtainable data and classifies a new data point based on its similarities to the current data. This algorithm can rapidly categorize original data into a precise group. The disadvantage was the High prediction time. DT (Decision Tree) was also considered as a popular supervised ML technique [27]. It was widely applied to many regression and classification problems. It always creates a tree-like path to the expected output. The main aim of DT was to develop a training approach that can forecast the target variables with simple decision learning rules inferred from the training information. A DT was mainly composed of three basic units: A decision node: It was usually used to describe the test attribute; A branch or edge: It indicates the outcome of the test attribute; Leaf: It can also be called the answer node. Different algorithms have been created to ensure DT's development and application to classification tasks. ID3 and C4.5 algorithms were possibly the most used. One of the most commonly used graphical structures for describing and managing unknown details was (NB) Naïve Bayes networks [34]. Two components define Bayes networks: A directed acyclic graph (DAG) was a graphical component where the vertices signify events, and the edges signify relationships between events. NB was considered as a very basic Bayes network made up of DAGs. It has only one root node, known as the parent node. The classification process can be achieved by treating the parent node as a secret parameter. Because it was a special form, that can only calculate the conditional probabilities in the presence of a training collection. If the network has been quantified, Baye's rule may be used to identify every new entity by adding values to its attributes. ES (Ensemble Stacking) was an ensemble learning which was widely used in machine learning to boost efficiency by sorting individual algorithms on a diverse array of regression and classification tasks. Stacking was a heterogeneous ensemble approach that can be both efficient and effective.

In recent times, it has been commonly used in ML competitions. It is similar to a super multi-layer perception. Every layer contains one or more models. Many methods were used to tackle a binary classification problem in machine learning, including Bayesian process algorithms, regression algorithms, kernel-based algorithms, decision tree algorithms, clustering algorithms, etc. Stacking could easily be equipped with various classifiers or regression approaches to boost effectiveness and simplification in a single model. A temporal sequential classification problem was exploited in [11] in predicting the early withdrawal of students. This was carried by tapping the actionable intelligent data power in the form of interactional activities of students with online educational system. The freely available OULA dataset was utilized here by adopting deep LSTM model. The implemented LSTM model outperforms the baseline approaches including artificial neural networks and regression approach. The accuracy rate of 97.25% was obtained in predicting the students' performance, 92.79% of precision and 85.92% of recall was obtained. The decision making process was found to be the major issue here. The risk of course dropping by the students were identified in [13] and the solutions of data absence were presented from preceding courses. The self-learner concept was presented which constructs ML models from the data created during the present course. It adopts the information regarding previously submitted assessments which leads to imbalanced data. The problems were specified through classification task and handled the data imbalance problems. Better results of precision, recall and F1 measure can be attained but the time complexities are more which leads to degraded performance. The ML algorithms were used in [14] to determine less engaged students in social science course and to estimate the effects of engagement over student's performance. Several attributes including final outcomes, education level, overall click towards virtual learning environment activities and assessment score. The outcomes of ML approaches including decision tree and gradient boosted classifiers have attained improved performance with respect to accuracy and kappa estimation. The recall of 93.4% and accuracy of 85.91% was attained by decision tree whereas gradient boosted tree model had obtained 91% of recall and 86.43% of accuracy. Consideration of lesser attributes was found to be the major limitation in this research. Table 1 describes the comparison of related works.

The daily student activities were explored in [12] over virtual learning environment whereas the daily student activities were utilized for success prediction as pass or fail. To incorporate this, the binary information was shown. The anonymized data activity was shown to group the students and diverse types of students were analyzed on the basis of daily binarized activity. The design implications and privacy of this research findings were discussed against the circumstantial data reduction principle and general data protection regulation of European Union. When the student information and daily activities are given as the input, the F1 score of 91.32%, precision of 93.33%, recall of 89.50% and accuracy of 87.98% was attained.

The dynamic effect of demographic characteristics was explored on academic results in [28] in online learning environment. The dynamic impact of six demographic characteristics with the utilization by utilizing the sample gathered from learners who comes under diverse disciplines. The neighborhood poverty level, appropriate education and region were found to signify strong predictors of entire learning results. The institutional support implications were describing through tailored approach for offering a more personalized support system for students. An approach of pre-course task was proposed in [20] for student performance prediction in case of traditional classroom-teaching circumstances that refers to detect at risk students for every course before the beginning. For representing the student sample and correlation utilization among courses, the multi-instance multi-label issue was considered. On consideration of data scarcity problem, an effective multi-task learning approach was

Table 1 Review on existing methods

Author	Technique used	Performance Metrics (%)	Limitations
Jaiswal et al. [15]	Entity-aware capsule network	Accuracy	Failed to provide better results and Computational complexity is high
Minn et al. [21]	Bayesian Knowledge Tracing (BKT)-LSTM	RMSE- 0.31 for Algebra dataset	Continuous monitoring of student's performance is not possible
Okubo et al. [23]	RNN	Accuracy	Performance evaluated on a small dataset
Jaques et al. [16]	Multimodal autoencoder	RMSE-0.134	Failed to provide better results and complex to analyse.
Katarya et al. [17]	ML approaches	Accuracy, precision and recall	Computational complexity and less information related approaches.
Ramakrishna et al. [29]	Data mining based classification model	Accuracy-95.5, Error rate-4.5	Failed to analyse learning skills of student
Raut et al. [27]	Decision tree approach	Response time-7.8 seconds	The accuracy performance is degraded due to huge accumulation of data
Tripathi et al. [34]	Naive bayes classification model	Accuracy- 92.16	The execution time of student performance prediction is more.
Hassan et al. [11]	Deep LSTM	Accuracy-97.25 Precision- 92.79 Recall-85.92	Complex decision making process in student prediction
Hlosta et al. [13]	ML models like naive bayes, random forest and XG Boost	Better results of precision, recall and F1 measure	Increased time complexities and degraded performance
Hussain et al. [14]	ML approaches like decision tree and gradient boosted classifiers	Recall-91 Accuracy- 86.43%	Consideration of lesser attributes due to lack of significant data
Heuer et al. [12]	Student activity assessment based on virtual learning environment	F1 score- 91.32, Precision-93.33%, Recall-89.50 and Accuracy-87.98%	Grouping of students based on anonymized data activity is not proper
Rizvi et al. [28]	Tailored approach based personalized support system	Overall accuracy-75.04	The learning outcomes are not effectively predicted.
Ma et al. [20]	Multi-instance multi-label issue based prediction model	Overall prediction accuracy-85.06	Attribute and parameter consideration are considerably less
Abubakar et al. [1]	Random forest algorithm	Precision-79.66 Recall-66.66 F-measure-70.66	Time consumption for exact prediction is more

proposed for predicting the student performance from diverse specialties over a combined architecture. The highest average prediction accuracy was found to be 85.06% but the attribute and parameter consideration are adversely less. The performance of students was predicted in [1] by contemplating the performance of student's interaction across e-learning environment. The ML approach called random forest algorithm was used in this research for prediction process. The outcomes are compared with k-nearest neighbor and decision tree approaches whereas the proposed model was found to be better. More important attributes that reveal student performance were used for accurate prediction. Effective prediction of student

performance with high robustness can be obtained through this research but the time consumption to predict exactly is more.

Student performance prediction has gained much attention in the field of research recently. Many of the research work failed due to a few limitations in their work. Various researchers have developed many methods to evaluate the performance of students. Some of the methods or algorithms are NB, KNN and DT. The increasing amount of datasets related to the student is the main reason behind the failure of student performance prediction. As the database increases, the complexity of the prediction process also increases. Therefore, it is very important to develop a method for student performance prediction considering many datasets. Learning analytics is considered a powerful tool for analyzing and predicting student performance. Many of the research work in this field requires more datasets for the analysis process. The dataset comprises many student attributes and student cohorts. Analysis of student performance considering this large dataset creates complications in the process. Therefore, finding an alternative method to solve this problem is very important. Nowadays, technology has rapid growth, so our country is developing daily. To continue this scenario in the future, it is essential to mould the students in the correct way at the correct time. The importance of performance analysis of students holds higher attention over here. If students' performance is analyzed regularly, early detection of their weak and strong parts can be possible. It makes the tutor give proper direction to their students. Moreover, better students' performance also helps the institution raise their dignity; since most educational institutions are like business centers, they always seek to make more profit day by day. A better future for a student is the main need for early detection of student performance.

3 Proposed methodology

Many obstacles exist to success, like fear. Therefore, the learning process analysis can be understood by predicting the student attention level as high, medium and low categories using different approaches in machine learning. In this work, the Hybrid ***Transient Search Capsule Network Based on Deep Auto Encoder (TSCNDE)*** was introduced to detect student performance.

- Initially, the educational dataset of students is collected and pre-processed to filter out duplicate records.
- The feature extraction process utilizes the ***deep neural network*** to study the different features of student performance. A ***Fuzzy equilibrium optimizer*** is used for the attribute selection process and is very helpful to avoid the complexity of a bunch of features in the learning process.
- The ***TSCNDE*** method introduces the fuzzy rules, and the fuzzy system creates a feature matrix. For student academic performance prediction, the fuzzy output is passed to the transient search Capsule Network-based Deep Auto Encoder. This algorithm is very helpful to minimize the maximum quantity of features without dropping its original property. In addition, the overall computation time and memory space are lowered. Moreover, in this combination, a hybrid model is applied for obtaining an initial solution. Models are tested with test data, and the obtained results are assessed on parameters like specificity, precision, accuracy, and sensitivity.
- The accuracy is calculated using confusion matrix because it is a good metric for analyzing unbalanced data. The time complexity can also be reduced using this method. The

TSCNDE method is implemented using python. The block diagram of the proposed TSCNDE technique is shown in Fig. 1.

3.1 Data pre-processing

It is the initial and most significant process in predicting student performance. It is a data mining technique that results in efficient data for further processing. The OULA dataset signifies the massive amount of data, and it is very essential to perform valuable extraction of data namely information about student in LM (learning management) system, and interactive logs seen in the online learning programs. The amount of raw data comprises of redundant or irrelevant data. Data pre-processing aims to transform the dataset into suitable form representing free of missing values, well-structured, and complete. Here, the steps followed in data pre-processing are Data cleaning, Attribute selection and Transformation of data.

- **Data cleaning**: This is the initial step in pre-processing which removes the irrelevant parts in the data. Noise, missing data etc., are considered irrelevant parts. The unreasonable or incomplete data has to be patched, deduplicated, removed or corrected to enhance the data quality.
- **Attribute Selection**: The efficiency of the proposed model can be reduced with the presence of so many feature attributes. Hence, selection of attributes play a pivotal role that represents a significant impact on performance modelling.
- **Data Transformation:** In order to obtain input data in an appropriate form, cleaned data are forwarded to the transformation stage. The transformed output contains a large amount of data, which is very difficult for processing. Therefore, reducing the given data into a processable form is very important. This stage is known as the data reduction in pre-

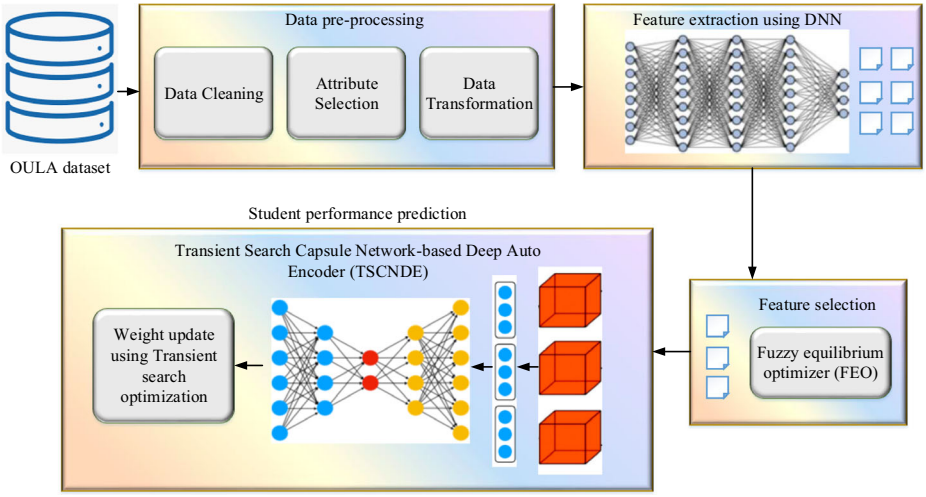


Fig. 1 Block diagram of the TSCNDE method

processing. It is the final pre-processing stage, and the output obtained from the data reduction stage is forwarded to the feature extraction process.

3.2 Feature extraction

Feature extraction is a technique of extracting valuable information from a dataset that contains redundant information. As a result, the dataset will be reduced with a limited number of information by eliminating the redundant data [13]. This may increase the speed of further data processing and helps to reduce the complexity of the system. The features from the OULA dataset are extracted using the DNN (Deep Neural Network) model. DNN is one of the most extensively used and powerful deep learning algorithms. However, this DNN architecture includes an input layer that gets the data and forward to the *"hidden layers"* and after a magic trick the information can provided by the *output layer*. DNN mainly has three types of layers. They are the input layer (IP), output layer (OP), and hidden layers (HL) that exist between the input-output layers. As the figure shows, the input-output layers are single layers. Depending on the algorithm's complexity, the hidden layers may be single or multi-layered. Multiple nodes exist in each layer.

Here, the OULA dataset is used for the prediction of student performance. It contains eleven attributes related to students in which two attributes (code_ module, code_ presentation) are considered redundant. The remaining attributes (student_id, gender, region, qualification, age, disability, result, percentage) are extracted using DNN. The following parameters are considered for the extraction process. The input data given to the first (visible) layer defines the attribute information acquired by calculating the features as well as the hidden layer stated as,

$$\begin{aligned} E^1 &= \left\{E_1^1, E_2^1, \ldots, E_x^1, \ldots . E_3^1\right\}; 1 \le x \le 3 \\ B^1 &= \left\{B_1^1, B_2^1, .., B_o^1, \ldots, B_e^1\right\}; 1 \le o \le e \end{aligned} \tag{1}$$

Where, E_x^1 denotes the x^{th} visible neuron. The hidden layers are specified as B^1, B_o^1 indicate the o^{th} hidden neuron, the number of hidden neurons is represented as e. The HL nodes were designed to calculate the weighted value and a bias which is represented as:

$$Q_j = \sum_i^m k_i * w_{ij} + \theta_j \tag{2}$$

Where, the input data is k_i, θ be the bias and w_{ij} is the weighted link. Therefore, Q_j can be transformed with the following transfer function (sigmoid):

$$F\left(Q_j\right) = \frac{1}{1 + e^{-Q_j}} \tag{3}$$

Next, the network training procedure starts with the weights initialization via choosing a set of random values. During the training process the network weights are adjusted accordingly:

$$\Delta w_{ij} = -n \frac{\theta E}{\partial w_{ij}} \tag{4}$$

Where, E signifies the error, and the learning rate is denoted as n. Here, the state activation function can be denoted as M. All operations in DNN can be indicated using the following equations. The input and output operation is given by,

$$i_t = M\left(k_t a_t^i + w_t h_{t-1} + \theta_i\right) \tag{5}$$

$$o_t = M\left(k_0 x_t^i + w_0 h_{t-1} + \theta_0\right) \tag{6}$$

Where, θ is the bias vector, i_t ***and*** o_t are the input-output states and w is the updated weight from the hidden vector.

3.3 Feature selection using fuzzy equilibrium optimizer (FEO)

FEO is used for the attribute selection process and is very helpful in avoiding the complexity of a bunch of features in the learning process. The feature selection process by applying fuzzy rules can be described as follows.

Let X be data for selecting the features, and it can be formulated as follows.

$$X = \left\{ x_d \middle| d = 1, 2, 3, \ldots\ldots\ldots, n \right\} \tag{7}$$

Where n specifies the total amount of patterns, $x_d \in R^a$ and a represents the number of features f_i. Let x_d^a be the value of a^{th} feature f_a of patterns x_d. Then, we can define each pattern of x_d the labelled dataset X by using a vector as follows.

$$x_d = \left[x_d^1, x_d^2, x_d^3, \ldots., x_d^a\right] \tag{8}$$

Let F denotes a set in which each element in F represents a feature. That is,

$$F = \{f_1, f_2, f_3, \ldots\ldots, f_a\} \tag{9}$$

Here use a membership function H to project the dataset into a fuzzy space. The membership function H can be represented as follows.

$$H = \left\{ \beta_{11}, \beta_{12}, \beta_{13}, \ldots.\beta_{1p}, \beta_{21}, \beta_{22}, \beta_{23}, \ldots, \beta_{2q}, \beta_{a1}, \beta_{a2}, \beta_{a3}, \ldots.\beta_{ar} \right\} \tag{10}$$

Where β_{ij} represents the j^{th} fuzzy set of i^{th} feature. p, q, r in the definition of the membership, set indicates the positive real numbers, which are used to represent the cardinality of the fuzzy set of the first, second…. and the p^{th} feature, respectively. Let F_x be the projected set of fuzzy of the dataset X. It can be defined as follows.

$$F_x = \left(x_d, \beta(x_d)\right) \tag{11}$$

Where, $\beta(x_d)$ is a vector which is represented using Eq. (12).

$$\beta(xd) = \left[\beta_{11}\left(x_d^1\right), \beta_{12}\left(x_d^2\right), \ldots, \beta_{1p}\left(x_d^1\right), \beta_{21}\left(x_d^2\right), \beta_{22}\left(x_d^2\right), \ldots.\beta_{a1}\left(x_d^a\right), \beta_{a2}\left(x_d^a\right), \ldots.\beta_{ar}\left(x_d^a\right)\right] \tag{12}$$

Let $|f_i|$ denotes the number of fuzzy sets for each f_i feature. Then the fuzzy set β_{ij} can be defined as follows.

$$\beta_{ij} : x_d^i \rightarrow [0, 1]; \forall i \in \{1, 2, 3, \ldots a\} \wedge \forall j \in \{1, 2, \ldots |f_i|\} \wedge k \in \{1, 2, 3, \ldots\ldots\ldots n\} \tag{13}$$

The resolution of the membership function is quantified by using the value $\sum_{i=1}^{a} |f_i|$. An optimal combination of a fuzzy set (β_{ij}) determines the fuzzy selection process. Let $j(.)$ be a function used to evaluate a subset of a fuzzy set and α be all possible combinations of a fuzzy subset. Then fuzzy feature selection is achieved by obtaining a fuzzy set subset $H_{optimal}$ that should satisfy the following equation.

$$j\left(H_{optimal}\right) = E(j(H_i)), \forall H_i \subseteq \alpha, \alpha = 2^H \tag{14}$$

Where, E is the minimum or maximum operator. The value of $H_{optimal}$ can be improvised using an ***Equilibrium Optimizer (EO)***. The required optimization process for the feature selection can be described as follows.

3.3.1 Parameter optimization using EO

The below mentioned four stages describe the mathematical modelling of EO.

Step 1: Initialization

EO utilizes a group of elements, each of which signifies a concentration vector. This concentration vector holds the improvised solution of $H_{optimal}$. The random generation of the initial concentration vector is according to the following equation.

$$\begin{gathered} v_{ci} = c_{\min} + (c_{\max} - c_{\min}) * s_i \\ i = 1, 2, \ldots n \end{gathered} \tag{15}$$

Where, the concentration vector is represented by v_{ci}, the upper bound dimension in the problem is determined by $c_{\min}$, whereas the lower bound dimension is determined by $c_{\max}$. Equation 15, s_i indicates random vector with [0,1] interval, and n is the total number of particles present inside the group as population. The value of s_i is in between [0, 1].

Step 2: Equilibrium pool and candidates (E_{pc})

EO also tries to achieve a better optimization result like all optimization algorithms. It always searches for the system's equilibrium state. After attaining the state of equilibrium, it forces to move toward the near-optimal solution of the optimization problem. EO doesn't know the concentration level to attain equilibrium during the optimization process. Therefore, it is forced to assign five particles. Among the five particles, four particles are the best ones in the population, and the extra one is the average of these four particles.

Further exploitation and exploration of the operator are carried out with the help of these five equilibrium particles. The selected five particles are stored as vectors, generally known as equilibrium pools. The following equation indicates the typical representation of the equilibrium pool. Algorithm 1 describes the steps involved in parameter optimization using the EO approach.

$$\overrightarrow{E}_{pc} = \left[\overrightarrow{E}_{pc(1)}, \overrightarrow{E}_{pc(2)}, \overrightarrow{E}_{pc(3)}, \overrightarrow{E}_{pc(4)}, \overrightarrow{E}_{pc(avg)} \right] \tag{16}$$

Algorithm 1 Parameter optimization using EO

Initialize the population of a particle (i= 1, 2, 3…….. n)

Set the fitness value of four particles in the equilibrium pool, $\vec{E}_{pc}$ with a large value.

Set the values of parameters

$While\ (n_{it} < \tau_{\max})$

For each i^{th} particle

Calculate the fitness value of the particle $i\ \vec{B}_f(\vec{E}_{pc(i)})$

If $\left((\vec{B}_f(\vec{E}_{pc(i)}) < (\vec{B}_f(\vec{E}_{pc(1)}))\right)$

Set $\vec{E}_{pc(1)}$ with $\vec{E}_{pc(i)}$ and $\vec{B}_f(\vec{E}_{pc(1)}\ with\ \vec{E}_{pc(i)})$

$else\ If\left(\vec{B}_f(\vec{E}_{pc}) > \vec{B}_f(\vec{E}_{pc(1)}\ and\ (\vec{B}_f(\vec{E}_{pc(i)}) < \vec{B}_f(\vec{E}_{pc(2)})\right)$

Set $\vec{E}_{pc(2)}$ with $\vec{E}_{pc(i)}$ and $\vec{B}_f(\vec{E}_{pc(2)}\ with\ \vec{E}_{pc(i)})$

$else\ If\left(\vec{B}_f(\vec{E}_{pc}) > \vec{B}_f(\vec{E}_{pc(2)}\ and\ (\vec{B}_f(\vec{E}_{pc(i)}) < \vec{B}_f(\vec{E}_{pc(3)})\right)$

Set $\vec{E}_{pc(3)}$ with $\vec{E}_{pc(i)}$ and $\vec{B}_f(\vec{E}_{pc(3)}\ with\ \vec{E}_{pc(i)})$

$else\ If\left(\vec{B}_f(\vec{E}_{pc}) > \vec{B}_f(\vec{E}_{pc(3)}\ and\ (\vec{B}_f(\vec{E}_{pc(i)}) < \vec{B}_f(\vec{E}_{pc(4)})\right)$

Set $\vec{E}_{pc(4)}$ with $\vec{E}_{pc(i)}$ and $\vec{B}_f(\vec{E}_{pc(4)}\ with\ \vec{E}_{pc(i)})$

$end\ if$

$end\ for$

$\vec{E}_{pc(avg)} = (\vec{E}_{pc(1)} + \vec{E}_{pc(2)}, + \vec{E}_{pc(3)} + \vec{E}_{pc(4)})/4$

The equilibrium pool $\vec{E}_{pc} = [\vec{E}_{pc(1)}, \vec{E}_{pc(2)}, \vec{E}_{pc(3)}, \vec{E}_{pc(4)}, \vec{E}_{pc(avg)}]$

Accomplish the memory saving

Assign the value of τ using $\tau = \left(1 - \frac{n_{it}}{\tau_{\max}}\right)^{\left(d_o * \frac{n_{it}}{\tau_{\max}}\right)}$

For each i particle

Choose one candidate from the $\vec{E}_{pc}$

Generate random vector γ

Construct $\vec{B}_f$ using $\vec{B}_f = e^{-\vec{\gamma}(\tau - \tau_0)}$

Update the concentration

$end\ for$

n_{it}^{++}

$end\ while$

Step 3: updating the concentration

EO usually has a balance between diversification and intensification. Let $\overrightarrow{\gamma}$ represents a random vector that lies in the interval [0, 1]. Then the expression for the fitness function can be explained as follows.

$$\overrightarrow{B}_f = e^{-\overrightarrow{\gamma}(\tau-\tau_0)} \tag{17}$$

Where,τ signifies the present iteration, and the initial value is represented by τ_0. $\overrightarrow{B}_f$ represents an exponential term. As the amount of iterations rises, the value of τ decreases according to the following equation.

$$\tau = \left(1-\frac{n_{it}}{\tau_{\max}}\right)^{\left(d_o * \frac{n_{it}}{\tau_{\max}}\right)} \tag{18}$$

Where, n_{it} represents the total number of iterations, τ and $\tau_{\max}$ represents the current and maximum value of iterations, respectively. To control the capability of intensification, a constant parameter d_0 is used.

Step 4: Parameter optimization

The optimization of the parameter is achieved by updating the concentration. The following equation describes the parameter optimization process.

$$H_{opt} = \overrightarrow{E}_{pc} + \left(\overrightarrow{E}-\overrightarrow{E}_{pc}\right) * \overrightarrow{B}_f + \left(\frac{1}{\gamma * V}\right) * \left(1-\overrightarrow{B}_f\right) \tag{19}$$

Where, γ is a random vector in between [0, 1]. $V = 1$ and F is an exponential term.

3.4 Feature classification

Feature classification is a process of recognizing a pattern by categorizing a huge number of data into different classes. Feature classification is always done after the feature selection process. The selected features are categorized according to some parameters in the feature classification stage. There are many methods used to undergo the classification process. The classification process ensures more accurate results. Here, the feature classification process is achieved with the help of the Transient Search Capsule Network-based Deep Auto Encoder (TSCNDE). The fuzzy output is passed to the TSCNDE for student academic performance prediction.

3.4.1 Capsule network _Deep autoencoder

A capsule generally indicates a vector in which its length denotes the probability of the existence of an entity. Capsules are arranged as a group of neurons to help analyze the whole features by recognizing each feature related to students. The orientation of the vector signifies the properties of the entity. Figure 2 describes the typical architecture of a capsule network-based deep autoencoder.

As Fig. 3 depicts, a capsule network mainly has four parts. They are the Convolution layer, Primary caps layer, Digit caps layer, and decoder. Here, the last part (Digit Caps and decoder) is replaced using the Deep Autoencoder (DAE). Let the input features applied to the capsule

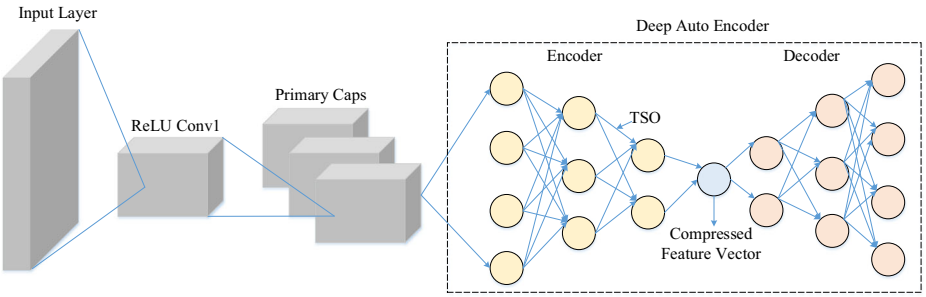


Fig. 2 Architecture of Capsule Network-based Deep Auto Encoder

network be denoted as i_j, and the corresponding output is denoted as o_j. The following equation represents the output operation of the capsule network.

$$o_j = \frac{\left\|i_j\right\|^2}{1 + \left\|i_j\right\|^2 \left\|i_j\right\|} \tag{20}$$

The outcome of the capsule network is composed of a set of vector values related to student features, commonly known as initialization parameters (IP). Due to some limitations, these parameter does not provide an accurate result. Therefore, another optimization algorithm is known as Transient Search Optimization (TSO) approach is adopted here to optimize the output parameters of the capsule network.

3.4.2 Parameter optimization using TSO algorithm

TSO is a meta-heuristic algorithm based on physics [25]. Here, TSO is used to optimize the output parameter of the capsule network, which is done according to the transient behavior of switching circuits. This TSO algorithm can be modelled based on the following characteristics such as search-agents initialization among search area lower and upper bounds, best solution searching (exploration

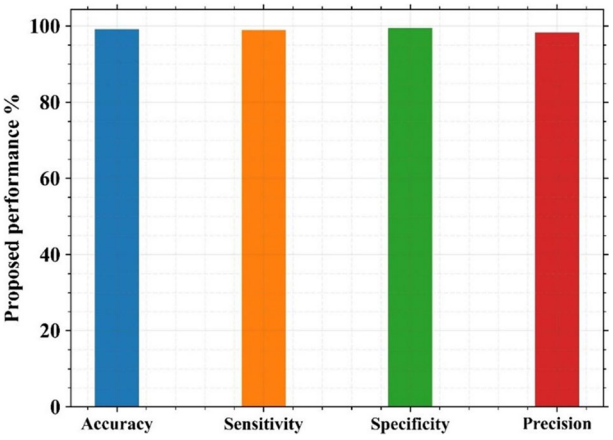


Fig. 3 Performance results of TSCNDE method in parameters like Accuracy, Sensitivity, Specificity, and Precision

phase), reaching the best (exploitation phase) solution. The algorithm to implement the TSO process is as follows. The parameter optimization of TSO is mentioned in Algorithm 2.

Algorithm 2 Parameter optimization using TSO

Initialize the population and the best positions
Evaluate the maximum number of iterations in the population
While $l < l_{\max}$
    Update the values of O_j using equation (22)
do for all populations
    Update the population
end do
    Calculate the number of iterations of all new population
    Update the best value if the iteration value is less than the previous value.
$l = l + 1$
end while
Output the optimized value

Initially, the search-agents of TSO are randomly initialized and generated as,

$$Z = l_b + ran \times (u_b - l_b) \tag{21}$$

Next, the TSO exploration performance is motivated by the 2nd order RLC circuits oscillation around zero whereas the TSO exploitation was characterized by the exponential decaying of the 1st order discharge. The mathematical model of TSO algorithm with exploitation as well as exploration can be expressed as,

$$Z_{l+1} = \begin{cases} Z_l^* + \left(Z_l - Z_l^* \cdot C_1\right) e^{-T} & r_1 < 0.5 \\ Z_l^* + e^{-T}[\sin(2\pi T) + \cos(2\pi T)] \left| Z_l - Z_l^* \cdot C_1 \right| & r_1 \geq 0.5 \end{cases} \tag{22}$$

$$T = 2 \times v \times r_2 - v \tag{23}$$

$$C_1 = v \times k \times r_3 + 1 \tag{24}$$

$$v = 2 - 2\left({}^{l}/_{L_{\max}}\right) \tag{25}$$

Where, *ran* represents the random number, l_b and u_b specifies the lower and upper bound, random coefficients are C_1 and T, v specifies the variable that varies from 2 to 0, random numbers are r_1, r_2, r_3. The number of maximum iterations is $L_{\max}$, the search agents position is Y_l, the iteration number is signified as l, best position is Y_l^* and k represents the constant number as (k = 0,1,…). In order to optimize o_j the following Eq. (22) in TSO is considered. This output is fed as the input of the Deep Auto Encoder.

DAE represents a feed-forward neural network, and it gives copies of input as the output. As the figure shows, DAE is a system having fully connected network layers. DAE mainly has

two layers. The encoder section transforms the data with high dimensional into the lower dimensional representation at the bottleneck layer. High dimensional features are obtained at the decoder by converting the bottleneck features. The following Eq. (26) represents that AE (autoencoder) can realize the identity function as,

$$D^{'} = h_{w,b}(D) \approx D \tag{26}$$

Where, the input is represented as D, the bias is b and weight is denoted as w. Next, the AE encodes the input D into a compact form as,

$$S = \eta(wD + b)\eta(x) = \frac{1}{1 + \exp(-x)} \tag{27}$$

Where, S denotes the compressed form which is decoded to output $D^{'}$ with mapping function expressed as,

$$D^{'} = \eta\left(Sw^{'} + b^{'}\right) \tag{28}$$

The process can be iterated till $D \approx D^{'}$ and the outcome of final encoder signifies the high-level, compact form of input data. Here, the maximum features are minimized, and the original property is preserved.

4 Result and discussion

The result of the TSCNDE Transient search capsule network-based deep auto encoder is described as follows. The student performance is evaluated based on the OULA dataset [19]. It contains data related to students' previous performance and personal information. Python is the tool used for implementation. The experimental settings are described in Table 2. Section 4.1 gives a brief description of the OULA dataset.

4.1 OULA dataset

The dataset used in this work is OULA which stands for Open University Learning Analytics dataset (https://analyse.kmi.open.ac.uk/open_dataset). This dataset comprises different features related to students. Some of the features are clickstream history, demographics of students, information regarding assessment submission etc. It contains details about 32,593 students over a nine-month course duration in a raw tabular form. The dataset is collected between 2014

Table 2 Experimental configurations

Parameter	Value
Processor	Intel (R) Core (TM) i3–3220 CPU @ 3.30 GHz
RAM	8GB
System type	64 bit Operating System
EO_Generation Probability (GP)	0.5
Random numbers	$r_1 = r_2 = [0, 1]$
TSO	$a$decreases from 2 to 0, k=0

to 2015. It also contains information regarding the student's interaction with the VLE environment. Table 3 indicates the details of the OULA dataset.

4.2 Experimental results

Here, the TSCNDE method is assessed using evaluation measures such as precision, accuracy, specificity, and sensitivity. Because all four measures analyzed here are considered to be very effective for the evaluation of different methods. The following Table 3 shows the evaluation result of TSCNDE with other existing methods such as KNN, DT, NB and ES. The initial solution is achieved through this model. The subsequent Eqs. (29), (30), (31), (32) represents the calculation of parameters like accuracy, sensitivity, specificity and precision.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{29}$$

$$sensitivity = \frac{TP}{TP + FN} \tag{30}$$

$$specificity = \frac{TN}{FP + TN} \tag{31}$$

$$precision = \frac{TP}{TP + FP} \tag{32}$$

In these Eqs. (29, 30, 31, 32), TN and TP indicate the number of true negative and true positive performance predictions. FN and FP represent the false negative and false positive performance predictions. FPTN indicates the number of false positives derived from negative samples in the classification.

Table 4 depicts the comparison graph in terms of different ML techniques. Here, four ML techniques are executed and used for the comparison process: KNN, DT, NB, and ES. KNN indicates K-Nearest Neighbor, which is a classification and regression technique. Likewise,

Table 3 Dataset Details

	StudentInfo	StudentAssessment	StudentVIe	StudentRegistration
Rows	32,593	173,912	10,655,280	32,593
Columns	code_module code_presentation id_student gender region highest_education imd_band Age_band Num_of_prev_attempts Studied_credits disability Final_result	id_assessment id_student date_submitted is_banked score	code_module code_presentation id_student id_site date sum_click	code_module code_presentation id_student date_registration date_unregistration

Table 4 Performance comparison of TSCNDE method with existing ML methods

Parameter	TSCNDE (proposed method)	KNN	DT	NB	ES
Accuracy (%)	**99.2**	98.2	97.8	98.2	98.7
Precision (%)	**99.8**	70	59	64	79
Sensitivity (%)	**98.9**	98.1	97.9	98.4	98.7
Specificity (%)	**98.7**	92.1	89.9	86.6	94.8

The bold entries reflects the proposed results.

DT is the Decision Tree classifier, and ES is the Ensemble Stacking classifier, commonly used for most classification processes. NB is the Naïve Bayes classifier, considered probabilistic and helps achieve a higher accuracy level. All four methods used here for the comparison process are considered very efficient to produce better accuracy. At the same time, due to some limitations in some other parameters, they don't give much better results.

The performance of the TSNDE method is graphically represented in Fig. 3. The horizontal axis corresponds to different evaluation parameters considered here for performance prediction, and the vertical axis corresponds to the result of the TSCNDE method.

Figure 3 demonstrates the performance of the TSCNDE method. Various parameters like accuracy, sensitivity, specificity and precision were analyzed. By calculating accuracy of the TSCNDE method is obtained as 99.2%, and it is higher than all other existing methods. It indicates the value is closer to the measured performance of the students. The sensitivity of the TSCNDE method is 98.9%. The sensitivity of the TSCNDE method is slightly higher than other existing methods such as ES and NB. It shows higher performance than the DT method. After evaluation, the specificity of the TSCNDE method is 98.7%. In addition, a confusion matrix for performance evaluation is also used here.

The performance result using the confusion matrix is depicted in Fig. 4. The confusion matrix is a distinct type of layout holding two dimensions, including actual and predicted occurrences. With the help of such a table layout, the conception of the performance procedure

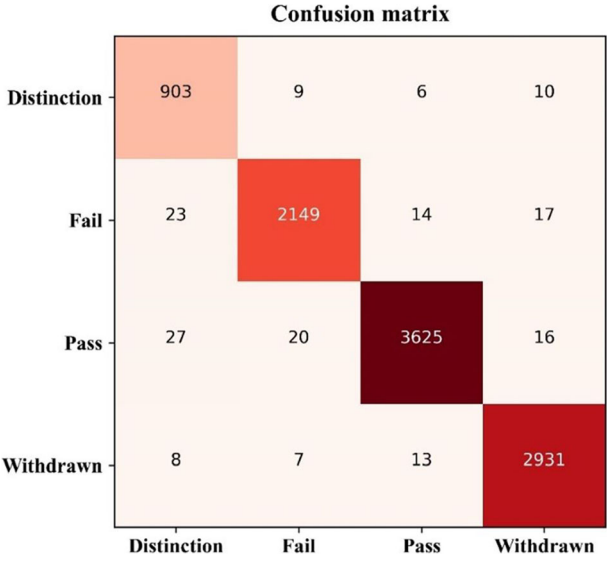


Fig. 4 Performance result using a confusion matrix

in deep learning can be made possible. The confusion matrix is also called the error matrix. Every element in the column signifies the actual occurrences in this matrix, and elements in the row signify predicted occurrences (or vice versa). The confusion matrix shown in Fig. 4 summarizes distinction, fail, pass and withdrawn. It shows the mutual relationship between real class parameters and the predicted class parameters. The number of students who got distinctions, failed, passed, and withdrawn is 903, 2149, 3625 and 2931. Likewise, the performance related to students can effectively be predicted using the confusion matrix. Figure 5 represents the comparison of the TSCNDE method with existing ML methods.

Figure 5 compares our TSCNDE model with other existing KNN, DT, NB, and ES methods. The comparison results are taken in four evaluation parameters: precision, accuracy, specificity, and sensitivity. As the figure depicts, the sensitivity of the TSCNDE technique is good. It shows better performance than the DT method. It has slightly improved performance in sensitivity as compared with the KNN and NB methods. The precision of the TSCNDE method is 99.8%. It shows higher performance than the DT, NB, and KNN method. All these three methods have a precision value below 70%. The ES method has almost 80% precision in student performance prediction. By analyzing all the results related to precision, it can summarise that the TSCNDE method has better performance related to precision.

The TSCNDE gives 99.2% accuracy in performance results. As the figure depicts, all other methods are less accurate than our TSCNDE method. The specificity of the TSCNDE method is 98.7%. This specificity value is much better than the KNN, DT, and NB (92.1%, 89.9%, 86. 6%) methods. The TSCNDE method shows a slightly improved value in specificity than the

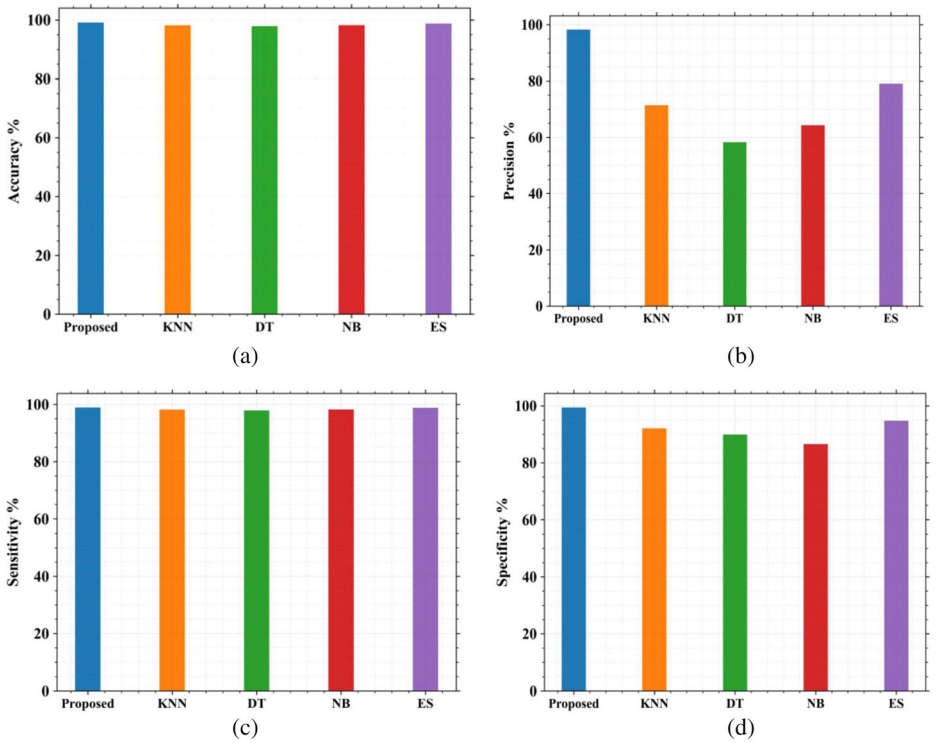


Fig. 5 Performance comparison with ML techniques

Table 5 Performance comparison with DL methods

Method	Accuracy (%)	Precision (%)
TSCNDE (proposed method)	**99.2**	**99.8**
LSTM	96.48	96.14
RNN	93.46	93.9
DBN	94.69	93.26
Autoencoder	92.89	91.79

The bold entries reflects the proposed results.

ES (94.8%) method. In short, the TSCNDE method gives better performance in accuracy, precision, sensitivity and specificity than the other existing methods.

4.3 Comparison with deep learning methods

Table 5 represents the performance comparison of the TSCNDE technique with some existing DL methods. Three methods are chosen to compare the performance of the TSCNDE technique. LSTM indicates the Long Short Term Memory. It is a DL method used for the different classification processes. This method gives better results, but the operating cost and the complexity of such systems are considered very high. DBN (Deep Belief Network) and RNN stands for Recurrent Neural Network, widely used for different prediction processes. Similarly, Autoencoders are a neural network structure typically used for different prediction processes.

Table 5 shows the result of the TSCNDE technique with the existing DL approaches such as LSTM, RNN and Autoencoder. Table 6 illustrates the comparative analysis on OULA dataset.

Various methods mentioned in the graph for student performance prediction. All three methods give better results in some parameters and show poor performance due to limitations. Figure 6 indicates the comparison graph of TSCNDE with some existing deep learning methods.

Figure 6 shows the comparison graph of the TSCNDE method in terms of accuracy and precision with existing DL methods such as LSTM, RNN and auto-encoder. The graph shows that the accuracy of the TSCNDE method is 99.2%. It is much better than the other DL methods such as LSTM (96.48%), RNN (93.46%), DBN (94.69%) and Autoencoder (92.89%). The precision value of the TSCNDE method is 99.8%. The corresponding value of LSTM, RNN, DBN and Autoencoder is 96.14%, 93.9%, 93.26% and 91.79%, respectively. That is, TSCNDE performs better than other DL methods. Figure 7 indicates the result of the TSCNDE method using a ROC curve.

Table 6 Comparison on OULA dataset

Author Name	Technique Used	Performance (%)
Proposed	**TSCNDE**	**Accuracy (99.2%)**
Hassan et al. [11]	LSTM	Accuracy (97.25%)
Rizvi et al. [28]	DT	Accuracy (83.14%)
Heuer and Breiter [12]	SVM	Accuracy (90.85%)
Hussain et al. [14]	ML algorithms	Accuracy (88.52%)
Hlosta et al. [13]	ML algorithms	F1-score (71.31%)

The bold entries reflects the proposed results.

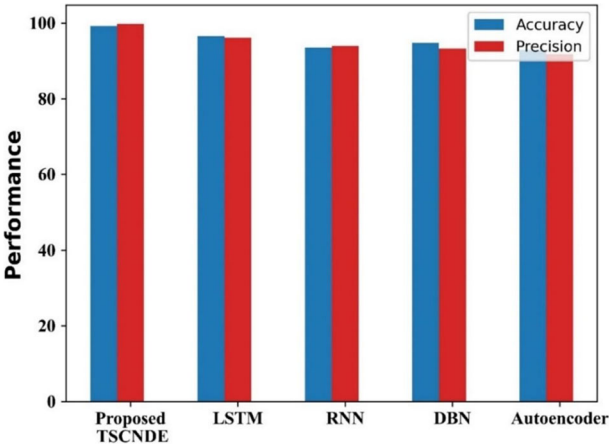


Fig. 6 Performance comparison with DL methods

Figure 7 represents the ROC curve of the TSCNDE method, which stands for Receiver Operating Characteristics. Here, it is used to analyze the performance of the TSCNDE classification approach. The X-axis signifies the False Positive Rate in the ROC curve, and the Y-axis signifies the True Positive Rate. The performance evaluation is carried out with the help of various threshold values. It shows the ability of the classification method to differentiate classes. As the value of the curve increases, the model's performance also increases.

4.4 Discussion

Nowadays, due to the emergence of technologies the educational organizations are distinctive and play a pivotal role in the country's development. Education changes the lives of people,

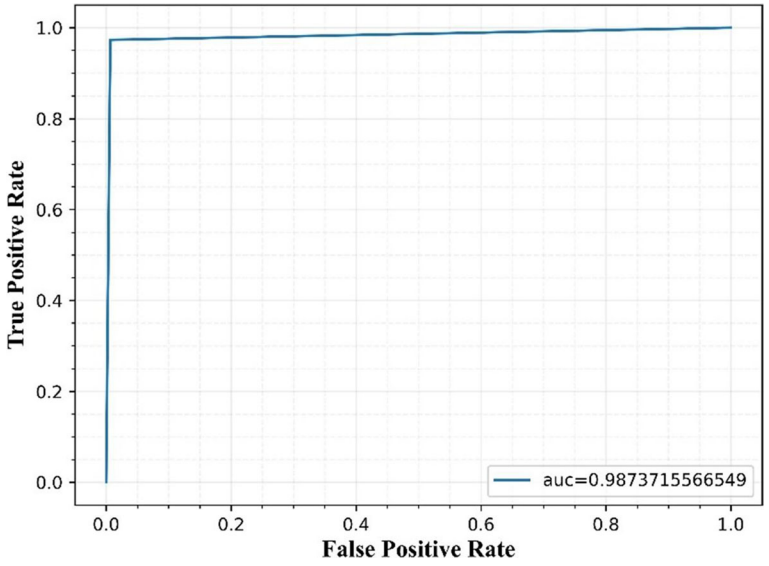


Fig. 7 ROC curve

societies and the world. Now the formal education system goes beyond like Web-based Education System, Online Education System, Workshops, Seminars etc. Most of the existing methods be unsuccessful with the bulk data stored in educational databases. The major goal of the proposed TSCNDE based deep learning framework is the performance prediction of students with the success obtained from student learning (SL). One of the most essential parts of education is the student's academic performance. Effective monitoring of student's academic performance is suitable to identify the students who perform low. The best suited method is the student performance prediction that supports the low performing students in the initial stage. Mining the educational data aims to extract the interesting patterns from OULA (Open University Learning Analytics) dataset using deep learning models. This work perform an effective student performance prediction using transient search capsule network based on the deep Autoencoder (TSCNDE). With the proposed (TSCNDE) model, the performance of the students can be categorized as Distinction, Fail, Pass and Withdrawn. Accurate student performance prediction plays a vital role in improving the teaching efficiency, promoting personalized learning, solving the problem of a student dropping out etc. Moreover, due to the significant progress in Deep Learning (DL) models, the proposed TSCNDE model has ensued in the effective prediction of data. Early prediction of the student's performance can effectively minimize the risk of failure. The results on the OULA dataset is compared with machine and deep learning algorithm to evaluate the merits of the proposed system. The metrics used for computing are accuracy, precision, sensitivity and specificity. The results showed the significance of the proposed DL model in early prediction of the students' academic performance. This proposed model forecasts the success of students and utilized as the base for early intervention on potential failure of students to attain the learning purposes. The presented TSCNDE model aims to analyze the use of both machine learning and deep learning techniques to predict the achievement of student systematically. The existing approaches on predicting the success of SL is very crucial to each one engaged in determining the most appropriate strategies that help the students to execute the learning task. Machine learning models are significantly used but these models fail to handle massive amount of data. The proposed hybrid TSCNDE model resolve the problem of detecting students with poorer academic/educational performance and discover the students' performance and obtained best outcomes.

5 Conclusion

Predicting student performance is one of the major requirements in recent years. Consistent early-stage predictions of a student's future performance could be critical to simplifying timely educational involvement during a course. Many research works have been done to evaluate the students' performance, and due to the rapid growth of datasets related to students, accurate results are not obtained. Hence the limitations are overcome in the proposed research work by utilizing the advanced techniques of machine learning to achieve good performance analysis of students. The TSCNDE work is implemented using different machine learning techniques, and the student performance is evaluated and compared with recently developed methods. Using this TSCNDE method, the academic problems related to students can be solved. The attributes used for the prediction process are highly accurate, and they provide better results in the measurement of student performance than the existing methods. Assessing all the features of a student's curriculum can reduce almost all the problems faced by many students and

institutions. The TSCNDE technique is compared with the existing approaches, and the result obtained possesses high accuracy. The comparison result shows the efficiency of our TSCNDE method. The TSCNDE method is 99.2% accurate in forecasting student performance and analyzing the PYTHON simulation tool. The corresponding precision value of the proposed TSCNDE method is 99.8%, which means it can give precise results regarding student performance prediction. Also, the TSCNDE method is about 98.9% sensitive and 98.7% specific in student performance prediction. This early prediction of student performance improves the educational position of the current system. The limitation of the proposed TSCNDE method is that only single dataset and limited performance metrics are used for predicting the student's performance. This impacts the performance of the proposed system. Therefore, a lot of data can be taken into consideration to obtain much better results. In future, this research work can be prolonged by utilizing more than two datasets for student performance prediction.

Authors contributions All authors have equal contributions in this work.

Data availability Data sharing not applicable to this article.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate All the authors involved have agreed to participate in this submitted article.

Consent to publish All the authors involved in this manuscript give full consent for publication of this submitted article.

Conflict of interest Authors declare that they have no conflict of interest.

References

1. Abubakar Y, Ahmad NBH (2017) Prediction of students' performance in e-learning environment using random forest. Int J Innov Comput 7(2)
2. Andolsek KM (2016) Improving the medical student performance evaluation to facilitate resident selection. Acad Med 91(11):1475–1479
3. Asselman A, Khaldi M, Aammou S (2021) Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Environ 1–20
4. Badugu S, Rachakatla B (2020) Students' performance prediction using machine learning approach. In: Data engineering and communication technology. Springer, Singapore, pp 333–340
5. Bansal M et al (2021) Transfer learning for image classification using VGG19: Caltech-101 image data set. J Ambient Intell Human Comput
6. Bansal M, Kumar M, Kumar M (2021) 2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors. Multimed Tools Appl 80(12):18839–18857
7. Chakraborty P, Yousuf MA, Rahman S (2021) Predicting level of visual focus of Human's attention using machine learning approaches. In: Proceedings of international conference on trends in computational and cognitive engineering. Springer, Singapore, pp 683–694
8. Cheung LL, Kan AC (2002) Evaluation of factors related to student performance in a distance-learning business communication course. J Educ Bus 77(5):257–263

9. Giri A, Vignesh M, Bhagavath V, Pruthvi B, Dubey N (2016) A placement prediction system using k-nearest neighbors classifier. In: 2016 second international conference on cognitive computing and information processing (CCIP). IEEE, pp 1–4
10. Guillén-Gámez FD, Mayorga-Fernández MJ (2020) Identification of variables that predict teachers' attitudes toward ICT in higher education for teaching and research: a study with regression. Sustainability 12(4):1312
11. Hassan S-U, Waheed H, Aljohani NR, Ali M, Ventura S, Herrera F (2019) Virtual learning environment to predict withdrawal by leveraging deep learning. Int J Intell Syst 34(8):1935–1952
12. Heuer H, Breiter A (2018) Student success prediction and the trade-off between big data and data minimization. DeLFI 2018-die 16. E-learning Fachtagung Informatik
13. Hlosta M, Zdrahal Z, Zendulka J (2017) Ouroboros: early identification of at-risk students without models based on legacy data. In: Proceedings of the seventh international learning analytics & knowledge conference, pp 6–15
14. Hussain M, Zhu W, Zhang W, Raza Abidi SM (2018) Student engagement predictions in an e-learning system and their impact on student course assessment scores. Computational intelligence and neuroscience 2018:1–21
15. Jaiswal AK, Tiwari P, Garg S, Shamim Hossain M (2021) Entity-aware capsule network for multi-class classification of big data: a deep learning approach. Futur Gener Comput Syst 117:1–11
16. Jaques N, Taylor S, Sano A, Picard R (2017) Multimodal auto encoder: a deep learning approach to filling in missing sensor data and enabling better mood prediction. In: 2017 seventh international conference on affective computing and intelligent interaction (ACII). IEEE, pp 202–208
17. Katarya R (2019) A review: predicting the performance of students using machine learning classification techniques. In 2019 third international conference on I-SMAC (IoT in social, Mobile, analytics and cloud) (I-SMAC) IEEE 36-41
18. Kistner S, Rakoczy K, Otto B, Ewijk C D-v, Büttner G, Klieme E (2010) Promotion of self regulated learning in classrooms: investigating frequency, quality, and consequences for student performance. Metacogn Learn 5(2):157–171
19. Kuzilek J, Hlosta M, Zdrahal Z (2017) Open university learning analytics dataset. Sci Data 4(1):1–8
20. Ma Y, Cui C, Jun Y, Guo J, Yang G, Yin Y (2020) Multi-task MIML learning for pre-course student performance prediction. Frontiers of Computer Science 14(5):1–10
21. Minn S (2020) BKT-LSTM: efficient student modeling for knowledge tracing and student performance prediction. arXiv preprint arXiv:2012.12218
22. Namoun A, Alshanqiti A (2020) Predicting student performance using data mining and learning analytics techniques: a systematic literature review. Appl Sci 11(1):237
23. Okubo F, Yamashita T, Shimada A, Ogata H (2017) A neural network approach for students' performance prediction. In: Proceedings of the seventh international learning analytics & knowledge conference, pp 598–599
24. Pujianto U, Prasetyo WA, Taufani AR (2020) Students academic performance prediction with k-nearest neighbor and C4. 5 on SMOTE-balanced data. In: 2020 3rd international seminar on research of information technology and intelligent systems (ISRITI). IEEE, pp 348–353
25. Qais MH, Hasanien HM, Alghuwainem S (2020) Transient search optimization: a new meta-heuristic optimization algorithm. Appl Intell 50(11):3926–3941
26. Rai S, Shastry KA, Pratap S, Kishore S, Mishra P, Sanjay HA (2021) Machine learning approach for student academic performance prediction. In Evolution in Computational Intelligence Springer, Singapore, 611–618
27. Raut AB, Nichat MAA (2017) Students performance prediction using decision tree. Int J Comput Intell Res 13(7):1735–1741
28. Rizvi S, Rienties B, Khoja SA (2019) The role of demographics in online learning; a decision tree based approach. Comput Educ 137:32–47
29. Sajja VR, Lakshmi PJ, Naik DSB, Kalluri HK (2021) Student performance monitoring system using decision tree classifier. In machine intelligence and soft computing, vol 2021. Springer, Singapore, pp 393–407
30. Shafi M, Mahboobe MR, Neyestani SE, Jafari M, Taghvaei V (2021) The quality improvement indicators of the curriculum at the technical and vocational higher education. Int J Instr 14(1):65–84
31. Shaheed K, Mao A, Qureshi I, Kumar M, Hussain S, Ullah I, Zhang X (2022) DS-CNN: a pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition. Expert Syst Appl 191:116288
32. Shaheed K, Mao A, Qureshi I, Abbas Q, Kumar M, Zhang X (2022) Finger-vein presentation attack detection using depthwise separable convolution neural network. Expert Syst Appl 198:116786
33. Shaheed K et al (2022) Recent Advancements in Finger Vein Recognition Technology: Methodology, Challenges and Opportunities. Inf Fus 79:84–109

34. Tripathi A, Yadav S, Rajan R (2019) Naive Bayes Classification Model for the Student Performance Prediction. In 2019 2nd international conference on intelligent computing. Instrum Control Technol (ICICICT) IEEE 1:1548–1553
35. Verma P, Sood SK, Kalra S (2017) Smart computing based student performance evaluation framework for engineering education. Comput Appl Eng Educ 25(6):977–991
36. Walia S et al (2021) Fusion of handcrafted and deep features for forgery detection in digital images. IEEE Access 9:99742–99755
37. Zeineddine H, Braendle U, Farah A (2021) Enhancing prediction of student success: automated machine learning approach. Comput Electr Eng 89:106903

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/365174053

Development of A Fully Faired Recumbent Bike using A Three-Piece Mold

Article *in* International Journal of Engineering and Technical Research · November 2022

DOI: 10.17577/IJERTV11IS100082

CITATIONS
0

READS
6

4 authors, including:

Hrishabh Jha
Delhi Technological University
2 PUBLICATIONS **1** CITATION

SEE PROFILE

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

Development of A Fully Faired Recumbent Bike using A Three-Piece Mold View project

All content following this page was uploaded by Hrishabh Jha on 06 November 2022.

The user has requested enhancement of the downloaded file.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 10, October-2022

Development of A Fully Faired Recumbent Bike using A Three-Piece Mold

Hrishabh Jha[1], Avhishek Jha[1], Kunal Nath[1], Rajesh Kumar[2]
[1] Department of Mechanical Engineering, Delhi Technological University, Delhi
[2] Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi, India

***Abstract*—Sustainable transportation is the ability to meet the mobility needs of a society without causing any damage to the environment and impairing the mobility needs of future generations. It comprises low and zero-emission, energy-efficient, affordable modes of transport, including electric and alternative-fuel vehicles. The project aims to create a fully faired recumbent to combat the energy crisis by providing a sustainable alternative to regular commuters over other conventional means of transportation. The primary goal is to develop an adaptable, technically feasible and economically successful HPV. The frame material of the bike is made of Chromoly Steel core, and it includes an adjustable Bottom Bracket Clamp to allow riders of different heights to ride comfortably. The flexibility to accommodate various riders makes the recumbent bike suitable as a commuter vehicle. The fairing is designed on SolidWorks and fabricated with Carbon Fiber using a 3-piece mold. Adherence to essential safety standards is assured through Roll-over Protection System Analysis, Structural Analysis, and Aerodynamic Analysis on ANSYS.**

Keywords—Sustainable Transportation; Recumbent; Velomobile; SolidWorks; ANSYS; Carbon Fiber; Roll-over Protection System

I. INTRODUCTION

Recumbents are human-powered vehicles, further enclosed with a fairing (aerodynamic shell) for improving aerodynamic performance and protection from weather and collisions. Any streamlined human powered vehicle consisting of a full fairing could be referred to as a velomobile [1,2].

The design goal of the vehicle was to create a state-of-the-art human powered vehicle, which can handle real life situations, enhancing vehicle safety and practicality, particularly as a form of everyday transportation and be a safer, fast, and lightweight vehicle. Diversification of recumbent geometries were studied as per the American Society of Mechanical Engineers (ASME) guidelines along with different types of bicycles and trikes, which helped determine a fully faired, low racer bike as the most suitable option to achieve the defined project goals.

The design/processes that were developed/adopted and used for vehicle:
a. Use of wooden jigs to determine the constraints upon which the frame and fairing were developed.
b. Use of blue foam for the box ribbing of the fairing.
c. Number of layers of carbon fiber to be used is decided upon the strength testing done in the previous year.
d. 3-piece mold technique was used.

The ASME HPVC 2019 constraints [3] were used and any measure to increase rider, team members and bystander's safety without compromising the performance, stability and ride quality of the vehicle guided Vehicle's Design.

TABLE 1. DESIGN SPECIFICATIONS FROM ASME CONSTRAINTS

	ASME Constraints	Project Goals
Safety	• Have a Roll-over Protection System (RPS) that will: ○ Produce less than 5.1 cm of total deformation of application of 2670 N of top load (At an angle 12° away from the front) ○ Produce less than 3.8 cm of total deformation of application of 1330 N side load. ○ No body contact and adequate abrasion resistance in the event of a fall and slide respectively. • Include safety harness that holds rider to vehicle during a crash • Include Head/Tail light, bell or horn, front/rear/side reflectors • Field of View of at least 180° • Free from sharp edges, open tubes, protruding screws or any other potential hazards	• The RPS at all points has a reasonable FOS (> 2.5) at all points and the von – misses stress is below the maximum tensile strength. • The safety attachment points stay intact in the event of emergency stop. • The steering system remains intact in the event of a crash • The horn is easily accessible. • No exposed wires and all batteries are located away from the rider. • Use commercially available seat harness
Performance	• Demonstrate stability at 5-8 km/hr for 30m (fast paced walking speed). • Brake from 25 to 0 km/hr in 6 m • Turn within an 8 m radius (26.2 ft) • Clear a speed bump with maximum height up to 5 cm. • Transport a parcel with dimensions 38x33x20 cm and a maximum mass of 5.5 kg.	• Total Weight less than 25kg • Top Speed of 90 kmph • Easy rider Ingress/Egress using Better Rider Comfort • Minimize the overall drag experienced by the fairing • Create a more efficient drivetrain for achieving higher top speed. • Maximum vehicle length less than 8.5 ft (2.6 m)
Rider	• Log a minimum of 30 minutes of vehicle riding experience before the event. • The rider must wear appropriate clothing and properly fitting helmets with fastened straps that CPSC Safety Standard for Bicycle Helmet (16 CFR Part 1203)	• Attend three workouts per weeks focused on improving stamina, strength and overall fitness. • Cycling sessions which simulate actual competition three times a week

II. METHOD

A House of Quality was done to determine factors to focus on. Drivetrain efficiency and Rider satisfaction were identified as the primary factors over other factors. This is decided based on the importance rating given to the metrics analogous to the project requirements. The other factors, which were of high priority, consisted of Field of view, ergonomics, and vehicle weight [4].

Product Design Specifications obtained for designing the vehicle obtained from the HOQ and ASME constraints are as follows. Based on the design criteria and Quality Function Deployment, the Pugh's concept selection technique is adopted [5].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 10, October-2022

Product Design Specifications obtained for designing the vehicle obtained from the HOQ and ASME constraints are as follows. Based on the design criteria and Quality Function Deployment, the Pugh's concept selection technique is adopted [5].

TABLE 2. PRODUCT DESIGN SPECIFICATIONS

Metric	Target Value
Coefficient of Drag	< 0.1
Field of View	225°
Turning Radius	< 5 m
Rider Satisfaction	9/10
Weight	30 kg
Cost	< INR 200,000
Rider Height Range	20 cm
Braking distance from 25 kmph speed	< 6 m
Low Speed Stability	Stable at 5 kmph for 30m

TABLE 3. RATING CRITERIA

Score	Quality
5	Excellent
4	Good
3	Average
2	Poor
1	Very Poor

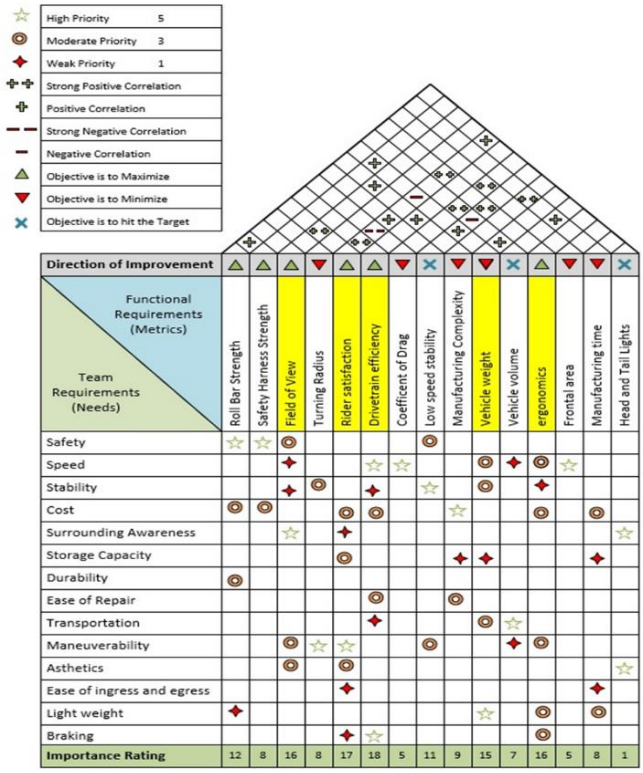


Fig. 1. House of Quality

A decision matrix was made with weighted average given to each category for critical design considerations.

TABLE 4. DECISION MATRIX FOR VEHICLE TYPE

Metric	Weightage (%)	Fully faired recumbent low racer	Partially Faired Low Racer	Trike	Upright Partially Faired	Tadpole (Recumbent)
Frame weight	14	4	3	3	5	2
Aerodynamics	17	5	4	3	1	3
Rider Comfort	15	4	4	4	3	4
Speed	15	5	4	3	2	2
Maneuverability	12	3	4	4	5	3
Prior Knowledge	13	5	4	3	5	2
Drivetrain Efficiency	14	4	3	3	5	3
Total	100	4.33	3.72	3.27	3.57	2.73

The vehicle configuration was decided after considered various possibilities for vehicle geometries and determining a score based on QFD corresponding to each category. A fully faired recumbent low racer configuration is concluded as the most effective based on all the above metrics.

A similar comparative study was done for the frame body using a variety of materials namely AISI 4130 Chromoly Steel, AISI 1020 Steel, Aluminium 6061, and Carbon Fiber. AISI 4130 Chromoly Steel was selected based on the decision matrix. The difference between total metric value of AISI 4130 and Carbon Fiber is insignificant; however, due to easy machining AISI 4130 Chromoly Steel is preferred.

TABLE 5. DECISION MATRIX FOR FRAME MATERIAL

Metric	Weightage (%)	Aluminum 6061	AISI 1020 Steel	AISI 4130 Chromoly Steel	Carbon Fiber
Strength/ Weight Ratio	18	4	1	4	5
Ease of Machining	21	3	4	4	3
Ease of Repair	12	3	4	4	3
Availability	9	4	5	3	3
Cost	7	3	5	3	2
Durability	11	3	4	4	3
Weight	13	4	1	4	5
Corrosion Resistance	4	4	3	4	5
Recyclability	5	4	4	3	3
Total	100	3.49	3.13	3.79	3.63

FWD has several advantages over RWD primarily shorter and more efficient drivetrain, lesser vibrations, lesser no. of idlers, compact drivetrain [6]. Although the major issue of "Heel Strike" is present in FWD, still it was considered because of other benefits. The issue of heel strike was tackled by changing the Bottom-bracket height and avoiding the use of derailleur hanger. The multichain drivetrain results in more gear ratio choices, but derailment and complexity of the drivetrain increases. Thus, a single chain drivetrain was selected.

The seat height was kept at 28 cm and seat back angle at 40 degrees to increase rider comfort. The seatback and headrest angles were decided based on rider inputs.

TABLE 6. DECISION MATRIX FOR DRIVETRAIN

Criteria	Weightage (%)	Single Chain	Multi Chain
Efficiency	40	5	3
Gear Reduction	30	1	2
Manufacturability	20	4	3
Cost	10	4	3
Total	100	3.5	2.7

TABLE 7. FRAME PARAMETERS

Parameter	Value
Front Wheel Diameter	0.497 m
Head Tube Angle	73.4°
Wheel Base	1.24 m
Trail	0.55 cm

The vehicle was designed on a Front Wheel Drive comprising of a single 60 teeth chainring and Shimano 8 speed cassette shifting mechanism [7,8]. A threaded Bottom bracket mechanism with an aluminium spider crank with crank arm length 165 mm was installed on the frame. Two idlers were used for proper routing and idler sizing was done to provide tension relief to the second idler which was directly routed to

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 10, October-2022

the derailleur. The drivetrain was selected as per market availability. Idler dimensions were optimized to prevent chain derailment.

Fig. 2. Recumbent bike frame

A fairing provides the added advantage of protecting the rider from dust, dirt, and other unwanted obstacles. High emphasis was given on the aerodynamic stability of the vehicle. The designing process involved using a Java Applet named JavaFoil, on which NPL ECH and NACA 6 series were considered for the sections of the vehicle [9,10]. The ECH 2866 met the needs of the team dimensionally for the top section of the fairing and ECH 4466 for the side section. These sections were further modified to achieve close to ideal fairing design.

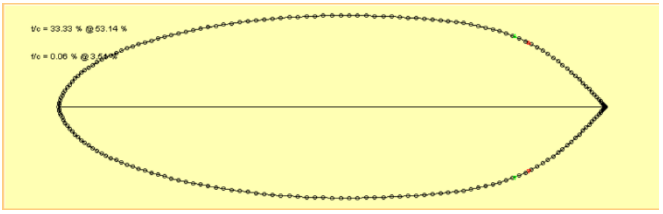


Fig. 3. ECH 2866 NACA 6 series design

A fully faired carbon fiber fairing using two female molds of glass fiber was made in which carbon fiber was laid up. The upper mold was then divided into two parts for partition of door and to reduce the seam length which led to a three-piece mold structure. This reduced the number of cuts for the door from four to one. CBF (car body filler) was used at places with severe depressions and the molds were sanded with different grit sizes of sandpaper ranging from 240 to 1200 to get the perfect surface finish to the molds for Carbon Fiber Layup. The fairing of vehicle was made from Carbon Fiber using vacuum bagging and conventional hand layup techniques.

Fig. 4. Fairing Mold

Fig. 5. Partition for door

Rib placement was given a special emphasis to form a structural member throughout the vehicle [11]. Ribs made of blue foam were strategically placed along the regions which must be cut such as the front wheel and windshield and along the RPS, where the rider's movement is not obstructed, and where the rider is to be seated. The fairing's removable top hatch was mainly for the ingress/egress of the rider. Even after removal of the top hatch, the rider was protected by the fairing RPS in the event of a crash. The rear hatch houses the rear braking, electronics components, and storage area.

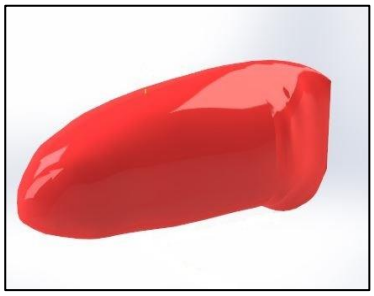


Fig. 6. Fairing CAD Design

III. ANALYSIS

A. Roll-over Protection System Analysis

Considering the safety of the rider to be of utmost priority, the vehicle incorporates a strategically designed roll bar to prevent the rider from direct contact with the road surface and to reduce the impact of any accidental collision. A four-point seat belt was used with attachment points to the side rod and the bottom of the RPS was chosen to constrain the rider's motion during an impact. The finite element model was developed using shell elements in the ANSYS 16.0 Composite Pre-Post (ACP) system [12]. Due to the rollover protection system (RPS) being also supported by the fairing, the Analysis of the RPS was performed on vehicle's fairing.

A top load of 2670 N was applied to the top of the roll bar, directed downwards at an angle of 12 degrees from the vertical towards the rear of the vehicle keeping the seat belt attachment points fixed and subject to a reactant force. When the vehicle was subjected to a top load, the load was distributed from the ground to the fairing (Integrated RPS), then to the side rods and the bottom of the frame (both containing the seat belt attachment point). Thereby, protecting body contact from the ground. There was no indication of permanent deformation, fracture, or delamination. The maximum elastic deformation as per the Analysis carried out was 1.0361 cm which was less than the maximum allowable deformation of 5.1 cm, meeting ASME specification. This gave a FOS value of 4.47.

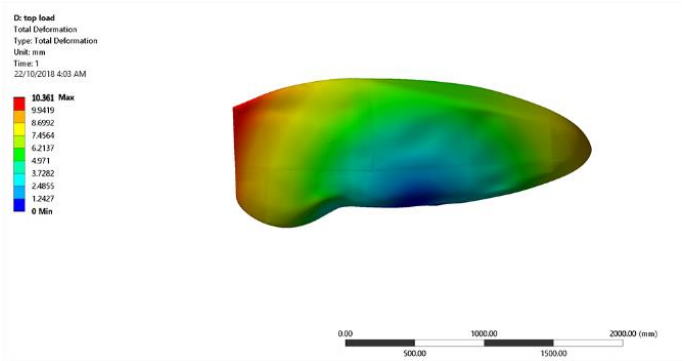


Fig. 7. Total Deformation under action of Top Load on RPS

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 10, October-2022

A side load of 1330 N was applied horizontally to the side of the roll bar at shoulder height keeping the seat belt attachment points fixed and subject to a reactant force. When the vehicle was subjected to a top load, the load was distributed from the ground to the fairing (Integrated RPS), then to the side rods and the bottom of the frame (both containing the safety harness attachment point). Thereby, protecting body contact from the ground. There was no indication of permanent deformation, fracture, or delamination. The maximum elastic deformation as per the Analysis carried out was 0.79356 cm which was less than the maximum allowable deformation of 3.8 cm, meeting ASME specification. This gave a FOS value of 4.46.

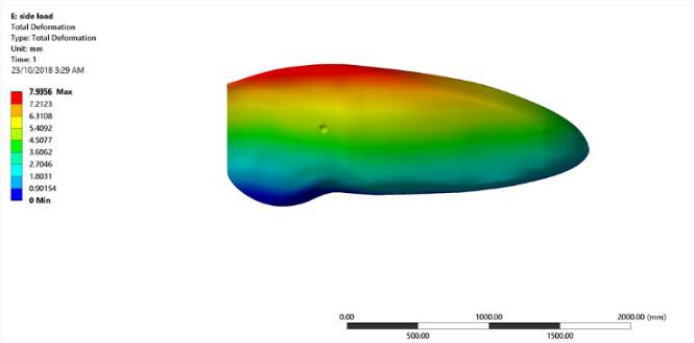


Fig. 8. Total Deformation under action of Side Load on RPS

B. Structural Analysis

Structural Analysis on the frame using ANSYS were carried out measuring the ability of the frame to withstand various stress scenarios which the vehicle may undergo during its normal operation. All the forces were assumed to be below the elastic limit of the material.

Various stresses considered on the bottom bracket include scenarios where the chain experiences large tension force, where there was sudden obstruction leading to considerable friction during the rotation of the pedals, etc. These situations were modelled on the AISI 4130 Chromoly Steel. An automated mesh was developed with an average orthogonal quality of 0.867. A Remote force of 600 N was applied along the Z axis at the (92,0,0) mm from the reference coordinate on the centre of bottom bracket as shown below. The maximum (Von-Mises) stress of 41.089 MPa was observed in the upper section of the Bottom Bracket Coupler showing that there was a FOS of 6.08.

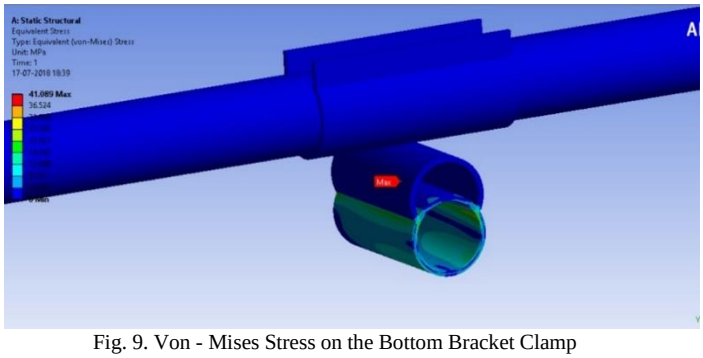


Fig. 9. Von - Mises Stress on the Bottom Bracket Clamp

A significant step to integrate a human-powered vehicle was to ensure that rider comfort was maintained even in rough terrains. The velomobile must be capable of withstanding minor obstacles such as road bumps (with heights up to 9 cm), gravel, sinks and small puddles. An automated mesh was developed with an average orthogonal quality of 0.812. A compressive force of 2802 N along head tube's axis and a bending force of 1201 N at coordinates (-137.62, -408.8, 0) along x-axis considering centre of the head tube as the reference point were applied on the head tube and the deflection produced was analysed in the event of a road bump. The frame displayed a total deformation of 0.144 mm and had a FOS greater than 1. The loading conditions were done according to ASTM F2273-11.

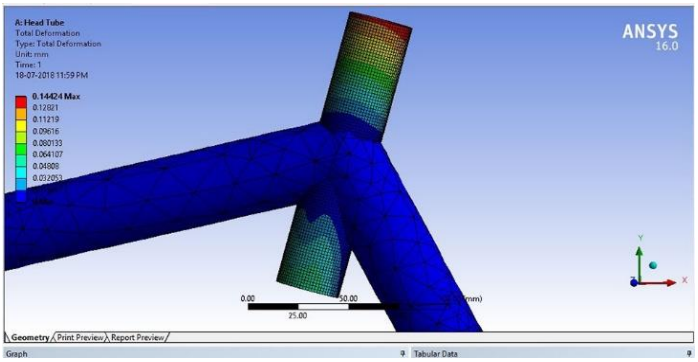


Fig. 10. Total Deformation after Equivalent Stress on the Head Tube

A load of 2500 N was applied towards the front side of the vehicle at the attachment point. The force 2500 N was designed such that the vehicle (weighing ~ 100kgs with rider) travelling at 20 m/s can be stopped within 0.2 seconds. The frame was modelled on ANSYS Space Claim, and the results were analysed on the Static Structural. An automated mesh has been developed with an average orthogonal quality of 0.91. The stress on the seat attachment point was observed to be 149.19 MPa and the total deformation of seat harness 0.9592 mm.

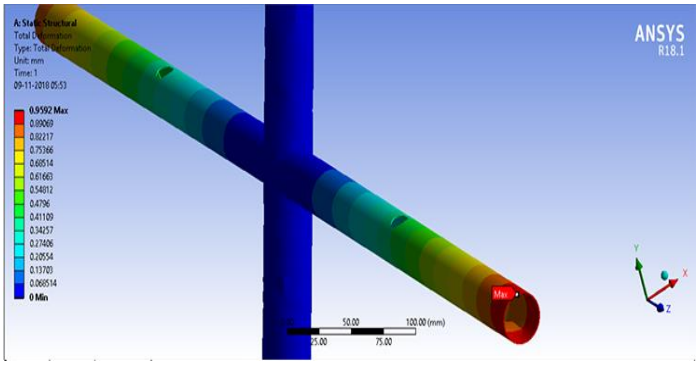


Fig. 11. Total Deformation after Equivalent Stress on the Seat Harness

C. Aerodynamic Analysis

The front wind analysis of the draft fairing was done in Ansys Fluent, and it yielded a drag coefficient of 0.119 (First Iteration) when the maximum speed was assumed to be 20 m/s (72 kmph). The average wind speed was considered as 4.167 m/s (15kmph). For CFD, high emphasis was given on the mesh, every mesh metric was checked so that any inaccuracy.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 10, October-2022

RANS based K-Omega Turbulence Model in Steady State, with all y+ wall treatment was chosen as it was not very memory-intensive and yields good convergence even with complex geometries [13]. The model also proved to be stable under high relaxation factors which helped in observing faster convergence without losses in accuracy. After visualization of the flow, an early flow separation was found. To tackle the issue, a guide curve along the rear end of the fairing was incorporated to guide the flow and letting the flow stayed attached for a longer period. The incorporation of the guide curve proved detrimental towards the fairing. The drag coefficient reduced significantly. In the final iteration, the drag coefficient came out to be 0.087 which was much lower than the 1st iteration.

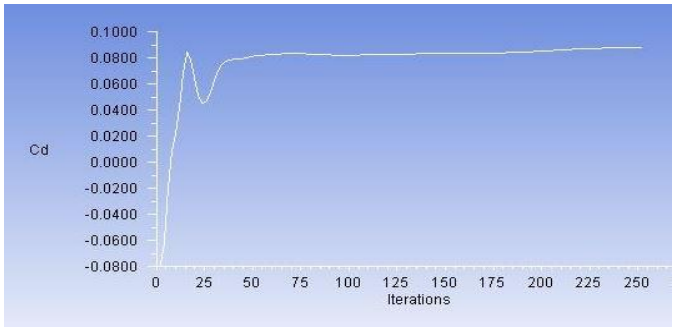


Fig. 12. Cd vs iteration curve

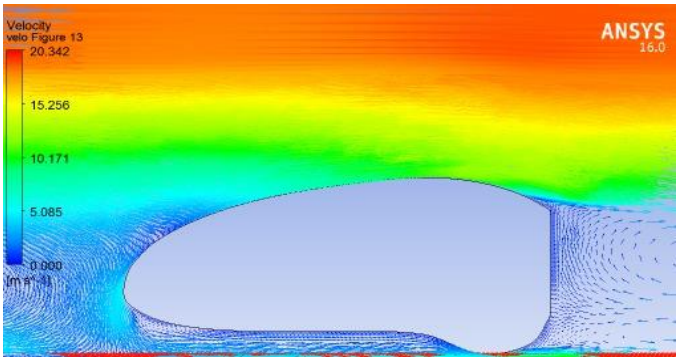


Fig. 13. Velocity contour in a 2.5D Analysis

For observing the large eddy structures in the wake region, the Spalart – Allmaras Detached Eddy Simulation was used with Implicit Unsteady being state of time [14]. Though more computationally intensive it helped understand the effect of cross wind which the vehicle will encounter. Changes were made to ensure that the pressure change across the fairing was significantly lower and the drag force came out to be 9.78 N.

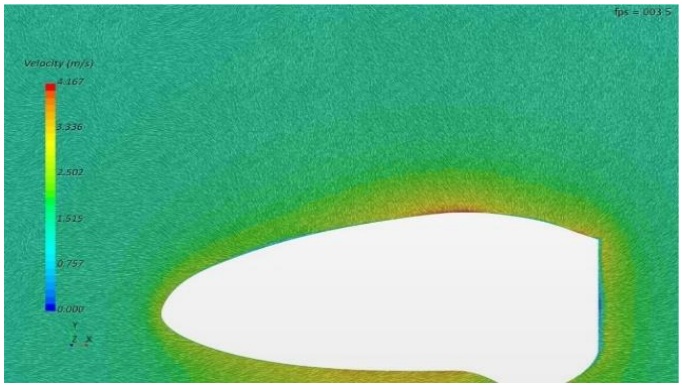


Fig. 14. Velocity Contour around a plane

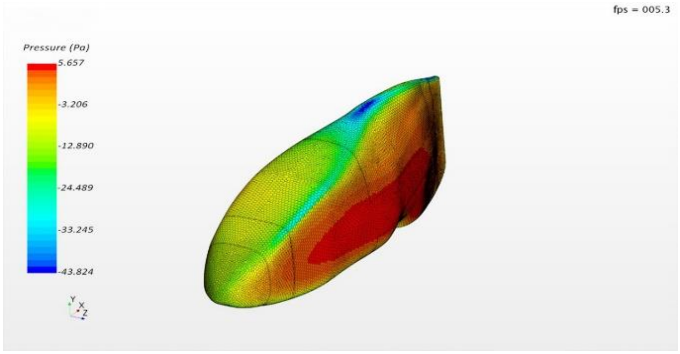


Fig. 15. Pressure distribution w.r.t. atmosphere Analysis

IV. TESTING

A. Heel Strike Testing

Metal prototype model of the preliminary design was made using AISI 1020 Steel to cross check the initial measurements obtained using the jig. Off the shelf (stock) components were used for making the prototype facilitating flexibility for optimization. The mock-up facilitated idler positioning and seat ergonomics, proper steering position and handlebar geometry. To examine the heel and toe curves of the tallest rider for the fairing measurements, the vehicle was aligned with the movement of the rider's leg parallel to the white board and the curve was traced based on the movement of the leg. Both the curves were measured in accordance with the Bottom Bracket of the frame. Both the curves were measured wrt different point of reference (i.e., BB location) depending on the orientation of the board with the vehicle. The heel min and toe max curves were traced and were considered during the fabrication of the fairing for considering the clearance values at the crucial points in the fairing.

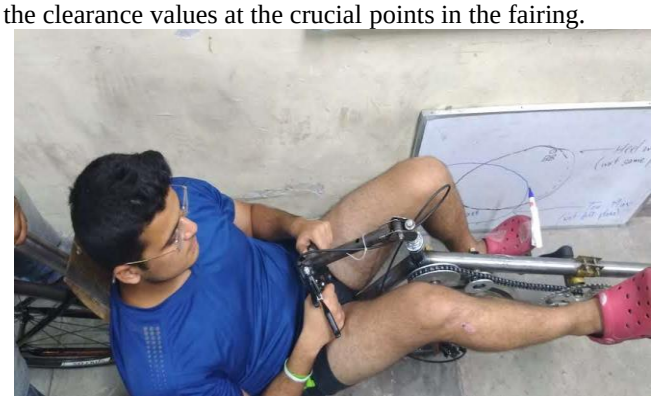


Fig. 16. Measurement of Heel and Toe Curves

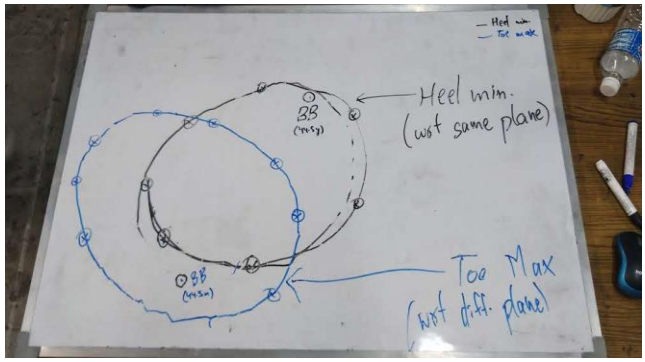


Fig. 17. Toe Max and Heel Min Curves

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 10, October-2022

B. Fairing Dimensions Testing

To ensure that the fairing can accommodate different riders comfortably, the tallest and shortest riders were made to sit on the bike, and three view photographs were taken. These photos were imported to SolidWorks on the front, top and right planes to give a basic layout of the Vehicle frame with the rider. All the measurements of the crucial points such as pedalling radius, height of rider with helmet, knees and toes, width of shoulders, etc. were plotted on planes whose distances were measured from the centre of front wheel.

The profiles were created, and the fairing was then lofted using these profiles. There was extensive use of Siemens NX 11.0 using its human modelling feature to check clearance at various points in the initial and final verification of the fairing model developed. The fairing was designed according to the vast range of height of riders so that the vehicle can accommodate different riders comfortably.

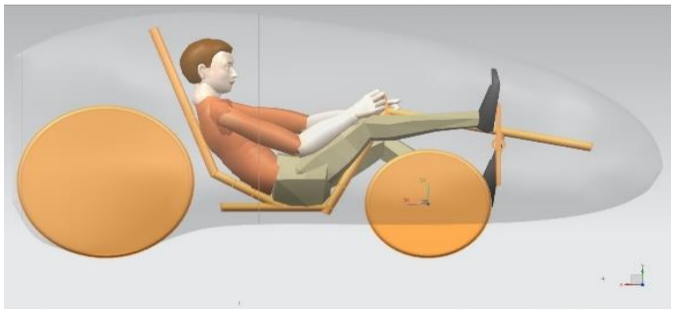


Fig. 18. Human Modelling on NX 11.0

Fig. 19. Frame Dimensions Testing

C. Windshield Material Testing

The windshield area was maximized to increase the field of view. The material was selected based on mold ability and mist formation as the criteria [15]. The molding was done using hot air gun over the glass fiber mold and the material was allowed to acquire the shape of the mold. Acrylic sheets were brittle and could shatter on impact and hence were unsafe to use considering rider's safety. Polycarbonate (Lexan) sheets with thickness 2 mm were selected as ideal material for the windshield.

TABLE 8. OBSERVATIONS OF WINDSHIELD MATERIALS

Material	Thickness	Observation
Polycarbonate (Lexan)	2 mm	Low Mold ability, No Mist
Polycarbonate (Lexan)	4 mm	Poor Mold Ability, Slight Mist
Acrylic (Plexiglass)	2 mm	Good Mold Ability, Slight Mist
Acrylic (Plexiglass)	4 mm	Low Mold Ability, Slight Mist

V. CONCLUSION

The vehicle easily met all the constraints set during the start of the project. Vehicle met the low-speed stability testing (< 5kmph), braking distance testing (25kmph to 0kmph in 6m), turning radius (<8m) requirements. More standard components are recommended instead of customized parts to improve replaceability and decrease cost. Results of the analysis done could also be verified through yarn tuft testing. The project was successful in creating a vehicle in the field of safe and effective Sustainable Transport. Possessing extensive adaptability to challenging operating conditions coupled with enhanced manoeuvrability and robustness makes this innovation a commercially viable alternative.

TABLE 9. COMPARISON OF DESIGN AND ACTUAL VALUES

Metric	Target Value	Estimated Value
Coefficient of Drag	<0.1	0.087
Field of View	225°	In Progress
Turning Radius	<5m	In Progress
Rider Satisfaction	9/10	10/10
Weight	25kg	27kg
Cost	<200,000	1,76,000
Rider Height Range	20cm	18cm
Braking distance from 25kmph speed	< 6m	In Progress
Low Speed Stability	Stable at 5kmph for 30m	In Progress

REFERENCES

[1] M. Ferrari et al., "Development of a hybrid human-electric propulsion system for a velomobile," Eighth International Conference and Exhibition on Ecological Vehicles and Renewable Energies, Monaco, pp. 1-8, 2013.

[2] F. Walle, The velomobile as a vehicle for more sustainable transportation, Royal Institute of Technology, Stockholm, Sweden, 2004.

[3] C. Priyadarsini, B. Khanna, and C. Raviteja, "Design and Analysis of Human Powered Vehicle," International Journal of Management, Technology and Engineering, vol. 9, pp. 3104-3113, January 2019.

[4] A. Olewnik, and K. Lewis, "Limitations of the House of Quality to provide quantitative design information," International Journal of Quality & Reliability Management, vol. 25, pp. 125-146, January 2008.

[5] G. Okudan, and S. Tauhid, "Concept selection methods – a literature review from 1980 to 2008," Int. J. Design Engineering, Vol. 1, pp. 243-277, 2008.

[6] M. Behera, and B. Behera, "Design Failure Modes and Effects Analysis of a Human Powered Recumbent Vehicle," International Journal of Engineering Research & Technology (IJERT), Vol. 5, pp. 27-32, April-2016.

[7] W. Patterson, The Chronicles of the Lords of the Chainring, California Polytechnic State University, 1998.

[8] K. Singh, M. Iyengar, and N. Iyengar, "Design and Analysis of Human Powered Hybrid Vehicle," International Journal of Mechanical Engineering and Technology (IJMET), Vol. 9, pp. 594-605, April 2018.

[9] P. Migliore, "Comparison of NACA 6-Series and 4-Digit Airfoils for Darrieus Wind Turbines," J. Energy, vol. 7, pp. 291-292, August 1983.

[10] R. Selvaraj et al., "Optimisation of swept angles for airfoil NACA 6-series," Int. J. Computer Aided Engineering and Tech., Vol. 9, pp. 229-240, 2017.

[11] R. Jones, Mechanics of Composite Materials, Virginia Polytechnic Institute and State University, 1998.

[12] F. Berg, R. Behling, and M. Helbig, "Rollover crash tests," International Journal of Crashworthiness, vol. 7, pp. 487-498, 2002.

[13] F. Menter, "Influence of Freestream Values on k omega Turbulence Model Predictions." AIAA Journal, Vol. 30, pp. 1657-1659, 2013.

[14] P. Spalart, Detached-Eddy Simulation, Annual Review of Fluid Mechanics, pp. 181-202, January 2009.

[15] Y. Peng et al., "Finite element modeling of crash test behavior for windshield laminated glass," Int. Journal of Impact Engineering, vol. 57, pp. 27-35, 2013.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

View publication stats

J. Inst. Eng. India Ser. B
https://doi.org/10.1007/s40031-022-00807-6

ORIGINAL CONTRIBUTION

DFIG-Based Wind Plant Coupled Series Compensated Transmission Line: Modeling and SSR Stability Analysis

Neevatika Verma[1] · **Narendra Kumar**[1]

Received: 3 January 2022 / Accepted: 12 September 2022
© The Institution of Engineers (India) 2022

Abstract Wind energy production is expanding and has become a terrific source of renewable energy. The installed capacity of wind energy has nearly doubled every 3 years during the past two decades of the twenty-first century. Wind farms typically are not close to the grid, so a series compensated transmission lines are required to improve the transfer capability of extant power transmission lines economically. Sub-synchronous resonance (SSR), which can cause torsional interactions that result in T-G shaft fatigue and a related protective mechanism, is always a risk with series compensation. A modified IEEE first benchmark model is employed for SSR analysis in this paper. Power transmission via a series compensated transmission line from a wind generator with a doubly fed induction generator (DFIG) base has been investigated. In this paper, the dynamics of the grid side converter (GSC) as well as rotor side converter (RSC) has utilized to model the DFIG. As a result, system dynamics may be greatly impacted. However, in previous paper for RSC and GSC, a constant voltage source model has been taken into consideration. Consequently, for some important parameters, these models are unable to replicate the dynamic effect of SSR. More accurate dynamics are reflected in the analysis by using GSC and RSC modeling. The effect of SSR under various operating conditions, such as wind speed changes and altering the series compensation, has been investigated using the eigenvalue analysis. Time-domain simulations have been conducted using MATLAB to verify the analysis.

Keywords Eigenvalue · Series compensation · Sub-synchronous resonance (SSR)

Abbreviations

SSR	Sub-synchronous resonance
DFIG	Doubly fed induction generator
GSC	Grid side converter
RSC	Rotor side converter
WPP	Wind power plant

List of Symbols

ρ	Air density (kgm^{-3})
V_{wi}	Wind speed (m/s)
R	Rotor radius (m) of wind turbine
A	Blade sweep area (m^2)
C_f	Design constant of wind turbine blades
λ	Tip speed ratio
Ω_m	Mechanical angular velocity
ω_{t}	Shaft speed of wind turbine
ω_r	Shaft speed of DFIG
T_g	Internal torque of the system
T_m	Electromagnetic torque of the generator
D_t	Damping coefficient of wind turbine
H_t	Inertia constant of wind turbine
H_g	Inertia constant of wind generator
D_{tg}	Damping coefficient of wind generator
K_{tg}	Shaft stiffness of the coupling between two masses
ω_e	Synchronous reference speed
$\psi_{\mathrm{qm}}, \psi_{\mathrm{dm}}$	Quadrature and direct axes air gap flux linkage
$i_{\mathrm{qr}}, i_{\mathrm{dr}}$	Quadrature and direct axes DFIG rotor current component
$i_{\mathrm{qs}}, i_{\mathrm{ds}}$	Quadrature and direct axes DFIG stator current component

✉ Neevatika Verma
nivetikaverma2311@gmail.com

[1] Department of Electrical Engineering, Delhi Technological University, New Delhi 110042, India

Published online: 28 October 2022

X_M	Magnetizing reactance
i_{0s}, i_{0r}	Stator and rotor zero sequence current components
$v_{\mathrm{qs}}, v_{\mathrm{ds}}$	Quadrature and direct axes components of the terminal bus voltage
$v_{\mathrm{qr}}, v_{\mathrm{dr}}$	Quadrature and direct axes components of the DFIG rotor voltage
v_{0s}, v_{0r}	Stator and rotor zero sequence voltage components
X_r, X_s	Rotor reactance and stator reactance
ω_b	Base Speed (2.pi.60 rad/s)
R_r	Rotor resistance
$X_{\mathrm{ls}}, X_{\mathrm{lr}}$	Stator and rotor leakage reactance
P_R	RSC power
P_G	GSC power
$v_{\mathrm{qg}}, v_{\mathrm{dg}}$	Quadrature and direct axes components of the grid side voltage
$i_{\mathrm{qg}}, i_{\mathrm{dg}}$	Quadrature and direct axes components of the grid side current
ω_s	Angular synchronous frequency
V_{dc}	Voltage of DC link
C_{dc}	Capacitor of DC link
R_l	Resistance of transmission line
i_l	Transmission line current
v_c	Capacitor voltage
v_s	Generator bus voltage
V_B	Infinite bus voltage
X_c	Series capacitor reactance
$V_{\mathrm{Bq}}, V_{\mathrm{Bd}}$	Quadrature and direct axes of the infinite bus voltage
X_l	Total reactance of transmission line
$i_{\mathrm{dl}}, i_{\mathrm{ql}}$	Components of transmission line current in direct and quadrature axes
X_{tg}	Reactance at gride side converter
$R_{\mathrm{RSC}}, L_{\mathrm{RSC}}$	Resistance and inductance at RSC
$R_{\mathrm{GSC}}, L_{\mathrm{GSC}}$	Resistance and inductance at GSC
Q_s	Delivering reactive power through a DFIG stator

Introduction

As renewable energy is the current requirement to meet the global power need without much impact on the environment. Wind power plays a significant role in fulfilling the power requirement. To accommodate the additional power, transmission infrastructural changes are frequently required for wind farms located far from load centers. From the economic considerations, series capacitor compensation of existing lines is typically recommended, over constructing new lines, to boost power transmission capacity and stability margin [1, 2]. On the other hand, fixed series capacitor compensation is a source of SSR phenomenon not only in new power plants but also in existing generating power systems and power electronic equipment [3, 4]. The primary source of the SSR phenomenon is the electrical resonance frequency, which is smaller in amplitude to the natural system frequency. It usually occurs if the natural mechanical frequency of the turbine matches with the electrical resonance frequency—the SSR oscillation results in the slip voltage of SSR frequency at the rotor circuit [5]. The reason for the shaft failure is due to the oscillating electromagnetic torque, which consequently results in the failure of the shaft [6]. This incident invites serious research on the SSR phenomenon in DFIGs and series compensated transmission lines [7–9].

The SSR phenomena were first discussed in 1937, and it resurfaced in the 1970s after two shaft collapses at the Mohave site received attention [10]. A decade ago, the potential relationship between wind farms and series capacitors has been noticed. Due to increasing levels of wind power, SSR occurrences have been observed in series-compensated DFIG-based wind power plants (WPP) in the south of Texas, the southwest of Minnesota, and the north of China (Hebei province) [11, 12]. Following the incidence, extensive research into the SSR problem among DFIGs and series-compensated power systems was conducted. According to reference [5], DFIGs are more susceptible to SSR in series compensated systems than fixed-speed wind turbine generators and wind generators connected via back-to-back converters.

Several previous research works discuss detailed mathematical modeling of a DFIG-based WPP [3, 6–10, 13].

However, the RSC, GSC dynamics have not been considered for the modeling and system analysis. The RSC and GSC have been considered to be a constant voltage source model [7, 14–18]. and a constant DC connection voltage among the RSC and the GSC. The RSC, as well as GSC, must be simulated and included in the system modeling for internal dynamic analysis as they can introduce significant impact in the overall system dynamics. The dynamics of DFIG, RSC and GSC can be included in the stability analysis to improve its accuracy.

Now, because this modeling being more accurate, the authors have included the dynamics of RSC and GSC for overall system modeling in this paper. By means of the small signal stability technique, the dynamic behavior of the system is assessed [19]. A detailed state-space mathematical model of the study system has been carried out for SSR analysis since SSR is a small signal stability issue, the system requires a state-space model. The eigenvalue analysis technique has been used to analysis the stability of the system and to obtain the various oscillation modes of the system. The eigenvalue approach is a popular and straightforward method of stability analysis. In eigenvalue analysis, the imaginary portion of the eigenvalues corresponds to the frequencies, and the real portion of the eigenvalues displays the stability of various oscillation modes. The frequency of various modes as well as stability change as a result of operating variables like wind speed, including changes in series compensation level. The significant information regarding the

damping of each mode and its oscillatory frequency can be achieved from the above analysis. The eigenvalues enable to know the various modes of oscillations available in the complex electrical system.

System Modeling

Instead of a synchronous generator combined with a series compensated transmission lines, the IEEE first benchmark model has been updated to investigate the SSR phenomena in DFIG-based wind energy conversion systems, as illustrated in Fig. 1 [7, 20]. A 161 kV series compensated long transmission line connects an infinite bus to a 100 MW DFIG-based WPP in this study. The DFIG has a 575 V rated voltage and a 60 Hz frequency. The 67 wind turbines that make up the 100-megawatt wind farm each having a 1.5 MW capacity. In the Appendix, parameters for both the aggregated model and the single model are given. The examined system depicted in Fig. 1 must be constructed using state-space equations, as will be discussed below, in order to investigate the effect of the SSR phenomenon on system stability.

1) *Modeling of wind turbine* A wind turbine's mechanical output torque is proportional to the wind speed and can be expressed as [14]:

$$Tm = \frac{\rho RAC_p V_{\text{wi}}^2}{2\lambda} \tag{1}$$

C_p denotes the blade power coefficients, which is determined by the blade pitch angle 'θ' and tip speed ratio as follow:

$$C_p = 0.5\left(\frac{RC_f}{\lambda} - 0.022 * \theta - 2.0\right)e^{-0.255\left(\frac{RC_f}{\lambda}\right)} \tag{2}$$

The tip speed ratio is expressed as:

$$\lambda = \frac{\Omega_m R}{V_{\text{wi}}} \tag{3}$$

Figure 2 depicts the link between wind speed, mechanical power and rotor speed, all these values in per unit system, for a 1.5 MW wind turbine when the pitch angle 'θ' is zero with 200ft rotor radius is assumed.

It can be developed a lookup table as shown in Table 1 that lists the various wind speeds and the rotor speed, maximum mechanical power, and shaft torque. It should be emphasized that in the small-signal analysis technique, Table 1 depicts the wind turbine model. This Table, by means of including a torque control loop, can obtain the desired optimum rotor speed.

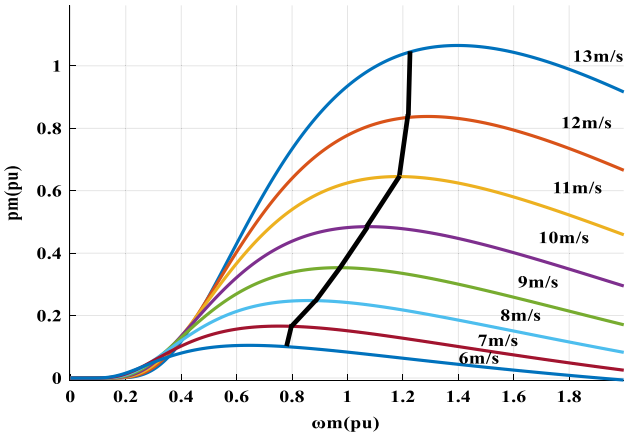


Fig. 2 Relationship between wind power wind speed and rotor speed

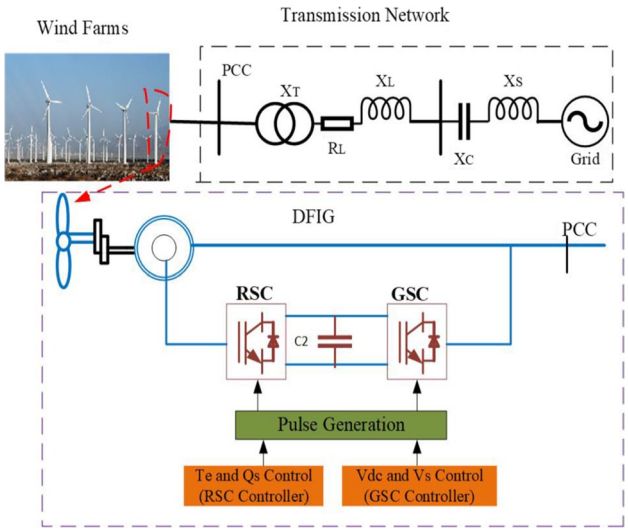


Fig. 1 IEEE first benchmark model of DFIG-based WPP

2) *Modeling of Shaft System* In power system analysis, the two-mass system frequently represents torsional dynamics. The first mass is represented by wind turbine shaft, which has a low speed, and a DFIG high-speed shaft system represents the second mass system; and a state-space representation of a two-mass system is as follows [14, 21]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\omega_t\\ \omega_r\\ T_g\end{pmatrix} = \begin{pmatrix}\frac{-D_t-D_{\text{tg}}}{2H_t} & \frac{D_{\text{tg}}}{2H_t} & \frac{-1}{2H_t}\\ \frac{D_{\text{tg}}}{2H_g} & \frac{-D_g-D_{\text{tg}}}{2H_g} & \frac{-1}{2H_g}\\ K_{\text{tg}}\omega_g & -K_{\text{tg}}\omega_e & 0\end{pmatrix}\begin{pmatrix}\omega_t\\ \omega_r\\ T_g\end{pmatrix} + \begin{pmatrix}\frac{T_m}{2H_t}\\ \frac{-T_e}{2H_t}\\ 0\end{pmatrix} \tag{4}$$

ω_t, ω_r and T_g are the state variables. T_m and T_e are the input side of the shaft system model. The constant of a two-mass model system is $D_t, H_t, H_g, D_{\text{tg}}$ and K_{tg}. The

optimal wind turbine torque T_m is obtained from Table 1 at any required wind speed, whereas T_e it could be written as air gap flux leakage and currents of the DFIG is expressed:

$$T_e = \psi_{\mathrm{qM}} i_{\mathrm{dr}} - \psi_{\mathrm{dM}} i_{\mathrm{qr}} \tag{5}$$

The air gap flux linkage is obtained as follows:

$$\psi_{\mathrm{qM}} = X_M \left(i_{\mathrm{qs}} + i_{\mathrm{qr}}\right) \tag{6}$$

$$\psi_{\mathrm{dM}} = X_M \left(i_{\mathrm{ds}} + i_{\mathrm{dr}}\right) \tag{7}$$

3) *Modeling of DFIG*

The modeling of DFIG with RSC, a dynamic model of the sixth order, is used. The model is characterized as [22]**:**

$$\dot{X} = A_D X_D + B_D U_D \tag{8}$$

where the state variables are denoted by X_D, and U_D denotes the inputs of DFIG, B_D denotes the constants of electrical parameter, and A_D represents the constants of state variables.

$$X_D = \left[i_{\mathrm{qs}}; i_{\mathrm{ds}}; i_{0s}; i_{\mathrm{qr}}; i_{\mathrm{dr}}; i_{or}\right] \tag{9}$$

$$U_D = \left[v_{\mathrm{qs}}; v_{\mathrm{ds}}; v_{0s}; v_{\mathrm{qr}}; v_{\mathrm{dr}}; v_{or}\right] \tag{10}$$

The parameters i_{0s}, i_{0r}, v_{0s} and v_{0r} are equal to zero because the system is balanced.

A_D and B_D are expressed as:

$$A_D = -B_D \begin{pmatrix} R_S & \frac{\omega_S}{\omega_b} X_S & 0 & 0 & \frac{\omega_S}{\omega_b} X_M & 0 \\ -\frac{\omega_S}{\omega_b} X_S & R_S & 0 & -\frac{\omega_S}{\omega_b} X_M & 0 & 0 \\ 0 & 0 & R_S & 0 & 0 & 0 \\ 0 & \frac{\omega_S - \omega_r}{\omega_b} X_M & 0 & R_r & \frac{\omega_S - \omega_r}{\omega_b} X_r & 0 \\ -\frac{\omega_S - \omega_r}{\omega_b} X_M & 0 & 0 & -\frac{\omega_S - \omega_r}{\omega_b} X_r & R_r & 0 \\ 0 & 0 & 0 & 0 & 0 & R_r \end{pmatrix}$$

$$B_D = -\omega_b \begin{pmatrix} X_S & 0 & 0 & X_M & 0 & 0 \\ 0 & X_S & 0 & 0 & X_M & 0 \\ 0 & 0 & X_{lS} & 0 & 0 & 0 \\ X_M & 0 & 0 & X_r & 0 & 0 \\ 0 & X_M & 0 & 0 & X_r & 0 \\ 0 & 0 & 0 & 0 & 0 & X_{lr} \end{pmatrix}^{-1} \tag{11}$$

Except for ω_b, ω_r and ω_s, which are in rad/sec, all parameters and variables are in per-unit.

4) *Modeling of DC link* A first-order model describes the dynamics of the capacitor in the dc link between the GSC and the RSC [22].

$$C_{\mathrm{dc}} V_{\mathrm{dc}} \frac{\mathrm{d}V_{\mathrm{dc}}}{\mathrm{d}t} = P_R - P_G \tag{12}$$

$$\mathrm{P}_R = \frac{1}{2}\left(v_{\mathrm{qr}} i_{\mathrm{qr}} + v_{\mathrm{dr}} i_{\mathrm{dr}}\right) \tag{13}$$

$$\mathrm{P}_G = \frac{1}{2}\left(v_{\mathrm{qg}} i_{\mathrm{qg}} + v_{\mathrm{dg}} i_{\mathrm{dg}}\right) \tag{14}$$

V_{dc} is the actual value, whereas $v_{\mathrm{qr}}, i_{\mathrm{qr}}, v_{\mathrm{dr}}, i_{\mathrm{dr}}, v_{\mathrm{qg}}, i_{\mathrm{qg}}, v_{\mathrm{dg}}, i_{\mathrm{dg}}$ are in per-unit system.

5) *Series Compensated Transmission line modeling* Generally, the dynamics of the series compensated model are ignored, but for the SSR analysis, it is essential. The transmission line can be electrically depicted as lumped series RLC circuit and adopts the synchronous reference frame to modeling the series compensated transmission line. Per phase dynamics equations are expressed as [7]:

$$R_l . i_l + L \frac{\mathrm{d}i_l}{dt} + v_c = v_s - V_B \tag{15}$$

$$C \frac{\mathrm{d}v_c}{\mathrm{d}t} = i_l \tag{16}$$

The series compensated transmission's line state-space representation is expressed as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} i_{\mathrm{ql}} \\ i_{\mathrm{dl}} \\ v_{\mathrm{cq}} \\ v_{\mathrm{cd}} \end{pmatrix} = \begin{pmatrix} -\frac{R_l}{X_l} & -\omega_e & -\frac{1}{X_l} & 0 \\ -\omega_e & -\frac{R_l}{X_l} & 0 & -\frac{1}{X_l} \\ X_c & 0 & 0 & -\omega_e \\ 0 & X_c & \omega_e & 0 \end{pmatrix} \begin{pmatrix} i_{\mathrm{ql}} \\ i_{\mathrm{dl}} \\ v_{\mathrm{cq}} \\ v_{\mathrm{cd}} \end{pmatrix} + \omega_b \begin{pmatrix} \frac{v_{\mathrm{qs}} - V_{\mathrm{Bq}}}{X_l} \\ \frac{v_{\mathrm{ds}} - V_{\mathrm{Bd}}}{X_l} \\ 0 \\ 0 \end{pmatrix} \tag{17}$$

The synchronous reference speed and the base speed both are 1pu or 377 rad/s.

6) *Modeling of DFIG converter* This study aims to look at both RSC and GSC converter controllers. In previous research, RSC and GSC converter controllers are considered constant voltage sources [7, 18, 22], while the dynamics of RSC and GSC may cause an enormous

Table 1 Mechanical power, torque, rotor speed (Pu), and wind speed

V_{wi}	7 m/s	8 m/s	9 m/s	10 m/s	11 m/s	12 m/s
ω_m	0.8	0.88	0.97	1.06	1.18	1.2
P_m	0.18	0.28	0.39	0.47	0.64	0.84
T_m	0.22	0.32	0.4	0.46	0.54	0.7

variation in a system model. The system model cannot be used to analyze the internal dynamics of the system if RSC and GSC are not taken into account. The following state-space equations governing the dynamics of the RSC-GSC converter.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} i_{\mathrm{dr}} \\ i_{\mathrm{qr}} \\ i_{\mathrm{dg}} \\ i_{\mathrm{qg}} \end{pmatrix} = \begin{pmatrix} -\frac{R_{\mathrm{RSC}}}{L_{\mathrm{RSC}}} & \omega_s & 0 & 0 \\ -\omega_s & -\frac{R_{\mathrm{RSC}}}{L_{\mathrm{RSC}}} & 0 & 0 \\ 0 & 0 & \frac{R_{\mathrm{RSC}}}{L_{\mathrm{GSC}}} & -(\omega_s - X_{\mathrm{tg}}) \\ 0 & 0 & -(\omega_s + X_{\mathrm{tg}}) & \frac{R_{\mathrm{GSC}}}{L_{\mathrm{GSC}}} \end{pmatrix} \begin{pmatrix} i_{\mathrm{dr}} \\ i_{\mathrm{qr}} \\ i_{\mathrm{dg}} \\ i_{\mathrm{qg}} \end{pmatrix} + \begin{pmatrix} \frac{1}{L_{\mathrm{RSC}}} & 0 & 0 & 0 \\ 0 & \frac{1}{L_{\mathrm{RSC}}} & 0 & 0 \\ 0 & 0 & \frac{1}{L_{\mathrm{GSC}}} & 0 \\ 0 & 0 & 0 & \frac{1}{L_{\mathrm{GSC}}} \end{pmatrix} \begin{pmatrix} v_{\mathrm{dr}} \\ v_{\mathrm{qr}} \\ v_{\mathrm{dg}} \\ v_{\mathrm{qg}} \end{pmatrix} \tag{18}$$

Vector-controlled techniques are adopted in this study [21]. The control loop is shown in Figs. 2 and 3. The value of optimal reference torque used for the RSC controller parameters is listed in Table 1. The value remains constant when the wind speed surpasses the rated speed. The reference torque is the torque that delivers the maximum rotational speed at the measured rotational speed when the wind speed is less than the rated speed. Using this lookup table, the wind turbine can extract the maximum amount of wind energy feasible.

7) *Integrated system modeling* To understand the complete system, it is useful to think of it as an integration of the various subsystems. Equation expresses the entire system model [7].

$$\left(\dot{X}\right) = \left(A_{\mathrm{sys}}\right)\left(X_{\mathrm{sys}}\right) + \left(B_{\mathrm{sys}}\right)\left(U_{\mathrm{sys}}\right) \tag{19}$$

$$\text{where } \left(X_{\mathrm{sys}}\right) = \left(\, X_{\mathrm{shaft}} \;\; X_{\mathrm{DFIG}} \;\; X_{\mathrm{tl}} \;\; X_{\mathrm{dc}} \,\right) \tag{20}$$

The system matrix of the entire system is given by

$$\left(A_{\mathrm{sys}}\right) = \begin{pmatrix} A_{\mathrm{shaft}} & 0 & 0 & 0 \\ 0 & A_{\mathrm{DFIG}} & 0 & 0 \\ 0 & 0 & X_{\mathrm{tl}} & 0 \\ 0 & 0 & 0 & A_{\mathrm{dc}} \end{pmatrix} \tag{21}$$

These interconnected subsystems are able to communicate with each other via state and control variables in the individual subsystems. Algebraic equations of $v_{\mathrm{ds}}, v_{\mathrm{qs}}, i_{\mathrm{ds}}$ and i_{qs} are used to link the controllers. Now, $v_{\mathrm{ds}}, v_{\mathrm{qs}}, i_{\mathrm{ds}}$ and i_{qs} expressed as:

$$v_{\mathrm{ds}} = v_{\mathrm{dg}} + X_{\mathrm{tg}} i_{\mathrm{qg}} \tag{22}$$

$$v_{\mathrm{qs}} = v_{\mathrm{qg}} + X_{\mathrm{tg}} i_{\mathrm{dg}} \tag{23}$$

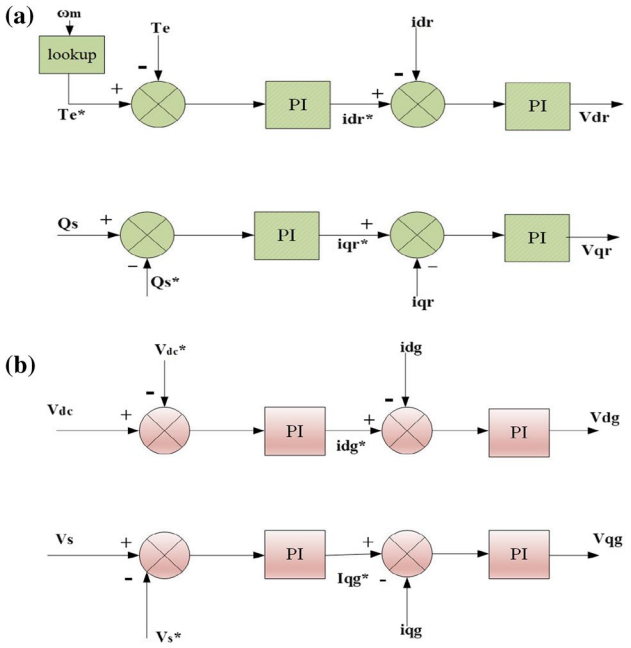


Fig. 3 **a** RSC, **b** GSC control loop

$$i_{\mathrm{ds}} = i_{\mathrm{dg}} - i_{\mathrm{dl}} \tag{24}$$

$$i_{\mathrm{qs}} = i_{\mathrm{qg}} - i_{\mathrm{ql}} \tag{25}$$

Developed torque and stator reactive power is given by:

$$T_g = X_m\left(i_{\mathrm{dr}} i_{\mathrm{qs}} - i_{\mathrm{ds}} i_{\mathrm{qr}}\right) \tag{26}$$

$$Q_s = \frac{1}{2}\left(v_{\mathrm{qs}} i_{\mathrm{ds}} - v_{\mathrm{ds}} i_{\mathrm{qs}}\right) \tag{27}$$

A state-space model is used to investigate the system's small signal stability, which leads to time-domain simulation analysis and eigenvalue computation.

Result and Discussion

To analysis the effect of the SSR on system stability eigenvalue analysis has been done. A simulation-based time-domain analysis has also been done to validate the findings of the eigenvalue analysis.

1) *Eigenvalue analysis* The eigenvalue approach is one of the most critical approaches for studying a dynamic system's small-signal stability. Based on the order of the system matrix, this analysis generates a collection of eigenvalues. The eigenvalues can be used to determine

a system's stability; if the real component of an eigenvalue is positive, the system is unstable; if the real part is negative, the system is stable. The frequency of oscillations is represented by the imaginary portion. The size of the system matrix (A_{sys}) is determined, which is the 22×22 size of the matrix, resulting in 22 eigenvalues. In eigenvalue analysis, chosen two case to investigate (i) different compensation levels of a transmission line and (ii) different wind speeds.

(i) Different compensation levels:

In order to identify the effect of different compensation levels to the sub-synchronous frequency at which system is to be unstable. Table 2 displays the set of eigenvalues from (λ_1–λ_{22}) for fixed wind speeds at 7 m/s and at a constant capacitive compensation level of 50%. At resultant, the eigenvalues (λ_5) and (λ_6) have found unstable at these operating circumstances, and at this condition, the system becomes increasingly unstable. The oscillating frequency at which the system is unstable at the sub-synchronous frequency which is responsible for the SSR. The remaining computed eigenvalues operate in a variety of ways. Electromechanical modes with eigenvalues of (λ_7–λ_{12}) are reported to be responsible for electromechanical system dynamics. (λ_1, λ_2) and (λ_3, λ_4) belong to the extremely high frequency and supersynchronous modes, respectively, because the frequency of these dynamics is higher than the system's, they are always stable. The (λ_{13}–λ_{22}) do not have any frequency terms and are classified as oscillatory as positive modes or non-oscillatory as a negative mode, with amplitudes which are variable as a function of compensation levels as well as wind speeds.

From above result, it has been observed that mode (λ_5, λ_6) is only unstable. Therefore, it is necessitating to check this mode at different levels of series compensation and to observe the pattern. Table 3 shows that at 30% of series compensation and below 30% of series compensation, the system becomes stable. The real part of the eigenvalue (λ_5, λ_6) decreases as the level of series compensation decreases, whereas the amplitude of oscillating frequency increases with the decrease in the series compensation. Above 30% of series compensation the system becomes tends to unstable mode due to decrement of positive damping coefficient. It is verified from eigenvalue analysis that the 30% of series compensation would be the safe value of compensation for which the stability of the system can be ensured without compromising its performance. A detailed eigenvalue analy-

Table 2 Eigenvalue analysis of the complete system model with 50% compensation and a wind speed of 7 m/s

Mode	Eigenvalue	Frequency (Hz)	Description
$\lambda_{1,2}$	$-1790 \pm j\ 889$	141	Very high freq
$\lambda_{3,4}$	$-19.2 \pm j568.2$	90.4	Sup-sync
$\lambda_{5,6}$	**$1.6 \pm j187.8$**	**29.9**	**SSR**
$\lambda_{7,8}$	$-52.4 \pm j7.9$	1.25	Electromechanical
$\lambda_{9,10}$	$-0.63 \pm j18.8$	2.86	Electromechanical
$\lambda_{11,12}$	$-6.41 \pm j8.6$	1.36	Electromechanical
$\lambda_{20,21}$	$-0.0 \pm j0.0$	0	PI controller
λ_{13}	-5.9	0	No oscillatory
λ_{14}	-7.4	0	No oscillatory
λ_{15}	-5.8	0	No oscillatory
λ_{16}	-5.6	0	No oscillatory
λ_{17}	-1.9	0	No oscillatory
λ_{18}	-48.2	0	No oscillatory
λ_{19}	-37.7	0	No oscillatory
λ_{22}	-0.0	0	No oscillatory

Bold value represents the "unstable mode"

Table 3 Eigenvalue analysis of mode 3($\lambda_{5,6}$) at varying series compensationx

%Compensation	Mode3 ($\lambda_{5,6}$)		Remarks
	Eigenvalue	Frequency (Hz)	
10	$-4.7 \pm j291.2$	46.36	Stable
20	$-3.02 \pm j256$	40.76	Stable
30	$-1.5 \pm j229$	36.46	Stable
40	**$0.01 \pm j207$**	32.96	Unstable
50	**$1.6 \pm j187.8$**	29.90	Unstable
60	**$3.16 \pm j170$**	27.07	Unstable
70	**$4.8 \pm j155.8$**	24.80	Unstable
80	**$6.4 \pm j142.3$**	22.65	Unstable

Bold value represents the "unstable mode"

sis of the complete system model with 30% series compensation at a 7 m/s wind speed is shown in Table 4 and observed that all the system modes are in stable condition.

(ii) Different wind speeds.

In order to identify the effect of different wind speeds to the sub-synchronous frequency at which system is to be unstable. Considering the typical wind speed of 7 m/s, 9 m/s and 11 m/s and the fixed series compensation level of 30%, 40%, 50% and 60% as an example, as compensation increases, the real part of eigenvalue (λ_5, λ_6) grows in size, and the system becomes increasingly unstable, as shown in Table 5. However, as the wind speed increases, the real part of mode (λ_5, λ_6) very slightly decreases. Above this result, it can be concluded that the wind speed has an insignificant impact on the SSR. The system stability is observed too dependent on the percentage of compensation offered to the transmission line while fairly insensitive to change in wind speed.

2) *Time-domain analysis* To verify the eigenvalue analysis results, MATLAB/Simulink has been employed for time-domain analysis. Based on the time-domain result depicted in Figs. 4 and 5 on increasing the series compensation level, the system loses its stability. At a constant wind speed of 7 m/s, Fig. 4 displays the electrical power (Pe) dynamic response at 30%, 40%, 50%, and 60% of series compensation values. It has been noticed that above 30% of series compensation, the dynamic response of the active power of the system becomes unstable. At 30% of series compensation, the response tends to stabilize. The second case shown in Fig. 5 considers a specific wind speed at 9 m/s with the same various compensation values and obtained the result that after varying the wind speed, there is a slight improvement in their damping.

The positive damping decreases as the compensating level rises. The above result reveals that as the compensation level is raised, the SSR frequency rises, while the damping decreases at resultant increased amplitude of oscillation. Whereas, with increased wind speeds, the sub-synchronous oscillation frequency and amplitude of oscillation decrease somewhat. These data support the theory that the dominant mode becomes slightly unstable when wind speed decreases.

Table 4 Eigenvalue analysis of the complete system model with 30% compensation and a wind speed of 7 m/s

Mode	Eigenvalue	Frequency (Hz)	Remark
$\lambda_{1,2}$	−1794 ± j887.	141.3	Stable
$\lambda_{3,4}$	−16.1 ± j524.8	83.6	Stable
$\lambda_{5,6}$	−1.5 ± j229.2	36.5	Stable
$\lambda_{7,8}$	−49.1 ± j10.9	1.73	Stable
$\lambda_{9,10}$	−0.63 ± j18.8	2.99	Stable
$\lambda_{11,12}$	−4.56 ± j10.2	1.62	Stable
λ_{13}	−7.51	0	Stable
λ_{14}	−5.9	0	Stable
λ_{15}	−5.7	0	Stable
$\lambda_{16,17}$	−2.2 ± j0.4	0.06	Stable
$\lambda_{18,19}$	−0.01 ± j0.07	0.01	Stable
λ_{20}	−0.00	0	Stable
λ_{21}	−48.2	0	Stable
λ_{22}	−37.7	0	Stable

Table 5 Eigenvalue analysis of mode 3($\lambda_{5,6}$) at varying wind speed and % of series compensation

Compensation (%)	Wind speed	Mode3 ($\lambda_{5,6}$)		Remark
		Eigenvalue	Frequency (Hz)	
30	7 m/s	−1.5 ± j229.23	36.50	Stable
	9 m/s	−1.54 ± j229.20	36.49	Stable
	11 m/s	−1.6 ± j229.11	36.48	Stable
40	7 m/s	**0.01 ± j207**	32.96	Unstable
	9 m/s	−0.04 ± j206.9	32.94	Marginally stable
	11 m/s	−0.12 ± j 206.8	32.92	Stable
50	7 m/s	**1.6 ± j187.8**	29.90	Unstable
	9 m/s	**1.5 ± j187.7**	29.88	Unstable
	11 m/s	**1.4 ± j187.5**	29.85	Unstable
60	7 m/s	**3.2 ± j170.9**	27.21	Unstable
	9 m/s	**3.1 ± j170.73**	27.18	Unstable
	11 m/s	**2.9 ± j170.4**	27.13	Unstable

Bold value represents the "unstable mode"

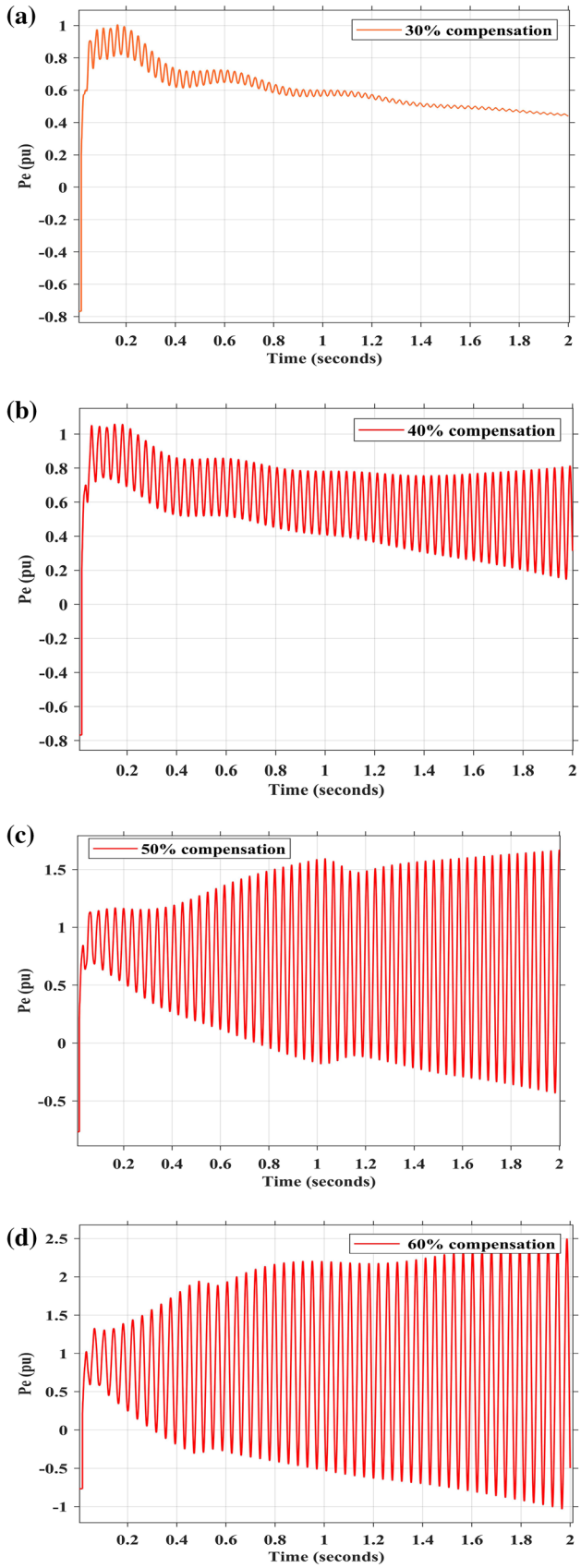


Fig. 4 Electrical power dynamics response at **a** 30%, **b** 40%, **c** 50% and **d** 60% compensation at fixed wind speed 7 m/s

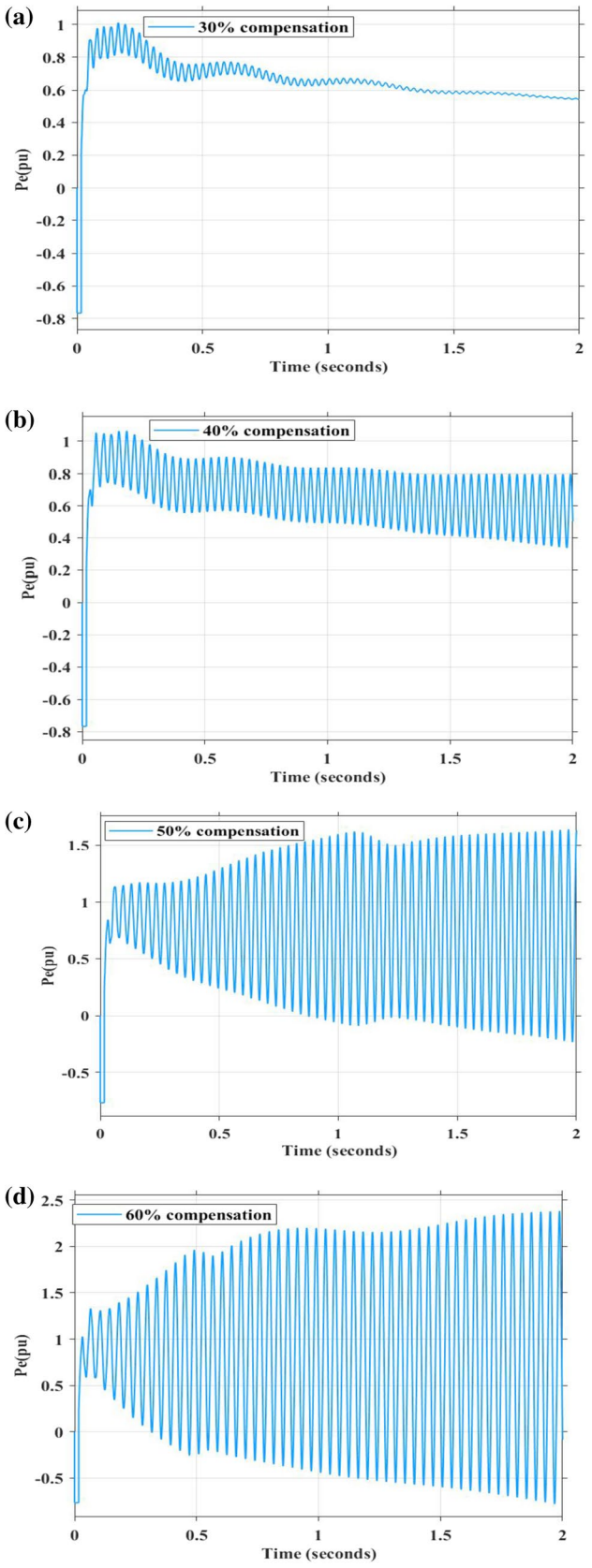


Fig. 5 Electrical power dynamics response at **a** 30% **b** 40%, **c** 50% and **d** 60% compensation at fixed wind speed 9 m/s

Conclusion

For the small-signal stability investigation, a detailed mathematical model of the 100 MW grid-connected DFIG-based WPP based on the IEEE first benchmark model, as well as a dynamic model of the RSC and GSC converters, has been done. The numerous modes of oscillations present in the integrated system can be easily discovered using eigenvalue analysis. The stability of the various oscillating modes can be identified by analyzing the nature of eigenvalues. In MATLAB/Simulink, oscillations present in the electromagnetic torque or active power produced for various compensation levels are investigated, and the eigenvalue analysis is validated. Accordingly, in this paper, SSR analysis has been carried out using an eigenvalue technique for various wind speeds and series capacitance compensation levels and got different oscillation modes, i.e., SSR mode, super-synchronous mode, electromechanical mode, and some non-dominant modes have been explored. The time-domain analysis based on the nonlinear simulation has been employed to validate the eigenvalue outcome. The following conclusion has been obtained:

- The real component magnitude of the eigenvalue increases as the fraction of series compensation is raised, and the system becomes increasingly unstable.
- The increase in wind speed velocity slightly decreases the amplitude of the real part of eigenvalue, resulting in a slightly better stable system.

This suggests that the series compensation plays a vital role in the system's stability due to sub-synchronous resonance instead of variation in the wind speed. Nonetheless, the 30% of series compensation could be the safe compensation value for which the system's stability can be ensured without compromising its performance.

Funding No funding has been received for this work.

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.

Appendix

See Tables 6, 7, 8 and 9.

Table 6 DFIG parameters

Rated power	1.5 MW	100 MW
Rated voltage	575 V	575 V
Rated frequency	60 Hz	60 Hz
R_S	0.023	0.023
X_{ls}	0.18	0.18
R_r	0.016	0.016
X_{lr}	0.16	0.16
X_M	2.9	2.9
DC-link voltage	1150 V	1150 V
DC-link capacitor	10000 µF	67*10000 µF

Table 7 Transmission line parameters

Transformer ratio	575 V/161 kV
Base power	100 MVA
R_l	0.02
X_l	0.5
X_C at 50% of series compensated line	64.8 Ω
X_T	0.14
X_{sys}	0.06

Table 8 Drive train parameters

H_t	4.32 s
H_g	0.685 s
D_t	0
D_g	0
D_{tg}	1.5 p. u
K_{tg}	1.11

Table 9 PI parameters of DFIG converter

K_{P1}	0.2	K_{i1}	0.4
K_{P2}	4	K_{i2}	100
K_{P3}	0.01	K_{i3}	0.04
K_{P4}	4	K_{i4}	100
K_{P5}	0.1	K_{i5}	2
K_{P6}	0.08	K_{i6}	0.5
K_{P7}	0.1	K_{i7}	2
K_{P8}	0.08	K_{i8}	0.5

References

1. M. Khayyatzadeh, R. Kazemzadeh, Sub-synchronous resonance damping using high penetration PV plant. Mech. Syst. Signal Process. **84**, 431–444 (2017)
2. M. Rahmani, G. Vinasco, M.J. Rider, R. Romero, P.M. Pardalos, Multistage transmission expansion planning considering fixed

series compensation allocation. IEEE Trans. Power Syst. **28**(4), 3795–3805 (2013)
3. G. Li et al., Analysis and mitigation of subsynchronous resonance in series-compensated grid-connected system controlled by a virtual synchronous generator. IEEE Trans. Power Electron. **35**(10), 11096–11107 (2020)
4. N. Kumar, S. Gupta, N. Singh, Restraining SSR with CDRPF-signal supported static var system. Sustain. Energy Grids Netw. **16**, 136–144 (2018)
5. S. Kumar, N. Kumar, Alleviation SSR and low frequency power oscillations in series compensated transmission line using SVC supplementary controllers. J. Inst. Eng. Ser. B **98**(3), 255–266 (2017)
6. V. Jain, N. Kumar, Designing of supplementary controller for STATCOM for damping of oscillations in power system. J. Inst. Eng. Ser. B **99**(5), 419–429 (2018)
7. L. Fan, R. Kavasseri, Z.L. Miao, C. Zhu, Modeling of DFIG-based wind farms for SSR analysis. IEEE Trans. Power Deliv. **25**(4), 2073–2082 (2010)
8. J. Shair, X. Xie, S. Member, J. Yang, J. Li, H. Li, Adaptive damping control of subsynchronous oscillation in DFIG-based wind farms connected to series-compensated network. IEEE Trans. Power Deliv. **8977**, 1–13 (2021)
9. G. Li, Y. Chen, S. Member, A. Luo, S. Member, Y. Wang, An inertia phase locked loop for suppressing sub-synchronous resonance of renewable energy generation system under weak grid. IEEE Trans. Power Syst. **36**(5), 4621–4631 (2021)
10. A. Adrees, *Risk Based Assessment of Subsynchronous Resonance in AC/DC Systems*. (2017)
11. J. Shair, X. Xie, L. Wang, W. Liu, J. He, H. Liu, Overview of emerging subsynchronous oscillations in practical wind power systems. Renew. Sustain. Energy Rev. **99**(September 2018), 159–168 (2019)
12. G. Kanghui, W. Feng, X. Zhang, Sub-synchronous interactions in power systems with wind turbines: a review. IET Renew. Power Gener. Spec. **13**(1), 4–15 (2018). https://doi.org/10.1049/iet-rpg.2018.5199
13. J.B. Ekanayake, L. Holdsworth, X.G. Wu, N. Jenkins, Dynamic modeling of doubly fed induction generator wind turbines. IEEE Trans. Power Syst. **18**(2), 803–809 (2003)
14. Y. Lei, A. Mullane, G. Lightbody, R. Yacamini, Modeling of the wind turbine with a doubly fed induction generator for grid integration studies. IEEE Trans. Energy Convers. **21**(1), 257–264 (2006)
15. H.A. Mohammadpour, A. Ghaderi, E. Santi, Analysis of subsynchronous resonance in doubly-fed induction generator-based wind farms interfaced with gate – controlled series capacitor. IET Gener. Transm. Distrib. **8**(May), 1998–2011 (2014)
16. A.S. Trevisan et al., Analysis of low frequency interactions of DFIG wind turbine systems in series compensated grids. Electr. Power Syst. Res. **191**(Feburary 2020), 106845 (2021)
17. J. Shair, X. Xie, W. Liu, X. Li, H. Li, Modeling and stability analysis methods for investigating subsynchronous control interaction in large-scale wind power systems. Renew. Sustain. Energy Rev. **135**(August 2020), 110420 (2021)
18. M. Abdeen et al., A recent analytical approach for analysis of subsynchronous resonance in doubly-fed induction generator-based wind farm. IEEE Access **9**, 68888–68897 (2021)
19. Y. Xu, M. Zhang, L. Fan, Z. Miao, Small-signal stability analysis of type-4 wind in series-compensated networks. IEEE Trans. Energy Convers. **35**(1), 529–538 (2020)
20. P.S.E.C. Ieee, IEEE subsynchronous resonance task force of the dynamic system performance working group power system engineering. IEEE Trans. Power Appar. Syst. **PAS-96**(5), 1565–1572 (1977)
21. F. Mei, B. Pal, Modal analysis of grid-connected doubly fed induction generators. IEEE Trans. Energy Convers. **22**(3), 728–736 (2007). https://doi.org/10.1109/TEC.2006.881080
22. L. Fan, S. Member, C. Zhu, S. Member, Modal analysis of a DFIG-based wind farm interfaced with a series compensated network. IEEE Trans. Energy Convers. **26**(4), 1010–1020 (2011)

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

Springer Nature or its licensor 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.

symmetry

MDPI

Article

Distributed Network of Adaptive and Self-Reconfigurable Active Vision Systems

Shashank * and Indu Sreedevi

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India
* Correspondence: shashank.ece.dtu@gmail.com or shashank_2k17phdec05@dtu.ac.in

Abstract: The performance of a computer vision system depends on the accuracy of visual information extracted by the sensors and the system's visual-processing capabilities. To derive optimum information from the sensed data, the system must be capable of identifying objects of interest (OOIs) and activities in the scene. Active vision systems intend to capture OOIs with the highest possible resolution to extract the optimum visual information by calibrating the configuration spaces of the cameras. As the data processing and reconfiguration of cameras are interdependent, it becomes very challenging for advanced active vision systems to perform in real time. Due to limited computational resources, model-based asymmetric active vision systems only work in known conditions and fail miserably in unforeseen conditions. Symmetric/asymmetric systems employing artificial intelligence, while they manage to tackle unforeseen environments, require iterative training and thus are not reliable for real-time applications. Thus, the contemporary symmetric/asymmetric reconfiguration systems proposed to obtain optimum configuration spaces of sensors for accurate activity tracking and scene understanding may not be adequate to tackle unforeseen conditions in real time. To address this problem, this article presents an adaptive self-reconfiguration (ASR) framework for active vision systems operating co-operatively in a distributed blockchain network. The ASR framework enables active vision systems to share their derived learning about an activity or an unforeseen environment, which learning can be utilized by other active vision systems in the network, thus lowering the time needed for learning and adaptation to new conditions. Further, as the learning duration is reduced, the duration of the reconfiguration of the cameras is also reduced, yielding better performance in terms of understanding of a scene. The ASR framework enables resource and data sharing in a distributed network of active vision systems and outperforms state-of-the-art active vision systems in terms of accuracy and latency, making it ideal for real-time applications.

Keywords: active vision; self-adaptation; self-reconfiguration; smart camera network

check for updates

Citation: Shashank; Sreedevi, I. Distributed Network of Adaptive and Self-Reconfigurable Active Vision Systems. *Symmetry* **2022**, *14*, 2281. https://doi.org/10.3390/sym14112281

Academic Editor: Zhixun Su

Received: 29 September 2022
Accepted: 25 October 2022
Published: 31 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

A smart camera network (SCN), defined by Reisslein et al. [1], is a real-time distributed embedded system that is configured to perform computer-vision tasks by processing sensor data obtained from a plurality of cameras via cooperative sensing. The smart camera network is generally deployed to perform complex computer-vision tasks that require more than one camera to extract visual information. Visual surveillance of large areas [2], complex sports analytics [3], situation cognizance [4], tele-immersion, automated driverless vehicles [5], ambient assisted living [6], computer-vision-based disaster management, robotic vision, and other applications rely on SCNs to achieve desired functionalities. The smart camera network aims to co-operatively fuse sensor data to develop scene understanding and further associates scene understanding with a control system configured to achieve the overall functionality of a computer vision system.

The advancement in computer-vision applications has taken a tremendous leap in the last decade. According to a survey reported in [2], the growth of computer-vision applications is estimated at a compound annual growth rate (CAGR) of 7.6% from 2020

to 2027. Computer vision systems [7] aim to leverage human efforts by developing an understanding of events through the processing of data obtained from a number of sensors (hereinafter interchangeably referred to as "cameras" or "camera sensors"). Further, based on understanding, most of the smart computer vision systems aim to enable the operation of control systems to deliver a desired functionality automatically without any human interaction. An ideal computer vison system should extract data from the activities detected by the sensors, generate an understanding based on processing of the data, and provide optimal functionality through an appropriate control system.

Computer vision systems [7] rely on the processing of digital images or videos obtained from one or more cameras to obtain an understanding of an activity or an event. Understanding of the activity or event highly depends on the quality of data in terms of information about objects of interest. To obtain optimum information about an OOI from captured data, the cameras must be configured to capture OOIs with the highest possible resolution. In some respects, it can be concluded that the objects of interest must be captured as close to the center of the camera's field of view as possible.

A computer vision system [7] needs to process the data to identify an OOI and critical activities in the scene to derive scene understanding, based on which the cameras can be reconfigured to capture the OOI in the center of the FOV. Thus, data processing and camera calibration are interdependent. A system capable of reconfiguring the parameters of the cameras to manipulate the viewpoints of the cameras in order to investigate the environment and obtain better information from it is known as an active vision system. As reconfiguration of the cameras and the processing performance of the active vision system are interdependent, designing advanced active vision systems employing a network of cameras co-operatively working towards a desired functionality in real time is very challenging.

Traditional model-based active vision systems have limited computational resources, due to which they can only identify activities in known environments for which they are designed and fail badly to identify and understand new activities in unforeseen conditions. Active vision systems employing artificial intelligence (AI) manage to tackle unforeseen conditions; however, due to iterative training processes, they are not reliable for real-time applications. The impacts of unforeseen conditions and uncertainties in computer vision systems are presented in [2].

To address the abovementioned problem, this article showcases an adaptive self-reconfiguration (ASR) framework for active vision systems operating co-operatively in a distributed blockchain network. The ASR framework facilitates active vision systems to share information about their learning towards new activities and unforeseen environments with other systems in the distributed network. The information shared can be utilized by any system in the distributed blockchain network to tackle an identical condition and thus saves a lot of time which would otherwise have been taken up with iterative model training. Further, as the learning duration is reduced, the duration of the reconfiguration of the cameras is also reduced, thus yielding better performance in terms of understanding a scene.

To develop a good understanding of the abovementioned reconfiguration problem and its impact on the performance of an active vision system, this article primarily provides a detailed discussion of the challenges in developing an active vision system at different operational levels. We further provide an extended survey of various systems and methods proposed for addressing the challenges. Additionally, this article highlights a trend in the state-of-the-art systems and methods proposed to address the active vision challenges and showcases a common threat for most of the contemporary solutions. This article further provides definitions of the concepts of self-reconfiguration and self-adaptation and presents a role for self-reconfiguration and self-adaptation in the enhancement of the performance of active vision systems. Finally, this article provides an adaptive self-reconfiguration (ASR) framework to enhance the performance of active vision systems and their applications.

A computer vision system generally includes one or more camera sensors, software or circuitry for data processing, and a control system. The camera sensors are configured for co-operative sensing in the operating environment and gathering raw sensor data. The raw

sensor data are fused to obtain pre-processed visual information that is utilized for visual processing. The visual processing includes several steps to process and furnish the sensor data and obtain a visual understanding. A control signal is generated based on the visual understanding and is transmitted to the control system. Based on the control signal received, the control system actuates one or more components of the system such that the desired functionality of the system is achieved. The abovementioned process is depicted in Figure 1.

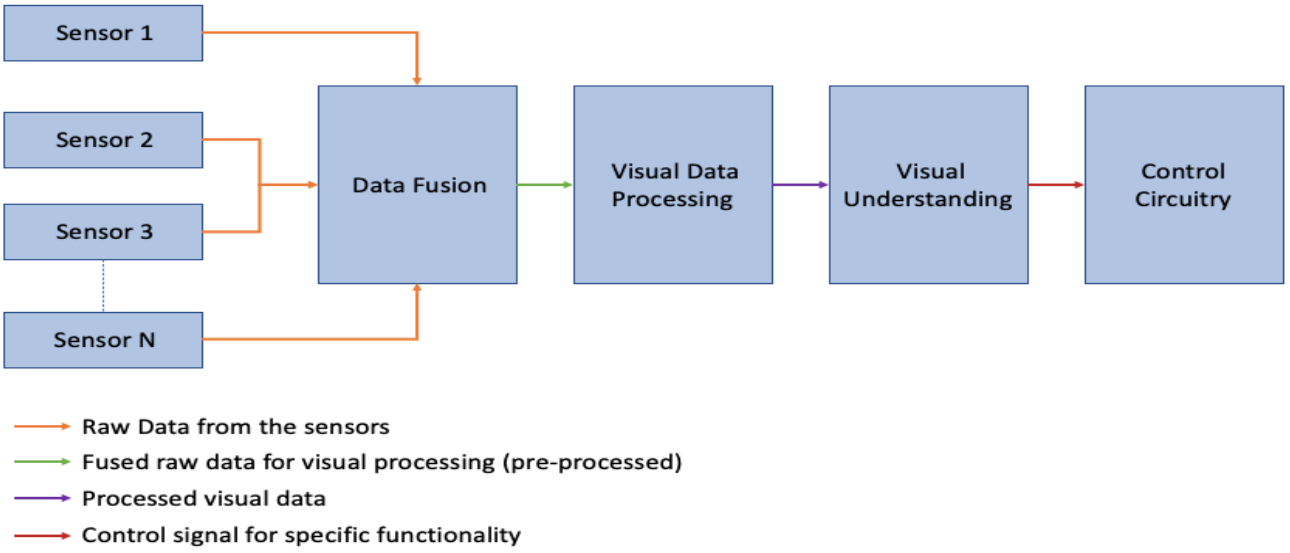


Figure 1. A general process flow of a computer vision system employing a smart camera network.

A multi-camera active vision system [8] (hereinafter interchangeably referred to as an "active vision system") is basically a computer vision system that employs an SCN with the capability of altering the viewpoints of the sensor nodes, thereby yielding better data for processing. The operation of active vision systems can be differentiated into two levels: a deployment level (i.e., for data extraction) and a processing level (i.e., where the visual processing takes place to obtain an understanding). The challenges associated with the deployment of an active vision system employing one or more SCNs can also be classified into two categories. Challenges of the first type [9] are associated with reconfiguration of the sensor nodes to obtain sensor data bearing optimal information, relative to each camera's resource limitations. Challenges of the second type [2] are associated with the data-processing level of operation, occurring at the time of processing the data to develop an understanding of a scene while keeping the computational complexity as low as possible. Thus, an efficient active vision system demands efforts at both hardware and software levels of deployment. Detailed reviews of the challenges at the data-processing level and the sensor-calibration level of active vision systems are presented in [2] and [9], respectively. A multi-tier taxonomy of challenges for active vision systems employing camera networks on the basis of operational classification is shown in Figure 2.

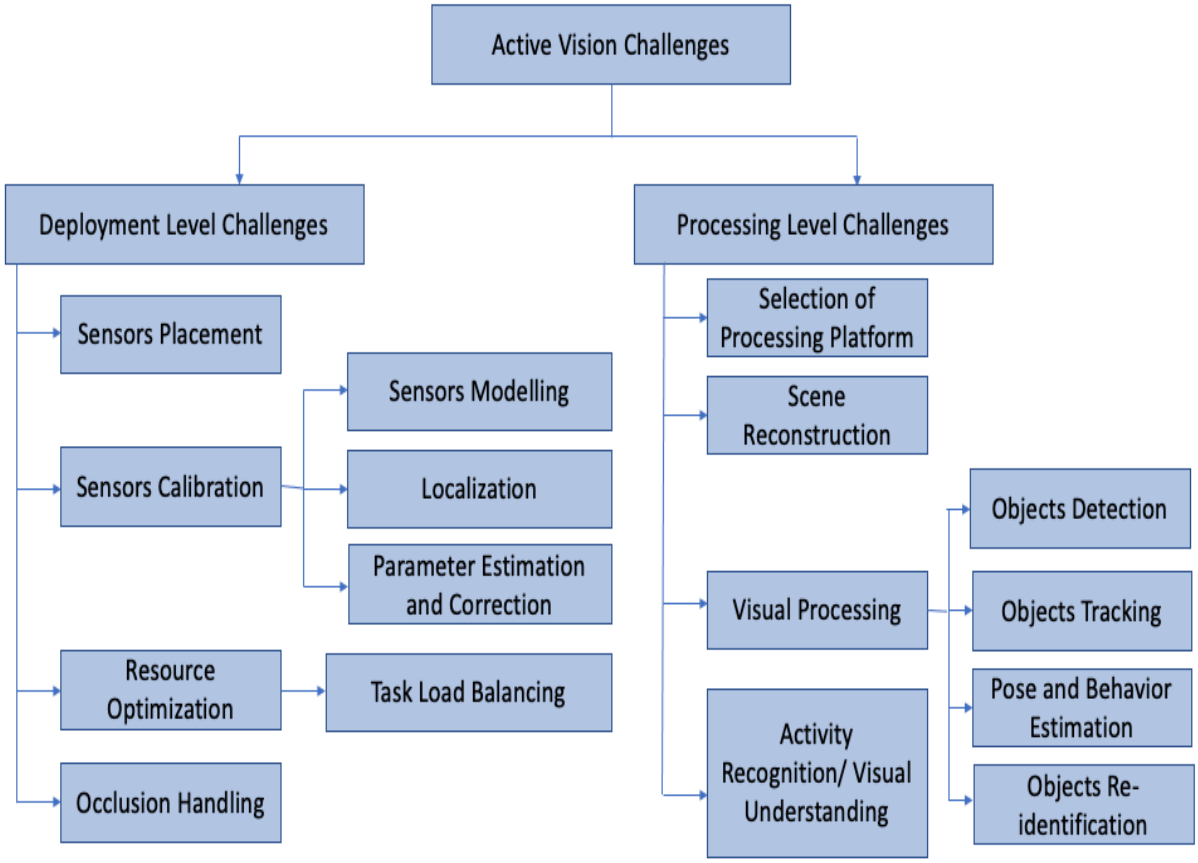


Figure 2. Taxonomy of challenges in active vision applications employing camera networks.

2. Challenges

A smart camera network (SCN) enhances the availability of raw data and thus the chances of deriving better information; however, it also increases the computational complexity of the active vision system of which it is a part. The configuration space of each sensor depends on the intrinsic parameters (such as optical center, focal length, etc.) and the extrinsic parameters of the sensor (such as rotation and translation, etc.), which require calibration [8].

Apparently, an active vision system utilizing a single fixed camera has much lower computational complexity and thus faces fewer challenges at the deployment level than an active vision system employing a network of mobile pan–tilt–zoom (PTZ) cameras. However, an active vision system employing multiple camera sensors can have an occlusion-avoidance capability that may be lacking in an active vision system employing a single camera for sensing.

2.1. Deployment-Level Challenges

Due to limited resources with respect to most of the sensor nodes and the high computational complexity of active vision systems, it becomes challenging to allocate tasks to the nodes. Additionally, optimization of the resources by each sensor node is equally important. Further, designing a sensor architecture and accurate sensor placement according to the overall objective of the system is another challenge at the deployment level. Furthermore, calibration of the sensors of the SCN and occlusion handling by the system also add to the challenges at the deployment level.

2.1.1. Sensor Placement

Generally, cameras in SCNs are placed with overlapping FOVs to reproduce the entire operating environment. However, in some cases, with limited resources and larger operational areas, it becomes very difficult to place cameras with overlapping FOVs. The

placement of cameras has a direct impact on the quality and quantity of data available for processing. For example, if an object is captured in the center of a camera's FOV, the quality of the data, and thus the visual information available from the data, is much higher than if the object is captured at the edges of the camera's FOV. In addition, camera placement must ensure maximum coverage of events and thus camera placement becomes a critical part of the deployment of SCNs. For the above reasons, sensor placement is critical to the deployment of an SCN.

2.1.2. Calibration

The configuration space of an SCN includes the internal and external camera parameters of the sensor nodes or cameras of the SCN. It is critical to change the configuration space of each camera, considering the available and used resources. When the target changes its position in the field of view (FOV) of a sensor or moves from the FOV of one sensor to another, the configuration of the sensor must be changed by modifying its parameters so that data can be acquired to yield optimal information (i.e., by keeping the object as close to the center of the FOV as possible).

Some SCN designs also emphasize minimizing resource utilization by inactivating sensors if no activity is detected over a specified period of time, which alters the dynamic topology of the network. Thus, dealing with such dynamically changing situations is very challenging. At the operational level, calibration challenges are further classified as follows.

- *Sensor modelling*

The parameters of the configuration space of each sensor depend on the type of sensors used in the SCN. The sensor model provides information about the configuration space as well as the available resources, such as power, bandwidth, and the overall quality of service (QoS), possessed by the sensors. Therefore, sensor modelling plays a crucial role in active vision systems.

- *Localization*

When an object moves away from the FOV of a sensor, it is important to hand over the object to another sensor in the SCN. Since the topology of the SCN is dynamic, the network must be configured to receive information about the active nodes and their relative positions at all times. This information is obtained through the active localization of sensor nodes in the SCN. Localization plays an important role in the system's deployment, as it aids in determining the dynamic relative positions of active sensor nodes in the SCN.

- *Parameter estimation and correction*

Real-time calibration of the active sensors participating in the SCN is required to capture the objects of interest (OOIs) with maximum information. Therefore, parameter estimation and correction play important roles in active vision systems employing SCNs.

2.1.3. Resource Optimization and Task-Load Balancing

Most of the present-day active vision systems require sensor mobility and thus rely on batteries and wireless communication. Due to limited resources, optimum resource utilization becomes critical. Resource optimization can be achieved in two ways, either by using low-power-consuming components at the nodes or by using the sensor nodes smartly. Low-power-consuming components are generally expensive and thus add to the overall expense of deploying an SCN. In a pre-established network, employing new low-power-consuming components may lead to having to discard the entire system and build a new one from the scratch. On the other hand, using the sensors smartly (only when required) does not necessarily require changing the existing hardware or components in the SCN. However, getting a machine to learn when to activate while responding in real time can be challenging for a system. An outcome of a dynamically changing state of the nodes in the SCN is dynamic task-load allocation to the active nodes in the SCN.

The network topology changes dynamically as the sensors switch from active to inactive states. Additionally, in order to avoid deviating from the SCN's main goal (i.e.,

obtaining sensor data with high-quality visual information), the system's overall functionality is dynamically distributed amongst the nodes in the form of task loads. As a result, the SCN must compute the dynamic topology in real time, determine the nodes' localizations, and simultaneously distribute the task load among the active nodes in the SCN.

2.1.4. Handling Occlusions

The OOI can occasionally become obscured while being observed by a sensor, resulting in low-quality data due to the occlusion blocking the OOI. One way to deal with this situation is to hand over the OOI to the subsequent nearest sensor node. However, finding a next-best node for the OOI in real-time also creates challenges. Moreover, it becomes difficult if any nearby nodes are unavailable or if the OOI is outside the field of view (FOV) of any nearby nodes. There are a few prediction-based approaches that rely on computation; however, these approaches lack accuracy and may result in missing critical data, making them unreliable.

2.2. Challenges in Data Processing

Processing sensor data to obtain understanding is the most critical part of any active vision system. Active vision systems utilize data from multiple sources (i.e., multiple sensor nodes) and therefore require highly complex computational capabilities. The visual processing becomes more challenging when an application has to deliver results in real time. The data-processing challenges of an active vision system are discussed below.

2.2.1. Selection of a Processing Platform

The choice of processing platform is as important as the algorithms used to process the sensor data. The processing requirements of active vision are selected based on the system's overall functionality, which is typically determined by two factors: processing time and functional complexity. For instance, the architecture of a system that requires complex computations in real time might be more complex and therefore more expensive than one with a relaxed processing window for applications with simpler computations.

2.2.2. Scene Reconstruction

The data obtained by a number of active sensors need to be synchronized to obtain useful information. An individual sensor's data need to be stitched together in such a way that they result in the determination of an action performed by the OOI. The dynamically changing network topology makes this even more challenging.

2.2.3. Data-Processing Challenges

The active vision system relies on visual-processing algorithms to achieve the overall functionality of the system. Data-processing challenges include detecting an event or activity by the OOI and further deriving an understanding about the detected event or activity. Various visual-processing challenges in an active vision system are discussed below.

- *Object detection*

While detecting an OOI, an active vision system divides the acquired sensor data into a foreground and a background. It should be noted that in an active vision system, object detection and data extraction by sensors are interdependent, making it difficult for any active vision system to accurately detect the OOI. In addition, one or more objects in the observed scene may be performing multiple tasks, increasing the likelihood of problems. Variations in viewpoints, shifts in lighting conditions, occlusions, and similar factors often lead to issues with object detection.

- *Object classification and tracking*

Detected objects undergo classification to distinguish one object from another. Object classification requires the selection of appropriate methods for pattern recognition, clustering, and data segmentation. There are several methods for pattern recognition, clustering,

and segmentation and there are associated advantages and challenges. Therefore, the selection of an appropriate method for object classification is very critical. Further, information about the activity performed by the object(s) participating in an event under observation is extracted by tracking the OOI(s) in consecutive frames. Deciding tracking features based on the activity of an OOI can be challenging because the OOI perceived by the sensor in successive frames is likely to have different viewpoints (i.e., the shape of the OOI perceived by the sensors may appear variable).

- *Object Re-identification*

Many active-vision applications rely on relating activities of an OOI observed in non-consecutive frames. For instance, an understanding of an event can be obtained by comparing the actions of an OOI seen in various camera FOVs and at various times. For the system to reach a conclusion, it must re-identify an already identified OOI and correlate different activities of the OOI performed at different times.

- *Pose and behavior estimation*

Consecutive frames are subjected to spatio-temporal evaluations of consecutive frames to obtain pattern and pose information from the OOI tracked in the scene. An accurate relationship between the activities and the changes in the pose of the OOI is required to relate the visual information to the corresponding understanding of the activity. Further, the active vision system must be configured to estimate variation in the pose of the OOI and thus derive the behavior of the OOI based on the pattern of the poses.

2.2.4. Activity Recognition and Understanding

Activity detection and recognition follows two paradigms: static and dynamic. An activity can be detected by analyzing a single frame, known as static recognition, and requires only spatial evaluation of the frame. Dynamic approaches require the evaluation of multiple consecutive frames, utilizing scene reconstruction (also known as dynamic recognition), and thus require spatio-temporal computational capabilities. Static recognition processes do not require pose estimation. Dynamic recognition, on the other hand, is generally used to solve complex problems where the information is obtained by pose estimation, such that the pose information is subsequently processed.

It is worth noting that the process flow of an active vision system utilizing an SCN includes two basic components: one derived from the functionality of the SCN (i.e., for the dynamic calibration of parameters) and the other derived from the computer vision system (i.e., for visual processing). In addition, it must also be noted that both the abovementioned components are inter-related co-operatively.

2.3. System-Level Challenges

In an active vision system employing an SCN to extract sensor data, the calibration of parameters of the sensor nodes participating in the SCN is highly dependent on the detection and interpretation of the sensor data by the active vision system. Further, the effectiveness and performance of the active vision system (i.e., the data processing) depend on the quality of data extracted by the SCN nodes. Thus, it can be inferred that the functionality of the SCN and the data processing system are interdependent. Moreover, real-time decision making is even more challenging for such an interdependent system.

In some respects, an active vision system can be considered a type of computer vision system wherein the system is configured to make alterations in the configuration space of the sensor(s) employed by the system in real time in a manner pre-defined by way of design protocols. However, an active vision system is not capable of handling most unforeseen challenges because it is protocol-driven and therefore cannot make independent decisions. Thus, the system-level challenges mentioned above are very critical and hard to overcome. If the system is configured with self-reconfigurable properties, it can overcome the abovementioned challenges. If such a system is configured with adaptive capabilities, this may further

facilitate an enhancement in performance through active learning. The self-reconfiguration and adaptive properties of an active vision system are discussed later in Section 5.

3. Existing Solutions

To understand the development and progress at each operational level of active vision systems, we have performed an extended search survey for the last two decades. The results were generated with the keywords "active vision", "computer vision", "smart camera network", AND "the name of challenge", as discussed in Section 2, using the Google Scholar search tool.

3.1. Sensor Placement

The two major challenges to be addressed while deploying an SCN in reference to sensor placement are the maximization of the surveillance area and the handling of non-overlapping FOVs (i.e., managing handover). Indu et al. [10] and Zhang et al. [11] proposed methods for sensor placement in networks aiming to maximize the surveillance areas covered by the sensors. Silva et al. [12] proposed a system for coordination among the sensor nodes of an unmanned aerial vehicle (UAV) network for efficient surveillance. The systems and methods proposed in [10–12] present camera-placement solutions for optimized functionality; however, they lack flexibility in architecture and prioritization of surveillance space. An activity-based prioritization of the surveillance area is proposed in [13] by Jamshed et al., addressing the abovementioned concern. Vejdanparast [14] addressed the camera-placement problem for maximizing the area of surveillance by enhancing the fidelity of each camera in the network. Wang et al. [15] proposed Latin-Hypercube-based Resampling Particle Swarm Optimization (LH-RPSO) based on a camera placement algorithm for IoT devices and networks.

Redding et al. [16] proposed object handover based on multiple features, such as Zernike moments, scale-invariant feature transformation, gray-level co-occurrence matrices, color models, etc., using cross-matching for non-overlapping FOVs. In [17], Esterele et al. presented a method for the generation of an online real-time vision graph for the handover of information in a decentralized network with non-overlapping FOVs. The method proposed in [17] demonstrated that no prior knowledge of nodes is required and that it is easy to add or remove nodes from the network. Lin et al. [18] proposed an active handover control for real-time handover of single objects using multiple PTZ cameras, using the shortest distance rule and spatial relations. The method proposed in [18] proposed the readiness of a receiving camera before handover. A year-wise representation of the abovementioned sensor placement techniques along with their advantages is shown in Table 1.

Table 1. Evolution of sensor-placement techniques.

Ref.	Year	Methodology	Advantages
[16]	2008	Online system for tracking multiple people in an SCN with overlapping and non-overlapping views	Development of a larger, more capable, and fully automatic system without prior localization information
[10]	2009	Genetic algorithm	Maximum coverage of users; Defined priority areas with optimum values of parameters; The proposed algorithm works offline and does not require camera calibration; Minimizes the probability of occlusion due to randomly moving objects
[17]	2011	Ant-colony-inspired mechanism used to grow the vision graph during runtime	Generates a vision graph online; Increased autonomy, robustness, and flexibility in smart camera networks

Table 1. *Cont.*

Ref.	Year	Methodology	Advantages
[18]	2012	Approach to construct the automatic co-operative handover of multiple cameras for real-time tracking	Tracking a moving target quickly and keeping the target within the viewing scope at all times
[11]	2015	Novel model with non-uniformly distributed detection capability (DC)	Orientation of each visual sensor can be optimized through a least-squares problem; More efficient with an averaged relative error of about 3.4%
[13]	2015	Node-level optimal real-time priority-based dynamic scheduling algorithm	Portable system with ease of access in hard-to-access areas
[12]	2017	Coordination of embedded agents using spatial coordination on strategical positioning and role exchange	Persistent surveillance with dynamic priorities
[14]	2020	Novel decomposition method with an intermediate point of representation	Low computational expense; Higher fidelity of the outcomes
[15]	2020	Latin-Hypercube-based Resampling Particle Swarm Optimization (LH-RPSO)	LH-RPSO has higher performance than the PSO and the RPSO; LH-RPSO is more stable and has a higher probability of obtaining the optimal solution

3.2. Calibration

The major challenges in the calibration of nodes participating in an SCN concern sensor (camera) modelling, localization, and parameter estimation and correction. Some of the proposed solutions addressing various calibration challenges are discussed hereinbelow.

3.2.1. Camera Modelling

Some basic models for camera calibration are the thin-lens camera model, the pinhole camera model (linear-perspective projection model), the orthographic projection model, the scaled orthographic projection model, and the para-perspective projection model.

The thin-lens model is a linear calibration model that accounts for the effects of translation and rotation relative to a view plane. The pinhole model later introduced the effect of linear perspective projection; however, it has high computational complexity.

To overcome the high computational complexity of the pinhole model, Hall et al. [19] proposed a much simpler and computationally efficient linear model based on 3D affine transformation with linear perspective projection. The abovementioned linear models did not perform well, as they were unable to account for non-linear distortion, which was addressed by improving 3D affine transformation with non-linear perspective projection models by Tsai et al. in [20], Toscani's non-linear calibration model in [21], and by Wang et al. in [22].

3.2.2. Localization

Camera localization helps estimate the relative positions, orientations, and poses of active nodes in a network. Identifiers or markers in the form of lines, points, features, cones, circles, spheres, etc., are commonly used for the localization of nodes in a camera network.

Such identifiers are commonly used in unknown network environments but can also be used in known environment for improved accuracy and better scene mapping. Utilizing perspective points as markers, the Perspective-n-Point (PnP) algorithm can be used in a known network environment. Simultaneous localization and mapping (SLAM), as presented in [23,24], and structure from motion (SFM), as presented in [25], can be used for dynamically changing network environments. The SFM technique in [25] is based on human vision perseverance to estimate a 3D scene using 2D image data by combining image motion information with frame data. The Monte Carlo method in [26] uses a particle filter for localization and recursive Bayesian estimation for sorting and sampling.

Montzel et al. [27] proposed a distributed energy-efficient camera network localization method using sparse overlapping in 2004. In [28], Brachmann and Rother proposed 6D

pose estimation using an end-to-end localization pipeline. The geometric localization obtained via the head-to-foot location (poles) of pedestrians using an estimated distribution algorithm (EDA) in [29] can be utilized for self-calibration of the nodes in a network. A year-wise representation of the abovementioned localization techniques along with their advantages is shown in Table 2.

Table 2. Evolution of localization techniques.

Ref.	Year	Methodology	Advantages
[26]	1999	Online system for tracking multiple people in an SCN with overlapping and non-overlapping views	Development of a larger, more capable, and fully automatic system without prior localization information
[27]	2004	Sparse overlapping	Better energy efficiency and able to cope with networking dynamics
[23]	2006	SLAM	Locally optimal maps with computational complexity independent of the size of the map
[24]	2006	SLAM	Locally optimal maps with computational complexity independent of the size of the map
[29]	2016	Estimated distribution algorithm (EDA)	Accurate estimation of the features of moving objects (person)
[25]	2017	SFM	Better ambiguity handling in 3D environments
[28]	2018	6D pose estimation using an end-to-end localization pipeline	Efficient, highly accurate, robust in training, and exhibits outstanding generalization capabilities

3.2.3. Parameter Estimation and Correction

Zheng et al. [30] proposed a focal-length estimation method using parallel particle swarm optimization (PSO) with low time complexity and efficient performance. Führ and Jung [31] proposed a self-calibration method for the surveillance of pedestrians in a static camera network using a projection matrix obtained from non-linear optimization of an initial projection matrix obtained after pole extraction. The information in the projection matrix was used for the localization of cameras in the network. In [32], Yao et al. proposed a self-calibration model for dynamic multi-view cameras using golf and soccer datasets based on a field model.

Li et al. [33] proposed a greedy-descent-optimization-based parameter-estimation and scene-reconstruction framework for camera–projector pairs for self-calibration. A network of such systems can be used for efficient tele-immersion applications. Janne and Heikkilä [34] proposed a self-reconfiguration solution for a camera network with focal-length estimation using homography from unknown planar scenes. In [35], Tang et al. proposed a simultaneous distortion-correction self-configuration method using an evolutionary optimization scheme on an estimated distribution algorithm (EDA) for tracking and segmentation. A year-wise representation of the abovementioned parameter estimation techniques along with their advantages is shown in Table 3.

3.3. Resource Optimization: Topology Estimation and Task-Load Balancing

As the network topology changes, the task load needs to be altered to ensure that the overall functionality of the system is achieved. Marinakis and Dudek [36] proposed a system to estimate the topology of a visual network in the form of a weighted directed graph using statistical Monte Carlo expectation and sampling models. Hangel et al. in [37] addressed the problem of topology estimation for a large camera network and proposed a window-occupancy-based method as a solution. The method in [37] required a lot of assumptions and could not handle large numbers of data. Detmold et al. [38] proposed a topology-estimation method capable of handling data from a large number of nodes in the network by scalable collective stream processing using an exclusion algorithm in distributed clusters of nodes. The method proposed by Detmold et al. [38] is similar to the

decentralized processing scheme. Clarot et al. [39] proposed an activity-matching-based network topology for distributed networks. Zhou et al. [40] proposed topology estimation by means of a statistical approach in a distributed network environment, utilizing identity and appearance similarity.

Table 3. Parameter-estimation techniques.

Ref.	Year	Methodology	Advantages
[31]	2015	Projection matrix obtained from non-linear optimization	Better accuracy
[32]	2016	Field model	Automatic estimation of camera parameters with high accuracy
[33]	2017	Greedy descent optimization	Stable and robust automatic geometric projector camera calibration with high accuracy; Efficient in tele-immersion applications
[34]	2017	Homography from unknown planar scenes	Highly stable
[30]	2018	Parallel particle swarm optimization (PSO)	Low time complexity and efficient performance
[35]	2019	Evolutionary optimization scheme on an EDA	Capability of reliably converting 2D object tracking into 3D space

In [41], Farrel and Davis proposed network topology estimation for decentralized data processing. Zhu et al. [42] proposed a centralized processing approach for topology discovery using pipeline processing of lightning variations. In [43], Goutam and Misra proposed a trust-based topology management system for a distributed camera network. Tan et al. [44] proposed a method for topology estimation using blind distance as a parameter. In [45], Li et al. proposed topology estimation using Gaussian and mean cross-correlation functions for a distributed camera network. A year-wise representation of the abovementioned topology-estimation techniques along with their advantages is shown in Table 4.

Table 4. Topology-estimation techniques.

Ref.	Year	Methodology	Advantages
[36]	2005	Monte Carlo expectation maximization and sampling	Minimum effects of noise and delay
[37]	2006	Window-occupancy-based method	Efficient and effective way to learn an activity topology for a large network of cameras with a limited number of data
[38]	2007	Exclusion algorithm in distributed clusters	High scalability
[40]	2007	Statistical approach in distributed network environment	Robustness with respect to appearance changes and better estimation in a time varying network
[41]	2008	Decentralized data processing	Robustness with respect to variable appearance and better scalability
[39]	2009	Activity-based multi-camera matching procedure	Flexible and scalable
[42]	2015	Pipeline processing of lightning variations	Automated tracking and re-identification across large camera networks
[43]	2015	Trust-based topology management system	Higher average coverage ratio and average packet delivery ratio
[44]	2018	Blind-area distance estimation	Finer granularity and high accuracy
[45]	2018	Gaussian and mean cross-correlations	Better target tracking under a single region and better interference in multi-view regions

An efficient computer vision system differentiates the overall functionality into a number of small tasks to optimize the system's functionality. The task load for each active

node depends on its local state, orientation, and available resources. In some respects, resource utilization is related to task-load balancing at each active node in the SCN. Kansal et al. [46] presented a distributed approach for adaptive task-load assignment on the basis of available energy from the network environment, which significantly improved the lifetime of the system. Rinner et al. [47] proposed a heterogeneous multiple mobile-agent-based task-allocation framework utilizing a distributed multi-view camera network. Later, Rinner et al. [48] presented an updated approach to allocate tasks for traffic surveillance, proposing clustered surveillance areas. In [49], Karuppiah et al. proposed a hierarchy-based automatic resource allotment and task-load balancing algorithm using fault tolerance based on activity density for a distributed network. Dieber et al. [50] proposed expectation-maximization-based task-load assignment to optimize monitoring performance, with efficient resource utilization. Dieber et al. [51] extended their work on task-load balancing with market-based handover for real-time tracking with optimized resource utilization. In [52], Christos et al. proposed a market-based bidding framework for multi-task allocation for a distributed camera configuration. A year-wise representation of the abovementioned task-load balancing techniques along with their advantages is shown in Table 5.

Table 5. Task-load balancing techniques.

Ref.	Year	Methodology	Advantages
[46]	2003	Method for distributed adaptive task-load assignment	Better resource efficiency
[47]	2005	Multiple-mobile-agent-based task-allocation framework	Selective operation of the tracking algorithm to reduce the resource utilization
[48]	2005	Multiple-mobile-agent-based task-allocation framework	Selective operation of the tracking algorithm to reduce the resource utilization
[49]	2010	Hierarchy-based automatic resource allotment	Robust tracking
[50]	2011	Expectation-maximization-based approximation	Efficient approximation method for optimizing the coverage and resource allocation
[51]	2012	Market-based handover	Improved quality of surveillance with optimized resources
[52]	2016	Market-based handover	Improved quality of surveillance with optimized resources

3.4. Occlusion Handling

Occlusion-handling approaches either aim at handing over the OOI to the next-best sensor node or predicting the occluded part of the OOI and reproducing it virtually to obtain the missing visual information. Occluded objects in the camera field can result in loss of activity information and thus compromise the functionality of the control system utilizing the understanding provided by the computer vision. Wang et al. [53] proposed occlusion estimation at each point of a scene flow field with patch-match optimization utilizing feature consistency and smoothness regularization as performance parameters in space with an improved red–green–blue dense model. In [54], Quyang et al. proposed a framework based on a part-based deep model for pedestrian detection. The proposed model is capable of estimating information loss due to occlusion in the form of errors in detector scores, using visibility of parts as a parameter.

Shahzad et al. [55] proposed multi-object tracking with effective occlusion handling by modelling the foreground using a K-means algorithm, where the object information is associated after occlusion using a statistical approach. Rehman et al. [56] proposed clustering based on a variational Bayesian method and multi-object tracking based on concepts of attractive and repulsive forces depending upon Euclidean distances between objects, utilizing a social force model to avoid the effects of occlusion. In [57], Chang et al. proposed a convolutional-neural-network (CNN)-based tracking system with sparse coding for the pre-training of a network with the capability of handling occlusion effectively

for the surveillance and classification of vehicles. Zhao et al. [58] proposed an adaptive background formulation based on a Gaussian model for occlusion handling and object tracking in a coarse-to-fine-manner without-affecting-appearance model present in the system. In [59], Liu et al. proposed a distraction-aware tracking system based on a 3D mean-shift algorithm, capable of altering its appearance model and occlusion handling by utilizing depth information of the OOI. A year-wise representation of the abovementioned occlusion handling techniques along with their advantages is shown in Table 6.

Table 6. Occlusion-handling techniques.

Ref.	Year	Methodology	Advantages
[53]	2015	Patch-match optimization	Reduced computational complexity by large displacement motion
[54]	2015	Part-based deep model	Handles illumination changes, appearance change, abnormal deformation, and occlusions effectively
[56]	2015	Social force model	Improved tracking performance in the presence of complex occlusions
[55]	2016	K-means algorithm and statistical approach	Cost-effective in terms of resources (memory and computation)
[58]	2017	Gaussian model for occlusion handling	Handles appearance changes and is capable of dealing with complex occlusions
[57]	2018	CNN	High performance with a limited labelled training dataset
[59]	2018	Distraction-aware tracking system	Effective and computationally efficient occlusion handling

3.5. Selection of a Processing Platform

Selection of a platform to develop a computer vision system is as critical as designing or selecting algorithms specific to the functionality. Most commonly, the processing platforms may either be software-based, as with a central processing unit (CPU) or a graphical processing unit (GPU), or hardware-based, as with a field-programmable gate array (FPGA) and application-specific integrated circuits (ASICs). The selection of the platform depends on the requirements of processing capabilities, result accuracy, flexibility, timeliness, and resource utilization. A comparative evaluation of the selection of processing platforms for computer vision systems is presented by Feng et al. [60].

Systems that require flexible functionality usually prefer CPU- or GPU-based processing platforms; however, the efficiency of such systems is low. On the other hand, ASICs and FPGA are used for systems that require high efficiency and better and faster computations; however, these systems lack operational flexibility.

Hørup et al. [61] presented a comparative analysis of general-purpose computations performed by CPUs and GPUs in computer vision systems. Guo et al. [62] proposed a fast and flexible CPU-based computation system for human pose estimation. Tan et al. [63] proposed a fast yet flexible deep-learning-based computer vision system utilizing a GPU. Irmak et al. [64], Costa et al. [65], and Carbajal et al. [66] proposed FPGA-based computer vision systems, whereas Xiong et al. [67] presented an ASIC-based computer vision system with enhanced operational flexibility. To obtain the advantages of the two kinds of platform, a hybrid model, i.e., a platform with a hardware–software combination such as the one presented in [68], can also be utilized for computation.

The state-of-the-art research on the selection of processing platforms aims to serve the ends of efficiency and flexibility of computation. CPU- and GPU-based systems aim to improve efficiency and computational speed, whereas systems based on FPGA and ASICs are intended to enhance operational flexibility.

3.6. Scene Reconstruction

The sensors in the SCN obtain raw sensor data from their respective FOVs. The raw data are then fused together, utilizing spatio-temporal information (obtained through the frame count and relative localization of sensors) associated with the data. The integration

of sensor data into a virtual environment or a scene is called scene reconstruction. R. Szeliski [69] proposed a novel volumetric scene-reconstruction method using a layered structure and multiple depth maps. Martinec et al. [70] proposed 3D reconstruction using an uncalibrated image dataset and a pipelining approach to detect regions of interest (ROIs) and match them using random sample consensus (RANSAC). Peng et al. [71] addressed the network geometry-estimation problem utilized for scene reconstruction and proposed two-view geometry estimation using a local-structure-constraint-based L2-estimation–local-structure-constraint (L2E-LSC) algorithm.

For efficient scene reconstruction, effective point matching is imperative. Brito et al. [72] compared different state-of-the-art point correspondence methods, such as Scale Invariant Feature Transform (SIFT), Fast Retina Keypoints (FREAK), Oriented Fast and Rotated Brief (ORB), Binary Robust Invariant Scalable Keypoints (BRISK), and Speeded-Up Robust Features (SURF). Milani [73] introduced localization-based reconstruction for a heterogeneous network. Aliakbarpour et al. [74] reviewed different scene-reconstruction methodologies and presented a reconstruction method using parametric homography. Wang and Guo [75] presented a reconstruction method using plane primitives of an RGB-D frame. Ma et al. [76] proposed a mesh-reconstruction system using an adaptive octree division algorithm for point-cloud segmentation and mesh relabeling and reconstruction for scene reconstruction. Ichimru et al. [77] presented 3D scene reconstruction using a CNN under water, utilizing transfer learning with a bubble dataset to avoid distortions.

3.7. Data Processing

The overall objective of a computer vision system is to detect an activity or event of interest, perform processing on the data containing information about the event, develop an understanding of the event, and generate an action by way of a control unit based on the understanding. The visual processing includes a number of stages. Recent advancements in the most common stages of visual processing in computer vision systems are discussed hereinbelow.

3.7.1. Object Detection

The first step towards the visual processing of data in a computer vision system is detection of the OOI (also referred to as the foreground). Some traditional object-detection methods are the Viola and Jones technique [78], scale-invariant feature transformation (SIFT) [79], HOG-based detection [80], optical flow [81,82], and background subtraction [83]. Most of the recent computer vision systems involve machine-learning-based object detection, including neural-network-based object detection [84], "you only look once" (YOLO) [85], region proposals (R-CNN) [86], single-shot refinement neural networks (RefineDet) [87], Retina-Net [88], and single-shot multi-box detectors (SSDs) [89]. The recent neural-network-based object-detection methods provide much better accuracy as compared to the traditional detection methods, but they are highly dependent on the training data. A survey of the evolution of detection techniques in computer vision from probabilistic-prediction approaches to advanced machine-learning approaches is presented in [2].

Some challenges in object detection arise due to dynamic illumination changes, the movement of objects, and occlusions. Roy and Ghosh [90] proposed an adaptive background model (a histogram min–max bucket) using a single sliding window to add adaptability to the background detection. The adaptive background model used a median-finding algorithm incorporated to handle dynamic illumination changes. Bharti et al. [5] proposed an adaptive real-time occlusion-handling kernelized correlation framework for UAVs capable of updating location and boundary information based on the confidence values of the tracker. Min et al. [91] proposed a multiple-object-detection approach using pixel lifespan to blend ghost shadows to the background and a classifier based on a state vector machine (SVM) and a convolutional neural network (CNN) to avoid occlusions.

Detecting an object in a moving FOV is a tedious task and requires substantial approximations. Wu et al. [92] proposed an effective computational model to solve this problem. They evaluated a coarse foreground using singular-value decomposition and

reconstructed the background using the foreground information obtained through a fast in-painting technique. Mean-shift segmentation was used for further refinement of the foreground. Hu et al. [93] presented a tensor-based approach to detect mobile objects without changing the scene dynamics. For initial foreground detection, saliently fused sparse regularization was used and tensor nuclear norms were utilized to handle background redundancy. The foreground was further improved using a 3D locally adaptive regression kernel, which was used to compute spatio-temporal variations. A year-wise representation of the abovementioned object detection techniques and their advantages is shown in Table 7.

Table 7. Object-detection techniques.

Ref.	Year	Methodology	Advantages
[83]	1989	Background subtraction	Low computational complexity
[78]	2001	Viola and Jones technique	Low processing latency with high detection rate
[80]	2005	HOG-based detection	Precise object detection and classification
[79]	2012	Scale-invariant feature transformation	Efficient detection and localization of duplicate objects under extreme occlusion
[81]	2013	Optical flow	Accurate detection of moving objects
[86]	2014	Region proposals (R-CNNs)	High accuracy and precision for object detection
[92]	2015	Background subtraction and mean shift	Refined and precise foreground detection
[85]	2016	"You only look once"	Low latency multi-object detection
[89]	2016	Deep-neural-network-based SSD	Prediction-based detection for variable shapes of objects
[93]	2016	Tensor flow	Detection of mobile objects in FOVs
[84]	2017	Neural network	Multi-object detection with variable shapes
[90]	2017	Adaptive background subtraction model	Better accuracy as compared to traditional background subtraction
[91]	2017	State-vector machine and CNN-based classifier	Multiple-object-detection approach to detect ghost shadows and avoid occlusions
[82]	2018	Optical flow	Accurate detection of moving objects
[87]	2018	Single-shot refinement neural network	High detection accuracy
[5]	2018	Kernelized correlation framework	Real-time occlusion handling
[88]	2019	Retina-Net	Balanced detection performance in terms of latency, accuracy, and precision of detection

3.7.2. Object Classification and Tracking

The OOI can be classified after detection on the basis of one or more appearance parameters [94] called features that classify the object based on color, texture, shape, pixel motion, etc. Some basic features for object representation presented in [95] are points, shapes, and silhouettes or contours. Conventionally, object-classification methods [96] can be categorized as decision-based, statistical-probability-based, and soft-computing-based techniques. Some common decision-based classification methods are decision trees [97,98] and random forests [99]. Bayesian classification [100–102], discriminant analysis [103], logical regression [104], and nearest-neighbor [105] approaches use statistical probability for classification. State vector machines [106], multi-layered perceptrons [107], and neural networks [108,109] use soft computing for classification.

One of the major challenges in object tracking is distortion of the OOI. Villiers et al. [110] proposed real-time inverse distortion for distortion correction. A lot of methods for distortion correction and calibration use properties of vanishing points, as was first proposed in [111]. A distributed algorithm proposed by Caprile et al. [35] illustrated the utilization of tracking waking humans as poles to derive vanishing points for radial distortion

correction and self-calibration. Radial distortion correction was addressed in [112,113] by estimating the center of distortion. Huang et al. [114] proposed linear-transformation-based radial distortion correction, whereas Zhao et al. [115] used a pipelined process for radial distortion correction. Methods for the correction of radial as well as tangential distortion were proposed in [116–118]. Yang et al. [119] proposed estimation and correction of perspective distortion utilizing depth information. Color-calibration theory, discussed by Finlayson et al. in [120], has been used to address the challenge of optical distortion. In [121], Wong et al. presented a color-calibration approach using a multi-spectral camera model.

Another challenge in object tracking is to obtain the best possible information for each object of interest in the scene while dealing with multiple OOIs in the scene. One of the factors affecting the performance of this task is the motion of the camera nodes to capture the objects precisely. The motion blur sometimes increases to such an extent that the effect of moving the camera in the environment is nullified. Han et al. [122] proposed a motion-aware tracker to address the abovementioned problem by filling in the tracking fragments caused by occlusion or blur. Meinhardit et al. [123] proposed a former tracking system to address the challenges of multi-object tracking.

3.7.3. Object Re-Identification

Object re-identification encounters multiple challenges through occlusions, false object detection due to ghost shadows [124], illumination changes, and change in viewpoints. Zhang et al. [125] proposed an adaptive re-identification framework for spatio-temporal alignment utilizing Fisher vector learning to address illumination changes in the re-identification of OOIs. Yang et al. [126] used logical determinant metric learning to tackle re-identification through different camera views to overcome occlusions in the re-identification of objects. Multiple features, when fused together, improve re-identification capabilities; however, the role of each feature and the weights are critical. Geng et al. [127] proposed a feature-fusion method based on weighted-center graph theory to obtain the role of each feature in re-identification. Yang et al. [128] presented re-identification using partial information which can be utilized with occluded data.

3.7.4. Pose and Behavior Estimation

Pose-estimation methods utilize models to associate the patterns of postures, poses, or shapes of OOIs detected and track them to obtain meaningful information, thus yielding understanding. The major challenges in pose and behavior estimation are: the selection of a pose-estimation model and the association of information with a sequence of poses obtained through tracking an OOI. Pose-estimation methods can either be model-based and may use kinematic modelling, planar modelling, or volumetric modelling, or they can be model-free. Kinematic models rely on tracking the movements of points on the OOI, planar models rely on contours, whereas volumetric models rely on changes in the volume distributions of OOIs tracked over time. Kinematic models are easy to process but are not reliable. Planar and volumetric models are more reliable but have more computational complexity. Addressing the challenge of low computational complexity and high accuracy in pose and behavior estimation, Chen et al. [129] proposed an anatomically aware 3D pose-estimation model for human behavior analysis. Staraka et al. [130] proposed a kinematic skeletal-model-based pose-estimation method with real-time and accurate behavior estimation.

3.8. Visual Understanding

Systems achieve visual understanding by relating one or more pose behaviors and performing spatio-temporal analyses of the patterns of poses corresponding to events in a scene. Campbell et al. [131] utilized phase-space constraints to depict human motion. Oren et al. [132] used single-frame wavelet templates for pedestrian detection. Image captioning [133], manuscript reviewing in the medical field [134] and academia [135] are some applications of static activity recognition. Nguyen et al. [136] used multi-objective optimization for real-time activity monitoring. In [3,137], dynamic recognition was utilized

for sports analysis. Xiang et al. [138] proposed a system capable of making decisions for multiple-object tracking, whereas Wu et al. [4] proposed a dynamic activity recognition system for smart homes. In [139], Laptev et al. proposed a state-vector-machine-based abnormal human activity recognition system.

4. Contemporary Solutions

From the above, it can be observed that there has been a shift from model-based approaches to artificial-intelligence (AI)-based methods. We have also considered some of the earlier publications to understand the concepts of some of the recently proposed models. Further, it has also been observed that the emerging research fields in active vision systems are multi-object detection and tracking, occlusion handling, and sensor reconfiguration.

Artificial intelligence (AI) drives down the time taken to perform a task. It enables multi-tasking and eases workloads for existing resources. AI further facilitates decision making by making the process faster and smarter. For these reasons, most of the state-of-the-art systems addressing the abovementioned challenges rely on one or more artificial-intelligence (AI)-based approaches [140–156]. According to market research reported in [157], AI in computer vision has a 45% compound annual growth rate, which is the growth rate of research in the area. The main reason to switch from traditional model-driven approaches to artificial-intelligence-based systems is the high level of accuracy AI systems can provide as compared to the former. It has also been observed that most of the AI-based approaches are dependent on machine-learning (ML)- or deep-learning (DL)-based models.

Some of the recently proposed systems utilize the concepts of traditional model-based methods along with modern AI-based approaches to generate highly accurate hybrid systems [147,151], while some systems [153] capable of addressing the challenges through AI-based models utilize the basic principles of traditional model-based methods. Some of the state-of-the-art AI-based systems addressing the challenges of self-reconfiguration faced by active vision systems are presented in Table 8.

ML-based systems highly rely on training datasets to develop operational models. To make an ML-based system adaptive, AI models need to learn the unforeseen by deriving information from experience. Such systems fail drastically in centralized networks, as the nodes are trained on different datasets based on their specific experiences of distinct events in their surroundings.

Thus, the biggest challenge for adaptive systems is the sharing of information about events between the nodes in networks such that each event in the network of networks can be handled with equal accuracy by each node.

ML systems further suffer from visual attacks [158], the most common of which are adversarial attacks [159,160] which misguide the classifier and thus impact the accuracy of the system. One study [161] showcases the effects of adversarial attacks on the performance of machine-learning-based approaches. The systems are attacked by visual attacks which make small alterations in the weights of the classifier, thus misleading the classifier and reducing the accuracy of classification. Over a period of time, such attacks result in drastic degradation in the performance of the system. Refs. [162,163] present detailed surveys on the types of visual attacks and the methods proposed to detect and mitigate such attacks. Visual attacks [158] can be targeted if the effect of the attacks is predicted correctly by the model; however, in most scenarios, such attacks introduce random noise and thus are very difficult to reverse, causing permanent degradation in the classifier's performance over time. To overcome the problems of data loss and degradation of system performance due to such attacks, ML systems must be incorporated into a distributed network, such that the data are distributed throughout the entire network. Such a network is capable of detecting an attack in the initial stage itself and providing back-up for any kind of data loss to each participating node in the network.

Table 8. State-of-the-art AI-based approaches addressing reconfiguration and active-vision challenges.

Ref.	Challenge Addressed	AI-Based Approach Used
[140]	Camera calibration	Convolutional neural network (CNN)
[141]		Neural network
[142]	Parameter estimation	Convolutional neural network (CNN)
[143]		Deep neural network (DNN)
[144]	Pose estimation	Neural network
[145]	Object detection	Modified CNN
[146]	Object tracking	Residual neural network
[147]		Deep CNN and Kalman filter
[148]		Deep neural network (DNN)
[149]		CNN and deep sort
[150]		Deep-learning-based CNN
[6]	Activity detection	Slow–fast CNN
[151]		Neural network and strider algorithm
[152]	Object re-identification	CNN
[153]		Sparse graph-wavelet-based CNN
[154]	Object re-identification and occlusion handling	Deep-neural-network-based transfer learning
[155]	Localization	CNN
[156]		Neural network

5. Self-Adaptation and Self-Reconfiguration

The challenges discussed above have received solutions designed to tackle particular problems in well-known settings. However, none of the aforementioned approaches addresses the challenge with unforeseen conditions optimally due to unexpected changes in the environment. To overcome this limitation, an active vision system must be capable of understanding the changes in its environment and reconfiguring the parameters of the active sensors participating in the system on its own. Further, active vision systems must also enable the sharing of information with each other to handle unforeseen changes in a better way.

An active vision system capable of self-adjusting its configuration space is called a "*self-reconfigurable*" system. The ability of an SCN to adapt to such changes and reconfigure its parameters for optimized performance in dynamic and unforeseen conditions is called self-reconfiguration of the SCN [164]. Such a self-reconfigurable SCN for UAVs used for surveillance was proposed by Leong et al. in [165]. A comparative survey presented by Natarajan et al. [166] illustrates a number of self-reconfiguration models for computer vision systems with multiple active nodes participating in data extraction. Martinel et al. [167] proposed the distributed self-reconfiguration of an SCN to address the vehicle re-identification problem utilizing deep-learning models.

The capability of a system to share information to learn and adapt can be achieved through "*self-adaptation*". Self-adaptation is the ability of a network (in this case, specifically an SCN) to enhance its performance by enabling the network to update the configuration space of nodes, active participants, protocols, and functional algorithms. Some systems configured for self-adaptation require learning from past experience or the sharing of performance statistics from different parts of the network and opt for the settings which worked best for similar events.

Self-adaptation can be achieved through self-expression and self-awareness capabilities in a system. A system having self-awareness (SA) possesses knowledge of its state;

self-awareness gives the system the ability to share its state in the form of parameters and overall quality of service (QoS). Additionally, each node in the self-adaptive system possesses the ability to question its current state, investigate alternative configurations for better QoS, and change its state using active learning.

Rinner et al. [168] used a market-based approach for self-awareness and proposed six major steps for self-adaptation, namely, resource monitoring, object tracking, topology learning, object handover, strategy selection and objective formation. Lewis et al. [169] classified events as explicit and implicit and discussed the privacy, extent, and quality of self-adaptation. Lewis and Chandra [170] discussed formal models for self-adaptation and the application of self-adaptation in systems with artificial intelligence, conceptual systems, engineering, automotive systems, computing, etc. Wang et al. [171] discussed methods of self-adaptation with capabilities of online learning. In [172], Ali et al. proposed an auto-adaptive multi-stream architecture using multiple heterogeneous sensors with pipelined switches between processing states and ideal states to reduce power, using an FPGA implementation that demonstrated inter-frame adaptation capability with a relatively low overhead. Guettalfi et al. [173] proposed an architecture utilizing quality of service (QoS), resource estimation, a feedback mechanism, and state estimation for public and private self-awareness using actuators. Zhu et al. [174] and Lin et al. [175] proposed a self-adaptation-based person-re-identification system based on unsupervised learning. Wu et al. [176] and Rudolph et al. [177] proposed an adaptive self-reconfiguration framework for computer vision systems. Both frameworks proposed the sharing of information between the sensors in a network for its utilization. However, both adaptive self-reconfiguration frameworks were designed for centralized network configurations and thus possess limited scopes of learning.

An adaptive self-reconfigurable framework presented hereinbelow, when utilized for an SCN-enabled active vision system, can provide adaptive calibration of SCN sensor parameters in near real time. Further, in the data-processing part of the system, the framework can be utilized to adapt to the best specifications by learning from the experiences obtained from other sources and thus result in the minimization of re-iterative training for models that are specific to the development of scene understanding.

6. Adaptive Self-Reconfiguration Framework

Piciarelli et al. [8] proposed a dynamic reconfiguration framework for SCNs, as shown in Figure 3. The framework used a local state (f), resources (r), and QoS information (q) from a number of nodes to generate the overall state (F), overall resources (R), and overall quality (Q) of the system, utilizing an SCN. The dynamic reconfiguration framework showcased the functionality of a reconfigurator, which was configured to determine changes in parameters based on a resource model and the objectives of the system. The dynamic reconfiguration framework in [8] was designed for an independent SCN system working in a centralized environment and thus lacks adaptiveness to deal with unforeseen events. The framework needs to reconfigure itself from scratch if a new type of activity is discovered and thus is not suitable for real-time active-vision applications. Further, almost all the symmetric/asymmetric systems and frameworks discussed in Sections 3 and 4 lack adaptiveness and thus are likely to fail miserably in unforeseen conditions. Recent reconfiguration systems [177–179] based on the reconfigurator model of [8] have tried to improve the accuracy of detection; however, due to centralized networks of operation, they have limited scopes for learning and are not suitable to tackle unforeseen conditions in real time.

As discussed earlier, the reconfiguration of sensor nodes is highly reliant on the scene understanding gained by the processing of sensor data and vice versa. Due to this interdependency, it is very challenging for a system to reconfigure its configuration space while dealing with unforeseen situations. With prior knowledge of an event, scene understanding can be greatly improved and thus the reconfiguration of the sensor network.

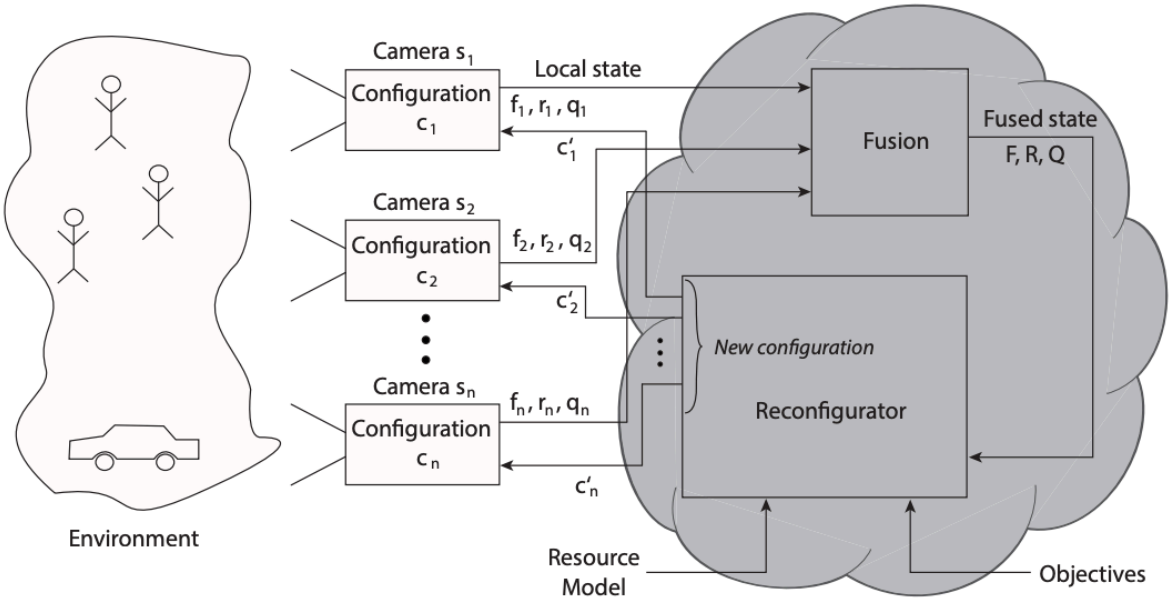


Figure 3. Self-reconfiguration framework as proposed in [8].

This article proposes an adaptive self-reconfiguration (ASR) framework to extend the scope of the framework proposed by Piciarelli et al. [8] to a number of networks, such that each network is configured to learn and reconfigure adaptively by utilizing the experiences and models of others to obtain the best visual understanding possible and thus perform optimal reconfiguration of the SCN. In spite of using a centralized reconfigurator, we propose a distributed network of systems comprising a number of datacenters or cloud servers to perform data computation and reconfiguration. For illustration, the architecture is inspired by the NEAR blockchain network, comprising a number of datacenters, such that each datacenter is capable of sharing datasets, performance parameters, and even trained models. Due to the symmetrical architecture of the distributed blockchain network, the proposed ASR system provides unbiased functionality for each smart camera network utilizing the network. This allows the model, datasets, and parameters to be used by any SCN to deal with an unforeseen situation if a similar condition has been encountered and dealt with by another SCN in the distributed network. Further, the ASR framework proposes the distribution of critical data throughout the blockchain network, providing data security, such that if an SCN is attacked by adversarial attack, the critical data can be retrieved. In some respects, the network of systems developed or deployed using the ASR framework utilizing a distributed network (blockchain) can be considered a self-adaptive system of active vision systems.

Model

An exemplary embodiment of the proposed ASR-framework-based architecture for deployment of the adaptive self-reconfiguration of active vision systems utilizing SCNs is shown in Figure 4. The framework consists of "m" number of active vision systems (AVSs) connected together in a distributed network. The architecture comprises "m" number smart camera networks coupled to each other by way of a distributed blockchain network "B". Sensed data and a local configuration space are associated with each sensor node of each SCN, which results in input data from each sensor being processed for reconfiguration by one or more datacenters in the blockchain network. Input data, local states, resources, and quality of service information from each sensor "s_i" are cumulatively passed to a fusion block of the SCN (represented by the set {Ci}) to obtain an overall system configuration space for the SCN. Each SCN has an overall functionality, defined by a number of objectives. The sensed data, the system's configuration space, and the SCN objectives are communicated as self-expression data of the SCN "E_i" to the blockchain network "B" for processing.

The blockchain network comprises "M" number of datacenters (cloud servers) that are configured to perform computations on the self-expression data (E_i) for each SCN. A datacenter is selected as per the objectives of the SCN, based on the resources available at the datacenter. The datacenter generates a model based on the objectives of the SCN and derives an understanding by way of the sensor's data using unsupervised artificial intelligence. Each activity and/or event determined by a model (by the datacenter) is assigned a pattern vector for identification, such that the pattern vector is generated based on the objectives and the understanding of the event by the model. The model further derives a best-possible configuration space corresponding to the event iteratively. Each pattern vector along with the model specifications is distributed to every datacenter in the blockchain network "B". Each event detected by any datacenter generates a pattern vector that is distributed throughout the entire blockchain "B". Based on the understanding of the detected activity, the selected datacenter generates self-awareness data (A_i) comprising a revised configuration space set $\{C_i^{'}\}$ for the SCN. Based on the revised configuration space set, each camera sensor (s_{ij}) is calibrated.

If an unforeseen condition is faced by another SCN in the network, the associated datacenter generates a pattern vector and compares the pattern vector with a pre-existing pattern vector in the blockchain network. Based on the comparison of the pattern vector with the pre-existing pattern vectors, a suitable datacenter is allotted and thus the latency of iterative processing is minimized. The improved configuration space is reverted back to the SCN through self-awareness data that are utilized by the reconfigurator to calibrate the parameters of each sensor of the SCN.

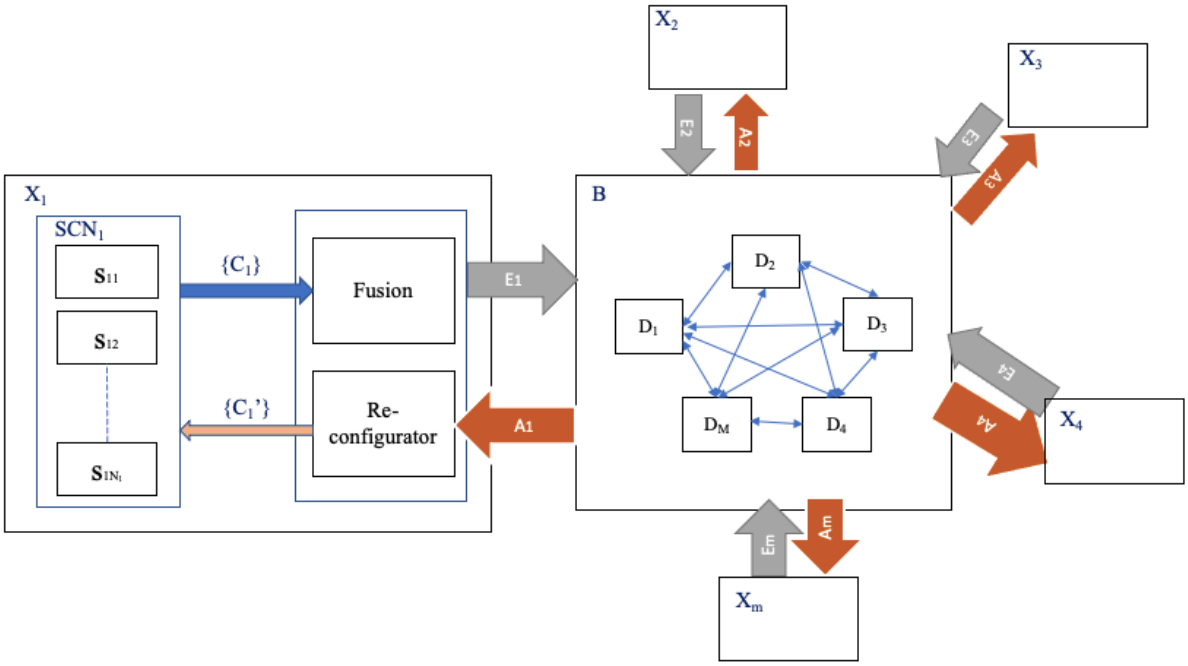


Figure 4. Adaptive self-reconfiguration framework for SCN-enabled active vision systems.

Notation used:
X_i: ith smart camera network;
s_{ij}: jth camera sensor of the ith active vision system;
$\{C_i\}$: Set of input data for the ith SCN;
$\{C_i^{'}\}$: Set of output data for the ith SCN;
E_i: Self-expression data for the ith SCN;
A_i: Self-awareness data for the ith SCN;
n_i: Number of camera sensors in the ith active vision system;
m: Number of active vision systems utilizing the ASR framework;

B: Distributed blockchain network comprising "M" number of datacenters;

D_k: kth datacenter in the distributed blockchain network.

To generate the pattern vector corresponding to each event, we propose an auto-encoder-based unsupervised learning model, such that, for each event, the sensor data received by a selected datacenter is converted to lower-dimensional data. The auto-encoder model determines a reconstruction error to restore the sensor data back to their original form. The reconstruction error is propagated to the auto-encoder to iteratively alter the weights of the model for training in order to minimize the reconstruction error. The data of weights of the trained model along with the objectives of the SCN are utilized to generate the pattern vector, such that the pattern vector is distributed throughout the blockchain. The datacenter further computes the reconfiguration parameters based on the understanding of the event and sends the reconfiguration parameters to the SCN in the form of self-expression data, which are utilized by the reconfigurator of the SCN to calibrate the parameters. If an identical event occurs in any other SCN in the blockchain, the selected datacenter generates a pattern vector for the event. The pattern vector is matched with the pre-existing pattern vectors in the blockchain, and the reconfiguration model of the closest pattern vector is shared with the SCN.

For an unbiased and seamless flow of operations, the distributed blockchain network protocols were designed based on consensus mechanisms. The datacenters are categorized as processor datacenter nodes and validator datacenter nodes. The categorization of the datacenters is based on the proof-of-active-participation (POAP) consensus mechanism [180]. The processor datacenter nodes generate a model and a pattern vector based on the sensor data, whereas the validator nodes validate the model and the pattern vectors. The processor nodes further share details of their resource consumption (bandwidth and computational resources, etc.) to obtain rewards. The validator nodes validate or authenticate the details about resource consumption, models, and pattern vectors generated by the processor nodes. The validator nodes further compare the pattern vector generated by the processor node with the pre-existing pattern vectors to determine the best suitable reconfiguration configurations. To ensure the unbiased functionality of the validator and the processor nodes, each participating processor node and validator node is required to raise a stake based on a proof-of-stake consensus mechanism [181]. Upon validation of the model and the reconfiguration, the processor nodes and the validator nodes receive their stake as well as a reward from the corresponding SCN network. A flow diagram of the process is shown in Figure 5.

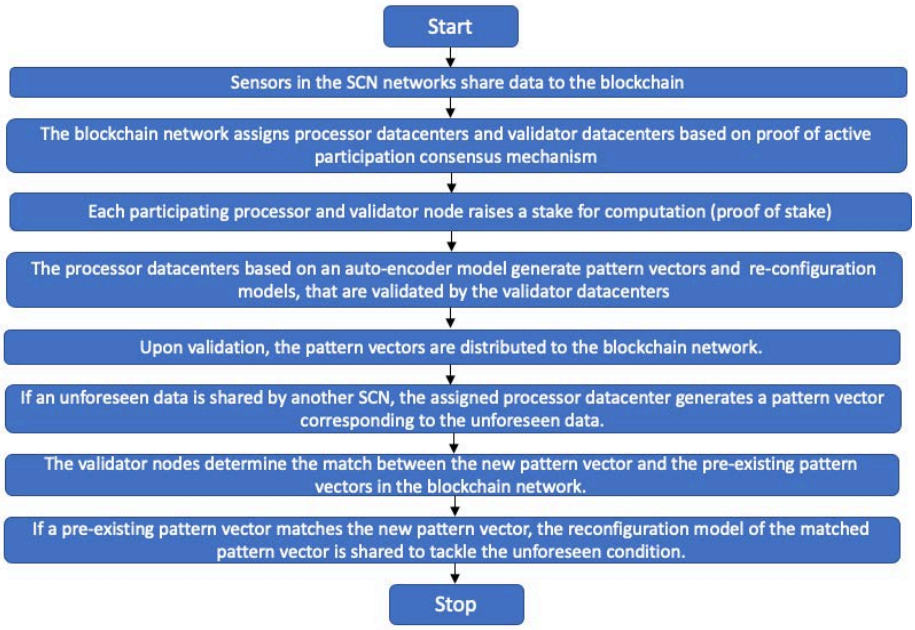


Figure 5. Flow diagram illustrating the adaptive reconfiguration using the ASR framework.

7. Results

For illustration of the functionality of the adaptive self-reconfiguration (ASR) framework, we utilized multiple surveillance video datasets and simulated results, comparing the centralized reconfiguration of [8] to our proposed distributed adaptive self-reconfiguration. Each frame of each video dataset was standardized to a resolution of 640 × 360 pixels. Activity maps were generated using the centralized approach presented in [8] and the proposed ASR, respectively, such that the activity maps updated with each frame added to an event. The performances of the systems were compared in terms of multi-object tracking accuracy (MOTA) results obtained for both activity maps. Further, due to the unavailability of resources necessary to develop a private blockchain network with datacenters, we utilized lower processing capabilities and thus, in spite of accurate processing latency, the results are represented with respect to training cycles (Ti) (each of 15 minutes duration); however, the processing capabilities of a datacenter are much higher and can be utilized to achieve results in near real time. We utilized standard regions with a convolutional neural network (R-CNN) model for multi-object detection by the centralized system of [8] as well as the distributed system based on the proposed ASR framework. All the results presented hereinbelow were derived using the MatLab Image Processing Toolbox. The activity map corresponding to the proposed ASR was further utilized to predict the upcoming event in the next frame. The SAR framework was utilized for multi-object detection by the exemplary embodiment method described in Section 6; however, it can be utilized to enhance any performance parameter, as illustrated in Section 2.

True-positive pixel count (TPC), false-positive pixel count (FPC), false-positive pixel-detection rate (FPR), true-negative pixel count (TNC), true-negative pixel-detection rate (TNR), and false-negative pixel count (FNC) were used as primary performance parameters to obtain MOTA values using Equation (1).

$$\text{MOTA (\%)} = (\text{Total count of pixels} - \text{falsely detected pixels}) \times 100 \quad (1)$$

7.1. Surveillance Dataset 1

A comparison of the performance of the system presented in [8] and the proposed ASR system, tested on surveillance dataset 1 in terms of MOTA, is presented in Table 9 and Figure 6. The predictions of the directions of vehicle movement in various random frames from surveillance dataset 1 based on the generated activity map are presented in Figure 7.

Table 9. Comparison of the performance for the system in [8] and the proposed ASR for surveillance dataset 1.

Pixels: 640 × 360		**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**
[8]	TPC	30,541	41,556	52,956	56,871	60,279	74,638	76,267
	TNC	18,719	20,268	28,514	33,400	39,277	35,098	41,905
	FPC	80,271	74,352	69,183	66,418	63,116	60,947	59,104
	FNC	1,00,869	94,024	79,747	73,711	67,728	59,717	53,124
	MOTA (%)	**21.38**	**26.92**	**35.36**	**39.18**	**43.21**	**47.63**	**51.29**
	TRAINING CYCLES TO OBTAIN ABOVE 80% MOTA: **18**							
ASR	TPC	63,018	69,217	72,141	76,238	78,908	79,519	86,211
	TNC	33,128	37,320	41,169	46,473	48,042	52,793	51,775
	FPC	47,982	43,755	40,073	36,117	36,431	31,824	29,273
	FNC	86,272	80,108	77,017	71,512	67,019	66,264	63,141
	MOTA (%)	**41.73**	**46.24**	**49.18**	**53.26**	**55.10**	**57.34**	**59.89**
	TRAINING CYCLES TO OBTAIN ABOVE 80% MOTA: **12**							

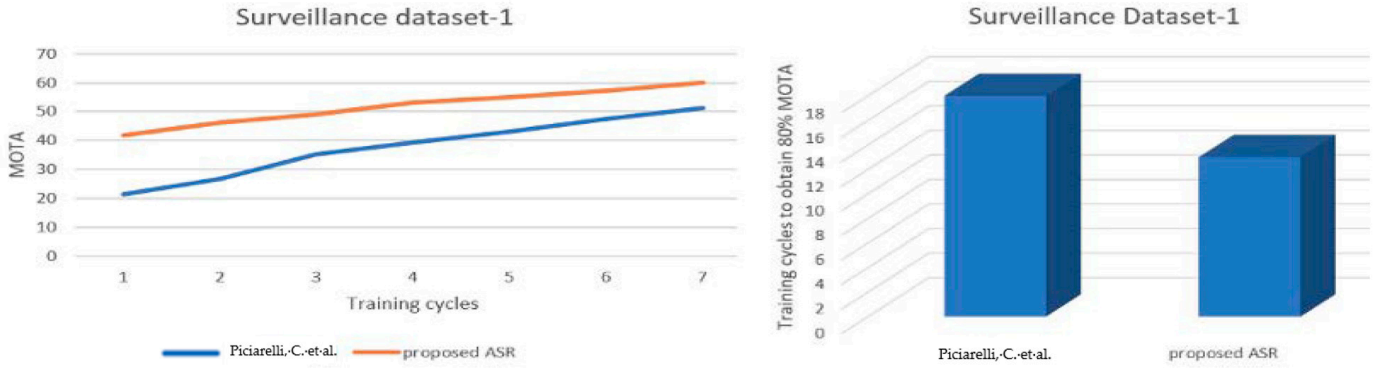


Figure 6. Comparison of the performances of the system in [8] and the proposed ASR for surveillance dataset 1.

Figure 7. Predicted motion of multiple objects in random frames from surveillance dataset 1.

7.2. Surveillance Dataset 2

A comparison of the performances of the system in [8] and the proposed ASR, tested on surveillance dataset 2 in terms of MOTA, is presented in Table 10 and Figure 8. The predictions of the directions of vehicle movement in various random frames from surveillance dataset 2 based on the generated activity map are presented in Figure 9.

Table 10. Comparison of the performances of the system in [8] and the proposed ASR for surveillance dataset 2.

		T1	T2	T3	T4	T5	T6	T7
[8]	TPC	23,211	27,324	29,841	33,266	36,421	39,972	41,101
	TNC	50,080	58,477	66,581	69,515	76,451	77,601	83,591
	FPC	89,233	85,161	82,686	79,957	75,277	72,098	69,035
	FNC	67,876	59,438	51,292	47,662	42,251	40,729	36,673
	MOTA (%)	**31.81**	**37.24**	**41.85**	**44.61**	**48.99**	**51.03**	**54.12**
	TRAINING CYCLES TO OBTAIN ABOVE 80% MOTA: **15**							
ASR	TPC	38,211	40,128	41,007	42,091	43,108	43,236	43,901
	TNC	70,860	77,652	83,017	84,952	90,317	92,884	96,666
	FPC	77,102	71,928	69,982	67,041	66,101	65,384	63,687
	FNC	44,227	40,692	36,394	36,316	30,874	28,896	26,146
	MOTA (%)	**47.34**	**51.12**	**53.83**	**55.14**	**57.91**	**59.08**	**61.01**
	TRAINING CYCLES TO OBTAIN ABOVE 80% MOTA: **11**							

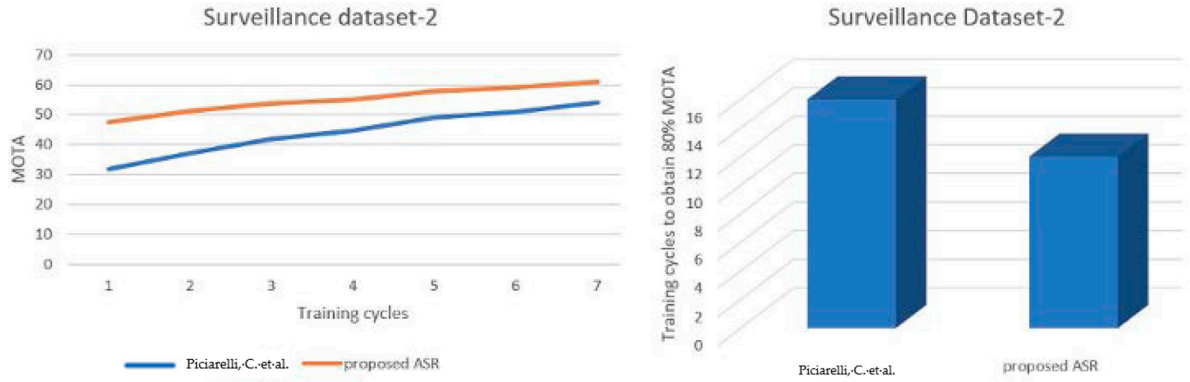


Figure 8. Comparison of the performances of the system in [8] and the proposed ASR for surveillance dataset 2.

Figure 9. Predicted motions of multiple objects in random frames from surveillance dataset 2.

A comparison of the performances of the system presented in [8] and the proposed ASR for multi-object tracking applications are represented in Figures 6 and 8. It must be noted that, as the reconfiguration is dependent on tracking accuracy, the reconfiguration presented by the distributed ASR model is better than that of the centralized reconfigurator model of [8].

8. Conclusions and Scope

The performance of active vision systems and reconfiguration of the sensors providing data to active vision systems are interdependent. However, the reconfiguration of the calibration space of an active vision system employing a network of sensors can be challenging. Most of the state-of-the-art active vision systems fail miserably in dealing with unforeseen conditions, as the reconfiguration model takes time to adapt to the new conditions to develop an understanding of an unforeseen event. Thus, reconfiguring such a system in real time is nearly impossible. Further, most active vision systems nowadays rely on artificial-intelligence-based models for the processing of sensor data to develop an understanding of an event. However, such models are prone to adversarial attacks and are thus threatened with data loss. Therefore, such systems cannot be relied on for making critical decisions in real time.

This article discusses the challenges at different operational levels in deploying an active vision system employing a camera network. This article presents a detailed description of the systems and methods proposed for addressing the challenges with respect to both data processing and reconfiguration, along with the state-of-the-art solutions for the same, proposing an adaptive self-reconfiguration (ASR) framework employing a blockchain-based distributed network for data processing and reconfiguration of the sensor network's configuration space. To make the understanding of the framework easier, this article has briefly defined the concepts of self-adaptation and self-reconfiguration in systems prior to the ASR framework.

The blockchain network of the ASR-framework-based architecture acts as a system of systems, connecting a number of smart camera networks together in a distributed

architecture. The blockchain further includes a number of datacenters configured to obtain data from the SCNs and perform data processing to obtain understanding about the events and activities in the scene. The datacenters further generate pattern vectors corresponding to the event or activity detected and distribute pattern vectors along with corresponding reconfiguration models to each datacenter in the distributed blockchain network, such that if a similar event is observed at any other SCN in the network, the reconfiguration model associated with a pre-existing pattern can be utilized for a much faster reconfiguration of the SCN. The blockchain network is founded on proof-of-stake and proof-of-active-participation consensus mechanisms to maintain an unbiased and smooth flow of operations between all the participating datacenters in the network. Further, due to the distributed architecture, critical data float in the blockchain network and thus the effect of adversarial attacks is minimized.

This article further compared the performance of a centralized active vision system with a distributed system based on the proposed ASR framework for multi-object tracking and showcased enhanced tracking performance in terms of multi-object tracking accuracy and low latency. The proposed framework was tested on a homogeneous environment, with some limitations and assumptions; however, the ASR framework aims to be developed for heterogeneous systems to enhance the scope of its applications in future.

Author Contributions: S.: Lead author of the manuscript (corresponding author), conceptualization and methodology, writing—original draft preparation, investigation, and editing; I.S.: Second author, research design, guidance, and reviewing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Training of both models (i.e., for [8] and the proposed ASR) was performed on a standard surveillance video dataset. The models were further tested on surveillance dataset 1 and surveillance dataset 2. The standard surveillance dataset, as well as surveillance dataset 1 and surveillance dataset 2, are non-restricted and free-to-use third-party datasets with 640 × 360-pixel resolutions. The datasets and the training and testing models can be accessed by clicking on this link: https://drive.google.com/drive/folders/1ZizuTfDZ2gT-ro1o2lAAo_SVsw4_MY-7 (accessed on 23 October 2022).

Acknowledgments: This work was carried out under the supervision of Indu Sreedevi in the Department of ECE, Delhi Technological University, New Delhi, India, and Shashank expresses immense gratitude to his guide and UGC for enlightening him throughout the process.

Conflicts of Interest: The authors declare that they have no conflicts 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 article.

References

1. Reisslein, M.; Rinner, B.; Roy-Chowdhury, A. Smart Camera Networks. *Computer* **2014**, *47*, 23–25.
2. Zhang, T.; Aftab, W.; Mihaylova, L.; Langran-Wheeler, C.; Rigby, S.; Fletcher, D.; Maddock, S.; Bosworth, G. Recent Advances in Video Analytics for Rail Network Surveillance for Security, Trespass and Suicide Prevention—A Survey. *Sensors* **2022**, *22*, 4324. [CrossRef] [PubMed]
3. Theagarajan, R.; Pala, F.; Zhang, X.; Bhanu, B. Soccer: Who has the ball? Generating visual analytics and player statistics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1749–1757.
4. Wu, C.; Khalili, A.H.; Aghajan, H. Multiview activity recognition in smart homes with spatio-temporal features. In Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras, New York, NY, USA, 31 August–4 September 2010; pp. 142–149.
5. Bharati, S.P.; Wu, Y.; Sui, Y.; Padgett, C.; Wang, G. Real-time obstacle detection and tracking for sense-and-avoid mechanism in UAVs. *IEEE Trans. Intell. Veh.* **2018**, *3*, 185–197. [CrossRef]
6. Agarwal, M.; Parashar, P.; Mathur, A.; Utkarsh, K.; Sinha, A. Suspicious Activity Detection in Surveillance Applications Using Slow-Fast Convolutional Neural Network. In *Advances in Data Computing, Communication and Security*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 647–658.
7. Hanson, A. (Ed.) *Computer Vision Systems*; Elsevier: Amsterdam, The Netherlands, 1978.

8. Piciarelli, C.; Esterle, L.; Khan, A.; Rinner, B.; Foresti, G.L. Dynamic reconfiguration in camera networks: A short survey. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *26*, 965–977. [CrossRef]
9. Jesus, T.C.; Costa, D.G.; Portugal, P.; Vasques, F. A Survey on Monitoring Quality Assessment for Wireless Visual Sensor Networks. *Future Internet* **2022**, *14*, 213. [CrossRef]
10. Indu, S.; Chaudhury, S.; Mittal, N.R.; Bhattacharyya, A. Optimal sensor placement for surveillance of large spaces. In Proceedings of the 3rd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), Como, Italy, 30 August 2009–2 September 2009; pp. 1–8.
11. Zhang, G.; Dong, B.; Zheng, J. Visual Sensor Placement and Orientation Optimization for Surveillance Systems. In Proceedings of the 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), Krakow, Poland, 4–6 November 2015; pp. 1–5.
12. Da Silva, L.C.; Bernardo, R.M.; De Oliveira, H.A.; Rosa, P.F. Multi-UAV agent-based coordination for persistent surveillance with dynamic priorities. In Proceedings of the International Conference on Military Technologies (ICMT), Brno, Czech Republic, 31 May–2 June 2017; pp. 765–771.
13. Jamshed, M.A.; Khan, M.F.; Rafique, K.; Khan, M.I.; Faheem, K.; Shah, S.M.; Rahim, A. An energy efficient priority based wireless multimedia sensor node dynamic scheduler. In Proceedings of the 12th International Conference on High-capacity Optical Networks and Enabling/Emerging Technologies (HONET), Islamabad, Pakistan, 21–23 December 2015; pp. 1–4.
14. Vejdanparast, A. Improving the Fidelity of Abstract Camera Network Simulations. Ph.D. Thesis, Aston University, Birmingham, UK, 2020.
15. Wang, X.; Zhang, H.; Gu, H. Solving Optimal Camera Placement Problems in IoT Using LH-RPSO. *IEEE Access* **2020**, *8*, 40881–40891. [CrossRef]
16. Redding, N.J.; Ohmer, J.F.; Kelly, J.; Cooke, T. Cross-matching via feature matching for camera handover with non-overlapping fields of view. In Proceedings of the 2008 Digital Image Computing: Techniques and Applications, Canberra, ACT, Australia, 1–3 December 2008; pp. 343–350.
17. Esterle, L.; Lewis, P.R.; Bogdanski, M.; Rinner, B.; Yao, X. A socio-economic approach to online vision graph generation and handover in distributed smart camera networks. In Proceedings of the 5th ACM/IEEE International Conference on Distributed Smart Cameras, Ghent, Belgium, 22–25 August 2011; pp. 1–6.
18. Lin, J.L.; Hwang, K.S.; Huang, C.Y. Active and Seamless Handover Control of Multi-Camera Systems With 1-DoF Platforms. *IEEE Syst. J.* **2012**, *8*, 769–777.
19. Hall, E.L.; Tio, J.B.; McPherson, C.A.; Sadjadi, F.A. Measuring curved surfaces for robot vision. *Computer* **1982**, *5*, 42–54. [CrossRef]
20. Tsai, R. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE J. Robot. Autom.* **1987**, *3*, 323–344. [CrossRef]
21. Faugeras, O.D. *The Calibration Problem for Stereo*; CVPR: Miami, FL, USA, 1986; pp. 15–20.
22. Weng, J.; Cohen, P.; Herniou, M. Camera calibration with distortion models and accuracy evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *10*, 965–980. [CrossRef]
23. Durrant-Whyte, H.; Bailey, T. Simultaneous localization and mapping: Part I. *IEEE Robot. Autom. Mag.* **2006**, *13*, 99–110. [CrossRef]
24. Durrant-Whyte, H.; Bailey, T. Simultaneous localization and mapping (SLAM): Part II. *IEEE Robot. Autom. Mag.* **2006**, *13*, 108–117. [CrossRef]
25. Özyeşil, O.; Voroninski, V.; Basri, R.; Singer, A. A survey of structure from motion. *Acta Numer.* **2017**, *26*, 305–364. [CrossRef]
26. Fox, D.; Burgard, W.; Dellaert, F.; Thrun, S. Monte carlo localization: Efficient position estimation for mobile robots. *Am. Assoc. Artif. Intell.* **1999**, *1999*, 1–7.
27. Mantzel, W.E.; Hyeokho, C.; Richard, G.B. Distributed camera network localization. In Proceedings of the Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 7–10 November 2004; Volume 2, pp. 1381–1386.
28. Brachmann, E.; Rother, C. Learning less is more-6d camera localization via 3d surface regression. In Proceedings of the Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4654–4662.
29. Tang, Z.; Lin, Y.S.; Lee, K.H.; Hwang, J.N.; Chuang, J.H.; Fang, Z. Camera self-calibration from tracking of moving persons. In Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Maxico, 4–8 December 2016; pp. 265–270.
30. Zheng, C.; Qiu, H.; Liu, C.; Zheng, X.; Zhou, C.; Liu, Z.; Yang, J. A Fast Method to Extract Focal Length of Camera Based on Parallel Particle Swarm Optimization. In Proceedings of the 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 9550–9555.
31. Führ, G.; Jung, C.R. Camera self-calibration based on nonlinear optimization and applications in surveillance systems. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *27*, 1132–1142. [CrossRef]
32. Yao, Q.; Sankoh, H.; Nonaka, K.; Naito, S. Automatic camera self-calibration for immersive navigation of free viewpoint sports video. In Proceedings of the 18th International Workshop on Multimedia Signal Processing (MMSP), Montreal, QC, Canada, 21–23 September 2016; pp. 1–6.
33. Li, F.; Sekkati, H.; Deglint, J.; Scharfenberger, C.; Lamm, M.; Clausi, D.; Zelek, J.; Wong, A. Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping. *IEEE Trans. Comput. Imaging* **2017**, *3*, 74–83. [CrossRef]
34. Heikkila, J. Using sparse elimination for solving minimal problems in computer vision. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 76–84.

35. Tang, Z.; Lin, Y.S.; Lee, K.H.; Hwang, J.N.; Chuang, J.H. ESTHER: Joint camera self-calibration and automatic radial distortion correction from tracking of walking humans. *IEEE Access* **2019**, *7*, 10754–10766. [CrossRef]
36. Marinakis, D.; Dudek, G. Topology inference for a vision-based sensor network. In Proceedings of the 2nd Canadian Conference on Computer and Robot Vision (CRV'05), Victoria, BC, Canada, 9–11 May 2005; pp. 121–128.
37. Van Den Hengel, A.; Dick, A.; Hill, R. Activity topology estimation for large networks of cameras. In Proceedings of the IEEE International Conference on Video and Signal Based Surveillance, Sydney, Australia, 22–24 November 2006; p. 44.
38. Detmold, H.; Van Den Hengel, A.; Dick, A.; Cichowski, A.; Hill, R.; Kocadag, E.; Falkner, K.; Munro, D.S. Topology estimation for thousand-camera surveillance networks. In Proceedings of the 1st ACM/IEEE International Conference on Distributed Smart Cameras, Vienna, Austria, 25–28 September 2007; pp. 195–202.
39. Clarot, P.; Ermis, E.B.; Jodoin, P.M.; Saligrama, V. Unsupervised camera network structure estimation based on activity. In Proceedings of the 3rd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), Como, Italy, 30 August–2 September 2009; pp. 1–8.
40. Zou, X.; Bhanu, B.; Song, B.; Roy-Chowdhury, A.K. Determining topology in a distributed camera network. In Proceedings of the IEEE International Conference on Image Processing, San Antonio, TX, USA, 16–19 September 2007; Volume 5, p. V-133.
41. Farrell, R.; Davis, L.S. Decentralized discovery of camera network topology. In Proceedings of the 2nd ACM/IEEE International Conference on Distributed Smart Cameras, Palo Alto, CA, USA, 7–11 September 2008; pp. 1–10.
42. Zhu, M.; Dick, A.; van den Hengel, A. Camera network topology estimation by lighting variation. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA), Adelaide, Australia, 23–25 November 2015; pp. 1–6.
43. Mali, G.; Sudip, M. TRAST: Trust-based distributed topology management for wireless multimedia sensor networks. *IEEE Trans. Comput.* **2015**, *65*, 1978–1991. [CrossRef]
44. Feigang, T.; Xiaoju, Z.; Quanmi, L.; Jianyi, L. A Camera Network Topology Estimation Based on Blind Distance. In Proceedings of the 11th International Conference on Intelligent Computation Technology and Automation (ICICTA), Changsha, China, 22–23 September 2018; pp. 138–140.
45. Li, Z.; Wang, J.; Chen, J. Estimating Path in camera network with non-overlapping FOVs. In Proceedings of the 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 10–12 November 2018; pp. 604–609.
46. Kansal, A.; Srivastava, M.B. An environmental energy harvesting framework for sensor networks. In Proceedings of the 2003 International Symposium on Low Power Electronics and Design, Seoul, Korea, 25–27 August 2003; pp. 481–486.
47. Bramberger, M.; Quaritsch, M.; Winkler, T.; Rinner, B.; Schwabach, H. Integrating multi-camera tracking into a dynamic task allocation system for smart cameras. In Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, Como Italy, 15–16 September 2005; pp. 474–479.
48. Bramberger, M.; Rinner, B.; Schwabach, H. A method for dynamic allocation of tasks in clusters of embedded smart cameras. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Waikolova, HI, USA, 12 October 2005; Volume 3, pp. 2595–2600.
49. Karuppiah, D.R.; Grupen, R.A.; Zhu, Z.; Hanson, A.R. Automatic resource allocation in a distributed camera network. *Mach. Vis. Appl.* **2010**, *21*, 517–528. [CrossRef]
50. Dieber, B.; Micheloni, C.; Rinner, B. Resource-aware coverage and task assignment in visual sensor networks. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 1424–1437. [CrossRef]
51. Dieber, B.; Esterle, L.; Rinner, B. Distributed resource-aware task assignment for complex monitoring scenarios in visual sensor networks. In Proceedings of the 6th International Conference on Distributed Smart Cameras (ICDSC), Hong Kong, China, 30 October–2 November 2012; pp. 1–6.
52. Kyrkou, C.; Laoudias, C.; Theocharides, T.; Panayiotou, C.G.; Polycarpou, M. Adaptive energy-oriented multitask allocation in smart camera networks. *IEEE Embed. Syst. Lett.* **2016**, *8*, 37–40. [CrossRef]
53. Wang, Y.; Zhang, J.; Liu, Z.; Wu, Q.; Chou, P.A.; Zhang, Z.; Jia, Y. Handling occlusion and large displacement through improved RGB-D scene flow estimation. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *26*, 1265–1278. [CrossRef]
54. Ouyang, W.; Zeng, X.; Wang, X. Partial occlusion handling in pedestrian detection with a deep model. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *26*, 2123–2137. [CrossRef]
55. Shehzad, M.I.; Shah, Y.A.; Mehmood, Z.; Malik, A.W.; Azmat, S. K-means based multiple objects tracking with long-term occlusion handling. *IET Comput. Vis.* **2016**, *11*, 68–77. [CrossRef]
56. Ur-Rehman, A.; Naqvi, S.M.; Mihaylova, L.; Chambers, J.A. Multi-target tracking and occlusion handling with learned variational Bayesian clusters and a social force model. *IEEE Trans. Signal Processing* **2015**, *64*, 1320–1335. [CrossRef]
57. Chang, J.; Wang, L.; Meng, G.; Xiang, S.; Pan, C. Vision-based occlusion handling and vehicle classification for traffic surveillance systems. *IEEE Intell. Transp. Syst. Mag.* **2018**, *10*, 80–92. [CrossRef]
58. Zhao, S.; Zhang, S.; Zhang, L. Towards occlusion handling: Object tracking with background estimation. *IEEE Trans. Cybern.* **2017**, *48*, 2086–2100. [CrossRef]
59. Liu, Y.; Jing, X.Y.; Nie, J.; Gao, H.; Liu, J.; Jiang, G.P. Context-Aware Three-Dimensional Mean-Shift with Occlusion Handling for Robust Object Tracking in RGB-D Videos. *IEEE Trans. Multimed.* **2018**, *21*, 664–677. [CrossRef]
60. Feng, X.; Jiang, Y.; Yang, X.; Du, M.; Li, X. Computer vision algorithms and hardware implementations: A survey. *Integration* **2019**, *69*, 309–320. [CrossRef]

61. Hørup, S.A.; Juul, S.A. *General-Purpose Computations*; Aalborg University: Aalborg, Denmark, 2011.
62. Guo, Y.; Liu, J.; Li, G.; Mai, L.; Dong, H. Fast and Flexible Human Pose Estimation with HyperPose. In Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 20–24 October 2021; pp. 3763–3766.
63. Tan, S.; Knott, B.; Tian, Y.; Wu, D.J. CryptGPU: Fast privacy-preserving machine learning on the GPU. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 24–27 May 2021; pp. 1021–1038.
64. Irmak, H.; Ziener, D.; Alachiotis, N. Increasing Flexibility of FPGA-based CNN Accelerators with Dynamic Partial Reconfiguration. In Proceedings of the 31st International Conference on Field-Programmable Logic and Applications (FPL), Dredsen, Germany, 1–3 September 2021; pp. 306–311.
65. Costa, A.; Corna, N.; Garzetti, F.; Lusardi, N.; Ronconi, E.; Geraci, A. High-Performance Computing of Real-Time and Multichannel Histograms: A Full FPGA Approach. *IEEE Access* **2022**, *10*, 47524–47540. [CrossRef]
66. Carbajal, M.A.; Villa, R.P.; Palazuelos, D.E.; Astorga, G.J. *Rubio Astorga; Reconfigurable Digital FPGA Based Architecture for 2-Dimensional Linear Convolution Applications*; Identitad Energetica: Madrid, Spain, 2021.
67. Xiong, H.; Sun, K.; Zhang, B.; Yang, J.; Xu, H. Deep-Sea: A Reconfigurable Accelerator for Classic CNN. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 4726652. [CrossRef]
68. Wei, L.; Peng, L. An Efficient OpenCL-Based FPGA Accelerator for MobileNet. *J. Phys. Conf. Ser.* **2021**, *1883*, 012086. [CrossRef]
69. Szeliski, R. Scene Reconstruction from multiple cameras. In Proceedings of the International Conference on Image Processing (ICISP), Vancouver, BC, Canada, 10–13 September 2000; Volume 1, pp. 13–16.
70. Micušık, B.; Martinec, D.; Pajdla, T. 3D metric reconstruction from uncalibrated omnidirectional images. In Proceedings of the Asian Conference on Computer Vision (ACCV'04), Jeju Island, Korea, January 2014.
71. Peng, L.; Zhang, Y.; Zhou, H.; Lu, T. A robust method for estimating image geometry with local structure constraint. *IEEE Access* **2018**, *6*, 20734–20747. [CrossRef]
72. Brito, D.N.; Nunes, C.F.; Padua, F.L.; Lacerda, A. Evaluation of interest point matching methods for projective reconstruction of 3d scenes. *IEEE Lat. Am. Trans.* **2016**, *14*, 1393–1400. [CrossRef]
73. Milani, S. Three-dimensional reconstruction from heterogeneous video devices with camera-in-view information. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Quebec, QC, Canada, 27–30 September 2015; pp. 2050–2054.
74. Aliakbarpour, H.; Prasath, V.S.; Palaniappan, K.; Seetharaman, G.; Dias, J. Heterogeneous multi-view information fusion: Review of 3-D reconstruction methods and a new registration with uncertainty modeling. *IEEE Access* **2016**, *4*, 8264–8285. [CrossRef]
75. Wang, C.; Guo, X. Plane-Based Optimization of Geometry and Texture for RGB-D Reconstruction of Indoor Scenes. In Proceedings of the International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018; pp. 533–541.
76. Ma, D.; Li, G.; Wang, L. Rapid Reconstruction of a Three-Dimensional Mesh Model Based on Oblique Images in the Internet of Things. *IEEE Access* **2018**, *6*, 61686–61699. [CrossRef]
77. Ichimaru, K.; Furukawa, R.; Kawasaki, H. CNN based dense underwater 3D scene reconstruction by transfer learning using bubble database. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikolova Village, HI, USA, 7–9 January 2019; pp. 1543–1552.
78. Viola, P.; Jones, M. Robust real-time object detection. *Int. J. Comput. Vis.* **2001**, *4*, 34–47.
79. Piccinini, P.; Prati, A.; Cucchiara, R. Real-time object detection and localization with SIFT-based clustering. *Image Vis. Comput.* **2012**, *30*, 573–587. [CrossRef]
80. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR'05), San Diago, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
81. Aslani, S.; Mahdavi-Nasab, H. Optical flow based moving object detection and tracking for traffic surveillance. *Int. J. Electr. Comput. Energetic Electron. Commun. Eng.* **2013**, *7*, 1252–1256.
82. Huang, J.; Zou, W.; Zhu, J.; Zhu, Z. Optical flow based real-time moving object detection in unconstrained scenes. *arXiv* **2018**, arXiv:1807.04890.
83. Tougaard, S. Practical algorithm for background subtraction. *Surf. Sci.* **1989**, *216*, 343–360. [CrossRef]
84. Rieke, J. Object detection with neural networks-a simple tutorial using keras. *Towards Data Sci.* **2017**, *6*, 12.
85. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the Computer Vision and Pattern Recognition, Las Vegas, USA, 27–30 June 2016; pp. 779–788.
86. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
87. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single-shot refinement neural network for object detection. In Proceedings of the Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4203–4212.
88. Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra r-cnn: Towards balanced learning for object detection. In Proceedings of the Computer vision and pattern recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 821–830.
89. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
90. Roy, S.M.; Ghosh, A. Real-time adaptive Histogram Min-Max Bucket (HMMB) model for background subtraction. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 1513–1525. [CrossRef]
91. Min, W.; Fan, M.; Guo, X.; Han, Q. A new approach to track multiple vehicles with the combination of robust detection and two classifiers. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 174–186. [CrossRef]

92. Wu, Y.; He, X.; Nguyen, T.Q. Moving object detection with a freely moving camera via background motion subtraction. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *27*, 236–248. [CrossRef]
93. Hu, W.; Yang, Y.; Zhang, W.; Xie, Y. Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition. *IEEE Trans. Image Processing* **2016**, *26*, 724–737. [CrossRef]
94. Parekh, H.S.; Thakore, D.G.; Jaliya, U.K. A survey on object detection and tracking methods. *Int. J. Innov. Res. Comput. Commun. Eng.* **2014**, *2*, 2970–2979.
95. Yilmaz, A.; Javed, O.; Shah, M. Object tracking: A survey. *Acm. Comput. Surv.* **2006**, *38*, 13. [CrossRef]
96. Du, C.-J.; Sun, D.-W. Object Classification Methods. In *Computer Vision Technology for Food Quality Evaluation*; Academic Press: Cambridge, MA, USA, 2016; pp. 87–110.
97. Ankerst, M.; Elsen, C.; Ester, M.; Kriegel, H.P. Visual classification: An interactive approach to decision tree construction. In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diago, CA, USA, 15–18 August 1999; pp. 392–396.
98. Anurag, S.; Han, E.; Kumar, V.; Singh, V. Parallel formulations of decision-tree classification algorithms. In *High Performance Data Mining*; Springer: Boston, MA, USA, 1999; pp. 237–261.
99. Schroff, F.; Criminisi, A.; Zisserman, A. Object Class Segmentation using Random Forests. In Proceedings of the British Machine Vision Conference, University of Leeds, Leeds, UK, 1–4 September 2008; pp. 1–10.
100. Bayes, T.; LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philos. Trans. R. Soc. Lond.* **1763**, *53*, 370–418.
101. Leung, K.M. *Naive Bayesian Classifier*; Polytechnic University Department of Computer Science/Finance and Risk Engineering: New York, NY, USA, 2007; pp. 123–156.
102. Kononenko, I. Semi-Naive Bayesian Classifier. In *European Working Session on Learning*; Springer: Berlin/Heidelberg, Germany, 1991; pp. 206–219.
103. Klecka, W.R.; Gudmund, R.I.; Klecka, W.R. *Discriminant Analysis*; Sage: New York, NY, USA, 1980; Volume 19.
104. Menard, S. *Applied Logistic Regression Analysis*; Sage: New York, NY, USA, 2002; Volume 106.
105. Hastie, T.; Tibshirani, R. Discriminant adaptive nearest neighbor classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 607–616. [CrossRef]
106. Durgesh, K.S.; Lekha, B. Data classification using support vector machine. *J. Theor. Appl. Inf. Technol.* **2010**, *12*, 1–7.
107. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [CrossRef] [PubMed]
108. Murtagh, F. Multilayer perceptrons for classification and regression. *Neurocomputing* **1991**, *2*, 183–197. [CrossRef]
109. Jmour, N.; Zayen, S.; Abdelkrim, A. Convolutional neural networks for image classification. In Proceedings of the International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet, Tinisia, 22–25 March 2018; pp. 397–402.
110. De Villiers, J.P.; Leuschner, F.W.; Geldenhuys, R. Centi-pixel accurate real-time inverse distortion correction. In Proceedings of the International Symposium on Optomechatronic Technologies, West Harbor, San Diago, CA, USA, 17–19 November 2008; Volume 7266, p. 726611.
111. Caprile, B.; Torre, V. Using vanishing points for camera calibration. *Int. J. Comput. Vis.* **1990**, *4*, 127–139. [CrossRef]
112. Wang, A.; Qiu, T.; Shao, L. A simple method of radial distortion correction with centre of distortion estimation. *J. Math. Imaging Vis.* **2009**, *35*, 165–172. [CrossRef]
113. Hartley, R.; Kang, S.B. Parameter-free radial distortion correction with center of distortion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1309–1321. [CrossRef]
114. Huang, K.; Ziauddin, S.; Zand, M.; Greenspan, M. One Shot Radial Distortion Correction by Direct Linear Transformation. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Abu Dhabi, Dubai, 25–28 October 2020; pp. 473–477.
115. Zhao, H.; Shi, Y.; Tong, X.; Ying, X.; Zha, H. A Simple Yet Effective Pipeline For Radial Distortion Correction. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Abu Dhabi, Dubai, 25–28 October 2020; pp. 878–882.
116. Wang, Y.M.; Li, Y.; Zheng, J.B. A camera calibration technique based on OpenCV. In Proceedings of the 3rd International Conference on Information Sciences and Interaction Sciences, Chengdu, China, 23–25 June 2010; pp. 403–406.
117. Lee, S.; Hong, H. A robust camera-based method for optical distortion calibration of head-mounted displays. *J. Disp. Technol.* **2014**, *11*, 845–853. [CrossRef]
118. Wang, Z.; Liu, M.; Yang, S.; Huang, S.; Bai, X.; Liu, X.; Zhu, J.; Liu, X.; Zhang, Z. Precise full-field distortion rectification and evaluation method for a digital projector. *Opt. Rev.* **2016**, *23*, 746–752. [CrossRef]
119. Yang, S.; Srikanth, M.; Lelescu, D.; Venkataraman, K. Systems and Methods for Depth-Assisted Perspective Distortion Correction. U.S. Patent 9,898,856, 20 February 2018.
120. Finlayson, G.; Gong, H.; Fisher, R.B. Color homography: Theory and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *41*, 20–33. [CrossRef]
121. Wang, H.; Yang, J.; Xue, B.; Yan, X.; Tao, J. A novel color calibration method of multi-spectral camera based on normalized RGB color model. *Results Phys.* **2020**, *19*, 103498. [CrossRef]
122. Han, S.; Huang, P.; Wang, H.; Yu, E.; Liu, D.; Pan, X. Mat: Motion-aware multi-object tracking. *Neurocomputing* **2022**, *476*, 75–86. [CrossRef]

123. Meinhardt, T.; Kirillov, A.; Leal-Taixe, L.; Feichtenhofer, C. Trackformer: Multi-object tracking with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–22 June 2022; pp. 8844–8854.
124. Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1337–1342. [CrossRef]
125. Zhang, W.; Ma, B.; Liu, K.; Huang, R. Video-based pedestrian re-identification by adaptive spatio-temporal appearance model. *IEEE Trans. Image Processing* **2017**, *26*, 2042–2054. [CrossRef]
126. Yang, X.; Wang, M.; Tao, D. Person re-identification with metric learning using privileged information. *IEEE Trans. Image Processing* **2017**, *27*, 791–805. [CrossRef]
127. Geng, S.; Yu, M.; Guo, Y.; Yu, Y. A Weighted Center Graph Fusion Method for Person Re-Identification. *IEEE Access* **2019**, *7*, 23329–23342. [CrossRef]
128. Yang, X.; Tang, Y.; Wang, N.; Song, B.; Gao, X. An End-to-End Noise-Weakened Person Re-Identification and Tracking with Adaptive Partial Information. *IEEE Access* **2019**, *7*, 20984–20995. [CrossRef]
129. Chen, T.; Fang, C.; Shen, X.; Zhu, Y.; Chen, Z.; Luo, J. Anatomy-aware 3d human pose estimation with bone-based pose decomposition. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 198–209. [CrossRef]
130. Straka, M.; Hauswiesner, S.; Rüther, M.; Bischof, H. Skeletal Graph Based Human Pose Estimation in Real-Time. In Proceedings of the BMVC, Dundee, UK, 29 August–2 September 2011; pp. 1–12.
131. Campbell, L.W.; Bobick, A.F. Using phase space constraints to represent human body motion. In Proceedings of the International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, 26–28 June 1995; pp. 338–343.
132. Oren, M.; Papageorgiou, C.; Sinha, P.; Osuna, E.; Poggio, T. Pedestrian detection using wavelet templates. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA, 17–19 June 1997; Volume 97, pp. 193–199.
133. You, Q.; Jin, H.; Wang, Z.; Fang, C.; Luo, J. Image captioning with semantic attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, GA, USA, 27–30 June 2016; pp. 4651–4659.
134. Wcg, P. Role of the manuscript reviewer. *Singap. Med. J.* **2009**, *50*, 931–934.
135. Polak, J.F. The role of the manuscript reviewer in the peer review process. *Am. J. Roentgenol.* **1995**, *165*, 685–688. [CrossRef]
136. Nguyen, H.; Bhanu, B.; Patel, A.; Diaz, R. VideoWeb: Design of a wireless camera network for real-time monitoring of activities. In Proceedings of the 3rd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), Como, Italy, 30 August –2 September 2009; pp. 1–8.
137. Ibraheem, O.W.; Irwansyah, A.; Hagemeyer, J.; Porrmann, M.; Rueckert, U. Reconfigurable vision processing system for player tracking in indoor sports. In Proceedings of the Conference on Design and Architectures for Signal and Image Processing (DASIP), Dresden, Germany, 27–29 September 2017; pp. 1–6.
138. Xiang, Y.; Alahi, A.; Savarese, S. Learning to Track: Online multi-object tracking by decision making. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4705–4713.
139. Laptev, I.; Caputo, B. Recognizing human actions: A local SVM approach. In Proceedings of the 17th International Conference on Pattern Recognition, Washington, DC, USA, 23–26 August 2004; pp. 32–36.
140. Duy, A.N.; Yoo, M. Calibration-Net: LiDAR and Camera Auto-Calibration using Cost Volume and Convolutional Neural Network. In Proceedings of the 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Jeju Island, Korea, 21–24 February 2022; pp. 141–144.
141. Cao, Y.; Wang, H.; Zhao, H.; Yang, X. Neural-Network-Based Model-Free Calibration Method for Stereo Fisheye Camera. *Front. Bioeng. Biotechnol.* **2022**, *10*, 955233. [CrossRef]
142. Chen, H.; Munir, S.; Lin, S. RFCam: Uncertainty-aware Fusion of Camera and Wi-Fi for Real-time Human Identification with Mobile Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 1–29. [CrossRef]
143. Dufera, T.T.; Seboka, Y.C.; Portillo, C.F. Parameter Estimation for Dynamical Systems Using a Deep Neural Network. *Appl. Comput. Intell. Soft Comput.* **2022**, *2022*, 2014510. [CrossRef]
144. Doula, A.; Sanchez Guinea, A.; Mühlhäuser, M. VR-Surv: A VR-Based Privacy Preserving Surveillance System. In Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts, New Orleans, NY, USA, 29 April–5 May 2022; pp. 1–7.
145. Pooja, C.; Jaisharma, K. Novel Framework for the Improvement of Object Detection Accuracy of Smart Surveillance Camera Visuals Using Modified Convolutional Neural Network Technique Compared with Global Color Histogram. *ECS Trans.* **2022**, *107*, 18823. [CrossRef]
146. Jiang, T.; Zhang, Q.; Yuan, J.; Wang, C.; Li, C. Multi-Type Object Tracking Based on Residual Neural Network Model. *Symmetry* **2022**, *14*, 1689. [CrossRef]
147. Jaganathan, T.; Panneerselvam, A.; Kumaraswamy, S.K. Object detection and multi-object tracking based on optimized deep convolutional neural network and unscented Kalman filtering. *Concurr. Comput. Pr. Exp.* **2022**, *34*, e7245. [CrossRef]
148. Deshpande, T.R.; Sapkal, S.U. Development of Object Tracking System Utilizing Camera Movement and Deep Neural Network. In Proceedings of the 2022 IEEE Region 10 Symposium (TENSYMP), Mumbai, India, 1–3 July 2022; pp. 1–6.
149. Praveenkumar, S.M.; Patil, P.; Hiremath, P.S. Real-Time Multi-Object Tracking of Pedestrians in a Video Using Convolution Neural Network and Deep SORT. In Proceedings of the ICT Systems and Sustainability, (ICT4SD), Goa, India, 23–24 July 2022; pp. 725–736.
150. Jhansi, M.; Bachu, S.; Kumar, N.U.; Kumar, M.A. IODTDLCNN: Implementation of Object Detection and Tracking by using Deep Learning based Convolutional Neural Network. In Proceedings of the 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 16–18 February 2022; pp. 1–6.

151. Barazande, J.; Farzaneh, N. WSAMLP: Water Strider Algorithm and Artificial Neural Network-based Activity Detection Method in Smart Homes. *J. AI Data Min.* **2022**, *10*, 1–13.
152. Wong, P.K.Y.; Luo, H.; Wang, M.; Cheng, J.C. Enriched and discriminative convolutional neural network features for pedestrian re-identification and trajectory modelling. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 573–592. [CrossRef]
153. Yao, Y.; Jiang, X.; Fujita, H.; Fang, Z. A sparse graph wavelet convolution neural network for video-based person re-identification. *Pattern Recognit.* **2022**, *129*, 108708. [CrossRef]
154. Mohana, M.; Alelyani, S.; Alsaqer, M.S. Fused Deep Neural Network based Transfer Learning in Occluded Face Classification and Person re-Identification. *arXiv* **2022**, arXiv:2205.07203.
155. You, C.; Zheng, H.; Guo, Z.; Wang, T.; Wu, X. Tampering detection and localization base on sample guidance and individual camera device convolutional neural network features. *Expert Syst.* **2022**, e13102. [CrossRef]
156. Karamchandani, S.; Bhattacharjee, S.; Issrani, D.; Dhar, R. SLAM Using Neural Network-Based Depth Estimation for Auto Vehicle Parking. In *IOT with Smart Systems*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 37–44.
157. AI In Computer Vision Market Research Report by Component (Hardware, Software), Vertical (Healthcare, Security, Automotive, Agriculture, Sports & Entertainment, and Others), and Region–Global Forecast to 2027. Available online: https://www.expertmarketresearch.com/reports/ai-in-computer-vision-market (accessed on 22 August 2022).
158. Andriyanov, N. Methods for preventing visual attacks in convolutional neural networks based on data discard and dimensionality reduction. *Appl. Sci.* **2021**, *11*, 5235. [CrossRef]
159. Wang, B.; Zhao, M.; Wang, W.; Dai, X.; Li, Y.; Guo, Y. Adversarial Analysis for Source Camera Identification. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 4174–4186. [CrossRef]
160. Zhang, C.; Benz, P.; Lin, C.; Karjauv, A.; Wu, J.; Kweon, I.S. A survey on universal adversarial attack. *arXiv* **2021**, arXiv:2103.01498.
161. Edwards DRawat, D.B. Study of Adversarial Machine Learning with Infrared Examples for Surveillance Applications. *Electronics* **2020**, *9*, 1284. [CrossRef]
162. Chakraborty, A.; Alam, M.; Dey, V.; Chattopadhyay, A.; Mukhopadhyay, D. A survey on adversarial attacks and defences. *CAAI Trans. Intell. Technol.* **2021**, *6*, 25–45. [CrossRef]
163. Akhtar, N.; Mian, A.; Kardan, N.; Shah, M. Advances in adversarial attacks and defenses in computer vision: A survey. *IEEE Access* **2021**, *9*, 155161–155196. [CrossRef]
164. SanMiguel, J.C.; Micheloni, C.; Shoop, K.; Foresti, G.L.; Cavallaro, A. Self-reconfigurable smart camera networks. *Computer* **2014**, *47*, 67–73.
165. Leong, W.L.; Martinel, N.; Huang, S.; Micheloni, C.; Foresti, G.L.; Teo, R.S. An Intelligent Auto-Organizing Aerial Robotic Sensor Network System for Urban Surveillance. *J. Intell. Robot. Syst.* **2021**, *102*, 33. [CrossRef]
166. Natarajan, P.; Atrey, P.K.; Kankanhalli, M. Multi-camera coordination and control in surveillance systems: A survey. *ACM Trans. Multimed. Comput. Commun. Appl.* **2015**, *11*, 1–30. [CrossRef]
167. Martinel, N.; Dunnhofer, M.; Pucci, R.; Foresti, G.L.; Micheloni, C. Lord of the rings: Hanoi pooling and self-knowledge distillation for fast and accurate vehicle reidentification. *IEEE Trans. Ind. Inform.* **2021**, *18*, 87–96. [CrossRef]
168. Rinner, B.; Esterle, L.; Simonjan, J.; Nebehay, G.; Pflugfelder, R.; Dominguez, G.F.; Lewis, P.R. Self-aware and self-expressive camera networks. *Computer* **2015**, *48*, 21–28. [CrossRef]
169. Lewis, P.R.; Chandra, A.; Glette, K. Self-awareness and Self-expression: Inspiration from Psychology. In *Self-Aware Computing Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 9–21.
170. Glette, K.; Lewis, P.R.; Chandra, A. Relationships to Other Concepts. In *Self-aware Computing Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 23–35.
171. Wang, S.; Nebehay, G.; Esterle, L.; Nymoen, K.; Minku, L.L. Common Techniques for Self-awareness and Self-expression. In *Self-Aware Computing Systems*; Natural Computing Series; Springer: Berlin/Heidelberg, Germany, 2016.
172. Isavudeen, A.; Ngan, N.; Dokladalova, E.; Akil, M. Auto-adaptive multi-sensor architecture. In Proceedings of the International Symposium on Circuits and Systems (ISCAS), Montreal, QC, Canada, 22–26 May 2016; pp. 2198–2201.
173. Guettatfi, Z.; Hübner, P.; Platzner, M.; Rinner, B. Computational self-awareness as design approach for visual sensor nodes. In Proceedings of the 12th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC), Madrid, Spain, 12–17 July 2017; pp. 1–8.
174. Zhu, Z.; Luo, Y.; Chen, S.; Qi, G.; Mazur, N.; Zhong, C.; Li, Q. Camera style transformation with preserved self-similarity and domain-dissimilarity in unsupervised person re-identification. *J. Vis. Commun. Image Represent.* **2021**, *80*, 103303. [CrossRef]
175. Lin, S.; Lv, J.; Yang, Z.; Li, Q.; Zheng, W.S. Heterogeneous graph driven unsupervised domain adaptation of person re-identification. *Neurocomputing* **2022**, *471*, 1–11. [CrossRef]
176. Wu, M.; Li, C.; Yao, Z. Deep Active Learning for Computer Vision Tasks: Methodologies, Applications, and Challenges. *Appl. Sci.* **2022**, *12*, 8103. [CrossRef]
177. Rudolph, S.; Tomforde, S.; Hähner, J. On the Detection of Mutual Influences and Their Consideration in Reinforcement Learning Processes. *arXiv* **2019**, arXiv:1905.04205.
178. Cai, L.; Ma, H.; Liu, Z.; Li, Z.; Zhou, Z. Coverage Control for PTZ Camera Networks Using Scene Potential Map. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 11–15 July 2022; pp. 1–6.
179. Suresh, S.; Menon, V. An Efficient Graph Based Approach for Reducing Coverage Loss From Failed Cameras of a Surveillance Network. *IEEE Sens. J.* **2022**, *22*, 8155–8163.

180. Liang, K. Fission: A Provably Fast, Scalable, and Secure Permissionless Blockchain. *arXiv* **2018**, arXiv:1812.05032.
181. Zhao, W. On Nxt Proof of Stake Algorithm: A Simulation Study. *IEEE Trans. Dependable Secur. Comput.* **2022**. Available online: https://ieeexplore.ieee.org/document/9837462 (accessed on 22 August 2022). [CrossRef]

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT).
India Habitat Centre, Lodhi Road, New Delhi, India. Sep 23-25, 2022

Dual Input Common LLC-Resonant Tank based off-Grid Charger

Vishal Verma
Electrical Engineering Department
Delhi Technological University
Delhi, India
vishalverma@dce.ac.in

Prajeet Shukla
Electrical Engineering Department
Delhi Technological University
Delhi, India
prajeetshukla95@gmail.com

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT) | 978-1-6654-9365-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBCONPT57482.2022.9938202

***Abstract*—The paper proposes novel topology "dual input common LLC-resonant tank-based charger for charging the 72V, 60Ah Lithium-ion battery. With four active switches, the topology is a twin half-bridge LLC circuit. This design eliminates the need for additional circuit components by allowing the two sources to share the same resonant tank for transferring energy to the load. The converter is operating in differential mode. Current controlled Variable frequency control using simple PI controller is employed to regulate battery charging at different line and load perturbations. A simulation study is done on 3.3kW proposed charger using MATLAB/SIMULINK to support the theoretical concept.**

Keywords—Dual input, LLC, charging, control

I. INTRODUCTION

The World is witnessing a paradigm shift towards eV/HeV/PHeV from fossil/fuel-based engines. The range anxiety is the biggest deterrent for their greater adoption[1], [2]. Since the quantum of high energy density batteries is limited by total volume/weight of the eV, bottleneck gets zeroed down to omnipresent charging stations and their fast-charging capabilities[3]–[5]. In developing countries like India where power infrastructure is not uniform throughout the country. The need for off grid EV chargers is essential for wider acceptance of EV/HEV's. Off grid PV fed EV chargers could be a possible solution to range anxiety for EV's in remote areas and rural areas, it could offer resilience against grid outages and could be used to deliver power during emergency response and natural disaster/ calamity.
Even, Big Solar farm could be installed for providing fast-charging solution. However, the output voltages of solar panels are usually low. Circumventing the issues associated with high turns ratio high frequency transformer two sets of PV panels can be used to lift the voltage of standard 380V dc bus.

Over the years LLC Converter have been seen as industry standard for its usage in fast charging stations due to high efficiency, high power density, wide load and line regulations and rugged operation exhibited by resonant circuits [1,4,6,8]. As stated by the authors in [6]–[13] interleaved, double resonant tank, parallel output, series input parallel output transformer connection are some of the topologies suggested for handling high current applications. When run at the same power level, interleaving and double resonant tank LLC-based DC-DC converters reduce resonant tank voltage and current stress. These converters, on the other hand, need two separate power transformers, which adds to the cost, size, and complexity of the device. As a result, the enormous component count and size of these typical interleaved and double resonant tanks, LLC based converters become the main disadvantages. To alleviate these drawbacks, this article proposes distributing power across the output using two isolated sources (could be two PV panel sets) and a common resonant tank. The major characteristics of the suggested converters are: i) a wide input voltage range of operation; ii) a low component count and complexity due to a single transformer; iii) Lower core loss due to leakage flux cancellation due to differential operation mode.

II. CIRCUIT DESCRIPTION

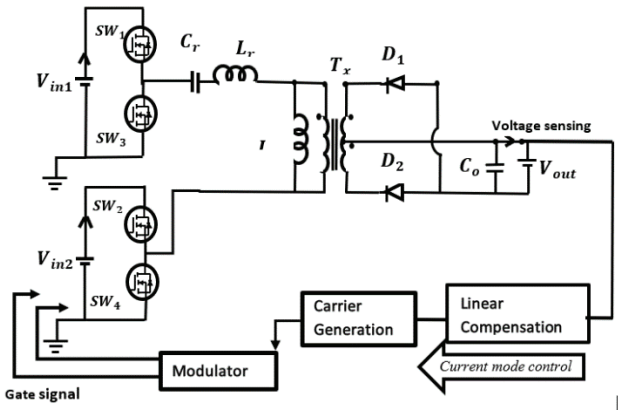


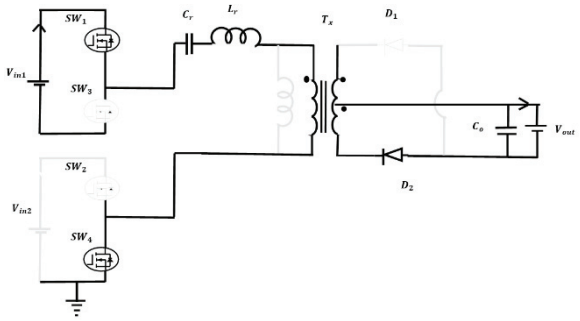
Fig.1 Dual Input common LLC- resonant tank-based charger

Fig.1 exhibits schematic of the proposed PV fed dual input resonant charger, which utilizes single resonant LLC tank to channelize energy in resonant mode, a mid-point tapped rectifier and output filter. This topology has six operating modes, during first three mode switch SW1 and SW3 are conducting whereas switch SW2 and SW4 are switched off and during next three mode switch SW3 and SW4 are conducting whereas SW1 and SW3 are switched off. Each pair of primary side (SW1, SW3 and SW2, SW4) switch operates at a duty of 50% for each half cycle of the converter's operation. Since direction of current in each half cycle reverses through magnetizing inductor, the topology essentially is operating in the differential mode. The differential mode current in the magnetizing inductor essentially reduces the radiated emissions in high power multilevel circuitry and in digital control circuitry. Primary side switches are switching at zero voltage condition throughout the converter operation and secondary side diodes are switched at zero current condition throughout the converter operation. The voltage stress across the primary switch is not more than the input voltage range since each source delivers 50% load. Similarly, the voltage stress

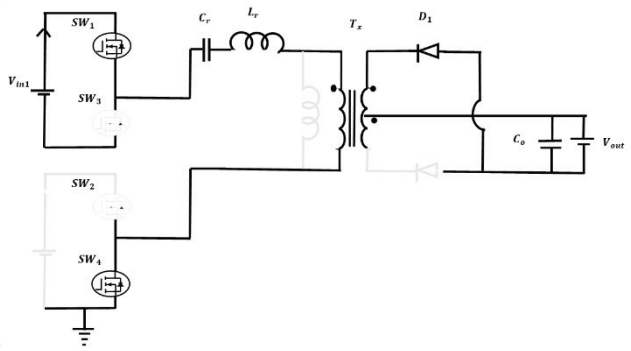
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:13:21 UTC from IEEE Xplore. Restrictions apply.

across the secondary side diode is twice the output voltage because of the center taped configuration at the output. The converter is operating in the constant current control mode with variable frequency control. Differential mode current flowing through magnetization leg further aids in the fast control using simple PI controller.

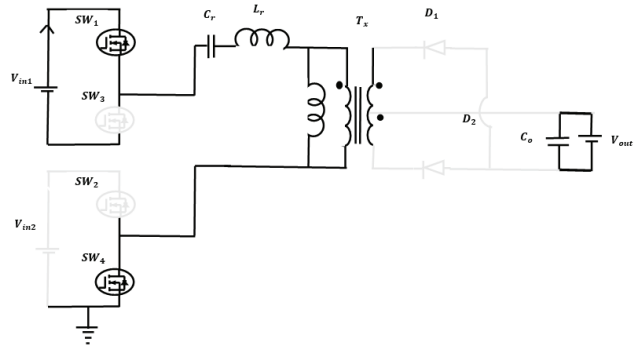
III. CIRCUIT OPERATION:



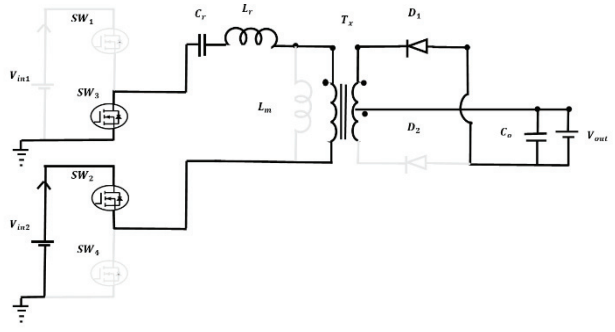
(a)



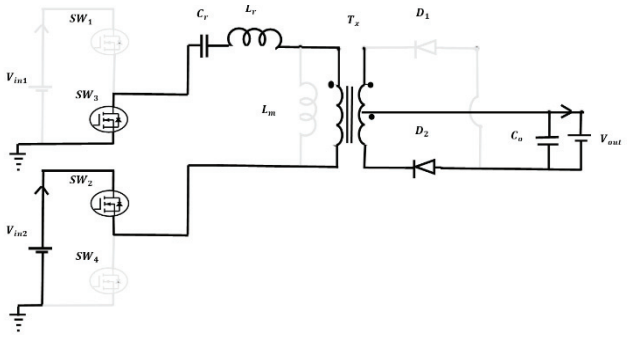
(b)



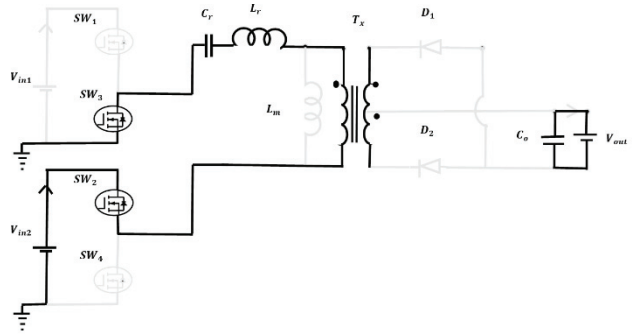
(c)



(d)



(e)



(f)

Fig.2 Operating modes of proposed dual input common resonant tank LLC charger, (a) Mode I, (b) Mode II, (c) Mode III, (d) Mode IV, (e) Mode V, (f) Mode VI

The modes of operation are summarized in the Table.1, for mode I, II and III switches SW_1 and SW_3 are conducting and for mode II, IV and VI switches SW_2 and SW_4 are conducting. For mode III and VI, L_m participates in the resonance and hence the characteristic impedance for mode III and VI are more than characteristic impedance of mode I, II, IV, V resulting into elliptical trajectory. The steady state trajectory is reflection of the resonant tank impedance.

TABLE I. PROPOSED CHARGER OPERATING MODES

Modes of operation	Resonant tank impedance Equation	Geometry
I	$\left[v_{Cr_N} - (1 - nV_{0N})\right]^2 + {i_{L_{rN}}}^2 = \left[V_{Cr_{0N}} - (1 - nV_{0N})\right]^2 + {I_{Lr_{oN}}}^2$	Circle with centre at (1-nV_{on}, 0)
II	$\left[v_{Cr_N} - (1 + nV_{0N})\right]^2 + {i_{L_{rN}}}^2 = \left[V_{Cr_{0N}} - (1 + nV_{0N})\right]^2 + {I_{Lr_{oN}}}^2$	Circle with centre at (1+nV_{on}, 0)
III	$(v_{Cr_N} - 1)^2 - \left(\frac{i_{L_rN}}{Z_0/Z_1}\right)^2 = \left[V_{Cr_{0N}} - 1\right]^2 + \left(\frac{I_{Lr0N}}{Z_0/Z_1}\right)^2$	Ellipse with centre at (1,0)
IV	$\left[v_{Cr_N} - nV_{oN}\right]^2 + {i_{L_{rN}}}^2 = \left[V_{Cr_{0N}} - nV_{0N}\right]^2 + {I_{Lr_{oN}}}^2$	Circle with centre at (nV_{on}, 0)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:13:21 UTC from IEEE Xplore. Restrictions apply.

V	$[v_{Cr_N} + nV_{0N}]^2 + {i_{L_rN}}^2 = [V_{Cr_{0N}} + nV_{0N}]^2 + {I_{Lr_{0N}}}^2$	Circle with centre at (-nV$_{on}$, 0)
VI	$(v_{C_{rN}})^2 - \left(\frac{i_{L_rN}}{Z_0/Z_1}\right)^2 = {V_{Cr_{0N}}}^2 + \left(\frac{I_{Lr0N}}{Z_0/Z_1}\right)^2$	Ellipse with centre at (0,0)

Gain Plot:

The DC gain characteristic of the LLC resonant converter is shown in Fig.2.5 for various values of the quality factor (Q), and it can be seen that there are multiple resonant frequencies depending on the value of quality factor, Q which is defined as the ratio between the characteristic impedance and the output load, may be calculated using Equation.

$$Q = \frac{\sqrt{L_r/C_r}}{R_e} \qquad (2.1)$$

where, Re, denotes the reflected output load resistance.

Furthermore, the areas of the DC gain characteristic with a positive gradient are suited for zero-current switching (ZCS) operation[14]–[16]. The designed converter will use metal-oxide field-effect transistors (MOSFETs) as the semiconductor switching device for which the ideal operating area is on the negative gradient. This is because ZVS is the most used soft-switching technique in MOSFET devices. ZVS lowers switching loss caused by diode recovery charge and semiconductor capacitance, which are typical in MOSFET switching devices. As the load varies the operating point of converter varies to provide the required step up or step-down voltage levels.

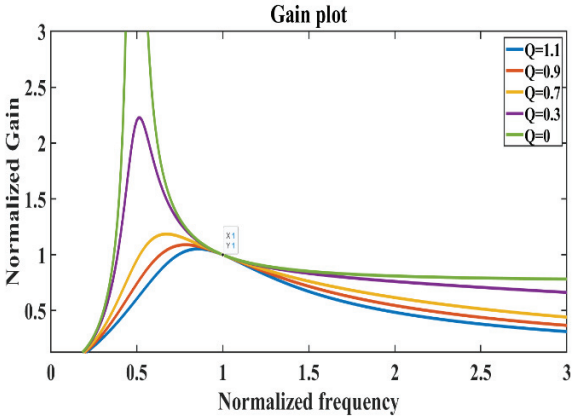


Fig.3 Gain Plot for designed DICR-LLC charger

IV. Control Scheme

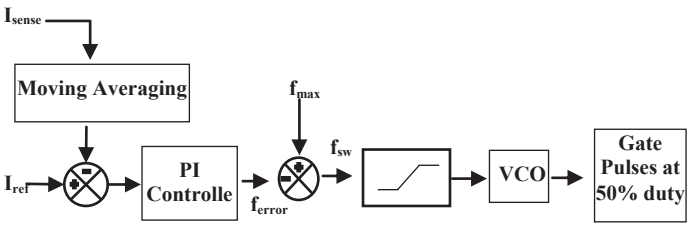


Fig.4 Block Diagram of proposed control scheme

Since, the relationship between frequency and control signal is linear, the proposed controller uses closed loop frequency control, allowing it to use a linear PI controller successfully while avoiding the complexity of DSP controllers. The control logic is depicted in a block diagram as illustrated in Fig.4. The output current is sensed and averaged (Iavg) which is further compared with the reference current value (Iref) to generate the error current signal. The PI controller is used to implement the steady state control. ferror component computed through PI controller is subtracted from the maximum frequency component which is taken to be twice the resonant frequency value. At the starting, converter is operated at this maximum frequency (fsw = 2fmax) which helps in reducing the transient current stress as at this frequency substantial impedance is being offered by the resonant tank. As steady state being reached the switching frequency transacts smoothly to the resonant frequency (fsw = fmax) where it is being operated at the highest efficiency.

V. Performance Evaluation:

To verify the performance of the proposed topology dual input common resonant LLC based charger, a simulation study is carried out on MATLAB/SIMULINK software. The Converter is shown charging the battery of 60Ah at 72V from two PV sources supplying 320-380V DC Bus at the resonant tank input. The parameters of resonant tank circuit are Lr=42.3µH, Lm=136µH and Cr=26.6n which is operating closely above the resonant frequency of 150kHz at full load condition. Fig.5 shows the key operating waveform of propose charger. Fig.5(a) and 5(b) shows the gate pulses for switches SW1-SW4 and SW2-SW3 operating at 50% duty ratio inclusive of dead time value of 170ns. Fig.5(c) shows the current waveform across resonant inductor (Ilr) and magnetization (Ilm). Fig.5(d) shows the sinusoidal voltage waveform across the resonant capacitor (Vcr) and square voltage waveform across resonant tank input. As both the sources are sharing power equally, symmetrical input current flows through both the sources as shown in Fig.6. Refereeing Fig.7 and Fig.8, indicates the ZVS and ZCS switching across the primary and secondary switches of proposed converter. The converter voltage is regulated by changing the switching waveform at different load conditions. The proposed charger is tested for different load condition and different line regulation as shown in Fig.10 and Fig.11. Fig.10 demonstrates the dynamic charging at full load, 50% light load and at 10% light load condition at Vin = 320V. It may be observed from the Fig.10 and Fig.11, the controller is able to smoothly transact from one operating load to another. Two cases are discussed here to validate the controller performance, first at Vin = 320V and second at Vin = 380V. For first case: Controller takes 0.5msec to reach

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:13:21 UTC from IEEE Xplore. Restrictions apply.

steady state condition of the charger at the full load condition, 0.6msec to settle from full load to 50% light load condition and 1.3msec to settle from 50% light load to 10% light load condition.

For second case: Controller takes 0.7msec to reach steady state at full load condition, 0.8sec to settle down to 50% light load and 2msec to settle down to 10% light load condition.

VI. RESULTS:

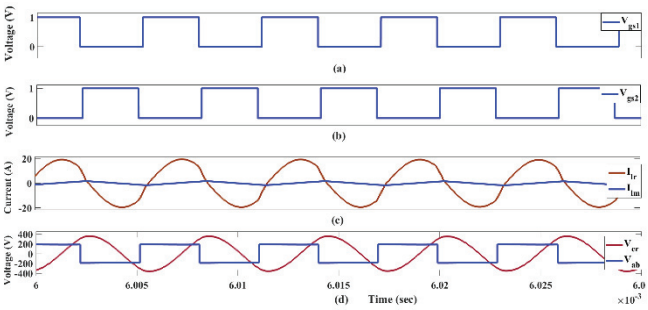


Fig.5 Key operating waveforms of the proposed DICR-LLC charger, (a) gate to source voltage across SW_1 and SW_4 (V_{gs1}), (b) gate to source voltage across SW_2 and SW_3 (V_{gs2}), (c) current across resonant inductor (I_{lr}) and magnetizing inductor (I_{lm}), (d) voltage at resonant tank input (V_{ab}) and across resonant capacitor (V_{cr}).

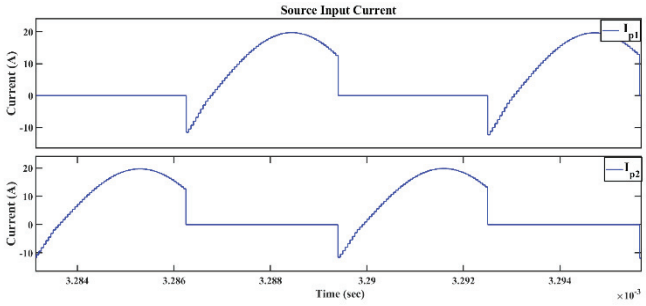


Fig.6 Source current waveform

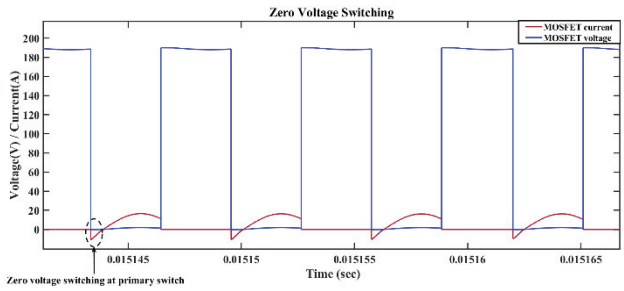


Fig.7 ZVS Soft switching at primary switches

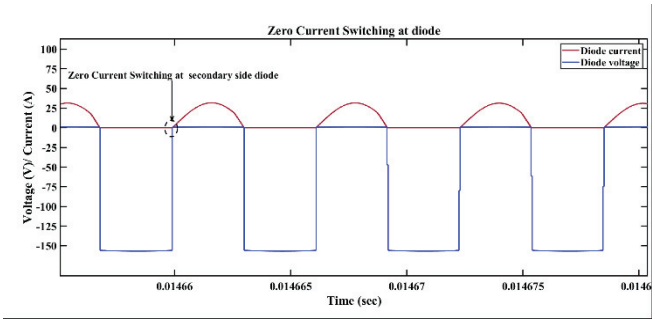


Fig.8 ZCS soft switching at secondary switches

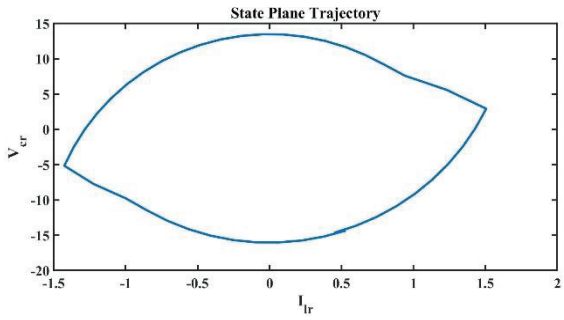


Fig.9 State plane trajectory at steady state

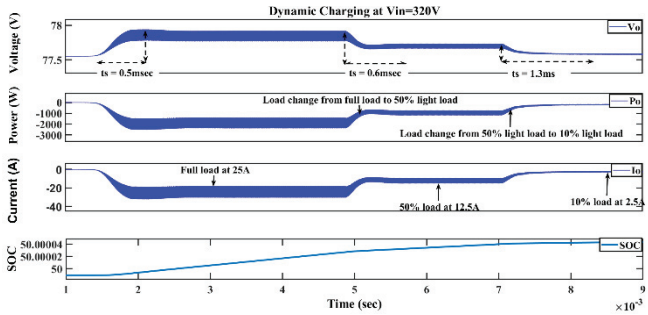


Fig.10 Dynamic Charging at full load, 50% light load and 10% light load at V_{in} = 320V

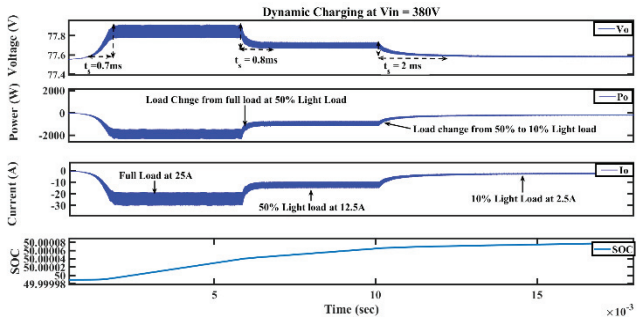


Fig.11 Dynamic Charging at full load, 50% light load and 10% light load at V_{in} = 380V

VII. CONCLUSION:

A new topology for the off-grid charging has been investigated. The observed performance of the dual input common resonant tank LLC based charger has demonstrated the dynamic charging at different load conditions and at different source condition. The demonstrated results conform the seamless control of tank circuit parameters using simple PI based linear control. A simulation study has been conducted and it has validated the effectiveness of the proposed topology. It has also been observed that the proposed topology has a fast response to load and line perturbations and is able to operate in fixed trajectory with soft switching at primary MOSFET switches and ZCS at secondary diode. At steady state operation, the switching frequency smoothly transacts near resonant frequency for achieving maximum efficiency and at light load condition the switching frequency increases to provide the required step-down operation. Simulation results of proposed resonant charger validates the theoretical analysis. The topology has the advantage of wide input voltage operation and dynamic charging at different load condition.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:13:21 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] J. Deng, S. Li, S. Hu, C. C. Mi, and R. Ma, "Design methodology of LLC resonant converters for electric vehicle battery chargers," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 4, pp. 1581–1592, 2014.

[2] F. Musavi, M. Craciun, D. S. Gautam, and W. Eberle, "Control strategies for wide output voltage range LLC resonant DC-DC converters in battery chargers," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1117–1125, 2014.

[3] J. Chen, C. Yang, S. Tang, and J. Zou, "A High Power Interleaved Parallel Topology Full-Bridge LLC Converter for Off-Board Charger," *IEEE Access*, vol. 9, pp. 157790–157799, 2021.

[4] Y. Wang, Z. Yan, M. Tian, and Y. Xiao, "Positive/negative pulse charging technology of battery based on PFC-LLC topology," in *Proceedings - 5th International Conference on Instrumentation and Measurement, Computer, Communication, and Control, IMCCC 2015*, Feb. 2016, pp. 1031–1036.

[5] D. Cittanti, M. Gregorio, E. Armando, and R. Bojoi, "Digital Multi-Loop Control of an LLC Resonant Converter for Electric Vehicle DC Fast Charging," in *ECCE 2020 - IEEE Energy Conversion Congress and Exposition*, Oct. 2020, pp. 4423–4430.

[6] X. Chen, S. M. Tayebi and I. Batarseh, "A Modified Dual-input LLC Converter for Standalone PV/Battery Power System," 2020 IEEE Applied Power Electronics Conference and Exposition (APEC), 2020, pp. 969-976.

[7] S. M. Tayebi, H. Hu, S. Abdel-Rahman, and I. Batarseh, "Dual-Input Single-Resonant Tank LLC Converter with Phase Shift Control for PV Applications," *IEEE Transactions on Industry Applications*, vol. 55, no. 2, pp. 1729–1739, Mar. 2019.

[8] X. Chen, S. M. Tayebi and I. Batarseh, "Design and Implementation of a Dual-Input LLC Converter with Semi-Active Rectifiers for PV Applications," 2019 IEEE Energy Conversion Congress and Exposition (ECCE), 2019, pp. 6404-6411

[9] S. M. Tayebi, H. Hu, O. Abdel-Rahman and I. Batarseh, "Design and analysis of a dual-input single-resonant tank LLC converter for PV applications," 2018 IEEE Applied Power Electronics Conference and Exposition (APEC), 2018, pp. 476-483.

[10] A. Mustafa and S. Mekhilef, "Dual Phase LLC Resonant Converter with Variable Frequency Zero Circulating Current Phase-Shift Modulation for Wide Input Voltage Range Applications," *IEEE Transactions on Power Electronics*, vol. 36, no. 3, pp. 2793–2807, Mar. 2021.

[11] S. Milad Tayebi, X. Chen, A. Bhattacharjee and I. Batarseh, "Design and Implementation of a Dual-Input LLC Converter for PV-Battery Applications," 2018 IEEE Energy Conversion Congress and Exposition (ECCE), 2018, pp. 5934-5940.

[12] A. A. Hussein, A. Alhatlani, S. Ghosh, and I. Batarseh, "A Double-Loop Maximum Power Point Tracking Algorithm for Dual-Input Phase-Shifted LLC Converter," in *ECCE 2020 - IEEE Energy Conversion Congress and Exposition*, Oct. 2020, pp. 3205–3209.

[13] Shih-Ming Chen, Yong-Hong Haung, Yi-Yuan Chung, Yi-Hsun Hsieh and Tsorng-Juu Liang, "A novel interleaved LLC resonant converter," IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, 2013, pp. 293-297.

[14] Y. H. Hsieh and F. C. Lee, "Small-Signal Dynamic and High-Bandwidth Design of LLC Resonant Converters," in *ECCE 2020 - IEEE Energy Conversion Congress and Exposition*, Oct. 2020, pp. 6136–6143.

[15] F. Musavi, M. Craciun, D. S. Gautam, W. Eberle, and W. G. Dunford, "An LLC resonant dc-dc converter for wide output voltage range battery charging applications," *IEEE Transactions on Power Electronics*, vol. 28, no. 12, pp. 5437–5445, 2013.

[16] S. K. Pidaparthy, B. Choi, and J. Jang, "Design and performance evaluation of digital control for LLC series resonant dc-to-dc converters," in *2014 International Power Electronics Conference, IPEC-Hiroshima - ECCE Asia 2014*, 2014, pp. 3638–3645.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:13:21 UTC from IEEE Xplore. Restrictions apply.

Advanced Production and Industrial Engineering
R.M. Singari and P.K. Kankar (Eds.)
© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE220724

Effect of Plasma on the Nucleation and Growth of Vertical Carbon Nanotube Array on 2-Dimensional Graphene

Vedant Chouksey[a,1], Meetali Choudhary [a], Agastya Sethi[a], Neha Jain[b] and Suresh C. Sharma [a]
[a] *Department of Applied Physics, Delhi Technological University, Delhi 110042, India*
[b] *Department of Applied Sciences (Physics), Maharaja Agrasen Institute of Technology, Delhi 110086, India*

Abstract. This paper studies the nucleation and growth of vertical carbon nanotube array on 2-dimensional graphene and elaborates the impact of plasma on the same. It observes an analytical model of growth of Carbon Nanotubes (CNT) on the wonder material, Graphene has been developed. The model includes the growth of the CNTs on graphene Plasma Enhanced Chemical Vapour Deposition (PECVD) technique where the plasma consists of hydrocarbon and hydrogen radicals and neutral molecules using the gas mixture of Argon, Hydrogen and suitable hydrocarbons. Further, using MATHEMATICA, the paper examines various trends such as relationship between radius of catalyst particle and time for different RF powers, the time variations of thickness of Graphene sheet for different RF powers, etc. It also discusses the results obtained after running the equations, with the results obtained experimentally. Such hybrids are crucial for making materials with a diverse set of wondrous electrical, chemical, physical and optical properties.

Keywords. Graphene, Carbon Nanotube, Graphene CNT Hybrid, PECVD

1. Introduction

Carbon is the sole material known to us that exists in all dimensions- zero, one, two, and three. There are several forms of carbon that have their own unique physical characteristics. The most popular of all forms of carbon are graphite and diamond. As a result of the structural arrangement of atoms in their respective forms, their physical properties differ.

Another category of lesser-known forms of carbon are fullerenes; cylindrical fullerenes, otherwise known as nanotubes; closed cage fullerenes, more commonly known as buckyballs or buckminsterfullerene, and spherical fullerenes. Some other forms of carbon with ongoing research are amorphous carbon, Q carbon, etc. But in this paper, our sole focus of attention will be on two forms of carbon in particular - Graphene and Carbon Nanotubes.

[1] Vedant Chouksey, Corresponding author, Department of Applied Physics, Delhi Technological University, Delhi 110042, India; E-mail: vedantchouksey.nd@gmail.com.

Carbon Nanotube, often abbreviated as CNT, is a one-dimensional allotrope of carbon. Rolled-up sheets of one layer carbon atoms are combined to make the cylindrical structures known as CNT. When their diameter is less than 1 nm, they are called single-walled carbon nanotubes, i.e., SWCNT. Whereas MWCNT i.e., Multi-Walled carbon nanotubes, comprise of two or more concentric and linked nanotubes. Their diameters can be more than 100 nm, their length - up to even micrometers.

As a result of their bonding together through sp^2 bonds, CNTs form strong bonds. Due to Van Der Waals's forces, they also have a tendency to rope together. Due to these factors, they are researched upon to develop high strength, low weight materials. They also have optical, thermal, electrical, and conductive properties that potentially have many applications.

Graphene, on the other hand, is a singular sheet of graphite. It is a two-dimensional allotrope of carbon. It is essentially a sheet in a hexagonal layout of carbon atoms. Graphene is widely considered a "wonder material" because of its diverse extraordinary properties. For instance, it has a very high tensile strength, excellent electrical conductivity, exceptional properties that are optical and thermal in nature. These make graphene a metaphorical gold mine and an extensive area of research ever since its discovery in 2004. It has a wide array of uses in computer chips, medical equipment, batteries, LCDs and OLEDs, etc.

The invention of CNT was led by years of prior work [1]. In the past few years, there has been an increasing interest in research related to carbon nanostructures of various dimensions like diamond films (3-Dimensional), graphene (2-Dimensional), carbon nanotubes (1-Dimensional), fullerenes (0 Dimensional), etc. Amongst all these allotropes, graphene has been particularly attractive for researchers due to its electric, optical and chemical properties [2].

Well, known methods for graphene synthesis include chemical vapour deposition, chemical exfoliation, mechanical exfoliation, etc. [3]. Chemical Vapour Deposition (CVD) in the presence of plasma is the most feasible technique amongst all the techniques. It meets the standards of graphene [4].

PECVD (plasma-enhanced CVD) has an advantage over all other methods of CVD [5]. It offers growth in low temperature and affords greater control over the nanostructures' shape, size, and position. It is also very versatile and allows the growth of nanostructure with specified properties [6].

Wang et al. [7] observed the acceleration of motion of hydrocarbon radicals and enhancement of the hydrocarbon ion deposition on the substrate exterior, all as a result of plasma. Plasma also provides conditions for the separation of CNFs - Carbon Nanoflakes. In-depth research work has been executed to figure out the influence of growth time, hydrogen gas, the temperature of an electron, feedstock gas ratio, and RF power on the dimensions, rate of growth, and structure of graphene sheet [8-13]. Nang and Kim [8] reported improvement in the structural quality of graphene film and a decrease in its thickness to a single layer of graphene when the growth time and plasma power increased.

Liu et al. [9] have noticed that an increase in RF power to 700 W from 500 W resulted in the deposited carbon material changing from 1D to 2D. When RF power was increased to over 800W, the growth ratio was reduced.

Jiang et al. [10] investigated the gas rate of flow's effect of gas flow rate on the development of graphene sheets. It was noted that in the presence of a particular CH_4 rate of flow and proportion of methane to hydrogen, the size of the particle of graphene enlarges with growth time, after which it gets saturated. Synthesized films are shown to

possess superb field emission properties. Levchenko et al. [11] fabricated graphene flakes onto a catalyst, Y-Ni, with the use of magnetic field enhanced plasma. They also studied and analysed the variation in the flake size of graphene on the parameters, density in plasma, surface temperature, and temperature of an electron.

Kim et al. [12] created a graphene film on the exterior of a nickel foil by the method of Microwave PECVD. The number of graphene layers was found to increase when the mixing ratio of H_2 and CH_4 is decreased. Chan et al. [13] demonstrated the superior quality of graphene film is fabricated by increasing the concentration of hydrogen by PECVD. Zhu et al. [14] observed that after the initial graphene sheets grew parallelly aligned to the substrate, the upper carbon sheets curled up because of a force exerted at the grain boundaries.

Terasawa and Saiki [15] explained and analysed the mechanism of the development of graphene on copper using the PECVD technique. Ghosh et al. [16] explained the impact of substrate on nanosheets growth. Zhao et al. [17] studied graphene thermal profiles and behavior of interfacial charge transport. Zangwill and Vvedensky [18] put forth a rate theory about the development of graphene onto metals. This showed that the growth of 2D islands of graphene is done by joining a cluster of 5 atom clusters. The previous steady state to the growth of graphene that is smallest in size is synthesized when 6 of the moving 5 atom clusters strike each other.

Mehdipour and Ostrikov [19] reviewed the growth and nucleation of graphene when the catalyst is a metal and the process CVD. The complete processes of the interaction of neutral gas with the surface were involved. The contribution of H_2 in the growth of graphene, where copper is the catalyst, is also described through this. Kim et al. [20] have used the Langmuir adsorption theory framework and 2D crystallization to model graphene growth and nucleation by the process of CVD on a copper surface. This model constructed by them identifies a whole range of parameters of growth, like gas pressure and temperature, which result in the coverage of the surface of graphene.

Wang et al. [21] create a back gate triode emission device with a basis in the CNTs field emission characteristics. These were fabricated using the Radio Frequency PECVD technique. Aksak et al. [22] showed that CNT diameter decreases as temperature increases. The CNT lengths directly increase with temperature.

Loffler et al. [23] optimized CNT growth by the method of PECVD. Subsequently, the impact of the parameters of growth was noted on the rate of growth of carbon nanotubes. This was reiterated at higher power, CNTs are better aligned however this comes at the cost of a high fetch rate due to ion bombardment.

Abdi et al. [24] looked into the impact that plasma powers have on CNT developed by the PECVD method. The effect of plasma power is inversely proportional to the grain size of the Nickel nanoparticle. As a result of this, nanotubes with smaller diameters were created. Cho et al. [25] have studied the impact of NH_3 on thermal CVD growth CNTs. An increase in carbon nanotube length was observed along with a diameter decrease along with the flow of NH_3.

Levchenko et al. [26] compared plasma aided processes to neutral flux deposition. They discovered plasma aided processes to be more optimal to control the aspect ratio of CNT. Mehdipour et al. [27] have investigated the effects of parameters like the thickness of sheath, the pressure of the gas, electron number density, etc., on the rate of growth of nanofibers of carbon. Ostrikov et al. [28] looked into the process involving plasma during the growth of nanostructures.

As it can be clearly observed that both Graphene and CNT have certain optical, electrical, and mechanical properties that make them of particular interest in a wide

array of research. Therefore, naturally, the existence of two carbon allotropes with wondrous properties leads to curiosity regarding their hybrids.

This leads to the primary object of study in this paper - a hybrid of 2-dimensional graphene and a one-dimensional carbon nanotube. This hybrid has great potential. It can improve electrical properties by decreasing agglomeration, which could not otherwise be done using graphene or CNT individually. Such hybrids of CNT and graphene have found applications in electron field emitters, lithium-ion battery electrodes, transparent conductors, etc.

This paper studies the nucleation and growth of carbon nanotubes on graphene and the effect of plasma on the growth of hybrid.

There is a fair amount of research on growth of graphene on CNT hybrids where parameters like gas pressure, the bias of substrate, input power, etc., are varied, and their effects are observed in a theoretical model. This paper focuses on the reverse structure, i.e., carbon nanotubes on graphene. Keeping the previous papers in mind, the following paper discusses the analytical model of the relatively little worked-upon CNT-graphene hybrid where carbon nanotubes are grown on graphene, which can have significant applications in making flexible nanoelectronics. Factors like RF power and gas pressure are varied, and changes in electron number density, electron temperature, defects, the thickness of graphene, the radius of the catalyst particle, and the radius of CNT tip are observed.

2. Analytical Model

The model in this paper can be divided into three stages. First is the growth of graphene. Following that is catalyst deposition. It is finally concluded by the growth of carbon nanotubes. The graphene sheet is grown in a plasma consisting of $Ar + H_2 + C_2H_2$.

This plasma is made up of electrons and two types of ions - Type 1, that is, acetylene ions ($C_2H_2^+$), and Type 2, that is, Hydrogen (H^+). The substrate is silicon (Si), and the catalyst is Copper (Cu). Copper makes a good catalyst in the case of graphene growth because of its low chemical affinity, less carbon solubility, and minimal precipitation of Cu from the bulk. This is a result of the completed electron shell of Cu. [29]

First stage of the process considers certain processes for the development, growth, and nucleation of the sheet of graphene in the presence of plasma as shown in figure 1 .

- The catalyst film is broken down into catalyst nanoparticles by the plasma power that is applied. This leads to the formation of nucleation centers locally [30]
- Hydrocarbon ions that are formed by the ionization of the gas go through the plasma sheath. Then following many processes like dehydrogenation and thermal dissociation, they get adsorb and diffuse and get attached to each other and make clusters of carbon atoms [31]
- The clusters of carbon clash, mix, diffuse and make graphene islands. The graphene islands further come together to form larger graphene islands.

After this, a thin nickel film is grown on the graphene. The nickel film is broken down into nanoparticles of nickel by heating. At high temperatures and pressure, the carbon nanotubes begin to grow on the graphene with the aid of nickel [32]. The Ni-Ni

bonds are the primary bonding type due to which cluster size increases. The binding energy also increases with an increase in cluster size. Ni and C have strong affinity due to strong pi bond formation [33].

Finally, CNT is grown on the graphene in a plasma mixture of Ar^+, H_2^+ and CH_4. There are neutral species and charged species in the plasma derived from CH_4 and H_2.

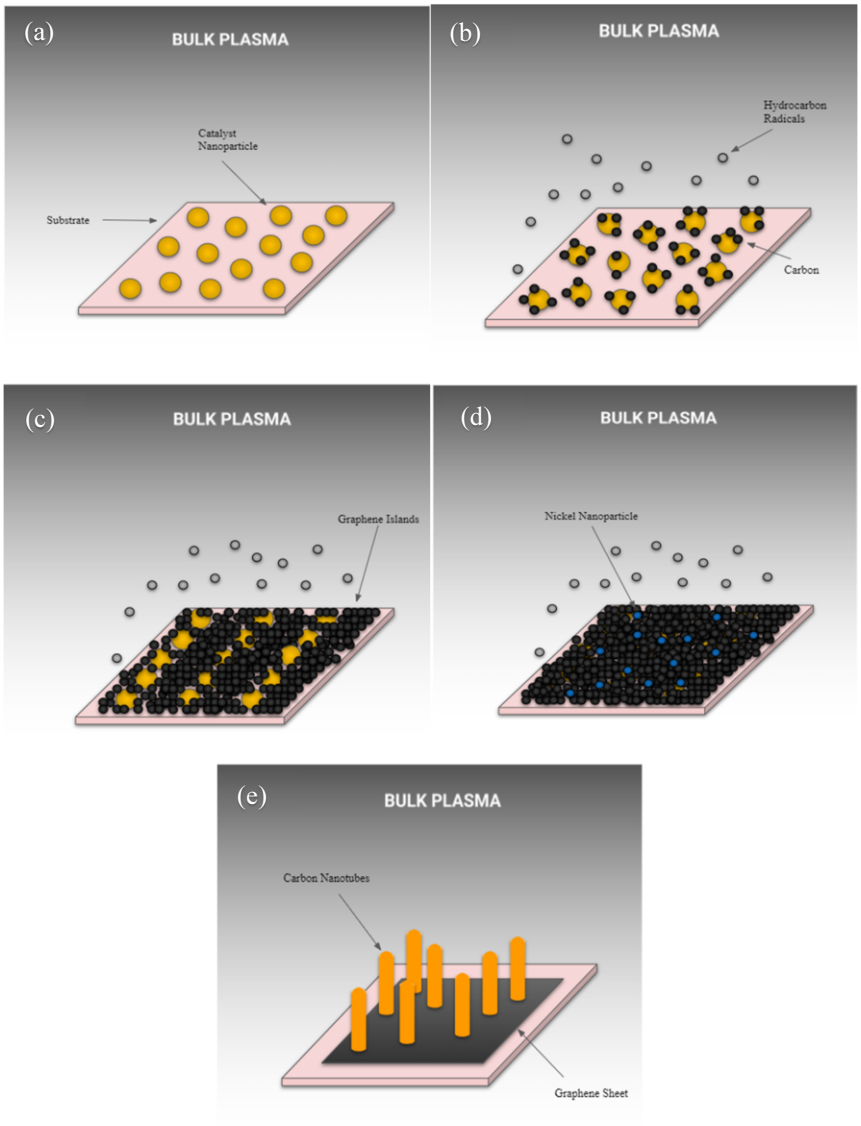


Figure 1. Schematic diagram of the growth process of growth of carbon nanotubes on graphene. (a) formation of catalyst nanoparticles on a substrate (b) graphene clusters growth on catalyzed substrate (c) graphene island growth leading to graphene sheet (d) Nickel catalyst deposition on graphene sheet (e) carbon nanotube growth on a catalyzed graphene sheet.

3. Equations

Equation (1) illustrates the energy balancing rate at the graphene surface. While the reaction with the plasma made of CH_4 and H_2 takes place, this surface is kept on the silicon substrate. Transferring of power to the surface of graphene by incident plasma species [Eq. (1)] results in graphene surface alteration and the creation of defects.

A_g, m_g, C_g, and q_g denote area, mass, specific heat, and density respectively. T_s is the graphene surface temperature, and the number of defects denoted by n_d, is formed per micron length in the surface of the graphene. Species that are positively charged or neutral of type j are denoted as J_{ij}.[34]

$$P_{in} = \partial / \partial t(m_g C_g T_s) = \rho_g A_g I_g C_g T_s \dot{n}_d / n_d$$
$$= [\sum_{J}^{1.2}[[Jij^{gtip} + J_{ij}^{gcurr})\varepsilon_{ij}^{C} + [Jij^{gtip} + J_{ij}^{gcurr})\varepsilon_{j}^{C}]] + (J_e^{gtip} + J_e^{gcurr})\varepsilon_e^{C}] + 3/2k_b T_s \sum_{J}^{1.2}(1-\Upsilon ij)(Jij^{gtip} + J_{ij}^{gcurr})$$
$$+[(v_0 \sigma_{ads} J_2 U_b) Ag] + \sum_{J}^{1.2} J_{ij} y_s (E_t)(1-\theta t) Ag$$
$$rg = a_g(1 + a_g / I_g)/2n_g \qquad (1)$$

As a result of the diffusion of the atoms of carbon at the edges of the islands formed by graphene, and the collection current due to ions and neutrals, the graphene sheet area increases. This is shown by Eq. (2).

Eq. (3) describes the etching by H_2 atoms, which causes a decrease in graphene sheet thickness. Type 2 flux of H_2 due to thermal dehydrogenation of type 1 ions is given by the 1st term, while type 2 ions which are adsorbed to the surface are given by the 2nd term. The disintegration of positively charged type 2 ions is the third term; the fourth term reflects ion-induced acetylene dissociation; and the 5th term denotes consolidation of H_2 ions due to acetylene ion's thermal dissociation. The sixth term refers to neutrals of acetylene adhering on the graphene sheet's surface.

$$ld(h^*t)/d\tau = (n_1 / v^* D_i \exp(-\{E_d + E_{imc}/k_b T_S\})\pi rg^2 / 2^*\pi^* r_g)^* I_{i\lg m} + (\gamma_{C2H2} I_{C2H2gn})^* M_{gn} / \rho_{gn} \qquad (2)$$

$$[h(\tau)+l]d(t)/d\tau = \{J_{i2}\exp(-\delta E_{sdh}/k_b T_s) + J_{i2}(1-\theta_t) + J_{i2} + \theta_{CH}(J_{i2} y_d + v_0 v \exp(-\delta E_t / k_b T_s))\}$$
$$^* h(\tau)/n_{i2} + \gamma_{C2H2}\pi r_{gn}^2 I_{C2H2gn}. \qquad (3)$$

CNT growth on the surface of the catalyst is constrained to reactive plasma of electrons, positively charged methyl ions (CH_3^+, CH_2^+). Denoted by ions A and ions B i.e. Hydrogen ions denoted as ions B, Substrate of silicon on which a Ni catalyst is positioned. The reactive plasma is made of Ar, H_2, CH_4 is taken into account, with CH_4 as the carbon source gas. The electric field caused by charge separation in this area is measured on the z-axis.

Continuity equation:

$$\partial n / \partial t + \nabla.(nu) = G - L \qquad (4)$$

The remaining terms of collision components constitute G and L .

Equation (4) reduces to

$$\nabla.(nu) = v_{iz} n_e \tag{5}$$

Second, ion momentum balance equation:

$$Mn_j u_j du_j / dz = en_j E - Mn_j v_{jn} u_j \tag{6}$$

M denotes ion mass of the plasma, u_j denotes fluid velocity, n_j denotes number density, u_{jn} denotes collision frequency, and the electric field is denoted by E.

The advancement of this is seen on nanoparticles of the catalyst, as shown in equations (7) and (8). CNT height is derived from Eq. (7), which depicts the growth of the cylindrical component, and the result of Equation (7) is used in Equation (8) to indicate CNT radius. The curved surface of the CNT tip is depicted in this equation. [35]

$$r_p d(2*\pi*h)/dt = (\{2*n_{ch} v \exp(-\delta Ei/kbTs) + 2\theta_{CH} j_{ia} y_d + 2j_{ia} + j_{ia}\sigma_{ads} j_{ib} + j_e v\} m_c + \{j_{ia}(1-\theta_t) + j_{ia}\sigma_{ads} j_H$$
$$+ j_{ia} v \exp(-\delta E_t / k_b T_s)\} m_{ia} * (D_s * 2/\rho_p * r_{ct}) * n_{iActcys} + \gamma_{CH3} \pi r_{ct}^2 n_{CH3ctcys} \tag{7}$$

$$d(\pi r_{ct}^2)/d\tau = \{j_{ib} v \exp(-E_b/k_b T_s) + j_{ib} v \exp(-\delta E_{th}/k_b T_s) + j_{ib}(1-\theta_t) + j_{ib} + \theta_{CH} * (j_{ib} y_d + vv_0 \exp(-\delta Ei/kbTs))\} h(\tau)/n_{ib} \tag{8}$$

4. Results and Discussion

This paper features a model based on the kinetics of plasma and plasma sheath structure which is used to grow carbon nanotubes on graphene. All equations related to the above model are solved using the Wolfram Application, 'MATHEMATICA'. The effects of variables like - substrate bias, total gas pressure, input power, and the composition of plasma, are explored by the Numerical and graphical solutions obtained through MATHEMATICA.

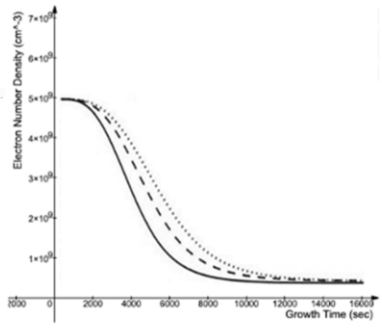


Figure 2. Plot of Electron Number Density with growth time at Pin= 300 W and Us = -50V

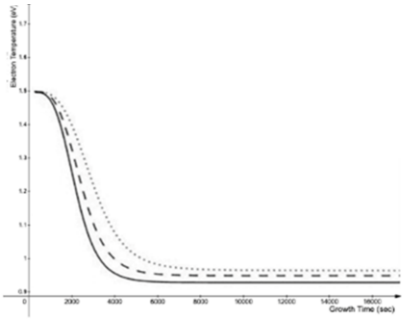


Figure 3. Variation of Electron Temperature vs growth time at Pin= 300 W and Us = -50V

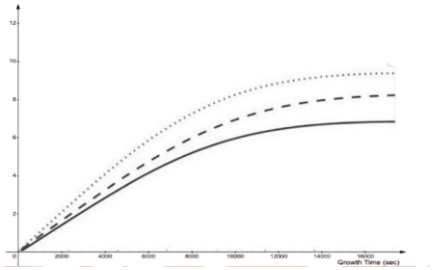


Figure 4. Variation of number of defects and growth time at Pin= 300 W and Us = -50V

Fig (2) shows the relationship of electron number density with growth time in plasma. Fig (3) shows the relationship between electron temperature and the growth time. [36] These graphs account for electron density increase rate by neutral species ionization, decay rate due to recombination of positive species and electrons, and collection current on the hybrid surface. The total gas pressure is set at U_s = -50V and P_{in}= 300 W. The electron temperature and electron number density both increase when the input power rises in constant pressure, electric field, and bias of the substrate.

Defects are formed on the surface of graphene due to plasma species that strike the surface. Neutral species formation, etching by hydrogen and ion sputtering lead to changes on the number density of surface defects. Fig (4) shows the relationship between the no. density of defects that are made per unit length and the growth time. The total gas pressure is set at U_s = -50V and P_{in}= 300 W. [37] We observe that when the pressure is constant, the number density of defects increases with time.

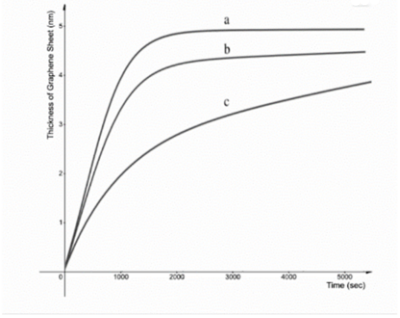


Figure 5. It represents the time variations of thickness of Graphene sheet against different Rf powers where a is for 50 W, b is for 100 W and c is for 200 W

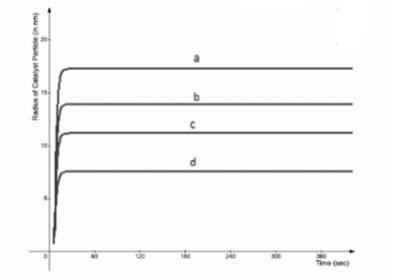


Figure 6. Relationship Between Radius of Catalyst Particle and time for different Rf powers where a is 50W, b is 100W, c is 150W and d is 200W

Fig. (5) shows the relationship between graphene thickness and time corresponding to 3 different RF powers where a is for 50 W, b is for 100 W, and c is for 200 W. We observe that as the RF power escalates, the graphene sheet thickness also drops. As the RF power rises in the plasma, the electric field grows stronger, which results in the increased bombardment of the graphene sheet by ions. The increased etching, which results in the decreased thickness of the sheet, is a result of the increased number density of ions of hydrogen. This changes the thickness of graphene sheet. [38]

Fig. (6) shows the relationship between the catalyst particle radius and time corresponding to 4 different RF Powers, where a is 50W, b is 100W, c is 150W and d is 200W. From the above graph, we can observe that the radius of the catalyst decreases with increasing RF power. Smaller and smaller nanoparticles are observed with increasing RF power [39]. The rise in plasma power leads to a stronger electric field in the plasma sheath; this results in an increased bombardment of ions on the catalyst particles. This etches the catalyst particles and reduces the radius. Instead, at the site of catalyst particles, the higher plasma power breaks down the hydrocarbon gas in the plasma. A change in the RF power can further change the Ni surface morphology. This, in turn, is found to affect diameter, rate of growth, and the density of the CNT. [40]

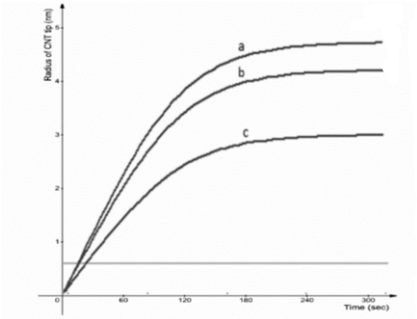


Fig 7. This figure shows the rate of change of radius of carbon nanotube tip with different temperatures

Fig. (7) shows the change in the radius of the carbon nanotube tip with time at various temperatures and ion density of type B ions. Here a is $n_{iB0} = 10^9 cm^{-3}$ and T_{i0} =1800 K, b is $n_{iB0} = 5 \times 10^{10} cm^{-3}$ and T_{i0} = 2000K and c is $n_{iB0} = 10^{11} cm^{-3}$ and T_{i0}= 2100K. It is observed that as the temperature of hydrogen ions and density of ions increases, the CNT radius decreases. This is a result of the bombardment of the tip of CNT by H^+. This bombardment etches the CNT tip, decreasing the radius. When the ion density and temperatures are sufficiently high, the hydrocarbons dissociate into hydrogen and carbon. This carbon adds to the CNT length and the hydrogen shapes the tip of the nanoparticle radius. Since the activation energy of H to Ni surface diffusion is more than the activation energy in the process of PECVD, i.e., 0.23 eV, we do not need to consider the possibility of hydrogen diffusion on the catalyst

5. Conclusion

The plasma kinetics and the growth of carbon nanotubes on graphene are studied and examined with the use of a descriptive analytical model. This model consists of the growth of graphene, the number density of defects of graphene, and the radius of the

Carbon nanotube tip while keeping charge neutrality kinetics plasma sheath interactions in view. Under controlled conditions, the different parameters are varied to obtain results. First, we observe the rise in the number density of electrons as the input power increases. Then we observe that the graphene sheet thickness decreases when bombarded with RF power is increased. The RF power also impacts the nickel catalyst, which in turn affects the CNT growth. Plots related to the electron number density and graphene defects are observed at different pressures. Higher RF power is observed to create a smaller radius of catalyst. The CNT tip radius is also found to reduce with the temperature and density of hydrogen ions.

CNT has incredible tensile strength and electric conductivity, whereas graphene has better thermal conductivity and tensile modulus. The combination of these two materials can give us some properties that would be useful in a myriad of industries. This is why different combinations and interactions of CNT-graphene hybrids are attempted. The growth of CNT on graphene can be particularly useful to get materials with unique field emission properties that can be used to design flexible nanomaterials. These particular type of hybrids possess the capacity to revolutionise the world of flexible nanomaterials

References

[1] Y. Wu, P. Qiao, T. Chong, and Z. Shen, Adv. Mater. 14, 64 (2002).
[2] C. N. R. Rao, A. K. Sood, K. S. Subrahmanyam, and A. Govindaraj, Angew. Chem. Int. Ed. 48, 7752 (2009).
[3] Y. H. Wu, T. Yu, and Z. X. Shen, J. Appl. Phys. 108, 071301 (2010).
[4] S. Vizireanu, S. D. Stoica, C. Luculescu, L. C. Nistor, B. Mituand, and G. Dinescu, Plasma Sources Sci. Technol. 19, 034016 (2010).
[5] I. Levchenko, M. Keidar, S. Xu, H. Kersten, and K. Ostrikov, J. Vac. Sci. Technol. B 31, 050801 (2013).
[6] K. Ostrikov, Plasma Nanoscience (John Wiley and Sons, (2008).
[7] B. B. Wang, K. Zheng, Q. J. Cheng, and K. Ostrikov, Appl. Surf. Sci. 325, 251 (2015).
[8] L. V. Nang and E.-T. Kim, J. Electrochem. Soc. 159, K93 (2012).
[9] W. Liu, T. Dang, Z. Xiao, X. Li, C. Zhu, and X. Wang, Carbon 49, 884 (2011).
[10] L. Jiang, T. Yang, F. Liu, J. Dong, Z. Yao, C. Shen, S. Deng, N. Xu, Y. Liu, and H.-J. Gao, Adv. Mater. 25, 250 (2013).
[11] I. Levchenko, O. Volotskova, A. Shashurin, Y. Raitses, K. Ostrikov, and M. Keidar, Carbon 48, 4570 (2010).
[12] Y. Kim, W. Song, S. Y. Lee, C. Jeon, W. Jung, M. Kim, and C.-Y. Park, Appl. Phys. Lett. 98, 263106 (2011).
[13] S. H. Chan, S. H. Chen, W. T. Lin, M. C. Li, Y. C. Lin, and C.-C. Kuo, Nanoscale Res. Lett. 8, 285 (2013).
[14] M. Zhu, J. Wang, B. C. Holloway, R. A. Outlaw, X. Zhao, K. Hou, V. Shutthanandan, and D. M. Manos, Carbon 45, 2229 (2007)
[15] T.-O. Terasawa and K. Saiki, Carbon 50, 869 (2012).
[16] S. Ghosh, K. Ganesan, S. R. Polaki, T. Mathews, S. D(2017).
[17] J. Zhao, M. Shaygan, J. Eckert, M. Meyyappan, and M. H. R€ummeli, Nano Lett. 14, 3064 (2014).
[18] A. Zangwill and D. D. Vvedensky, Nano Lett. 11, 2092 (2011).
[19] H. Mehdipour and K. Ostrikov, ACS Nano 6, 10276 (2012).
[20] H. Kim, E. Saiz, M. Chhowalla, and C. Mattevi, New J. Phys. 15, 053012 (2013).
[21] S. Wang, J. Wang, P. Miraldo, M. Zhu, R. Outlaw, K. Hou, X. Zhao, B. C. Holloway, D. Manos, T. Tyler, O. Shenderova, M. Ray, J. Dalton, and G. Mc Guire, Appl. Phys. Lett. 89, 183103 (2006).
[22] M. Aksak, S. Kir, and Y. Selamet, J. Optoelectron. Adv. Mater. Symposia 1, 281 (2009).
[23] R. Loffler, M. Haffner, G. Visanescu, H. Weigand, X. Wang, D. Zhang, M. Fleischer, A. J. Meixner, J. Fortagh, and D. P. Kern, Carbon 49, 4197 (2011).
[24] Y. Abdi, E. Arzi, and S. Mohajerzadeh, Eur. Phys. J. Appl. Phys. 44, 149 (2008).
[25] H. Jin Cho, N. Lee, I. Jang, H. Soo Uh, and J. Pyo Hong, J. Korean Phys. Soc. 50, 1848 (2007).
[26] I. Levchenko, K. Ostrikov, J. Khachan, and S. V. Valdimirov, Phys. Plasmas 15, 103501 (2008).

[27] H. Mehdipour, K. Ostrikov, and A. E. Rider, Nanotechnology 21, 455605 (2010).
[28] K. Ostrikov, E. C. Neyts, and M. Meyyappan, Adv. Phys. 62, 113 (2013).
[29] Riikonen, A. V. Krasheninnikov, L. Halonen, and R. M. Nieminen, J. Phys. Chem. C 116, 5802 (2012),
[30] K. Ostrikov, Plasma Nanoscience (John Wiley and Sons, (2008).
[31] H. Mehdipour and K. Ostrikov, ACS Nano 6, 10276 (2012).
[32] Bisht, Atul & Chockalingam, Sreekumar & Panwar, O.s & Kesarwani, A. & Singh, Bhanu Pratap & Singh, Vidya Nand. (2015).
[33] Wang Rao,Jonathan E. Mueller, Josef Anton, Qing Jiang, and Timo Jacob(2013).
[34] N.Gupta and S. C. Sharma , Physics of Plasmas 24, 073510 (2017)
[35] A.Tewari and S. C. Sharma , Physics of Plasmas 21, 063512 (2014)
[36] W. Takeuchi, H. Sasaki, S. Kato, S. Takashima, M. Hiramatsu, and M.Hori, J. Appl. Phys. 105, 113305 (2009).
[37] H. J. Cho, H. Kondo, K. Ishikawa, M. Sekine, M. Hiramatsu, and M. Hori, Carbon 68, 380 (2014).
[38] L. V. Nang and E.-T. Kim, J. Electrochem. Soc. 159, K93 (2012).
[39] Young Chul Choi, Young Min Shin, Young Hee Lee, Byung Soo Lee, Gyeong-Su Park (2000)
[40] N. Gupta, S. C. Sharma, and R. Sharma, Phys. Plasmas 23, 083509 (2016)

International Scholarly Research Network
ISRN Nanomaterials
Volume 2012, Article ID 809063, 6 pages
doi:10.5402/2012/809063

Research Article

Effect of Polyvinyl Alcohol on the Growth, Structure, Morphology, and Electrical Conductivity of Polypyrrole Nanoparticles Synthesized via Microemulsion Polymerization

Anurag Krishna,[1] Amit Kumar,[2] and Rajiv Kumar Singh[2]

[1] *Department of Applied Chemistry and Polymer Technology, Delhi Technological University, Bawana Road, New Delhi 110042, India*
[2] *National Physical Laboratory, Council of Scientific and Industrial Research, Dr. K. S. Krishnan Marg, New Delhi 110012, India*

Correspondence should be addressed to Rajiv Kumar Singh, rajivsingh@mail.nplindia.ernet.in

Received 6 July 2012; Accepted 27 August 2012

Academic Editors: A. Kajbafvala, B. Panchapakesan, A. Sorrentino, and C. Wang

Copyright © 2012 Anurag Krishna et al. 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.

Polypyrrole (PPy) nanoparticles were synthesized via microemulsion polymerization technique using sodium dodecyl sulfate as surfactant. Polyvinyl alcohol (PVA) was added as soft template during polymerization to modify the structure and properties of PPy nanoparticles. The synthesized materials namely, PVA-free and PVA added were characterized by Fourier transform infrared spectroscopy (FT-IR), X-ray diffraction (XRD), scanning electron microscopy (SEM), and DC electrical conductivity measurements. The sample synthesized in the presence of PVA has longer conjugation length as estimated from FT-IR investigation. Temperature dependence (4.2–300 K) of DC electrical conductivity measurement reveals that the PVA has a strong effect on the polymerization mechanism of PPy giving evidence of H-bonded assistance during polymerization leading to the synthesis of better ordered polymer. A growth mechanism has been proposed which explains the H-bonded assistance of PPy polymerization leading to enhanced structural ordering.

1. Introduction

Conjugated polymers have attracted considerable attention in the past few decades because of their potential application in electronic devices [1]. Their ease of processing and chemically tunable properties makes them useful for electronic, optoelectronic, electromechanical, and sensing device application [2, 3]. During the recent years conducting polymer nanostructures have received increasing attention from both fundamental research as well as application point of view. Conducting polymer nanostructures show high electrical conductivity, large surface area, short path lengths for the transport of ions, and high electrochemical activity as compared to its macrogranular structure or self-supporting films [4–7]. It can be synthesized by several approaches such as well-controlled solution synthesis [8, 9], soft-template methods [10], hard-template methods [11], and electrospinning technology [12]. Recently some conducting polymer nanowires and nanorods have been synthesized via hydrogen bonding [13].

Polypyrrole (PPy) is one of the most studied conducting polymers because of its good environmental stability, facile synthesis, ion exchange capacity, biocompatibility, and higher conductivity [14, 15]. It can be used in drug delivery, rechargeable batteries, supercapacitors, anhydrous electrorheological fluids, microwave shielding, and corrosion protection [16–18]. Soft template uses microemulsion polymerization which allows particles to transfer into spherical aggregates through the surfactant template for the production of PPy nanostructures. Conducting polymer nanoparticles (particle size) ~50–200 nm [19, 20] and ~20–50 nm [21] have been synthesized. It has been observed that the microemulsion polymerization has increased the yield of the PPy nanoparticles, the extent of the π-conjugation

Table 1: Peaks observed in FT-IR spectra of Sample A and B.

Absorption bands	Samples A (cm^{-1})	Sample B (cm^{-1})
C=C	1553.1	1538
C–C	1481.1	1453
C–N (in plane def.)	1280	1289
Vibration of pyrrole ring	1211	1156
N–H (in plane def.)	1044	1025
C–H out of plane vibrations	925	960
C–H wagging	787	764
N–H out of plane vibration	679	690

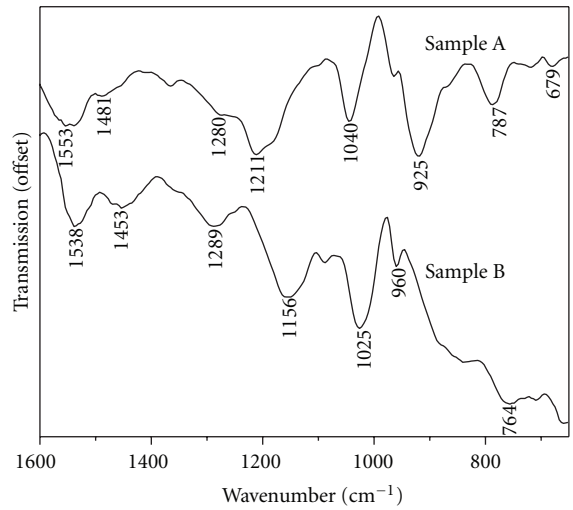


Figure 1: FT-IR spectra of PPy nanoparticles synthesized in the absence (sample A) and presence (sample B) of PVA.

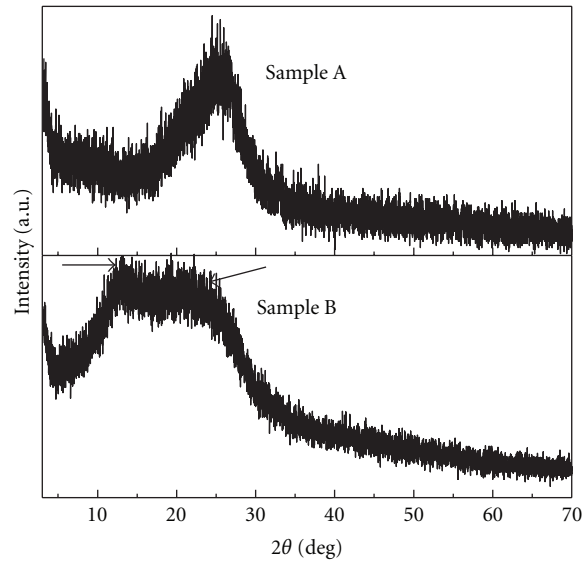


Figure 2: X-ray diffraction pattern of samples synthesized in the absence (sample A) and presence (sample B) of PVA.

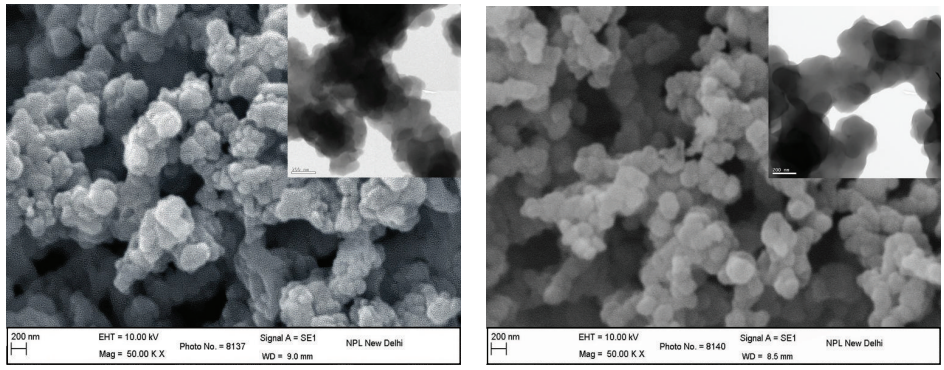
along the polymer backbone, and the ordered arrangement of the macromolecular chains. Song et al. 2004 have reported the synthesis of PPy nanoparticles doped with a variety of alkylbenzenesulfonic acid (ABSA) [22]. PPy doped with short alkyl chain showed higher conductivity than long alkyl chain. These achievements indicate that the microemulsion polymerization is powerful technique for the fabrication of polymer nanostructures.

In the present paper, we report the synthesis of PPy nanoparticles by microemulsion polymerization. The temperature-dependent (4.2 to 300 K) dc conductivity measurement along with structural investigations indicates that the addition of PVA during the polymerization affects the polymerization mechanism.

2. Experimental

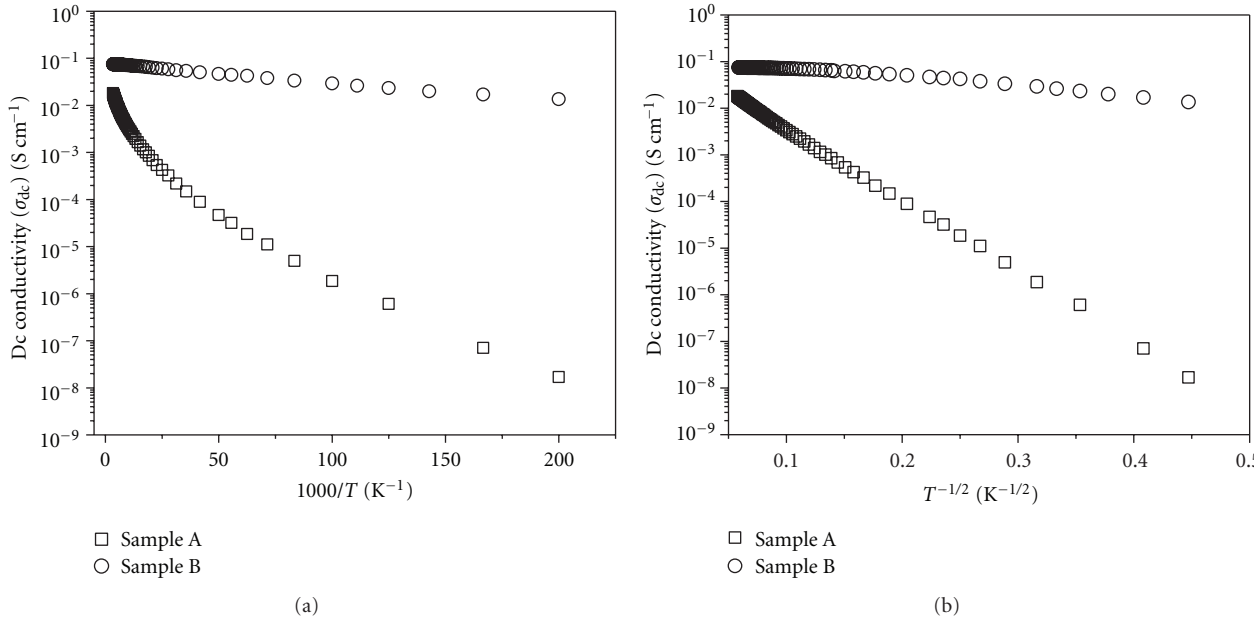
PPy nanoparticles were synthesized using doubly distilled 0.1 M pyrrole monomer (Fluka chemie) with 0.5 M ferric chloride ($FeCl_3$) (Sigma Aldrich). Sodium dodecyl sulfate (SDS) (Fluka) (0.03 M) was used as surfactant. The solution of SDS was prepared in distilled water (18 MΩ cm). Continuous stirring was carried out for 6 hrs at 275 K in inert atmosphere during polymerization. The black precipitate of PPy from the reaction mixture was filtered via vacuum filtration. The filtered precipitate was washed repeatedly with methanol and distilled water till clear and colorless filtrate was obtained. Finally, the PPy nanoparticles were dried for 6 hrs in vacuum oven at 333 K and was designated as sample A. PPy nanoparticles were also synthesized with the addition of polyvinyl alcohol (PVA) (5% by weight of pyrrole) in the abovementioned reaction mixture, and similar procedure was followed for polymerization and filtration and was designated as sample B.

Fourier transform infrared (FT-IR) spectra of both samples A and B were recorded using Smart orbit ATR-single reflection accessory of the Thermo Scientific Nicolet 5700 spectrometer, with a diamond crystal, taking four scans at a resolution of 4 cm^{-1}, and details of different vibration peaks are given in Table 1. X-ray diffraction of samples A and B was taken by Rigaku-make powder X-ray diffractometer for Cu-Kα (λ = 1.5404 Å) radiation. The scanning electron micrographs (SEMs) of samples A and B were taken using SEM model Zeiss EVO MA-10 microscope to investigate the surface morphology of the samples. A thin layer of gold was sputtered on the samples before loading in the microscope probe to nullify any charging effect. For dc electrical conductivity (σ_{dc}) measurement, circular pellets of both samples A and B having dia ~5 mm were made by using hydraulic press at $\sim 5 \times 10^8$ Pa. Conductivity measurements were performed by using four-probe technique. Keithley 238 high current source measuring unit was used for applying constant current and Keithley 6517A Electrometer/High Resistance meter was employed to measure the voltage. The thickness of the samples was accurately measured to calculate the bulk conductivity. Temperature-dependent (4.2–320 K) dc conductivity measurement was performed in liquid helium Dewar using Keithley's 224 Programmable Constant Current Source and 195 A Digital Multimeter.



(a) (b)

Figure 3: (a) SEM of PPy nanoparticles (sample A) and (b) SEM of PPy (sample B). Inset shows their corresponding TEM images.



(a) (b)

Figure 4: Plot of dc conductivity (σ_{dc}) versus (a) 1000/T and (b) $T^{-1/2}$ for both the samples in the temperature range 4.2–320 K.

3. Results and Discussion

A key property of a conducting polymer is the presence of conjugated double bond which gives rise to the electrical conductivity, as it allows the efficient transfer of electrons or positive charges along the polymer backbone. Increasing conjugation length and improving order throughout polymer chain will allow charge to migrate along a longer distance and hence enhances the conductivity. In the present work, the conjugation length and structural ordering of PPy nanoparticles have been influenced by the addition of polyvinyl alcohol (PVA) during the microemulsion polymerization.

Figure 1 shows the FT-IR spectra of samples A and B confirming the synthesis of polymers. Sample A shows the peaks at 1553 and 1481 cm^{-1} which correspond to C=C-stretching and C–C-stretching vibrations, respectively [23]. The peaks at 1280 and 1211 cm^{-1} are attributed to the C–N in plane deformation and vibrations of the pyrrole ring. The peak at 1040 cm^{-1} is attributed to N–H in plane deformation vibration [23, 24]. The peaks at 787 and 925 cm^{-1} correspond C–H wagging and C–H out of plane vibrations, respectively. The peak at 679 cm^{-1} is attributed to N–H out of plane vibration [24].

Sample B shows the peaks at 1538 and 1453 cm^{-1} which correspond to the stretching vibration of C=C and C–C, respectively [23]. The peaks at 1289 cm^{-1} correspond to C–N in plane deformation. The peak at 1156 cm^{-1} is attributed to the vibration of the pyrrole ring. The band of N–H in plane deformation is located at 1025 cm^{-1} while C–H out of vibration is found at 960 cm^{-1} [23, 24]. The peaks at 764 and 690 cm^{-1} corresponds to C–H wagging vibration and N–H out of plane vibration, respectively [24]. The characteristic peaks of PVA could not be observed in the spectrum, indicating that alcohol has been washed out from the final PPy nanostructures.

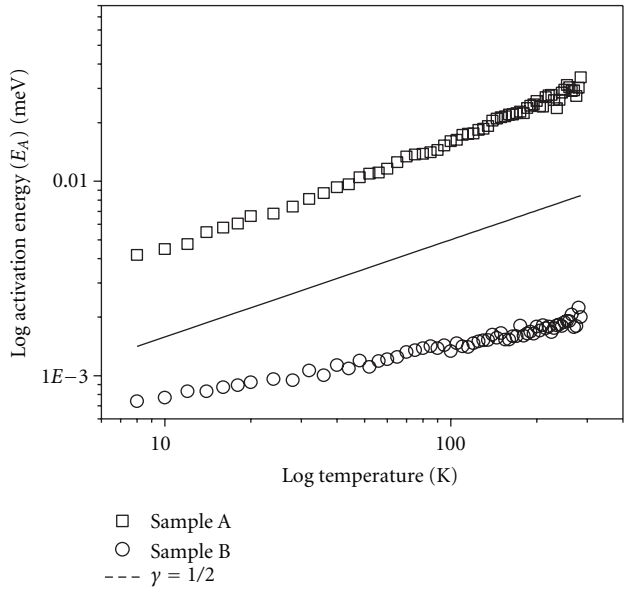


Figure 5: Plot of log E_A (E_A is the activation energy derived from Figure 4(a)) versus log T in the temperature range 8–290 K.

The FT-IR spectra for both samples A and B have all the characteristic peak of PPy, which suggest that it has been successfully synthesized. However, there are few differences between the spectra of samples A and B. The band of C=C stretching vibration located at 1553 cm^{-1} in the FT-IR spectrum of sample A synthesized in the absence of PVA shifts to 1538 cm^{-1} for sample B synthesized in the presence of PVA. Due to this shift the room temperature (~300 K) conductivity of sample B (~7.5×10^{-2} S/cm) is higher than sample A (~1.6×10^{-2} S/cm). The intensity of the C=C and C–C band in the two samples is different which can be linked with the conjugation length of PPy. The conjugation length is estimated from the ratio of peak intensity of C=C and C–C stretching vibrations [25]. The higher ratio represents longer effective π conjugation along the PPy chains and hence higher conductivity. PPy synthesized using PVA shows higher conjugation (~0.98) than the one synthesized without it (~0.91). This shows that the PVA considerably influenced the molecular structure of PPy. The hydrogen in the O–H group of PVA is capable of forming bond with the hydrogen in the N–H group of the pyrrole ring. These hydrogen bonds are known to be weak, and PVA is soluble in water, and quantity of PVA added was very small; therefore, these bonds get broken, and PVA is removed during the washing of the sample. FTIR confirms that this as no characteristic peak for PVA is observed.

Figure 2 shows the XRD pattern of the synthesized PPy nanoparticles with and without PVA. PPy nanoparticles synthesized in the absence of PVA (sample A) exhibit broad scattering peaks at 2θ value around 25.6°, which suggest that PPy is virtually amorphous. PPy nanoparticles synthesized in the presence of PVA (sample B) exhibit two peaks at 2θ values around 14.1° and 23.45° with almost same intensity which suggests that this sample B is less amorphous than sample A. It can be said here that sample synthesized in the presence of PVA is partially crystalline and has better ordering than sample synthesized in the absence of PVA. Higher ordering in PPy prepared in the presence of PVA may be attributed to the alignment given to the PPy chains via hydrogen bonding between N–H group of pyrrole ring and oxygen atom of PVA. The d spacing (interlayer spacing) value for the sample A is ~3.46 Å under whereas for sample B the values of d_1 and d_2 are ~6.28 Å and 3.79 Å, respectively.

The SEM micrograph in Figure 3(a) represents the general morphological features of the PPy nanoparticles obtained without using PVA (sample A) during the experiment process. The SEM image reveals the presence of globular particles with diameter ranging from 50 to 100 nm. These globular particles conglutinate together to form bigger aggregates. The size measured by particle size analyzer was found to around 80 to 100 nm (result not shown). Figure 3(b) shows the typical SEM images of PPy nanoparticles obtained when PVA was introduced to the experiment process (sample B). The SEM image reveals the existence of lamella disc-like structures with diameter ranging from 150 to 200 nm. The size measured by particle size analyzer was found to around 200 to 300 nm (result not shown). On careful observation it can be said that the sample looks well-ordered and the level of aggregation in this sample is very low as compared to the sample prepared in the absence of PVA. This suggests that the presence of the PVA in polymerization mixture has strongly influenced the morphology of synthesized PPy and is also responsible for better order and reduction in the aggregations. The inset in Figures 3(a) and 3(b) shows their respective transmission electron micrographs which further support the variation in particle size as well as a core-shell structure, the material of the shell being amorphous, and the core being formed by the more closely packed polymer macromolecules.

The variation of dc conductivity (σ_{dc}) for both the samples A and B is shown in Figure 4 as functions of (a) 1000/T, (b) $T^{-1/2}$ in the temperature range 4.2–300 K. The temperature-dependent variation of dc conductivity data indicates that the charge transport seems to occur by phonon-assisted hopping or thermally activated jumps between localized states [26–28]. According to Mott and Davis, the dc conductivity (σ_{dc}) can be written [27] as

$$\sigma_{dc} = \sigma_o \exp\left\{-\left(\frac{T_0}{T}\right)^{\gamma}\right\}, \quad (1)$$

where σ_0 is the infinite temperature conductivity, σ_{dc} is the conductivity at temperature T, and T_0 is the Mott's characteristic temperature that determines the thermally activated hopping among localized states at different energies and is also considered as a measure of disorder. We have $\gamma = 1/(d+1)$, where d is the dimensionality. The values of $d = 1, 2, 3$ signify the 1-dimensional, 2-dimensional, and 3-dimensional Mott's VRH. A plot of log σ_{dc} versus $(1/T)^{\gamma}$ should give a straight line for proper value of γ. It is evident from this figure that for both the samples that is, prepared in presence (sample A) and absence of PVA (sample B) exhibits the linear dependence of log σ_{dc} versus $T^{-1/2}$ (Figure 4(b)) and is better than that of log σ_{dc} versus $T^{-1/4}$ (result not

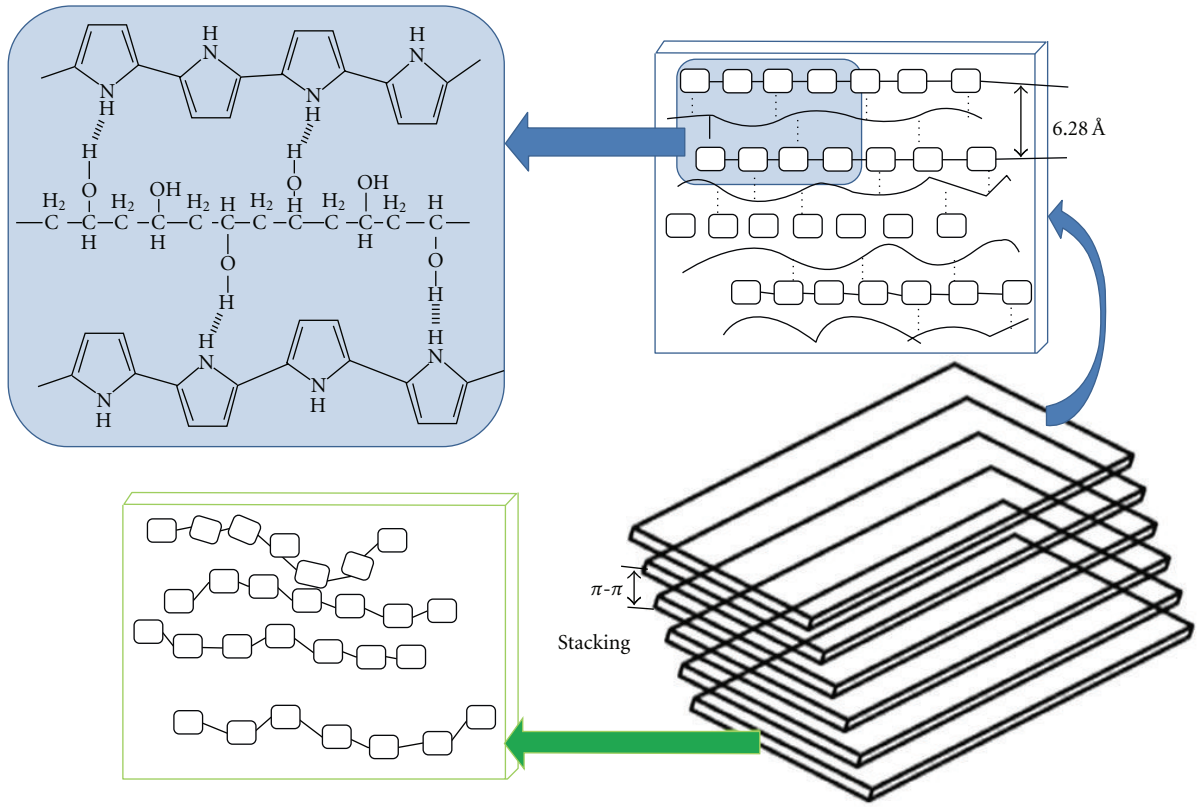


Figure 6: Mechanism of growth of PPy in presence of PVA showing the H-bonding assisted growth.

shown). The linearity factor of the PPy prepared in PVA is 0.99935 while for PPy prepared in absence of PVA is 0.96196. These linearity factors illustrate that the Mott's 1D-VRH mechanism of $T^{-1/2}$ type seems to be more appropriate [27, 29] for explaining the mechanism of dc conduction in both PPy samples.

The temperature region where (1) is valid should gives [26] the activation energy as

$$E_A = \frac{-d(\ln \sigma)}{d(1/k_B T)}, \tag{2}$$

where k_B is the Boltzmann's constant and E_A is the activation energy. Equation (2) can be correlated to the parameters of (1) and can be written as

$$E_A = \gamma k_B T_0 \left(\frac{T_0}{T}\right)^{\gamma-1}. \tag{3}$$

It is evident from (3) that a plot of log E_A versus log T should yield a straight line of slope = $(-(\gamma-1))$. To verify this, the activation energy (E_A) evaluated at different temperatures from Figure 4(a) has been plotted as log E_A versus log T in the temperature range 8–290 K and is shown in Figure 5. It can be clearly seen that the behavior of activation energy with temperature matches with a straight line corresponding to γ = 1/2 which is shown as a solid line in Figure 5. This gives a clear indication that the Mott's VRH mechanism of $T^{-1/2}$ type dominates the mechanism of charge transport in both PPy samples. Equations (1) and (3) give the slope T_0 from log σ_{dc} versus $T^{-1/2}$ plots (Figure 4) [27, 30] as

$$T_0 = \frac{B_0 \alpha^3}{k_B N(E_F)}, \tag{4}$$

where T_0 is the characteristic temperature, B_0 is dimensionless constant ~1.66 [27, 30], k_B is Boltzmann's constant, $N(E_F)$ is the density of states at the Fermi level, and α (=$1/r_P$) is the coefficient of exponential decay of the localized states involved in the hopping process.

Based on the experimental observations, a mechanism explaining the role of PVA in improving the order and conjugation length has been proposed in Figure 6. The hydrogen in the O–H group of PVA is capable of forming bond with the hydrogen in the N–H group of the pyrrole ring. Thus hydrogen bonding takes place between the N–H group of the Pyrrole ring and O–H group of the PVA molecule. Now, PVA acts as a bridging agent between the PPy chains and forms a temporary soft template which controls the growth of the PPy chains as shown in Figure 6. PVA forms linkage between the PPy chains via hydrogen bonding. These hydrogen bonds act as temporary soft template which is responsible for increasing the order of the PPy nanoparticles and hence increase in conjugation length and conductivity.

4. Conclusion

PPy nanoparticles synthesized by microemulsion polymerization in the presence of PVA show better ordering and conjugation length than those synthesized in absence of PVA. This study shows the importance of H-bonding assistance during polymerization and gives rise to a promising method to the controllable synthesis of other conducting polymers with high conductivity.

Acknowledgments

The authors are thankful to Director of National Physical Laboratory, New Delhi for encouragement. The authors thank Dr. Ramadhar Singh for helpful discussions.

References

[1] Z. Yao, H. W. C. Postma, L. Balents, and C. Dekker, "Carbon nanotube intramolecular junctions," *Nature*, vol. 402, no. 6759, pp. 273–276, 1999.

[2] K. Ramanathan, M. A. Bangar, M. H. Yun, W. Chen, A. Mulchandani, and N. V. Myung, "Individually addressable conducting polymer nanowires array," *Nano Letters*, vol. 4, no. 7, pp. 1237–1239, 2004.

[3] J. W. Gardner and P. N. Bartlett, "Brief history of electronic noses," *Sensors and Actuators B*, vol. B18, no. 1, pp. 211–220, 1994.

[4] J. Huang, S. Virji, B. H. Weiller, and R. B. Kaner, "Polyaniline nanofibers: facile synthesis and chemical sensors," *Journal of the American Chemical Society*, vol. 125, no. 2, pp. 314–315, 2003.

[5] H. Peng, L. Zhang, C. Soeller, and J. Travas-Sejdic, "Conducting polymers for electrochemical DNA sensing," *Biomaterials*, vol. 30, no. 11, pp. 2132–2148, 2009.

[6] R. J. Tseng, J. Huang, J. Ouyang, R. B. Kaner, and Y. Yang, "Polyaniline nanofiber/gold nanoparticle nonvolatile memory," *Nano Letters*, vol. 5, no. 6, pp. 1077–1080, 2005.

[7] N. R. Chiou, C. M. Lui, J. J. Guan, L. J. Lee, and A. J. Epstein, "Growth and alignment of polyaniline nanofibres with superhydrophobic, superhydrophilic and other properties," *Nature Nanotechnology*, vol. 2, pp. 354–357, 2007.

[8] D. Li and R. B. Kaner, "Shape and aggregation control of nanoparticles: not shaken, not stirred," *Journal of the American Chemical Society*, vol. 128, no. 3, pp. 968–975, 2006.

[9] N. R. Chiou and A. J. Epstein, "Polyaniline nanofibers prepared by dilute polymerization," *Advanced Materials*, vol. 17, no. 13, pp. 1679–1683, 2005.

[10] M. X. Wan, "A template-free method towards conducting polymer nanostructures," *Advanced Materials*, vol. 20, no. 15, pp. 2926–2932, 2008.

[11] M. Yang, J. Ma, C. L. Zhang, Z. Z. Yang, and Y. F. Lu, "General synthetic route toward functional hollow spheres with double-shelled structures," *Angewandte Chemie International Edition*, vol. 44, no. 41, pp. 6727–6730, 2005.

[12] M. K. Shin, Y. J. Kim, S. I. Kim et al., "Enhanced conductivity of aligned PANi/PEO/MWNT nanofibers by electrospinning," *Sensors and Actuators B*, vol. 134, no. 1, pp. 122–126, 2008.

[13] J. Zang, C. M. Li, S. J. Bao, X. Cui, Q. Bao, and C. Q. Sun, "Template-free electrochemical synthesis of superhydrophilic polypyrrole nanofiber network," *Macromolecules*, vol. 41, no. 19, pp. 7053–7057, 2008.

[14] Y. C. Liu and C. J. Tsai, "Enhancements in conductivity and thermal and conductive stabilities of electropolymerized polypyrrole with caprolactam-modified clay," *Chemistry of Materials*, vol. 15, no. 1, pp. 320–326, 2003.

[15] B. R. Saunders, R. S. Fleming, and K. S. Murray, "Recent advances in the physical and spectroscopic properties of polypyrrole films, particularly those containing transition-metal complexes as counteranions," *Chemistry of Materials*, vol. 7, no. 6, pp. 1082–1094, 1995.

[16] J. M. Pernaut and J. R. Reynolds, "Use of conducting electroactive polymers for drug delivery and sensing of bioactive molecules. A redox chemistry approach," *The Journal of Physical Chemistry B*, vol. 104, no. 17, pp. 4080–4090, 2000.

[17] K. Jurewicz, S. Delpeux, V. Bertagna, F. Béguin, and E. Frackowiak, "Supercapacitors from nanotubes/polypyrrole composites," *Chemical Physics Letters*, vol. 347, no. 1–3, pp. 36–40, 2001.

[18] J. W. Goodwin, G. M. Markham, and B. Vinent, "Studies on model electrorheological fluids," *The Journal of Physical Chemistry B*, vol. 101, no. 11, pp. 1961–1967, 1997.

[19] J. Jang and J. H. Oh, "Novel crystalline supramolecular assemblies of amorphous polypyrrole nanoparticles through surfactant templating," *Chemical Communications*, no. 19, pp. 2200–2201, 2002.

[20] Y. Liu and Y. Chu, "Surfactant-assisted synthesis of single crystal $BaWO_4$ octahedral microparticles," *Materials Chemistry and Physics*, vol. 92, no. 1, pp. 59–63, 2005.

[21] F. Yan, G. Xue, and M. Zhou, "Preparation of electrically conducting polypyrrole in oil/water microemulsion," *Journal of Applied Polymer Science*, vol. 77, no. 1, pp. 135–140, 2000.

[22] M. K. Song, Y. T. Kim, B. S. Kim, J. Kim, K. Char, and H. W. Rhee, "Synthesis and characterization of soluble polypyrrole doped with alkylbenzenesulfonic acids," *Synthetic Metals*, vol. 141, no. 3, pp. 315–319, 2004.

[23] Y. Liu, Y. Chu, and L. Yang, "Adjusting the inner-structure of polypyrrole nanoparticles through microemulsion polymerization," *Materials Chemistry and Physics*, vol. 98, no. 2-3, pp. 304–308, 2006.

[24] M. Omastová, M. Trchová, J. Kovarová, and J. Stejskal, "Synthesis and structural study of polypyrroles prepared in the presence of surfactants," *Synthetic Metals*, vol. 138, no. 3, pp. 447–455, 2003.

[25] V. P. Menon, J. Lei, and C. R. Martin, "Investigation of molecular and supermolecular structure in template-synthesized polypyrrole tubules and fibrils," *Chemistry of Materials*, vol. 8, no. 9, pp. 2382–2390, 1996.

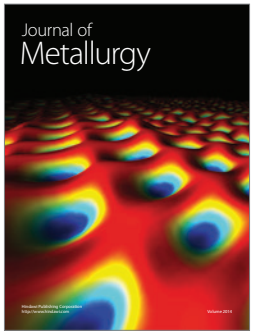
[26] R. K. Singh, A. Kumar, K. Agarwal et al., "DC electrical conduction and morphological behavior of counter anion-governed genesis of electrochemically synthesized polypyrrole films," *Journal of Polymer Science Part B*, vol. 50, no. 5, pp. 347–360, 2012.

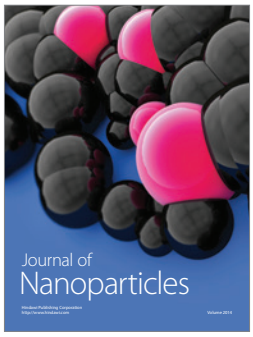
[27] N. F. Mott and E. A. Davis, *Electronic Processes in Noncrystalline Materials*, Oxford University Press, London, UK, 2nd edition, 1979.

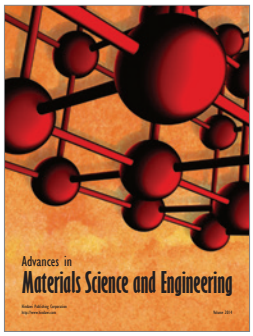
[28] R. K. Singh, A. Kumar, and R. Singh, "Mechanism of charge transport in poly(2,5-dimethoxyaniline)," *Journal of Applied Physics*, vol. 107, no. 11, Article ID 113711, 7 pages, 2010.

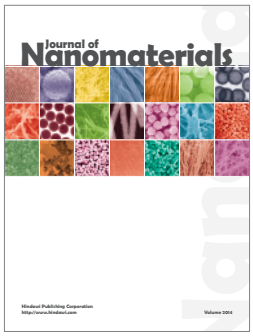
[29] Z. H. Wang, A. Ray, A. G. MacDiarmid, and A. J. Epstein, "Electron localization and charge transport in poly(o-toluidine): a model polyaniline derivative," *Physical Review B*, vol. 43, no. 5, pp. 4373–4384, 1991.

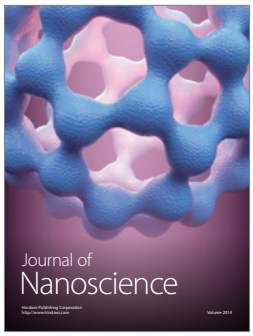
[30] D. S. Maddison and T. L. Tansley, "Variable range hopping in polypyrrole films of a range of conductivities and preparation methods," *Journal of Applied Physics*, vol. 72, no. 10, pp. 4677–4682, 1992.

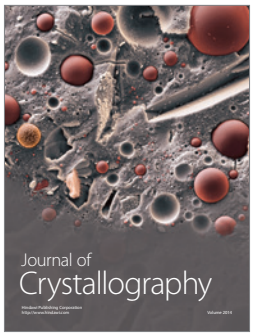












Hindawi

Submit your manuscripts at
http://www.hindawi.com

Wireless Personal Communications
https://doi.org/10.1007/s11277-022-10091-8

Effective Capacity Analysis over Cascaded Rayleigh Fading Channels with Arbitrary Correlation

Sandeep Kumar[1] · Puspraj Singh Chauhan[2] · Manpreet Kaur[3] · Tarun Lata[4]

Accepted: 11 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In contrary to Shannon's ergodic capacity, the delay-constrained effective rate is used to define the maximum data rate of the real-time applications in 5G and beyond networks. Moreover, the received signal envelope in urban micro-cells environment follows a cascaded Rayleigh distribution with correlated channels. The effective capacity (EC) performance of the delay-sensitive communication system over the cascaded Rayleigh fading channel with arbitrary correlation is studied. The closed-form mathematical expressions for the EC are derived in terms of the Meijer's G-function and the effect of different fading parameters on the effective throughput of the system is demonstrated. The simplified asymptotic expression for high signal-to-noise ratio (SNR) and low SNR regimes are provided to gain more insight into the system. The effect of different system parameters on the EC performance is also demonstrated. Simulation results are corroborated along with the numerical results to verify the correctness of the formulation. To the best of the authors' knowledge, the derived expressions for this channel model are new and have not been reported in the literature.

Keywords Cascaded channel · Arbitrary correlation · Urban micro-cells · Delay constraint · Real-time applications

1 Introduction

In the traditional communication systems, throughput was considered the only QoS parameter and ergodic capacity was used to characterize the effective data rate. Now, with the emergence of real-time applications like video conferencing, smart grid networks, and online interactive gaming, not only the throughput but also the delay-sensitivity of the communication system has gained the importance for the system designer [1]. When delay

✉ Manpreet Kaur
manpreettiet@gmail.com

[1] Central Research Laboratory, BEL, Ghaziabad, India

[2] Department of ECE, Pranveer Singh Institute of Technology, Kanpur, India

[3] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

[4] Department of Mathematics, Delhi University, Delhi, India

constraint is considered as a QoS [2], Shannon capacity is not able to characterize the effective rate of a system. To account for the delayed performance of the real-time systems, the effective rate was proposed in [3]. EC is the maximum rate of data transfer that can be supported by a wireless communication system with a guaranteed delay-bound QoS constraint. For this reason, many researchers have studied the EC performance of various delay-sensitive applications over different fading channels. In [4], the EC model over the Rayleigh fading channel is used for optimal channel allocation in the CRN, and in [5], this is applied for the performance evaluation of voice applications over the Ricean fading channel. The achievable capacity of the CRN in a delay-sensitive asymmetric fading channel is studied in [6], while a comprehensive survey of EC over the CRN for various multimedia applications is presented in [7]. Taking into consideration the multicast receiver cross-layer, EC analysis for the Nakagami-*m* fading channel is provided in [8]. In [9], the EC model over VoIP applications for the end-to-end routing and the link scheduling scheme for the multi-hop WSN is explored. In [10], authors have examined the attainable rate for multi-hop mobile networks with delay-bounds over VOIP applications. In order to improve the readability of this paper, we define abbreviations in Table 1.

Video applications are also challenging to analyse due to the strict delay bounds and burst flow of the networks. Regarding this, some excellent work using the concept of EC has been performed for the video streaming application over different fading channels. To enable high and low-speed video transmission, the authors in [11] have provided a QoS-driven resource allocation scheme to upper-bound the overall delay for the two-hop DF relay networks over the Rayleigh fading channel. The probability distribution of a delay process with video streaming support over the Rician fading channels is investigated in [12]. The EC concept is utilized in [13] to support real-time video streaming in a heterogeneous network, and the results are utilized for cooperative spectrum sensing to observe the QoS guarantee in CRNs. Similarly, in [14], a QoS-driven power allocation scheme in the CRN is proposed and the EC performance for different sensing scenarios in the

Table 1 Nomenclature

Acronyms	Definition
AF	Amplify and forward
CRN	Cognitive radio networks
DF	Delay-and-forward
EC	Effective capacity
FIFO	First-in-first-out
GNM	Generalized Nakagami-*m*
LOS	Line-of-sight
M2M	Mobile to mobile
MISO	Multiple input single output
NLOS	Non-LOS
PDF	Probability density function
QoS	Quality of service
RFID	Radio frequency identification
RVs	Random variables
SNR	Signal-to-noise ratio
V2V	Vehicle-to-vehicle
VoIP	Voice over IP
WSN	Wireless sensor networks

Nakagami-*m* fading environment is studied. End-to-end delay is an important consideration in wireless tele-ultrasonography medical systems, so in [15] the end-to-end EC-based delay distribution is estimated. In another application, the EC model is used to evaluate the support level of communication networks for smart grids irrespective of the network technologies [16].

Likewise, several multiplicative cascaded fading models have been an active area of research interest due to its application in the realistic environment, such as multi-hop communication channels, M2M communication channels, V2V communication system, keyhole channels and RFID pinhole scenarios. A survey of the application of cascaded fading channel in the realistic scenarios of 5G and beyond communication is provided in Table 2. As seen from the table, much of the previously published works have been devoted to study the statistical analysis of cascaded fading channels when the sub-channels (or clusters of scatterers) are assumed to be statistically independent.

In a cascaded fading channel, data transmission from the source antenna to the destination antenna is made possible through *N*-relays. In this scenario, the direct communication link is only viable between two adjacent levels, where each relay receives a faded version of the signal transmitted from the previous level and retransmits it to the next level. Therefore, the fading channels at each hop (between two adjacent levels) will exhibit dependent scatters (correlated) [31, 32]. Thus, the independence claim in this communication scenario is no longer valid and the impact of correlation on the wireless channel is to be considered. Moreover, the channel measurements in [33] show that the received signal envelope in urban micro-cells in 5G and beyond network environment follows a double Rayleigh distribution. However, to the best of our knowledge, studies that consider the effective rate analysis over the generalized cascaded Rayleigh fading channels with arbitrary correlation have not been reported in the literature. Motivated to fill this gap, we have analyzed the EC performance of the delay-sensitive communication system over the cascaded Rayleigh fading channel with arbitrary correlation. The simplified asymptotic expressions for high SNR and low SNR regimes are also provided to gain more insight into the system.

The remaining paper is organized as follows: Sect. 2 discusses the system and channel models in detail. Effective rate analysis over the generalized fading channel is presented in Sect. 3. In Sect. 4, the methodology for deriving the asymptotic expressions (high power and low power) of the EC is illustrated. Section 5, shows the variation of fading parameters and their impact on the EC performance. Finally, concluding points are given in Sect. 6.

2 System and Channel Model

2.1 Effective Throughput

To encourage the effective support of QoS in the next-generation wireless networks, it is essential to present a wireless channel in terms of connection-level QoS metrics. Using the physical-layer channel models, the performance evaluation of delay-limited applications can be complicated and unreliable in some cases [3]. With the support of EC, the channel can be modeled in terms of link-layer QoS metrics. To evaluate the performance of real-time applications of the wireless communication system under the QoS requirements such as delay, reliability, and efficiency, the EC concept is introduced in [3]. The EC model can be used to find the maximum data rate that a channel transmits

Table 2 Application of cascaded fading channel in the realistic scenarios

S. no.	References	Year	Channel model	Applications
1	[17]	2011	Cascaded Rayleigh/Nakagami-*m*	NLOS M2M cooperative communication system
			Double Rice/(multiple-LOS second-order scattering) MLSS	LOS M2M cooperative communication system
2	[18]	2016	N*Nakagami-*m*	V2V communication scenarios
3	[19]	2017	N*$\alpha - \mu$	M2M system employing AF relay
4	[20]	2019	Cascaded Nakagami-m, cascaded Rayleigh, 2*Nakagami-*m*, and the conventional cellular channel models	M2M communication, keyhole communication, multi-hop AF relaying, and RFID pinhole systems
5	[21]	2018	Cascaded $\alpha - \mu$	M2M & V2V system
7	[22]	2015	Rician/Rayleigh cascaded	Multi-cell RFID system
8	[23–25]	2015	N*Rayleigh fading channels	V2V communication scenarios
9	[26]	2018	double Nakagami-*m* fading channels	M2M communications, keyhole communication, and multi-hop relaying systems
12	[27]	2010	Cascaded generalized–K fading channel	MIMO keyhole system
13	[28]	2009	Cascaded GNM	Keyholes and multi-hop communication system where each relay terminal does not perform phase correction but amplifies the received signal with a constant gain
15	[29]	2006	Double-Rayleigh	The Keyhole effect due to real-world waveguides like tunnels or corridors will usually be very difficult to measure
18	[30]	2011	MISO Double Rayleigh Fading/RFID Channel	Passive RFID channel with multiple tag antennas

under the required delay constraint. So, EC can be used to analyze the performance of the system with delay bounds. The link-layer and physical channel model for analyzing the packet-switched networks are shown in Fig. 1.

The QoS guarantees heavily rely on the queueing model where the source traffic and the network service are matched using a FIFO buffer (queue). Thus, the queue prevents loss of packets that could occur when the source rate is more than the service rate, at the expense of increasing the delay. Consider a real-time system of bandwidth B where data packets are divided into frames having the same time duration T and with the QoS exponent θ as the delay factor. For the fading channel, the maximal arrival rate that can be supported under a statistical QoS constraint can be expressed by [34] as

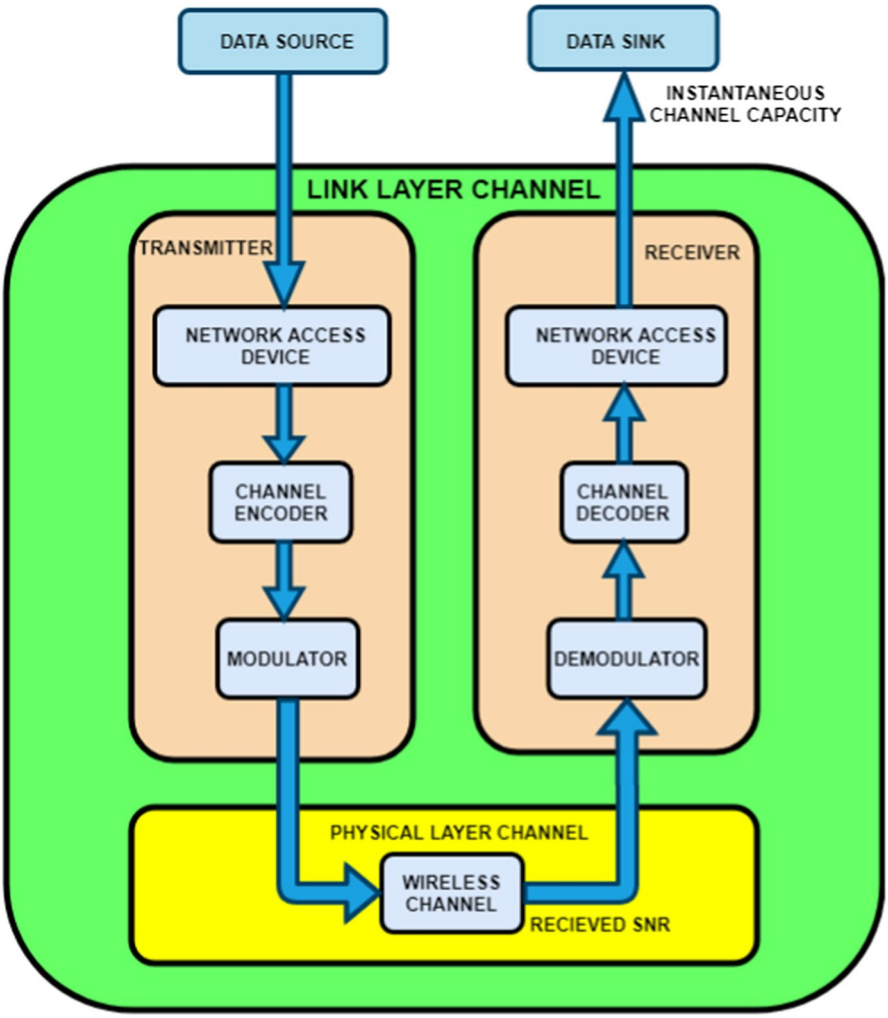


Fig. 1 The link-layer and physical channel model: basic components of packet-switched networks

$$R(\theta) = -\frac{1}{A}\log_2\left[E\left\{(1+\gamma)^{-A}\right\}\right] = -\frac{1}{A}\log_2\left[\int_o^\infty (1+\gamma)^{-A} f_\gamma(\gamma)d\gamma\right] \tag{1}$$

where $A = \theta TB/\log(2)$, $E\{.\}$ is the expectation operator and $f_\gamma(\gamma)$ is the PDF of the received SNR (γ) for the fading channel. For a queue of infinite buffer size, Q having a constant data rate with variable channel capacity, the delay exponent is mathematically represented as

$$\theta = -\lim_{q\to\infty}\frac{\log\left[\Pr\{Q > q\}\right]}{q} \tag{2}$$

Here q is length of the queue at a particular time instance and $\Pr\{.\}$ is the probability of occurrence of an event. The θ parameter specifies the maximum constant arrival rate that *the* system can sustain under certain QoS constraints. It is clear from the above equations that the EC depends upon the channel parameters and system parameters and when $\underset{\theta\to 0}{EC} = Shannon's Ergodic capacity$.

2.2 Cascaded Rayleigh Fading Channel with Arbitrary Correlation

Let $G = [G_1, G_2, ...G_N]$ be a vector of correlated complex-valued Gaussian RVs denoted by $N_c\left(0, 2\sigma_k^2\right)\forall k \in \{1, 2, ..., N\}$. The kth element of the vector may be represented in the form of the real and imaginary parts as

$$G_k = G_{X_k} + jG_{Y_k} \tag{3}$$

where $G_{X_k} = \sqrt{2}\sigma_k\left(\sqrt{1-\lambda_k^2}X_k + \lambda_k X_0\right)$, $G_{Y_k} = \sqrt{2}\sigma_k\left(\sqrt{1-\lambda_k^2}Y_k + \lambda_k Y_0\right)$ and the parameter $\lambda_k \in (-1, 1)$ is a correlation parameter with X_k and Y_k denoted by $N(0, 1/2)$. The cross-correlation coefficient between G_k and G_j for $k, j \in \{1, 2, ..., N\}$ and $k \neq j$ can be obtained as

$$\rho_{k,j} = \frac{E[G_k G_j^*] - E[G_k]E[G_j^*]}{\sqrt{Var[G_k]Var[G_j]}} = \begin{cases} 1 & k = j \\ \lambda_k\lambda_j & k \neq j \end{cases} \tag{4}$$

The product of the N correlated Rayleigh RVs that are not necessarily identically distributed is defined as

$$Z_N = \prod_{k=1}^{N}|G_k| \tag{5}$$

Using (3) in (5) and doing a lot of mathematical manipulation, the PDF of γ of N arbitrary correlated and non-identically distributed cascaded Rayleigh fading channel can be written using [35] as

$$f_\gamma(\gamma) = \frac{1}{\bar{\gamma}} \frac{P_N}{L_N} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \left[\prod_{j=1}^{N} \frac{1}{(k_j!)^2} \left(\frac{\lambda_j^2}{1-\lambda_j^2} \right)^{k_j} \right] \Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} G_{0,N}^{N,0} \left(\frac{\gamma}{\bar{\gamma}} \frac{P_N}{L_N} \middle| \begin{matrix} - \\ k_N, k_{N-1}, ..., k_1 \end{matrix} \right) \tag{6}$$

where $L_N = \prod_{j=1}^{N} (1-\lambda_j^2)$, $K_n = \sum_{i=1}^{n} k_i$, $\zeta_n = \sum_{i=1}^{n} \frac{\lambda_i^2}{(1-\lambda_i^2)}$ and $P_N = (-1)e^{-u} \sum_{n=1}^{N} \frac{d^n}{du^n} \left[\prod_{k=1}^{N} \left(1-\lambda_k^2+\lambda_k^2 u\right) \right] \Big|_{u=0}^{\infty}$, as defined in [35]. $G_{p\ q}^{m\ n}[.]$ is the Meijer G-function and $\Gamma(.)$,$(a!)$ are gamma function and factorial function respectively.

3 Effective Rate

We derive the mathematical formulation of the effective rate of the system over the non-identically distributed cascaded Rayleigh fading channel with arbitrary correlation. Using the relationship $(1+\gamma)^{-A} = \frac{1}{\Gamma(A)} G_{1\ 1}^{1\ 1} \left[\gamma \middle| \begin{matrix} (1-A) \\ 0 \end{matrix} \right]$ from Eq. (10) of [36], and substituting (6) into (1), the EC expression can be written as

$$R(\theta) = -\frac{1}{A} \log_2 \left[\begin{matrix} \frac{1}{\bar{\gamma}} \frac{P_N}{L_N} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \left[\prod_{j=1}^{N} \frac{1}{(k_j!)^2} \left(\frac{\lambda_j^2}{1-\lambda_j^2} \right)^{k_j} \right] \Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} \\ \times \int_0^{\infty} G_{1,1}^{1,1} \left(\gamma \middle| \begin{matrix} 1-A \\ 0 \end{matrix} \right) \times G_{0,N}^{N,0} \left(\frac{\gamma}{\bar{\gamma}} \frac{P_N}{L_N} \middle| \begin{matrix} - \\ k_N, k_{N-1}, ..., k_1 \end{matrix} \right) d\gamma \end{matrix} \right] \tag{7}$$

The integral in the above equation can be evaluated by applying equation (2.24.1.2) of [37] which yields

$$R(\theta) = -\frac{1}{A} \log_2 \left[\begin{matrix} \frac{1}{\Gamma(A)} \frac{P_N}{\bar{\gamma} L_N} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \left[\prod_{j=1}^{N} \frac{1}{(k_j!)^2} \left(\frac{\lambda_j^2}{1-\lambda_j^2} \right)^{k_j} \right] \Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} \\ \times G_{1,N+1}^{N+1,1} \left(\frac{P_N}{\bar{\gamma} L_N} \middle| \begin{matrix} 0 \\ k_N, k_{N-1}, ..., k_1, A-1 \end{matrix} \right) \end{matrix} \right] \tag{8}$$

With the aid of equation (9.31.5) of [38], (8) can be further simplified as

$$R(\theta) = -\frac{1}{A} \log_2 \left[\begin{matrix} \frac{1}{\Gamma(A)} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \left[\prod_{j=1}^{N} \frac{1}{(k_j!)^2} \left(\frac{\lambda_j^2}{1-\lambda_j^2} \right)^{k_j} \right] \Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} \\ \times G_{1,N+1}^{N+1,1} \left(\frac{P_N}{\bar{\gamma} L_N} \middle| \begin{matrix} 1 \\ k_N+1, k_{N-1}+1, ..., k_1+1, A \end{matrix} \right) \end{matrix} \right] \tag{9}$$

The EC of the special case, N uncorrelated cascaded Rayleigh fading channel can be obtained from (9) by substituting $\lambda_k = 0$ as

$$R(\theta) = -\frac{1}{A}\log_2\left[\frac{1}{\Gamma(A)}G_{1,N+1}^{N+1,1}\left(\frac{1}{\bar{\gamma}}\middle|\begin{matrix}1\\1,1,....,1,A\end{matrix}\right)\right] \tag{10}$$

4 Asymptotic Analysis

In this section, we present a detailed asymptotic analysis of the effective throughput in both high and low SNR regimes. The aim of providing asymptotic analysis for the system performance is to gain more physical insights and optimize system parameters accordingly. In addition, the asymptotic result reduces the analytical complexity while designing the system.

4.1 High SNR Approximation

High power analysis provides intuitive insight as how the system performance behaves with respect to the channel fading parameters at high SNR. The expression of EC at high SNR can be written using $(1+\gamma)^{-A} \cong \gamma^{-A}$ approximation in Eq. (1) as

$$R^{\infty}(\theta) = -\frac{1}{A}\log_2(I) \tag{11}$$

where $I = \int_0^{\infty} \gamma^{-A} f_{\gamma}(\gamma) d\gamma$, further putting Eq. (6) in the definition of I, we get

$$\begin{aligned} I = {} & \frac{1}{\bar{\gamma}}\frac{P_N}{L_N}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\left[\prod_{j=1}^{N}\frac{1}{(k_j!)^2}\left(\frac{\lambda_j^2}{1-\lambda_j^2}\right)^{k_j}\right]\Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} \\ & \times \int_0^{\infty}\gamma^{-A} G_{0,N}^{N,0}\left(\frac{\gamma}{\bar{\gamma}}\frac{P_N}{L_N}\middle|\begin{matrix}-\\k_N, k_{N-1}, ..., k_1\end{matrix}\right) d\gamma \end{aligned} \tag{12}$$

Using equation (24) of [36] and simplifying, the expression can be written as

$$R^{\infty}(\theta) = -\frac{1}{A}\log_2\left[\left(\frac{1}{\bar{\gamma}}\frac{P_N}{L_N}\right)^{A}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\left[\prod_{j=1}^{N}\frac{\Gamma(k_j+1-A)}{(k_j!)^2}\left(\frac{\lambda_j^2}{1-\lambda_j^2}\right)^{k_j}\right]\Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)}\right] \tag{13}$$

Similarly, the high power EC expressions for N uncorrelated cascaded Rayleigh fading channel can be obtained as

$$R(\theta) = -\frac{1}{A}\log_2\left[\left(\frac{1}{\bar{\gamma}}\right)^{A}\Gamma(1-A)^{N}\right] \tag{14}$$

4.2 Low SNR Approximation

In many communication networks, such as cellular networks, systems often operate in low-SNR situations. Hence it is beneficial to derive an approximation of (5) in the low-SNR regime. The effective rate performance for the fading channel is investigated for a low-power regime by considering that channel information is known to the receiver only where the transmitter may have either no channel information or imperfect channel information. By following the second-order Taylor expansion of the EC at $\rho \to 0^{+}$, as per below equation:

$$R(\rho,\theta) = R^{'}(0,\theta)\rho + R^{''}(0,\theta)\frac{\rho^2}{2} + o\left(\rho^2\right) \tag{15}$$

where $R^{'}(0,\theta)$ and $R^{''}(0,\theta)$ are the first and second-order derivatives of $R(\rho,\theta)$ with respect to average transmitting SNR,ρ at $\rho = 0$. Note that in the low-SNR regime, analysis based on ρ would lead to the misleading conclusions. The capacity analysis should be better conducted using the transmitted normalized energy per information bit $\left(E_b/N_0\right)$ rather than the average SNR ρ [39]. Following the same method as in [39], the effective rate in the low-SNR regime can be approximated by

$$R\left(\frac{E_b}{N_0}\right) \approx S_0 \log_2\left(\frac{E_b/N_0}{E_b/N_{0\min}}\right) \tag{16}$$

where

$$S_0 = \frac{-2\left(R^{'}(0,\theta)\right)^2 \ln(2)}{R^{''}(0,\theta)} \quad \text{and} \quad \left(\frac{E_b}{N_{0\min}}\right) = \frac{1}{R^{'}(0,\theta)} \tag{17}$$

where $\left(E_b/N_0\right)_{\min}$ is defined as the minimum energy required to transmit a bit which is achieved as the average transmit SNR, ρ approaches zero, whereas the wideband slope, S_0 is defined as the slope of the spectral efficiency curve which helps to study the energy efficiency at low but non-zero power levels. The first and second-order derivatives, $R^{'}(0,\theta)$ and $R^{''}(0,\theta)$ are defined by [40] as

$$R^{'}(0,\theta) = \frac{E[\gamma]}{\ln(2)} \quad \text{and} \quad R^{''}(0,\theta) = \frac{1}{\ln(2)}\left[A(E[\gamma])^2 - (A+1)E\left[\gamma^2\right]\right] \tag{18}$$

Putting (5) in the definition of $E[\gamma]$ and $E\left[\gamma^2\right]$, applying the relationship between hypergeometric function and Meijer G function and using (7.811.4) of [38] we get

$$E[\gamma] = \left(\frac{\overline{\gamma}L_N}{P_N}\right)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\left[\prod_{j=1}^{N}\frac{\Gamma(k_j+2)}{(k_j!)^2}\left(\frac{\lambda_j^2}{1-\lambda_j^2}\right)^{k_j}\right]\Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} \tag{19}$$

$$E\left[\gamma^2\right] = \left(\frac{\overline{\gamma}L_N}{P_N}\right)^2\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\left[\prod_{j=1}^{N}\frac{\Gamma(k_j+3)}{(k_j!)^2}\left(\frac{\lambda_j^2}{1-\lambda_j^2}\right)^{k_j}\right]\Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} \tag{20}$$

Putting the value of $E[\gamma]$ and $E\left[\gamma^2\right]$ from (18) and (19) into (17), the expressions for $R^{'}(0,\theta)$ and $R^{''}(0,\theta)$ can be obtained as

$$R^{'}(0,\theta)=\left(\frac{\overline{\gamma}L_N}{\ln(2)P_N}\right)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\left[\prod_{j=1}^{N}\frac{\Gamma(k_j+2)}{(k_j!)^2}\left(\frac{\lambda_j^2}{1-\lambda_j^2}\right)^{k_j}\right]\Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)} \tag{21}$$

$$R^{''}(0,\theta)=\frac{1}{\ln(2)}\left|\begin{array}{l} A\left\{\left(\frac{\overline{\gamma}L_N}{P_N}\right)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\left[\prod_{j=1}^{N}\frac{\Gamma(k_j+2)}{(k_j!)^2}\left(\frac{\lambda_j^2}{1-\lambda_j^2}\right)^{k_j}\right]\Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)}\right\}^2 \\ -(A+1)\left\{\left(\frac{\overline{\gamma}L_N}{P_N}\right)^2\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\left[\prod_{j=1}^{N}\frac{\Gamma(k_j+3)}{(k_j!)^2}\left(\frac{\lambda_j^2}{1-\lambda_j^2}\right)^{k_j}\right]\Gamma(K_N+1)(1+\zeta_N)^{-(K_N+1)}\right\} \end{array}\right| \tag{22}$$

Now putting (21) and (22) in (17), we get the expressions of $\left(E_b/N_0\right)_{\min}$ and S_0. Further, using (17) in (16), we get the final expression of EC for a low SNR regime. By following the same approach as used to get (14), we can evaluate the expression of low SNR EC for uncorrelated cascaded Rayleigh fading channel.

5 Numerical Discussion

In this section, we have presented the analytical results to study the impact of various system parameters on the performance of the effective rate. Simulation results are corroborated along with the numerical results to verify the correctness of the formulation. We are comparing the numerical results to the Monte Carlo simulations because it is a standard method of validation, that is well suited for communication systems, which relies on

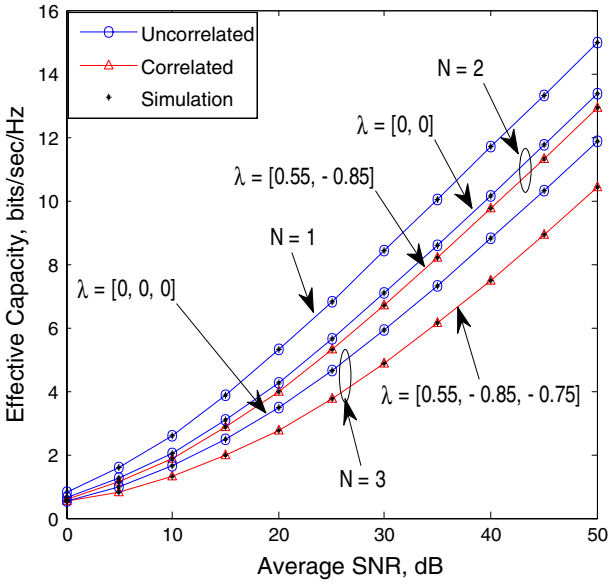


Fig. 2 Effective capacity versus average SNR for different values of cascading order $N=1, 2, 3$ and non-identical correlation parameters. For $N=2$, $\lambda=[0, 0]$, $[0.55, -0.85]$ and for $N=3$, $\lambda=[0, 0, 0]$ and $\lambda=[0.55, -0.85, -0.75]$

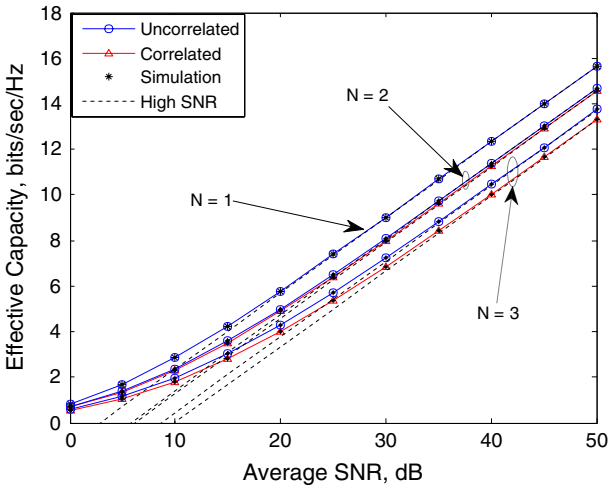


Fig. 3 Impact of cascading order on the asymptotic performance of the system for identical correlation parameters. For $N=2$, $\lambda=[0, 0]$, $[0.55, 0.55]$ and for $N=3$, $\lambda=[0, 0, 0]$ and $\lambda=[0.55, 0.55, 0.55]$

repeated random sampling to obtain the results. In the Monte-Carlo simulations, 10^6 random samples are generated, which are sufficient for the desired accuracy of the proposed methodology. Throughout this section, $\lambda = \left[\lambda_1 \lambda_2 ... \lambda_N\right]$ is used to represent the sets of $\{\lambda_i\}$ that are used in the derived results.

The variation of effective rate versus received SNR for different values of cascading order $N=1$, 2, 3 and non-identical correlation parameters are shown in Fig. 2. For $N=2$, $\lambda=[0, 0]$, $[0.55, -0.85]$ and for $N=3$, $\lambda=[0, 0, 0]$ and $\lambda=[0.55, -0.85, -0.75]$. It is evident that as the value of N increases, there is a significant decrease in the effective throughput of the system. Also, it concludes that the increase in the correlation value will decrease the capacity of the system. The asymptotic performance of the system in terms of high SNR approximation for different cascading orders and correlation parameters is depicted in Fig. 3. The following correlation parameters were used: $N=2$, $\lambda=[0, 0]$, $[0.55, 0.55]$ and for $N=3$, $\lambda=[0, 0, 0]$ and $\lambda=[0.55, 0.55, 0.55]$. As expected, increasing the cascaded level will degrade the system performance. The asymptotic

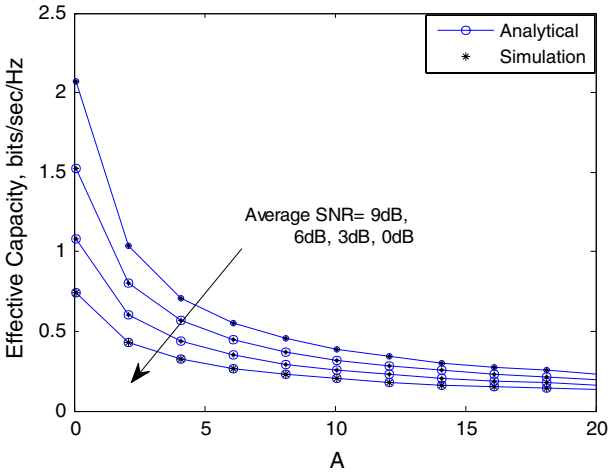


Fig. 4 Effective capacity versus A for different values of average SNR for $N=2$ and $\lambda=[0.55, -0.85]$

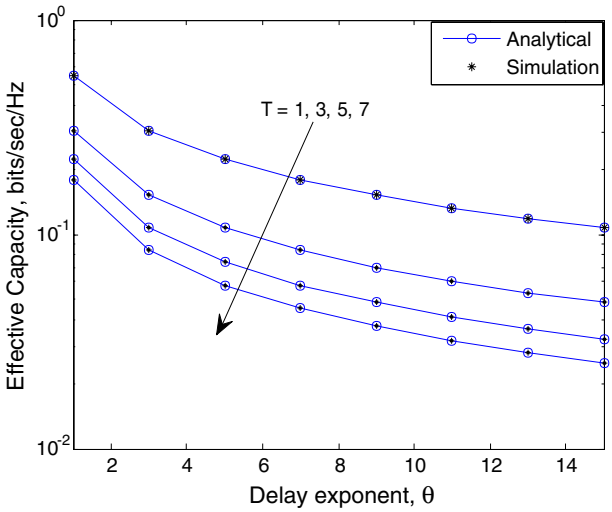


Fig. 5 Effective throughput versus the QoS exponent parameter for $N=3$ and $\lambda=[0.55, -0.85, -0.75]$

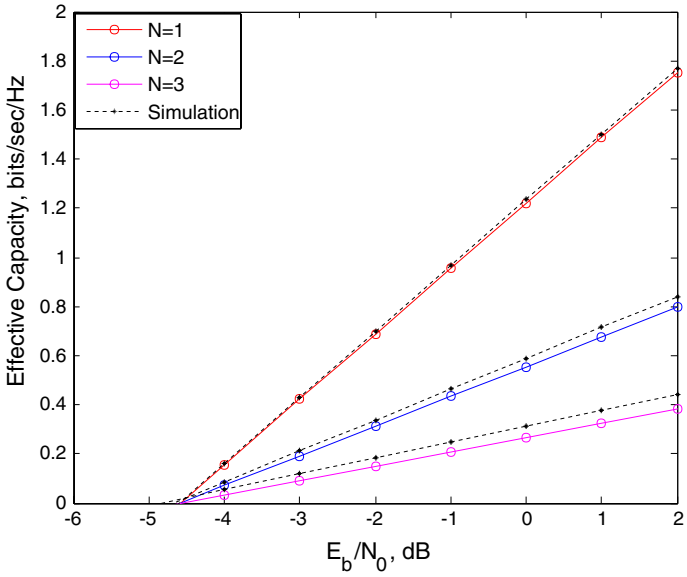


Fig. 6 Low SNR effective capacity versus average transmit SNR

plots merge quickly with the analytical plots for the lower value of cascading order and correlation parameters. The impact of average SNR on the EC as a function of A for $N=2$ and $\lambda=[0.55, -0.85]$ is shown in Fig. 4. It is noted that the system performance degrades with the increase in A value, and when the EC value is sufficiently small, a further increase in A can only have a slight impact on the effective throughput performance of the system.

The effective throughput versus the QoS exponent θ for $N=3$ and $\lambda=[0.55, -0.85, -0.75]$ is plotted in Fig. 5. Overall, the results show that the performance of the system degrades with a large QoS exponent, which implies that as the delay constraints become large, the less throughput system can handle. It is noted that the difference between the plots of delay constraints is larger at a higher value of T than at lower values of T. In Fig. 6, the results are drawn for effective capacity in the case of low SNR against E_b/N_0 for various values of cascading order and fixed values of $A=0.5$. From the figure, it can be inferred that with the decrease of cascading order, the effective rate of the system increases. The results produced here can be directly applied for designing the communication systems for real-time applications where the channel characteristics are modeled in the form of correlated cascading Rayleigh distribution.

6 Conclusions

We have derived the mathematical expression of the effective rate performance of the system over the cascaded Rayleigh fading channel with arbitrary correlation. The impact of system parameters on the EC's performance has been studied in detail. These results have extended and complimented the existing research on effective rate analysis. High-power and low-power asymptotic expressions of the maximum data rate have been provided to get more insights into the system. It is observed that the effect of cascaded levels and/or correlations between the channels on the performance of the effective throughput of the system is more pronounced than the other fading parameters of the channel. This is due to the fact that the signal transmitted over the cascaded channel experiences more severe fading than the conventional one-way channel. The validation of the proposed results has been performed through Monte-Carlo simulation. However, the practical measurement validation still needs to be carried out to find the application of the study for different practical scenarios. Further, parameter estimations of the proposed models through measurement data can be considered to gain more practical insight.

Author Contributions All the authors have equally contributed in this manuscript.

Funding No funding was received for this work.

Data Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Kaur, M., & Yadav, R. (2022). EC analysis of multi-antenna system over 5G and beyond networks and its application to IRS-assisted wireless systems. *Wireless Personal Communications, 124*, 1861–1881.
2. Kaur, M., & Yadav, R. K. (2021). Effective capacity analysis over Fisher–Snedecor F fading channels with MRC reception. *Wireless Personal Communications, 121*, 1693–1705.

3. Wu, D., & Negi, R. (2003). Effective capacity: A wireless link model for support of quality of service. *IEEE Transaction on Wireless Communications, 2*(4), 630–643.
4. Lien, S. Y., Tseng, C. C., Chen, K. C., & Su, C. W. (2010). Cognitive radio resource management for QoS guarantees in autonomous femtocell networks. In *IEEE international conference on communications (ICC)*, 2010.
5. Glisic, S., Nikolic, Z., Milosevic, N., & Pirinn, P. (2005). Effective capacity of advanced wireless cellular. In *IEEE 16th international symposium on personal indoor and mobile radio communications* (Vol. 4, pp. 2771–2780).
6. Qingxin, S., & Xiangquna, C. (2015). Effective capacity of cognitive radio systems in asymmetric fading channels. *The Journal of China Universities of Posts and Telecommunications, 22*(3), 18–25.
7. Piran, J., Pham, Q. V., Islam, S. R., Cho, S., Bae, B., Suh, D. Y., & Han, Z. (2020). Multimedia communication over cognitive radio networks from QoS/QoE perspective: A comprehensive survey. *Journal of Network and Computer Applications*. https://doi.org/10.1016/j.jnca.2020.102759
8. Du, X., & Zhang, Q. (2007). Cross-layer modeling for QoS-driven multimedia multicast/broadcast over fading channels in advances in mobile multimedia. *IEEE Communications Magazine, 45*(8), 62–75.
9. Wang, Q., Fan, P., Wu, D. O., & Letaief, K. B. (2011). End-to-end delay constrained routing and scheduling for wireless sensor networks. In *IEEE international conference on communications (ICC)*, 2011.
10. Chen, Y., & Darwazeh, I.: End-to-end delay performance analysis in IEEE 802.16j mobile multihop relay (mmr) networks. In *International conference on telecommunications* (pp. 488–492).
11. Du, Q., Huang, Y., Ren, P., & Zhang, C. (2011). Statistical delay control and QoS-driven power allocation over two-hop wireless relay links. In *IEEE global telecommunications conference—GLOBECOM*, *Houston, TX, USA*, 2011.
12. Chen, Y., & Darwazeh, I. (2013). An estimator for delay distributions in distributions in packet-based wireless digital communication systems. In *An estimator for d IEEE wireless communications and networking conference (WCNC)*, *Shanghai*, 2013.
13. Wang, X., & Zhang, J. (2017). Heterogeneous QoS-driven resource allocation over MIMO-OFDMA based 5G cognitive radio networks. In *IEEE wireless communications and networking conference (WCNC)*, *San Francisco, CA*, 2017.
14. Zhang, J., & Wang, X. (2017). Statistical QoS-driven cooperative power. In *Wireless communications and networking conference (WCNC), 2017* (pp. 1–6).
15. Chen, Y., Darwazeh, I., Philip, N., & Istepanian, R. (2013). End-to-end delay distributions in wireless tele-ultrasonography medical systems. In *IEEE global communications conference, GA, USA*, 2013.
16. You, M., Mou, X., & Sun, H. (2015). Effective capacity analysis of smart grid communication networks. In *International workshop on computer aided modelling and design of communication links and networks*, *Guildford, UK*, 2015.
17. Talha, B., & Patzold, M. (2011). Channel models for Mobile-to-Mobile cooperative communication systems: A state of the art review. *IEEE Vehicular Technology Magazine, 6*(2), 33–43.
18. Boulogeorgos, A. A., Sofotasios, P. C., Selim, B., Muhaidat, S., Karagiannidis, G. K., & Valkama, M. (2016). Effects of RF impairments in communications over cascaded fading channels. *IEEE Transactions on Communications, 65*(11), 8878–8894.
19. Bhowal, A., & Kshetrimayum, R. S. (2017). End to end performance analysis of M2M cooperative communication over cascaded a-μ channels. In *9th international conference on communication systems and networks (COMSNETS), Bangalore, India*, 2017.
20. Ata, S. O. (2019). Secrecy performance analysis over cascaded fading channels. *IET Communications, 13*(2), 259–264.
21. Kong, L., Kaddoum, G., & da Costa, D. B. (2018). Cascaded $a-\mu$ fading channels: reliability and security analysis. *IEEE Access, 6*, 41978–41992.
22. Bekkali, A., Zou, S., Kadri, A., Crisp, M., & Penty, R. V. (2015). Performance analysis of passive UHF RFID systems under cascaded fading channels and interference effects. *IEEE Transactions on Wireless Communications, 14*(3), 1421–1433.
23. Alghorani, Y., Kaddoum, G., Muhaidat, S., & Pierre, S. (2015). On the approximate analysis of energy detection over n*Rayleigh fading channels through cooperative spectrum sensing. *IEEE Wireless Communications Letters, 4*(4), 413–416.
24. Sofotasios, P. C., Mohjazi, L., Muhaidat, S., Al-Qutayri, M., & Karagiannidis, G. K. (2016). Energy detection of unknown signals over cascaded fading channels. *IEEE Antennas and Wireless Propagation Letters, 15*, 135–138.

25. Alghorani, Y., Kaddoum, G., Muhaidat, S., Pierre, S., & Al-Dhahir, N. (2016). On the performance of multihop-intervehicular communications systems over n*Rayleigh fading channels. *IEEE Wireless Communications Letters, 5*(2), 116–119.
26. Ata, S. O. (2018). Secrecy performance analysis over double Nakagami-m fading channels. In *41st international conference on telecommunications and signal processing (TSP), Athens, Greece*, 2018.
27. Peppas, K., Lazarakis, F., Alexandridis, A., & Dangakis, K. (2010). Cascaded generalised-K fading channel. *IET Communications, 4*(1), 116–124.
28. Yilmaz, F., & Alouini, M. (2009). Product of the powers of generalized Nakagami-m variates and performance of cascaded fading channels. In *GLOBECOM 2009: 2009 IEEE global telecommunications conference, Honolulu, HI, USA*, 2009.
29. Almers, P., Tufvesson, F., & Molisch, A. F. (2006). Keyhole effect in MIMO wireless channels: Measurements and theory. *IEEE Transactions on Wireless Communications, 5*(12), 3596–3604.
30. He, C., & Wang, Z. J. (2011). Closed-form BER analysis of non coherent FSK in MISO double Rayleigh fading/RFID channel. *IEEE Communications Letters, 15*(8), 848–850.
31. Griffin, J. D., & Durgin, G. D. (2007). Link envelop correlation in the backscatter channel. *IEEE Communication Letters, 11*(9), 735–737.
32. Chau, Y. A., & Huang, K. Y. (2012). On the second-order statics of correlated cascaded Rayleigh fading channels. *International Journal on Antenna & Propagation*. https://doi.org/10.1155/2012/108534
33. Erceg, V., Fortune, S. J., Ling, J., Rustako, A. J., & Valenzuela, R. A. (1997). Comparisons of a computer-based propagation prediction tool with experimental data collected in urban microcellular environments. *IEEE Journal on Selected Areas in Communications, 15*(4), 677–684.
34. Yadav, P., Kumar, S., & Kumar, R. (2020). Effective capacity analysis over α–κ–μ/Gamma composite fading channel. In *IEEE international conference on advances in computing, communication control and networking (ICACCCN), Greater Noida, India*, 2020.
35. Ghareeb, A., & Darwish, J. (2020). Statastics of cascaded Rayleigh fading channels with arbitrary correlation. *IET Communications, 14*(16), 2849–2857.
36. Adamchik, V., & Marichev, O. (1990). The algorithm for calculating integrals of hypergeometric type functions and its realization in REDUCE system. In *International conference on symbolic and algebraic computation, New York (USA)*, 1990.
37. Prudnikov, A. P., & Brychkov, Y. A. (1986). *Integrals and series volume 3, more special functions* (1st ed.). Gordon and Breach Science Publishers.
38. Gradshteyn, I., & Ryzhik, I. (2007). *Table of integrals, series and products* (7th ed.). Academic Press.
39. Matthaiou, M., & Zhong, C. (2012). Low-SNR analysis of MIMO Weibull fading channel. *IEEE Communications Letters, 16*(5), 694–697.
40. Yadav, P., Kumar, S., & Kumar, R. (2021). Analysis of EC over Gamma shadowed α–η–μ fading channel. In *IOP conference series: material science & engineering*, 2021.

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.

Sandeep Kumar received his B. Tech. in electronics and communication from Kurukshetra University, India in 2004 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. He received his Ph.D. from Delhi Technological University, Delhi, India in 2018. He is currently working as Member (Senior Research Staff) at Central Research Laboratory, Bharat Electronics Limited Ghaziabad, India. His research interests include the study of wireless channels, cognitive radio networks and 5G and beyond technologies etc. He has received various awards and certificates of appreciation for his research activities. He is also serving as a reviewer for IEEE, Elsevier and Springer journals etc.

Puspraj Singh Chauhan was born in UP, India. He received his Bachelor of Technology in Electronics and Communication Engineering from Uttar Pradesh Technical University, India in 2009, the M.Tech degree in Digital Signal Processing from Govind Ballabh Pant Engineering College, Pauri, Uttarakhand, India in 2011. He received the PhD. degree in Electronics and Communication Engineering from Govind Ballabh Pant Engineering College, Pauri, Uttarakhand, India in 2021. Currently he is working as Assistant Professor at Pranveer Singh Institute of Technology, Kanpur, U.P., India. He also received best paper award at UPCON 2018, Gorakhpur. His research interests include Wireless Communication and Propagation Channel Modeling.

Manpreet Kaur received her B. Tech. in Electronics and Communication Engineering from Punjab Technical University, Punjab, India in 2005 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. She is persuing her PhD. From Delhi Technological University, Delhi, India. She is working as Member (Senior Research Staff) at Central Research Laboratory, Ghaziabad, Bharat Electronics Limited. She has more than 15 years of experience in the field of research and development. She has received various awards and certificates of appreciation for her research activities. She is serving as a reviewer for IEEE, Elsevier and Springer journals etc. She has also served as a technical chair and member of technical programme committee of various reputed international conferences. Her research interest includes the wireless communication and cognitive radios networks and 5G and beyond technologies etc.

Tarun Lata was born in Delhi. She received her B. Sc., M. Sc. and Ph.D. in Mathematical Sciences all from the University of Delhi, Delhi, India in 2004, 2006 and 2021 respectively. She has worked as an assistant professor at Delhi Technical Campus, Bahadurgarh, Haryana, India from Mar. 2013 to Oct. 2013. She has also worked as a guest faculty in Delhi Technological University, Delhi, India from Sep. 2015 to Nov. 2016 and Aug. 2017 to Dec. 2017. She has received various awards and certificates of appreciation for her research activities. She is serving as a reviewer for Elsevier and Springer journals etc. She has also served as a technical chair and member of technical programme committee of various reputed international conferences. Her research interest includes Coding Theory and Information Technology etc.

Applied Physics A (2022) 128:1000
https://doi.org/10.1007/s00339-022-06147-y

Applied Physics A
Materials Science & Processing

Effects of Er^{3+} dopant on polymorphic phase boundaries and electrical properties of Sn^{4+}modified $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ multifunctional ceramics

Mohd Azaj Ansari[1,6] · A. R. James[2] · K. Zimik[3] · Renuka Bokolia[4] · R. S. Kundu[5] · K. Sreenivas[1]

Received: 25 July 2022 / Accepted: 14 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2022

Abstract
Ferroelectric ceramic compositions of $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ ($0 \le y \le 0.9$ mol%) modified with varying erbium content prepared by solid-state reaction method have been investigated for their structural, electrical, and light upconversion properties. Erbium (Er) dopant reveals a distinct change in the dielectric phase transitions and exhibit multifunctional properties. The coexistence of orthorhombic and tetragonal phases is identified using Rietveld refinement. Increasing Er content increases the tetragonal phase fraction, reduces the tetragonal distortion and inhibits the grain growth. Raman spectroscopy reveals preferential substitution of Er^{3+} for Ba^{2+} and shows additional modes at ~ 335, ~ 442, ~ 621, ~ 669, and ~ 786 cm^{-1}. Rhombohedral-to-orthorhombic-to-tetragonal phase transitions (T_{R-O} and T_{O-T}) shift to lower temperatures, and the tetragonal-to-cubic phase transition (T_{T-C}) remains unchanged as evidenced through dielectric and reversible heat capacity measurements. Increasing Er content widens the transition zone width ($T_{T-C} - T_{O-T} = \Delta T_{PTTR}$) from ~ 39 to 60 °C and extends the useful application range. Increasing Er content decreases DC resistivity (ρ_{dc}) from ~ 9.54 to 4.21×10^{11} Ω-cm, decreases remnant polarization from 2.64 to 1.85 μC cm^{-2}, and a maximum piezoelectric coefficient ~ 179 pC N^{-1} is observed for $y = 0.3$ mol%. The sensitivity of temperature dependent light upconversion reaches its maximum value of 0.00344 K^{-1} at ~ 450 K for $y = 0.5$ mol%.

Keywords Ceramics · Dielectric · Raman spectroscopy · Light upconversion

1 Introduction

Ferroelectric materials with the perovskite-type structure are compounds with $A^{2+}B^{4+}O_3^{2-}$ chemical formula (e.g., $BaTiO_3$ with Ba^{2+} and Ti^{4+} at the *A* and *B* sites, respectively). Among the different dopants or substituents used in modified-$BaTiO_3$ compositions, Sn^{4+} at the *B* site has shown the most promising influence such as the merging of all the three-phase transitions leading to a quasi quadrupole point at $x \sim 0.11$ in $BaSn_xTi_{1-x}O_3$ [1]. Enhancement in the piezoelectric properties ($d_{33} = 700$–920 pC N^{-1} at the quasi quadrupole composition ($x \sim 0.11$) has been reported at temperatures close to the phase transition temperature (~ 49 °C) [2], whereas, at room temperature Kalyani et al. [3], have reported high piezoelectric coefficient $d_{33} \sim 425$ pC N^{-1}, and high remnant polarization in $BaSn_xTi_{1-x}O_3$ with relatively low concentration levels of Sn ($x = 0.03$–0.06), when the powders were subjected to a specially devised poling technique, before compacting them into solid ceramic form. The improvements in the properties were ascribed to the coexistence of a predominant orthorhombic phase along with the

✉ Mohd Azaj Ansari
azaj.ansari@gmail.com

✉ K. Sreenivas
kondepudysreenivas@gmail.com

[1] Department of Physics and Astrophysics, University of Delhi, Delhi 11007, India

[2] Defence Metallurgical Research Laboratory (DMRL) Kanchanbagh, Hyderabad 500079, India

[3] Laser Science and Technology Centre (LASTEC) Metcalfe House, DRDO New Delhi, Delhi 110054, India

[4] Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

[5] Department of Physics, Guru Jambheshwar University of Science and Technology Haryana, Hisar, Haryana 125001, India

[6] Department of Physics, Keshav Mahavidyalaya University of Delhi, Delhi 110034, India

Published online: 24 October 2022

tetragonal phase. In recent years, coexistence of ferroelectric phases has been accounted for enhanced piezoelectric properties in $BaTi_{1-x}Zr_xO_3$ ceramics, and the large polarization observed in rare-earth (RE) ion doped 5.55%Nd-PMNT37 crystals [4, 5]. Thus, $BaSn_xTi_{1-x}O_3$ emerges as an interesting ferroelectric host material with multifunctional properties (ferro/piezo/pyroelectricity and electrocaloric) [6, 7], and the introduction of rare-earth ion Er^{3+} has provided optical functionality via light upconversion photoluminescence in $Ba_{1-x}Ti_{0.96}Sn_{0.04}O_3$ with $x = 0.5$ mol % Er [8].

In recent years, RE ion dopants in many ferroelectric compositions have shown interesting opto-mechanical effects [9], allowed non-contact detection of ferroelectric phase transitions [10], and provided good sensitivity for non-contact optical thermometry applications [11]. The interest in light sensitive rare-earth-ion-doped ferroelectric/piezoelectric host materials stems from the fact that ferroelectrics host materials provide (1) a low phonon energy, (2) allow crystal structure modifications with appropriate dopants, and (3) their functional (ferro/piezo/pyro) can be augmented with optical functionality [11]. The properties are largely governed by the relative size of the RE ions, their electronic configuration, dopant concentration, and the site selectivity in the $BaTiO_3$-based host lattice. Substitution at the *A* site is preferred by larger rare-earth ions (Nd and La), whereas amphoteric Er^{3+} (0.089 nm) has a tendency to substitute either at *A*, or at both *A* and *B* sites depending on the dopant concentration level [12]. However, the influence of Er^{3+} substitution at the *A* site in several ferroelectric compositions has presented mixed results. For example: (1) the Curie temperature either remains unchanged [13], or decreases [14], or increases [15–19] in several different compositions, (2) the remnant polarization mostly decreased, except in the case of $Ca_{1/3}Bi_{2/3})_{3-x}\ Er_xBi_2Ti_4O_{15}$ and $Sr_2Bi_4Ti_5O_{18}$ and [19, 20], (3) the piezoelectric coefficient, though scarcely investigated, has shown an improvement in the case of Er-doped BZT, BCZT, and KNNLN [13, 16, 21], and (4) changes in the electrical resistivity have not been investigated in detail. Furthermore, the dopants substituting at the *A* site do not show a similar trend in the shifts of the phase transition temperatures corresponding to the rhombo-to-ortho (T_{R-O}), ortho-to-tetra (T_{T-O}), and tetra-to-cubic (T_{T-C}) phase transitions of $BaTiO_3$. For example; (1) $Ba_{1-x}Ca_xTiO_3$ ($x = 0$–0.25) shows lowering of T_{R-O}, T_{T-O}, while T_{T-C} remains unchanged [22], (2) in $Ba_{1-z}Cd_zTiO_3$ ($z = 0$–0.25) all three transitions T_{R-O}, T_{T-O}, and T_{T-C} remain unchanged [23], (3) in $Ba_{1-x}Sr_xTiO_3$ ($x = 0$–0.60), T_{R-O}, T_{T-O}, and T_{T-C} are lowered [24], 4) $Ba_{1-z}La_zTi_{0.95}Sn_{0.05}O_3$ ($z = 0$–3.5) T_{R-O}, T_{T-O}, and T_{T-C} are lowered [25], (5) in $Ba_{1-x}Ti_{0.96}Sn_{0.04}O_3 + x$Er ($x = 0$–3.5 mol%) T_{R-O}, T_{T-O} are lowered while T_{T-C} remains unchanged [8].

Keeping in view, the possible enhancement in the multifunctional properties of $BaSn_xTi_{1-x}O_3$ for an optimum Sn content in the range (0.04–0.06 mol), it was of interest to further understand how the substitution of Er at the *A* site influences the coexistence of phases, and affects the ferroelectric phase transitions, and the resulting electrical properties. In the present work, $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ (BESn6T) compositions with varying Er content ($0 \le y \le 0.9$ mol%) have been studied. A preset value for Sn content ($x = 0.06$ mol) was considered because high piezo coefficients were reported for this specific Sn doping level [3], and, moreover, with Er^{3+}, the tetragonal-to-cubic phase transition temperature (T_{T-C}) is unaffected [8]. The main motivation for the present work was to examine the effects of Erbium dopant on the coexistence of phases, individual phase fraction, and study the shifts in the associated phase transitions. Changes in the microstructure dielectric properties, electrical resistivity, polarization, piezoelectricity, and temperature dependent light upconversion for optical thermometry have been studied.

2 Experimental procedure

Solid-state reaction method was used for the preparation of Er-modified $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ compositions ($0 \le y \le 0.9$ mol%) with fixed Sn content ($x = 0.06$) and varying Er content. Moisture-free high purity oxides (> 99.5%) from Sigma–Aldrich (India) were weighed in measured stoichiometric proportion [$(1 - y)\ BaO + yErO_{3/2} + 0.94TiO_2 + 0.06SnO_2 \rightarrow Ba_{1-y}Er_y\ (Sn_{0.06}Ti_{0.94})O_3$]. The powder mixture was mixed manually, calcined at 1200 °C for 3 h, and re-milled again in a high energy planetary ball mill for 6 h. The powder was pressed into ~ 1 mm thick circular pellets using a die of 10 mm diameter and a hydraulic press operated at (550 MPa), and finally sintered in air at 1350 °C for 5 h. The phase purity was examined using high-resolution X-ray diffraction (HR-XRD) namely the Bruker D-8 advanced X-ray diffractometer. A Jeol make (*JSM 66LV, JEOL*) scanning electron microscope was used to study microstructure. Raman spectra were recorded at room temperature using Invia model Renishaw fitted with a 785 nm laser. Sintered ceramics discs were polished, reannealed, and then electroded using silver paste which was cured at 600 °C for 30 min. Temperature-dependent dielectric measurements over a wide frequency range (1–10^6 Hz) and temperature (− 50 to 100 °C) were measured with a computer-controlled Novocontrol impedance analyzer. DC resistivity (ρ_{dc}) was determined from leakage current measurements using a 12 V battery after stabilizing the sample for nearly 30 min. Sintered ceramics were electrically poled at room temperature under a dc voltage of ~ 1.3 (kV/cm), and piezoelectric coefficient (d_{33}) was measured using Berlincourt d_{33} piezometer. *P–E* hysteresis loops were recorded at 50 Hz using the Sawyer Tower equipment manufactured by Marine India

Electrical Ltd. FTIR measurements were made in the ATR-mode using a Nicolet iS50FT-IR spectrometer, and light upconversion emission was measured using a Quanta master 8450-11 (Horiba) spectrometer equipped with a 0.98 µm infrared laser diode (400 mW).

3 Results and discussion

3.1 X-ray diffraction analysis

High-resolution X-ray diffraction (HR-XRD) measurements on sintered ceramic compositions Ba_{1-y} $Er_ySn_{0.06}Ti_{0.94}O_3$ (BESn6T) with fixed Sn ($x=0.06$) and varying Er dopant concentration ($0 \le y \le 0.9$ mol%) are shown in Fig. 1a. Single phase formation is confirmed, and no secondary phase is observed. Enlarged view of the diffraction peaks in the 2θ range 44–45.5° and 30.9–31.8° show variations in the intensity, diffuseness, and asymmetry in the peak profile. Upon de-convolution, the collective influence of orthorhombic (*Amm2*) + tetragonal (*P4mm*) phases of Sn-modified $BaTiO_3$ suggesting coexistence of the two phases was analyzed as shown in Fig. 1b, c [3, 26]. This was further confirmed through Rietveld refinement of the high-resolution X-ray diffraction patterns and the results are shown in Fig. 2a. The experimental and Rietveld refined patterns are compared for the composition with $y=0.1$ mol% of Er is shown in Fig. 2a, and the fitting parameters are shown in Table 1. It may be noted that when the tetragonal phase alone was considered, the fitted profile (red pattern) showed a clear mismatch with the experimental XRD data (blue pattern) as shown by the [trace-(i) in Fig. 2a]. A similar mismatch in the fitting [trace-(ii) in Fig. 2a] was encountered when the orthorhombic phase alone was considered. However, a good fit was obtained by considering the coexistence of both (tetra + ortho) phases as shown in the fitted data [trace-(iii) in Fig. 2a], and the refinement parameters showed acceptable goodness of fit [GOF = $(R_{wp}/R_p) \sim 1.07$ to 1.09] as shown in Table 2. For varying Er content in the range ($0 \le y \le 0.9$ mol%), the percentage fraction of the two coexisting phases was found to vary as shown in Fig. 2b. The tetragonal phase fraction increased constantly, and the orthorhombic phase decreased with increasing Er content ($0 \le y \le 0.9$ mol%). Accordingly, the peak intensity of the $(200)_T$ is found to increase with increasing erbium content (Fig. 1b), due to the increasing tetragonal phase fraction

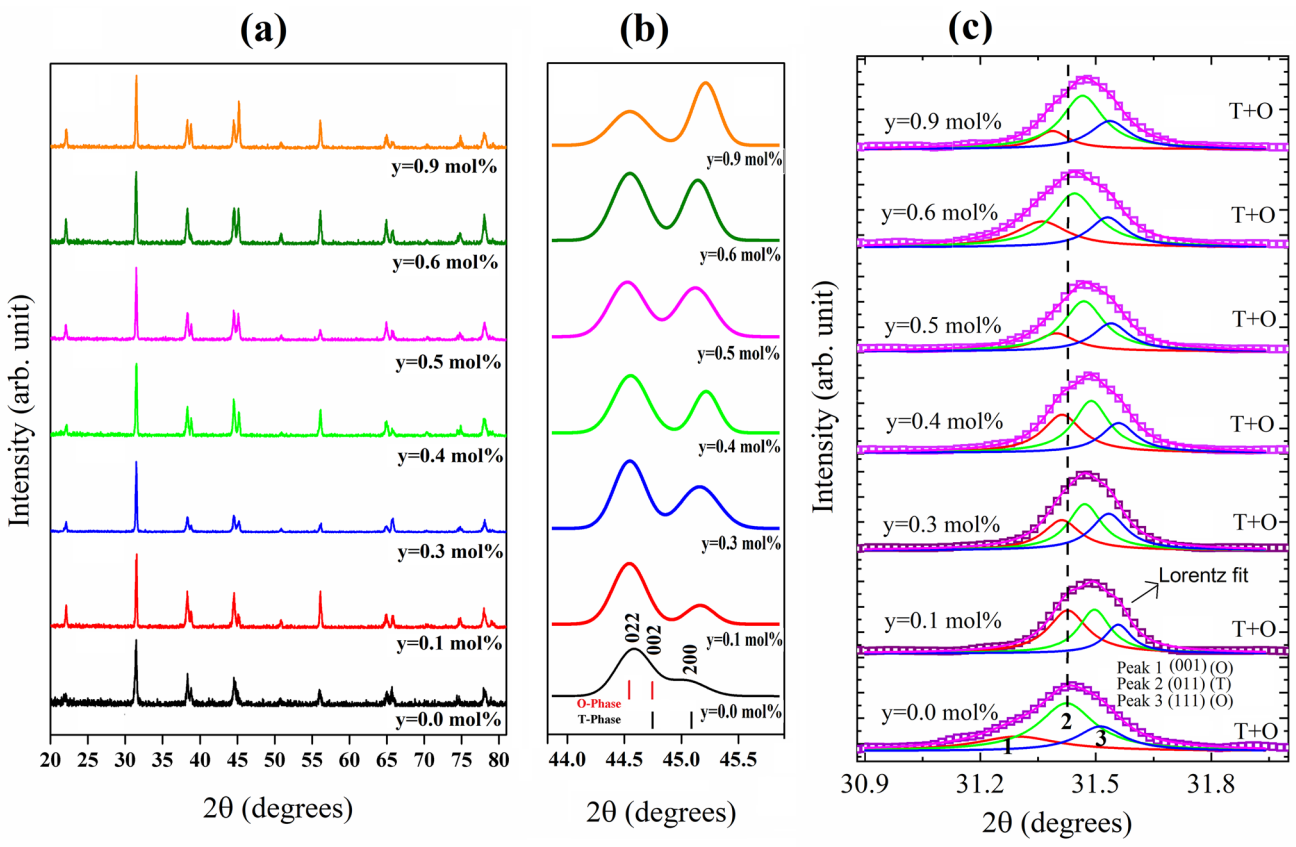


Fig. 1 **a** Room temperature HR-XRD pattern of $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ceramics, **b** enlarged view in the (2θ) range (44°–46°), **c** enlarged view of the de-convoluted diffraction peak (111) at $2\theta = 31.4°$

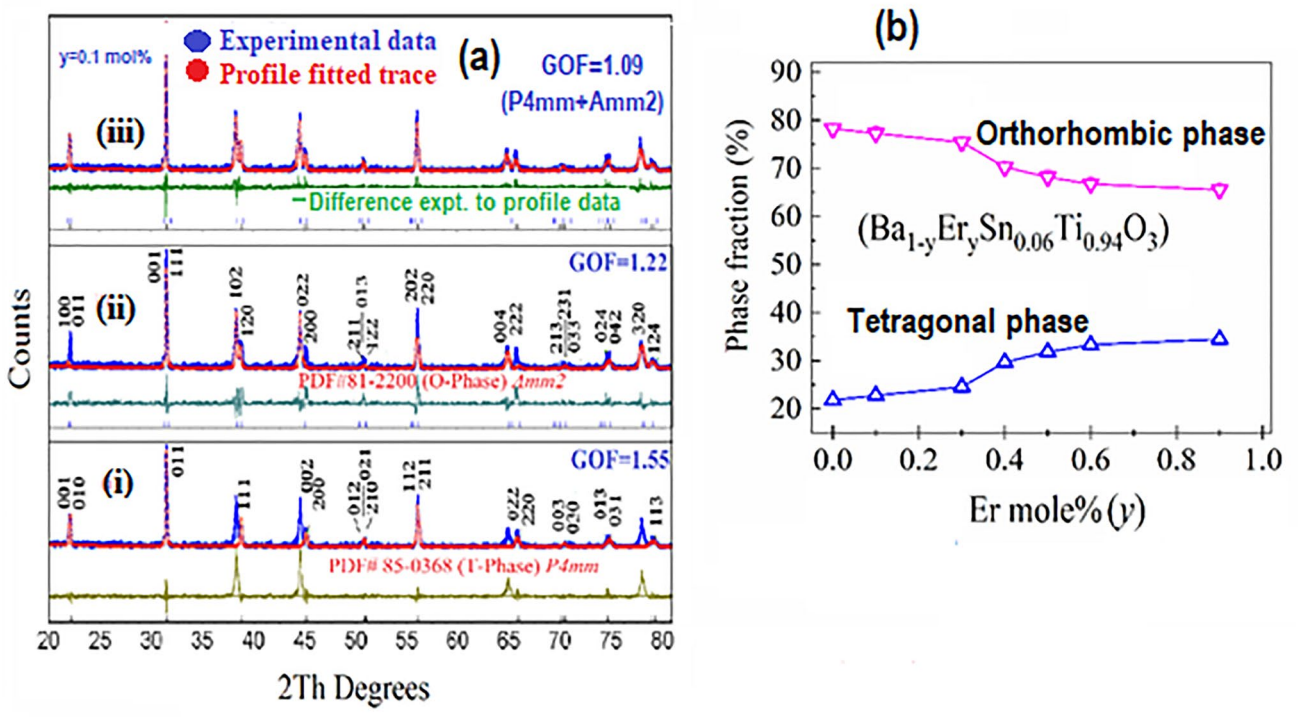


Fig. 2 **a** Rietveld refinement of XRD pattern of $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ composition with ($y = 0.1$ mol%) showing: (i) a mismatch in the fitting is shown, when *P4mm* tetragonal phase is considered during Rietveld refinement. (ii) A mismatch in the fitting shown, when *Amm2* orthorhombic phase is considered. (iii) A good-match between the exptl. and the fitted pattern when) Rietveld refinement is carried out under the two-phase model (*P4mm* + *Amm2*). **b** Variation in the phase fraction (tetra- and ortho-) with varying Er content (y) in $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ceramics

Table 1 Refined structural parameters and agreement factor for $y = 0.1$ mol% obtained after Rietveld refinement of X-ray diffraction using (*Amm2* + *P4mm*) phase coexistence model

Atoms	Space group: *P4mm*				Space group: *Amm2*			
	x	*y*	*z*	Atoms	*x*	*y*	*z*	Occ
Ba/Er	0.0000	0.0000	0.0000	Ba^{2+}/Er^{3+}	0.0000	0.0000	0.0000	1
Ti/Sn	0.5000	0.5000	0.4710	Ti^{4+}/Sn^{4+}	0.5000	0.0000	0.4950	1
O1	0.5000	0.5000	0.2400	O-1	0.0000	0.0000	0.4980	1
O2	0.5000	0.0000	0.5100	O-2	0.5000	0.2577	0.2371	1
a = b = 4.0024 (Å), c = 4.0285 (Å), V = 64.5359 (Å)^3, %Phase (*P4mm*) = 22.76				GOF(O + T) = (R_{wp}/R_p) = 1.09	a = 4.0629 (Å), b = 5.6848(Å), c = 5.7381(Å), V = 132.535(Å)^3, %Phase (*Amm2*) = 77.24			

from ~ 21.79 to 34.45%, while the intensity of the $(022)_O$ and $(002)_O$ reflections corresponding to the orthorhombic phase (Fig. 1b) decreased due to the decreasing orthorhombic phase fraction from ~ 78.21 to 65.55%. The diffraction peak (111) around 31.4° marked as peak-3 in Fig. 1c) shifts to higher angles with increasing Er. The substitution for Ba^{2+} at the *A* site having coordination number and ionic radius (CN = 12 and 0.161 nm) with smaller Er^{3+} ion (CN = 12, and 0.122 nm) results in the shrinkage of the lattice volume [27]. It is noted from Table 2 that the increasing Er content besides influencing the tetragonal/orthorhombic phase fraction, actually decreases the tetragonal distortion (*c*/*a*) more significantly in comparison to the orthorhombic distortion (*b*/*a*), and both tetragonal and orthorhombic unit-cell volumes decrease systematically.

3.2 Scanning electron microscopy and density

Increasing Er substitution in the prepared $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ ($0 \le y \le 0.9$ mol%) compositions a dense microstructure with non-uniform grain size is commonly seen in the micrographs as shown in Fig. 3a, f. The average grain size was obtained from the Lorentz fit of the histogram plot of grain size using Image-J software as shown

Table 2 Rietveld refinement of room temperature HR-XRD pattern using phase coexistence model with orthorhombic + tetragonal, goodness of fit (GOF), lattice parameters, volume, density, and average grain size

Erbium content (y) mol%	Phase	GOF	Phase percent (%)	Lattice constant (Å)			Tetragonal distortion	Orthorhombic distortion	Volume $(Å)^3$	Th. Density (%) ± 0.02	Average grain size(μm)
		(R_{wp}/R_p)		a	b	c	c/a	b/a			
0.0	Ortho	1.09	78.21	4.0651	5.7017	5.7465	1.0051	1.4025	133.19	98	25
	Tetra		21.79	4.0049		4.0255			64.553		
0.1	Ortho	1.08	77.24	4.0629	5.6848	5.7381	1.0065	1.3991	132.53	98	21
	Tetra		22.76	4.0024		4.0285			64.535		
0.3	Ortho	1.09	75.45	4.0512	5.6092	5.7643	1.0003	1.3845	130.99	98	15
	Tetra		24.55	4.0021		4.0036			64.464		
0.4	Ortho	1.09	70.23	4.0573	5.6097	5.7364	1.0042				
	Tetra		29.67	4.0023		4.0195		1.3826	130.56 64.389	98	14
0.5	Ortho	1.07	68.17	4.0535	5.5803	5.7321	1.0048	1.3766	129.66	98	12
	Tetra		31.82	3.9978		4.0173			64.208		
0.6	Ortho	1.09	66.72	4.0575	5.5508	5.7371	1.001	1.3680	129.21	98	10
	Tetra		33.28	3.9976		4.0017			64.198		
0.9	Ortho	1.08	65.55	4.0215	5.5638	5.6624	1.0012	1.3835	126.69	98	9
	Tetra		34.45	3.9975		4.0026			63.961		

in Fig. 3g, i, indicates that the average grain size decreases from ~ 25 to ~ 9 μm, and the theoretical density ~ 98 ± 0.01% remains the same for all the compositions ($0 \le y \le 0.9$ mol%). Normally, an increasing Sn content alone in BTO showed a reducing grain size in our earlier study [28], and an increasing Er content alone in BTO has resulted in reduced grain size [29]. In the present study with a fixed Sn content (0.06), an increasing Er content shows a reduction in grain size. At low Er concentration ($y = 0$–0.1), Raman spectroscopy reveals preferential substitution of Er at the *A* site, and at higher concentration levels ($y \ge 0.3$) due to the amphoteric nature of Er a possible substitution at the *B* site was noted. Increasing Er content appears to slow down the grain-boundary diffusion process and prevents grain growth, hence we observe a continuous reduction in the grain size (Table 2).

3.3 Raman spectroscopy

Figure 4a shows the room temperature Raman spectra recorded on $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ powders with increasing erbium content ($0 \le y \le 0.9$ mol%). As per the designed composition $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$, the smaller Er^{3+} ions (0.122 nm) would substitute for the larger Ba^{2+} ions (0.161 nm). However, Er^{3+} which is known to exhibit amphoteric behavior, can also occupy the *B* sites despite the mismatch in the ionic radii with Ti^{4+} (0.06 nm) and Sn^{4+} (0.069 nm) and promote oxygen vacancies in the modified compositions. However, at low concentrations of Er, it has been reported that a preference for Er substitution at the *A* sites occurs [27, 30]. For $y = 0$, the base $BaSn_{0.06}Ti_{0.94}O_3$ composition (Fig. 4a) exhibits the well-defined Raman modes at ~ 122, 160, 176, 190, 273, 303, 472, 517, 557, and 717 cm^{-1}, and the de-convoluted spectra (Fig. 4b) for $y = 0.9$ mol% shows the contribution from two weak modes at ~ 226 and 360 cm^{-1} distinctly whose signature is seen in all the Er-doped compositions (Fig. 4a). Furthermore, the Er^{3+}-doped compositions show distinctly the new Raman modes at ~ 335, 442, 621, 669, and 786 cm^{-1} appearing either as weak shoulders or humps (Fig. 4a), which evolve with increasing erbium content ($0.1 \le y \le 0.9$ mol%) [13, 14].

The additional modes arising due to Er dopant do not match with the standard modes assigned for Er_2O_3 and $ErTi_2O_7$, and thus discard the formation of any such secondary phases. The new modes may be related to (1) disorder created due to substitution of Er^{3+} in $BaSn_{0.06}Ti_{0.94}O_3$, (2) ionic imbalance between Er^{3+} and Ba^{2+} due to Er-O band for maintaining charge neutrality, and (3) uncontrolled substitution at the *A* and *B* sites in $BaSn_{0.06}Ti_{0.94}O_3$ compositions [14], or the formation of a trace amorphous phase with the presence of Er–O bonds [31, 32], and remain undetected through the X-ray diffraction analysis. The distinct feature in the Raman spectra Fig. 4a marked as (*a,b, c, d……q*), and the assignments of the identified modes are given in Table 3. The observed modes in Fig. 4a are grouped into three different zones I, II, and III. It is noted that the modes in zone I (100–400 cm^{-1}) relating to *A*-site substitution are constantly broadening, whereas in zone II (400–600 cm^{-1}) the modes are related to *B*–O bonds, and the zone III (600–800 cm^{-1}) shows the modes related to octahedra and the defect induced modes due to Er substitution. Figure 4b

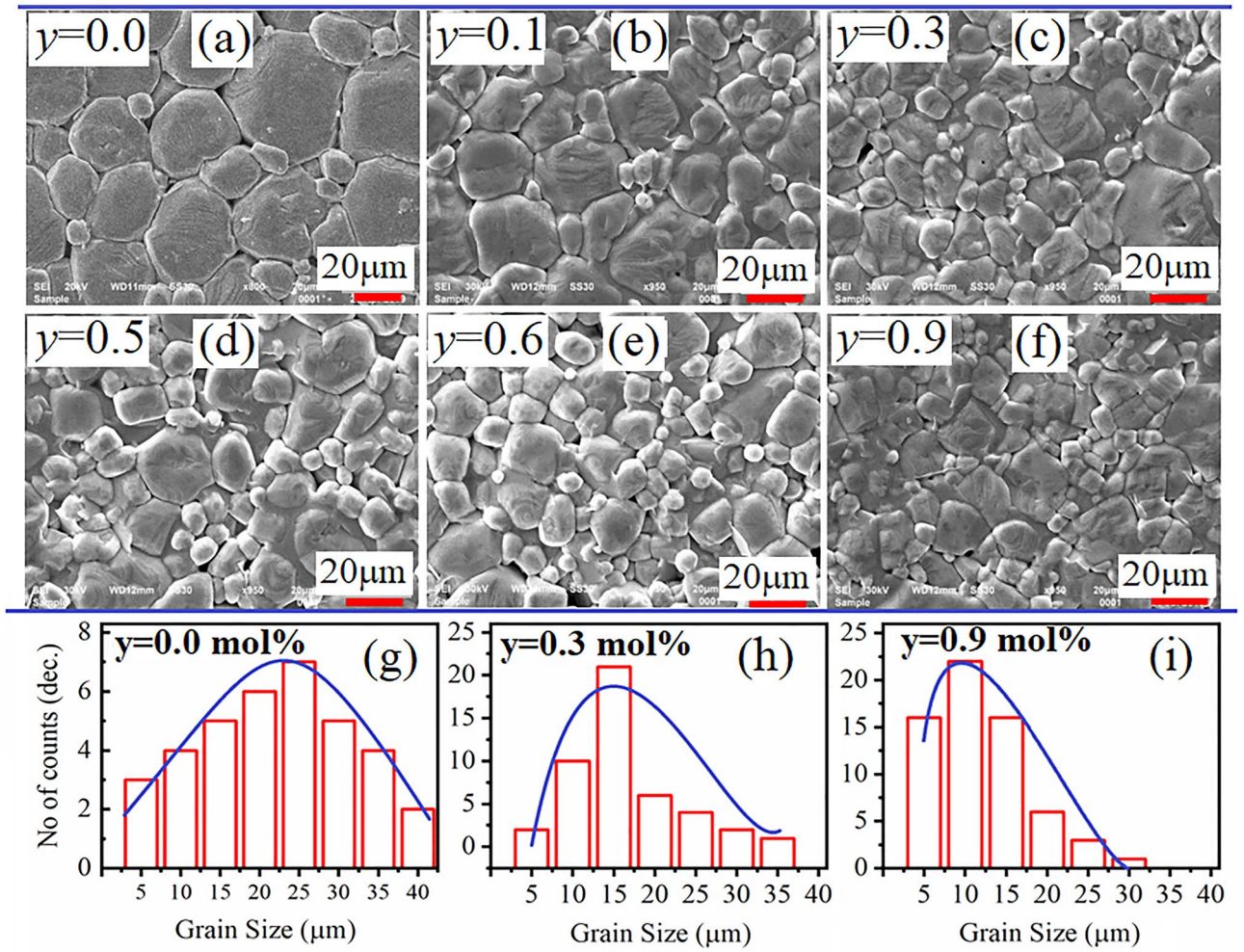


Fig. 3 **a–f** Scanning electron micrographs (SEM) of $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ceramics with varying Er content ($y = 0$–0.9 mol%), and (**g–i**) show the histograms for the grain size distribution

which shows the de-convoluted Raman spectra for the highest Er^{3+} content ($y = 0.9$ mol %), reveals 15 modes. The weak hump at ~ 160 cm^{-1} and the characteristic dip at ~ 176 cm^{-1} in Fig. 4a are absent as they disappear at the highest Er concentration (Fig. 4b).

3.3.1 Zone I (120–400 cm^{-1})

The broadening of the observed modes in zone I with increasing Er substitution indicates structural disorder, and corroborates with the fine changes in the (tetragonal/orthorhombic) distortion in the two coexisting phases as evidenced through HR-XRD. A weak mode at $a \sim 122$ cm^{-1} is expected (not clearly in Fig. 4a) due to possible coexistence of both TiO_6 and SnO_6 octahedra [28]. The weak hump at $b \sim 160$ cm^{-1} is due to constrained orthorhombic phase related to $[A_1^{(1)}(TO)]$ which weakens with increasing Er^{3+} content (Fig. 4a) [33], and is supported by the diminishing orthorhombic phase fraction detected from HR-XRD analysis. At $c \sim 176$ cm^{-1} the characteristic dip corresponds to the presence of a ferroelectric phase, and its continuous weakening with increasing Er content relates to the decreasing tetragonal and orthorhombic distortion (Table 2) [34]. The Raman modes at ~ 190, 226, and 273 cm^{-1} corresponding to $[E^{(1)}(LO), E^{(3)}(TO)]$ modes constantly broaden with increasing Er content [35, 36]. The mode $g \sim 303$ cm^{-1} which is noticeably quite sharp from the silent mode B_1 and the quasi-modes $[E^{(2)}(LO) + E^{(3)}(TO)]$ becomes less intense and slightly broadens due to the decreasing tetragonal distortion within the two coexisting phases (Fig. 2; Table2), and closely corroborates with the reducing tetragonal phase as seen in the enlarged range of HR-XRD pattern $2\theta = 44$–45.5°. The shoulder at $h \sim 335$ cm^{-1} seen specifically for the Er-doped compositions (Fig. 4a) gradually evolve with increasing with Er content, and corresponds to the increasing formation of Er–O bonds. The mode $i \sim 360$ cm^{-1} appearing as a weak shoulder does not vary with erbium content and is present in all Er-doped compositions as shown in (Fig. 4a).

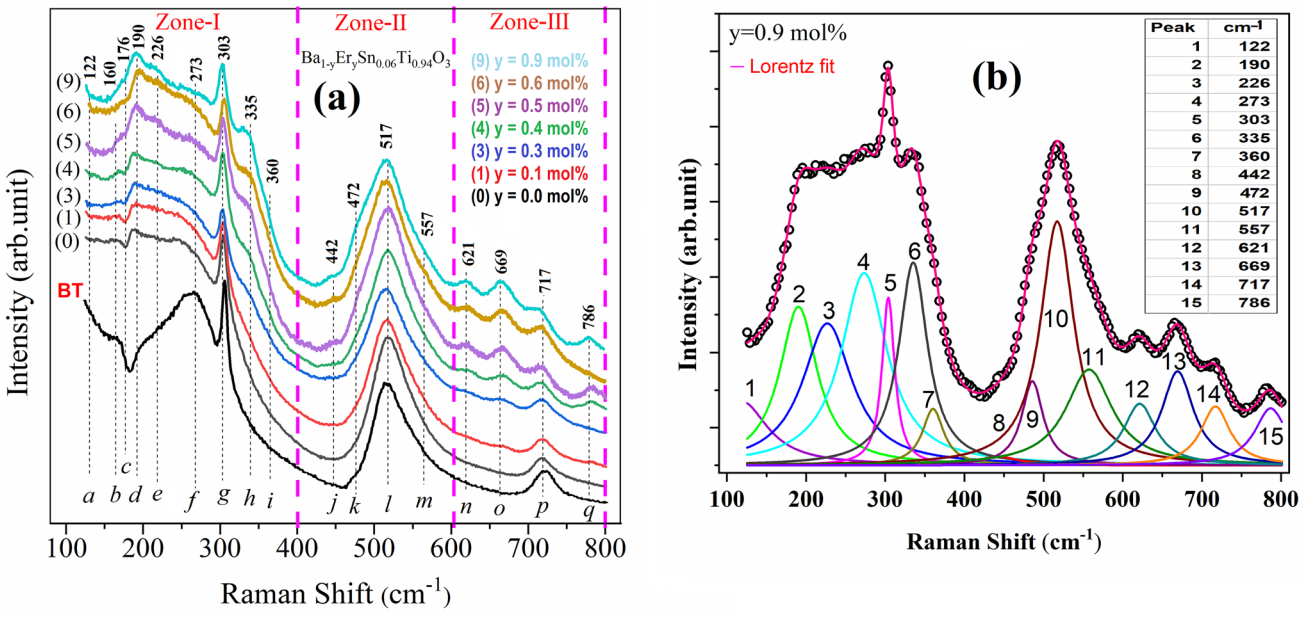


Fig. 4 **a** Room temperature Raman spectra recorded on $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ceramics. **b** De-convoluted Raman spectra for $y = 0.9$ mol%

Table 3 The assignment of identified Raman modes in $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ multifunctional ceramics with varying Er-doped ($0 \le y \le 0.9$ mol %) recorded at 300 K

Marked Raman modes	Raman modes (cm^{-1})	Assignment of Raman modes	References
a	122	Dip due to interface of $A_1^{(1)}(TO)$ phonos	[28]
b	160	$A_1^{(1)}(TO)$	[33]
c	176	$A^1(TO)$and$A^1(TO_2)$ of anti-resonance dip characteristic modes	[33]
d	190	$E^{(1)}(LO), E^{(2)}(TO)$	[34]
e	226	De-convoluted modes	[28]
f	273	$A_1^{(2)}(TO)$	[35]
g	303	B_1, $E^{(2)} + E^{(3)}(TO)$	[28]
h	335	Er–O bonds, abnormal modes	Present
i	360	De-convoluted modes	[28]
j	442	Er–O bonds, abnormal modes	Present
k	472	$A_1^{(1)}(LO) - E^{(1)}(LO)$ quasi-modes	[28]
l	517	$E^{(4)}(TO)A_1^{(3)}(TO)$	[36]
m	557	F_{1u} and E_g assignment	[37]
n	621	Er–O bonds, abnormal modes	Present
o	669	Er–O bonds, abnormal modes	Present
p	717	$A_1^{(3)}(LO) - E^{(3)}(LO)$ of $BaTiO_3$	[37, 38]
q	786	Er–O bonds, defect	Present

3.3.2 Zone II (400–600 cm^{-1})

The mode $l \sim 517$ cm^{-1} remains un-shifted, and becomes quite broad and asymmetric with increasing Er^{3+} content. The asymmetry develops due to the building up of the modes $j \sim 442$ cm^{-1} as a weak hump, and the appearance of the overlapping of the broadened modes at $k \sim 472$ and $l \sim 517$ cm^{-1} which were identified in the de-convoluted spectra. The small kink seen in the spectra of barium titanate (BT) in Fig. 4a—zone II, at ~ 472 cm^{-1} is due to the shared vibrations between Ti and O_6 octahedra and is identified with $[A_1^{(1)}(LO) - E^{(1)}(LO)]$ quasi-modes. It appears to be modified

due to the combined presence of Sn and Er substituting at the *B* site, and, therefore, appears clearly as a hump with increasing Er content, and is seen more clearly in the spectra shown in Fig. 4a for the increased Er content $y = 0.9$ mol% [34, 37].

3.3.3 Zone III (600–800 cm^{-1})

The Raman mode $p \sim 717$ cm^{-1} is associated with the breathing mode of TiO_6 octahedral in the ferroelectric phase assigned to $[A_1^{(3)}(LO) - E^{(3)}(LO)]$ of $BaTiO_3$ [36, 38], and is seen to be slightly distorted with increasing Er content, and suggests the possible substitution of Er at the *B* site. The two new modes marked as *n* and *o* appearing at ~ 621 and 669 cm^{-1}, respectively, become more prominent with increasing Erbium dopant, and the induced defect mode at ~ 786 cm^{-1} appears more significantly at the higher Er concentrations ($y = 0.5$ and 0.9 mol %).

3.4 Phase transitions

Effects of Er dopant on the three-phase transitions, rhombo–ortho (T_{R-O}), ortho–tetra (T_{O-T}), and tetra–cubic (T_{T-C}) have been analyzed using temperature dependent dielectric measurements and modulated differential scanning calorimetry (MDSC).

3.4.1 Dielectrics phase transitions

In Fig. 5a, the measured dielectric constant at room temperature (RT) shows a strong dispersion in the dielectric response $\varepsilon'(f)$ for $y = 0.4$–0.9 mol% due to the presence of trace amount of residual/defect phases undetected by XRD, and were unveiled only through Raman spectroscopy for compositions with high Er content. The observed dielectric dispersion in the low-frequency region could be understood in terms of well-known Maxwell–Wagner effect. In the high frequency region, the dispersion in $\varepsilon'(f)$ is reduced, and a comparison of the dielectric constant ε' at the highest frequency (1 MHz) as shown in the inset of Fig. 5a shows ε' increases from 2400 to 3200 up to $y = 0.3$ mol %, and thereafter it decreases with further increase in Er content. Such a variation in the measured ε' at RT is understood from the shift in the T_{O-T} phase transition to lower temperature as seen in Fig. 6a–h, and the measurement of ε' at RT corresponds to the dip in the $\varepsilon'(T)$ characteristics after the occurrence of the T_{O-T} phase transition. The dielectric loss factor of the prepared ceramics (tan $\delta \sim 0.025$) is low in the frequency range 1 kHz to 1 MHz with Er content ($y = 0.1$–0.3 mol%). In comparison, the compositions $y = 0.1$–0.3 mol% do not show the strong dispersion in *tan δ*, and the observed $\varepsilon'(f)$ response is more due to reduced electrode effects, because ε' is also showing a tendency towards

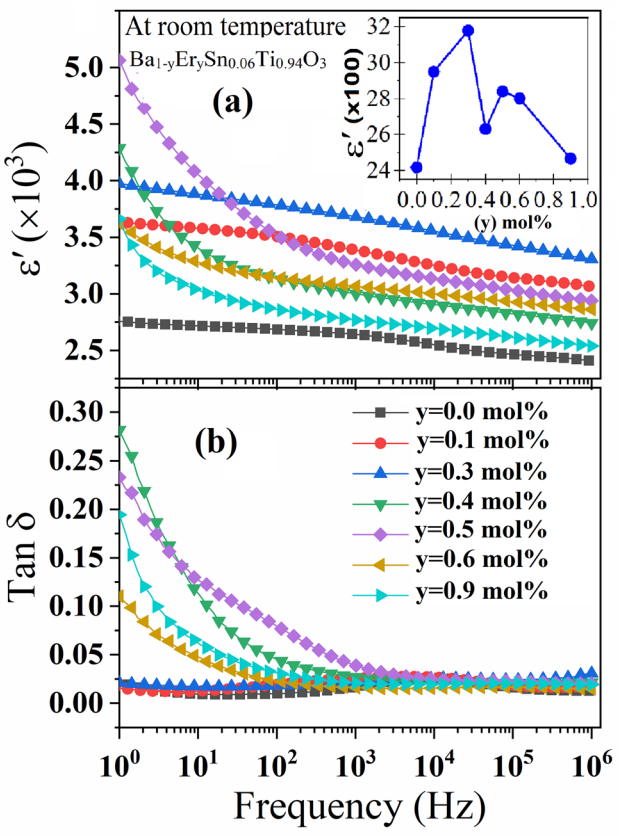


Fig. 5 Room temperature dielectric constant and Tan δ at 1 MHz with varying Er dopant in $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ($0 \le y \le 0.9$ mol%) ceramics

saturation at lower frequencies. A strong dispersion in the low-frequency region is evident more clearly in the dielectric loss tan(δ) characteristics for higher Er content. The measured dielectric dissipation factor tan $\delta \sim 0.025$ is low for the compositions $y = 0$ to 0.3 mol% which show relatively less dispersion as seen in Fig. 5b.

In Fig. 7, the dielectric phase transitions of the Sn-doped $BaTiO_3$ system, and the Er/Sn-doped $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ system show contrasting shifts in the phase transition temperatures. The Er/Sn-doped system shows an opposite trend, where T_{O-T} and T_{R-O} phase transitions shift towards lower temperatures, and the $T_{T-C} \sim 78$ °C phase transition is almost unaffected with increasing Er content. This is in contrast to the observations for Sn^{4+} substitution at the Ti^{4+} sites in the $BaTiO_3$ system where (T_{O-T} and T_{R-O}) shift to higher temperatures, and T_{T-C} shifts to lower temperatures (Fig. 7), and ultimately with increasing Sn, all the three-phase transitions tend to meet at the quadruple point for Sn ~ 0.11, which is well known in the reported literature [28]. In the present study, for a fixed Sn ~ 0.06 content (Fig. 7), the non-shifting of the phase transition $T_{T-C} \sim 78$ °C, with increasing Er^{3+} concentration in the range $y = 0.1$–0.9 mol% confirms the

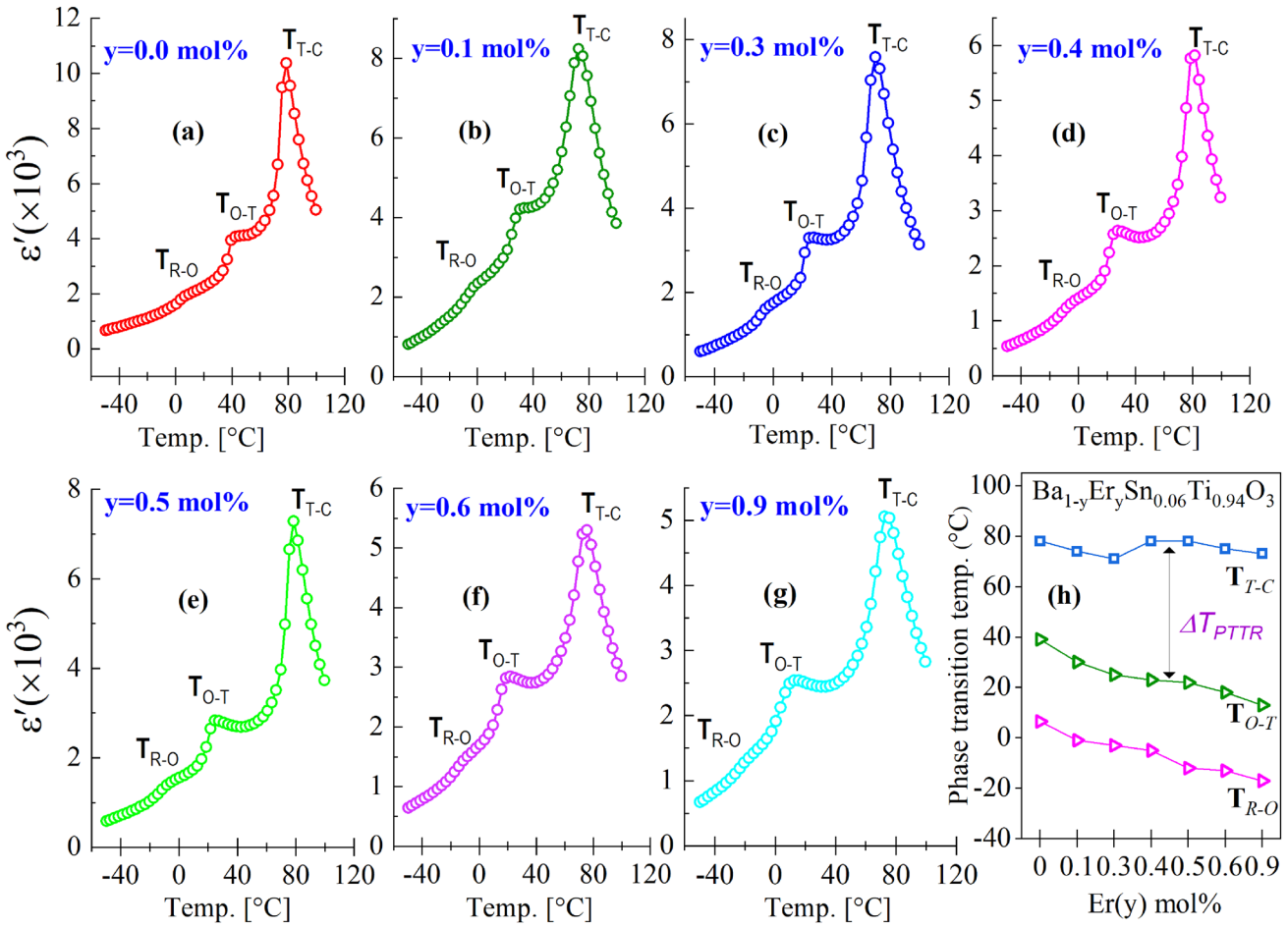


Fig. 6 **a**, **g** Temperature dependence of dielectric constant (ε′) with varying erbium dopant in $Ba_{1y}Er_yTi_{0.94}Sn_{0.06}O_3$ ($0 \le y \le 0.9$ mol%) and (**h**) shows the variation in the three-phase transition temperatures with varying erbium dopant in $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ($0 \le y \le 0.9$ mol%) ceramics

preferred substitution of Er^{3+} for Ba^{2+} at the *A* site. Shifting of T_{O-T} and T_{R-O} to lower temperatures as seen in the present study (Fig. 7) with increasing Er concentration was reported earlier in two other Er-doped systems: (1) Er-doped $BaTiO_3$, and (2) BZT-*x*Er but only at higher concentrations of Er content [12, 39]. However, in these earlier studies the results were left unexplained. Since Er^{3+} with its intermediate size is an amphoteric dopant and can substitute at either/both the cation sites (Ba^{2+} or Ti^{4+}) promoting oxygen vacancies, it is expected that even in the present study some substitution of Er^{3+} has probably occurred at the Ti^{4+} sites as well, and this is reflected in the small variations/fluctuations noticed in T_{T-C} transition in the range ~ 78 to 75 °C (Fig. 6h) which confirmed the possibility of some Er entering the *B* sites specifically at higher concentration of Er ($y = 0.06$–0.09 mol %), whereas at lower concentrations of Er dopant, we observe preferential substitution of Er^{3+} at the *A* sites. In the reported literature a high concentration of Er dopant in the BCZT-*x*Er, system with $x = 0.8$ mol % shifted the (T_{T-C}) to a significantly lower temperature with respect to the undoped system [40]. Dielectric constant and dissipation factor at room temperature (300 K), phase transition temperature, dielectric constant at Curie temperature ($\varepsilon'_{(Tm)}$) and the degree of diffuseness (γ) from modified Curie–Weiss ~ 1.23 to 1.38 for y (= 0.0–0.9 mol%) of Ba_{1-y} $Er_ySn_{0.06}Ti_{0.94}O_3$ ceramics at 1 MHz are given in Table 4 and suggests a diffused ferroelectric phase transition.

3.4.2 Modulated differential scanning calorimetry

The distinct shifts in the phase transition temperatures due to the addition of Er dopant in $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ ceramics were verified additionally by modulated scanning calorimetric (MDSC) measurements in the range (− 20 to 100 °C). The MDSC results in Fig. 8a, b show three exothermic pics, corresponding to T_{R-O}, T_{O-T}, and T_{T-C} phase transitions, respectively. With increasing Er, the T_{R-O} ~ (7 to − 15 °C) and T_{O-T} ~ (39–10 °C) decrease continuously and T_{T-C} ~ 78 °C almost remains same. Reversible heat capacity (*Cp*) of T_{R-O} (~ 5.25 to 4.96) and T_{O-T} (~ 5.64 to 5.18) and T_{T-C}

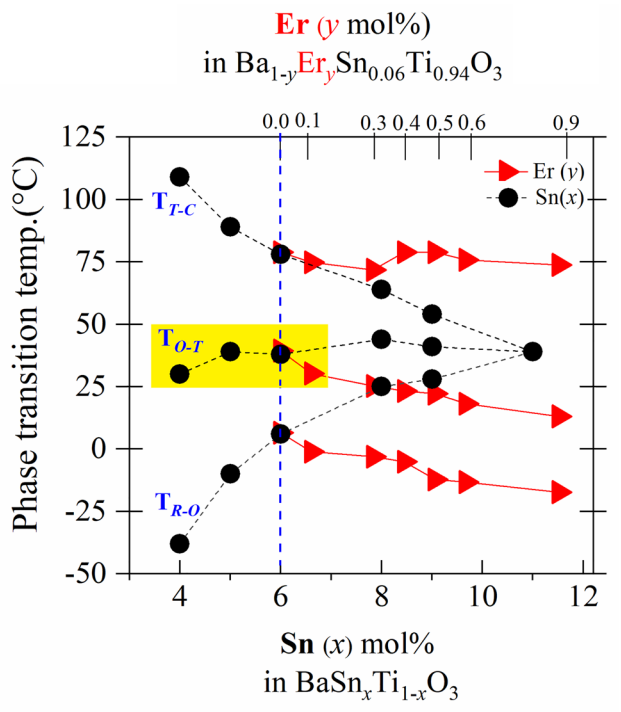


Fig. 7 Phase transition diagram of $BaSn_xTi_{1-x}O_3$ ceramics and $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ ceramics

(~6.13 to 5.16) mJ/°C is found to decrease continuously with increasing erbium content (Table 5; Fig. 9). The phase transition temperature zone (ΔT_{PTTR}) = ($T_{T-C} - T_{O-T}$) is found to increase, and is similar to the trend seen in the dielectric data shown in Fig. 8b. The enhanced phase transition temperature zone may be useful for application purposes around the room temperature (~ 30 °C) due to the coexistence of phases as compared to undoped $BaSn_{0.06}Ti_{0.94}O_3$ ceramics. Temperature and frequency-dependent dielectric constant show diffuse phase transition with varying erbium dopant $y = 0.0$, 0.1, 0.5, and 0.9 mol% and the degree of relaxation (ΔT_m) remains same indicating a diffused ferroelectrics phase transition as shown in Fig. 10a–d.

3.5 P–E hysteresis loop DC resistivity and piezoelectric coefficient

P–E hysteresis loops measured at RT (300 K) on $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ ceramics with varying erbium ($0 \le y \le 0.9$ mol%) dopant are shown in Fig. 11a, b. All the samples were subject to approximately the same value of maximum electric field ~ 15 kV/cm at 50 Hz to compare the saturation polarization at a constant applied field. The measured values of *Pr*, *Ps*, and coercive field *Ec* are given in Table 6. With increasing Er dopant, the measured *Ps, Pr,* and *Ec* values tend to decrease. A slightly higher rectangularity (*Pr/Ps*) ~ 0.24 is seen for $y = 0.4$ and 0.9 mol% in comparison to the lower values of *Pr/Ps* = 0.23, 0.222, 0.180, 0.22 for remaining compositions with varying Er content $y = 0$–0.6 mol% (Table 6). The observed variation in *Pr/Ps* reflects the non-centro symmetry for all the compositions expected from the varying phase percentage (*Amm2%* and *P4mm%*) only, because at RT (300 K), the XRD analysis established coexistence of phases only with varying Er content. All the compositions show decreasing *Ps, Pr,* and *Ec* values compared to undoped ($y = 0$). However, a reduction in the switchable polarization and poor retention at zero field as seen from decreasing remnant polarization *Pr* suggests the unsettling influence due to the high concentration of Er–O bonds. The high dc resistivity $\rho_{dc} \sim 10^{12}\ \Omega$ cm observed for $BaSn_{0.06}Ti_{0.94}O_3$ ceramics at room temperature is found to decrease to ~ $4.21 \times 10^{11}\ \Omega$ cm with increasing Er substitution in the range $y = 0.1$–0.9 mol%. The room temperature piezoelectric coefficient measured on poled samples without Er ~ 172 pC/N ($y = 0$), is found to increase to ~ 179 pC/N for $y = 0.3$ mol% and thereafter decreases with further increase in Er content from $y = 0.4$ to 0.9 mol% as given in Table 6.

3.6 Light upconversion luminescence

Room temperature (~27 °C) light upconversion emission spectra of Ba_{1-y} $Er_ySn_{0.06}Ti_{0.94}O_3$ ($0.1 \le y \le 0.9$ mol%) under the excitation of 980 nm exhibited two strong green emission

Table 4 Room temperature dielectric constant (ε′), dielectric loss (ε"), and dissipation factor (Tanδ), the phase transition temperatures (T_{R-O}, T_{O-T}, and T_{T-C}), $\Delta T_{PTTR} = (T_{T-C}) - (T_{T-C})$, dielectric constant at Curie temperature (ε′(Tm)) and the degree of diffuseness (γ) of Ba_{1-y} $Er_ySn_{0.06}Ti_{0.94}O_3$ ceramics at 1 MHz

Er^{3+} (*y*) mol%	ε'_{RT} @1 MHz	ε''_{RT} @1 MHz	$(Tan\delta)_{RT}$@ 1 MHz	T_{R-O} (°C)	T_{O-T} (°C)	T_{T-C} (°C)	ΔT_{PTTR} at $\varepsilon'_{(T)}$ (°C)	$\varepsilon'_{(Tm)}$ 1 MHz	(γ) 1 MHz
0	2715	46	0.017	6.5	39	78	39	10,378	1.32
0.1	2724	48	0.018	− 1.5	30	74	44	8236	1.34
0.3	2784	37	0.013	− 3.0	25	71	46	7588	1.39
0.4	3269	67	0.021	− 5.0	23	78	55	5761	1.32
0.5	3065	50	0.069	− 12	21	78	57	7283	1.32
0.6	2638	33	0.012	− 13	18	75	57	5293	1.36
0.9	2565	50	0.020	− 17	13	73	60	5052	1.36

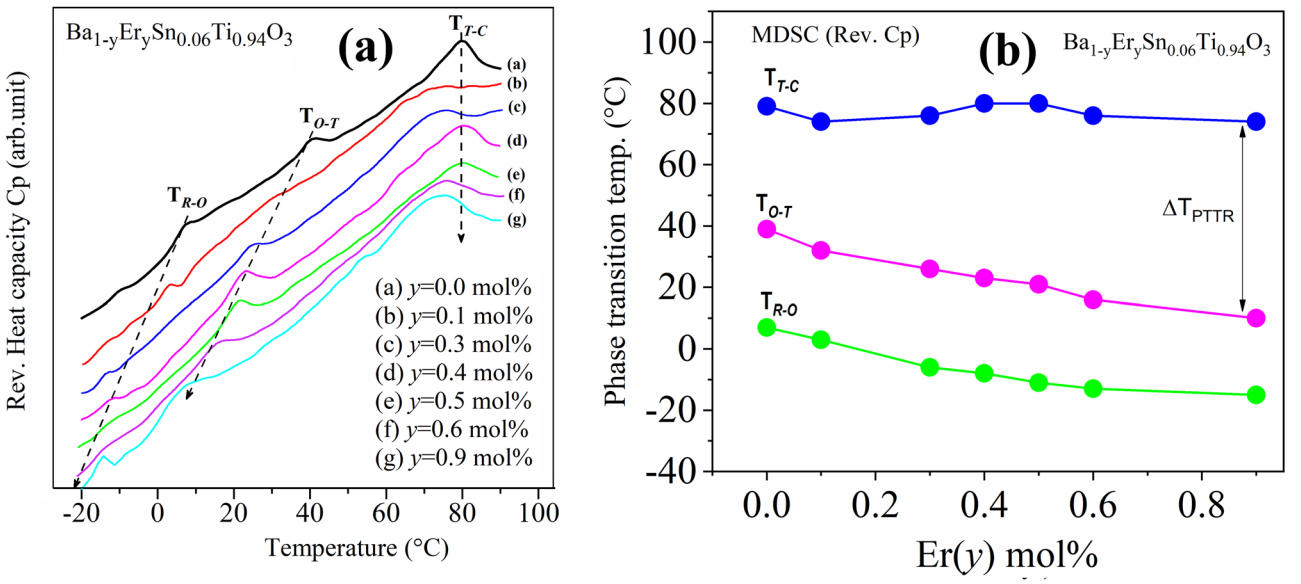


Fig. 8 **a** Temperature-dependent heat capacity (Rev. C_v) and **b** phase transition diagram with varying erbium dopant in $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ($0 \le y \le 0.9$ mol%)

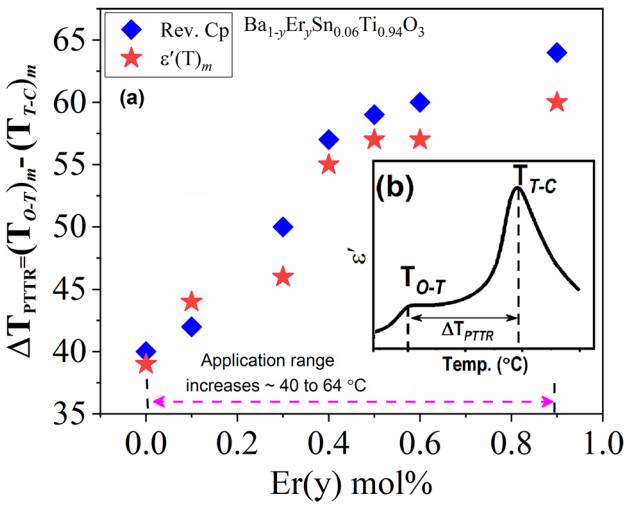


Fig. 9 Dielectric and heat capacity measurements showing the widening of the phase transition zone between T_{O-T} and T_{T-C} (ΔT_{PTTR}) for $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}$ ceramics

bands around ~ 524 nm, ~ 553 nm and one weak red emission band at ~ 657 nm as shown in Fig. 12a for all the compositions $y = 0.1$–0.9 mol%, and are related to the transitions ($^4H_{11/2} \rightarrow {}^4I_{15/2}$), ($^4S_{3/2} \rightarrow {}^4I_{15/2}$) and ($^4F_{9/2} \rightarrow {}^4I_{15/2}$), respectively. It is observed that the upconversion intensity increases with increasing Er^{3+}content up to $y = 0.5$ mol%. Above the critical concentration for quenching ($y = 0.5$ mol%) the interatomic distance of Er^{3+} ions become shorter, non-radiative energy transfer processes dominate, and accordingly the radiative emission is quenched [41]. The variation in the intensity of the upconverted emission intensity recorded in the temperature range 303–573 K for the $y = 0.5$ mol% (5Er-BSn6T) composition is shown in inset (b) of Fig. 12a. The variation of green upconversion emission at 524 and 553 nm for $y = 0.5$ mol% ceramics within the temperature range 303–573 K shows the position of the green bands does not change. The photoluminescence intensity ratio $R_{PL} = \left(\frac{I_{524}}{I_{553}}\right) = Cexp\left(-\frac{\Delta E}{kT}\right)$, where k = Boltzmann constant,

Table 5 Modulated differential scanning calorimetry (MDSC) measurements on the $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}$ ceramics annealed powder (quantity ~ 12 mg) showing phase transition temperature, phase transition temperature range ($\Delta T_{PTTR} = T_{T-C} - T_{O-T}$), and Rev. Heat capacity

Erbium	MDSC Phase transitions Temperature at Rev. C_p				MDSC Rev. Heat capacity		
(y) mol%	T_{R-O} ~ (℃)	T_{O-T} ~ (℃)	T_{T-C} ~ (℃)	ΔT_{PTTR} (℃)	C_p(R–O) ~ (mJ/℃)	C_p(O–T) ~ (mJ/℃)	C_p(T–C) ~ (mJ/℃)
0	7	39	79	40	5.25	5.64	6.13
0.1	3	32	74	42	4.57	4.87	5.22
0.3	− 6	26	76	50	4.02	4.08	4.55
0.4	− 8	23	80	57	4.10	4.43	4.82
0.5	− 11	21	80	59	4.20	4.67	5.44
0.6	− 13	16	76	60	4.52	4.89	5.46
0.9	− 15	10	74	64	4.96	5.18	5.61

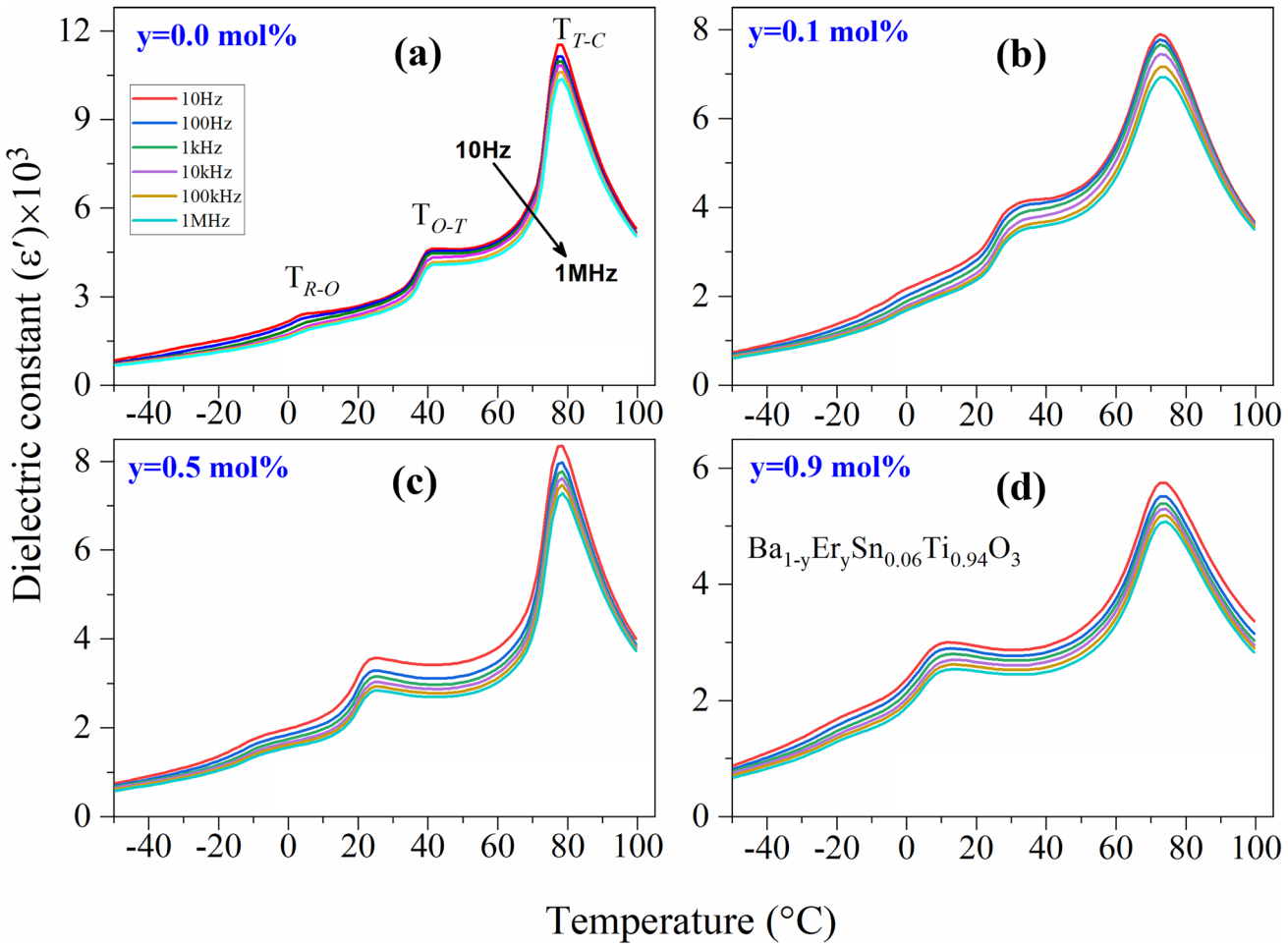


Fig. 10 **a–d** Frequency (10–1 MHz), and temperature (− 50 to 100 °C) dependence of dielectric constant (ε') of $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ for $y=0$, 0.1, 0.5 and 0.9 mol% ceramics

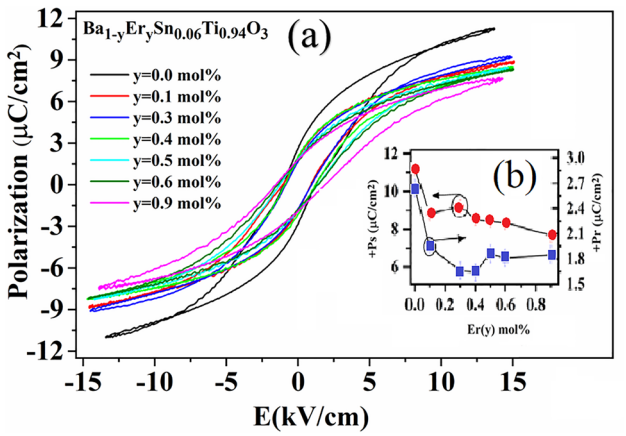


Fig. 11 **a**,**b** Room temperature *P–E* hysteresis loops of $Ba_{1-y}Er_yTi_{0.94}Sn_{0.06}O_3$ ceramics with varying Er ($0 \le y \le 0.9$ mol%)

and ΔE = energy gap between the $^4H_{11/2}$ and $^4S_{3/2}$. The ratio R_{PL} increases with increasing temperature as shown in Fig. 12c. The thermally coupled transitions due to Er doping enable the applicability of $Ba_{1-y}Er_ySn_{0.06}Ti_{0.94}O_3$ ($0.1 \le y \le 0.9$ mol%) ceramic for sensing the temperature optically based on the fluorescent intensity ratio (R_{PL}) method [42]. The absolute sensitivity (S_{OT}) can be defined as $S_{OT} = \frac{d(R_{PL})}{dT} = \left(\frac{\Delta E}{kT^2}\right) \times R_{PL}$. The variation of sensitivity as a function of temperature shown in Fig. 12d, exhibits a maximum sensitivity ~ 0.00344 K^{-1} around ~ 450 K. The measured temperature sensing performance of Ba_{1-y} $Er_ySn_{0.06}Ti_{0.94}O_3$ ceramics is of the same order as that observed for oxides, fluorides, and glasses [11, 33].

Table 6 Room temperature saturation polarization *Ps* (μC/cm^2), coercive field *Ec* (kV/cm), remnant polarization *Pr* (μC/cm^2), and dc resistivity ρ_{dc} (Ω-cm) of Ba_{1-y} Er_y $Sn_{0.06}Ti_{0.94}O_3$ ceramics

Erbium content (*y*) mol%	+*Ps* (μC/cm^2)	− *Ps* (μC/cm^2)	+*Pr* (μC/cm^2)	− *Pr* (μC/cm^2)	+*Ec* (kV/cm)	− *Ec* (kV/cm)	ρ_{dc} Ω-cm 10^{11}	+*Pr*/*Ps*	d_{33} (pC/N)
0.0	11.17	10.99	2.64	2.84	1.25	1.01	9.54	0.2363	172
0.1	8.80	8.76	1.96	1.83	0.93	1.09	8.53	0.2227	174
0.3	9.20	9.08	1.66	1.71	0.95	0.86	7.52	0.1804	179
0.4	8.53	8.22	2.07	2.10	1.20	0.92	6.12	0.2427	176
0.5	8.43	8.21	1.86	1.85	1.33	1.28	5.72	0.2206	173
0.6	8.33	8.16	1.81	1.82	1.37	1.38	5.12	0.2173	170
0.9	7.68	7.46	1.85	1.84	1.96	1.53	4.21	0.2408	168

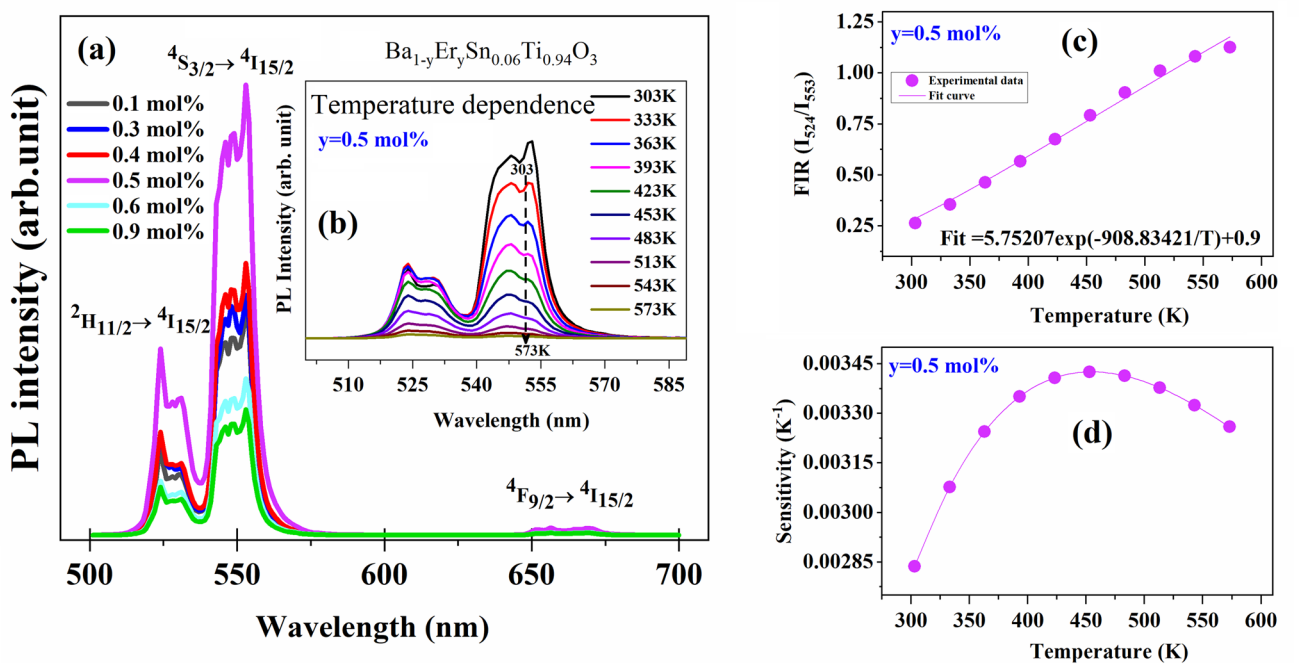


Fig. 12 **a** Room temperature light upconversion emission spectra with varying erbium content in Ba_{1-y} $Er_ySn_{0.06}Ti_{0.94}O_3$ ($0.1 \le y \le 0.9$ mol%). Inserted fig. **b** shows temperature dependent light upconversion for $y = 0.5$ mol%, and fig. **c**, **d** show the FIR and sensitivity plots respectively for $y = 0.5$ mol% erbium dopant

4 Conclusions

The coexistence of orthorhombic (*Amm2*) + tetragonal (*P4mm*) phases is confirmed by Rietveld refinement at RT (~ 300 K) using high-resolution XRD studies. Average grain size decreases from 25 to 9 μm with increasing erbium content. Raman spectroscopy highlights the presence of disorder due to Er substitution and shows additional ~ 335, 442, 621, 669, and 786 cm^{-1} due to Er-O bonds and the presence of a defect induced mode. With increasing erbium content T_{R-O} and T_{O-T} shift to lower phase transition temperature while T_{T-C} almost remains unchanged, and the width (ΔT_{PTTR}) between the two-phase transitions increases from ~ 39 to 60 °C. For $y = 0.3$–0.4 mol%, the shifts in the phase transition temperatures T_{O-T} close to room temperature (Table 4) provide coexistence of phases along the ortho-tetragonal polymorphic phase boundary (PPB), and thus increase the likelihood for enhanced properties. The room temperature dc resistivity decreases from ~ 10^{12} to 4.21×10^{11} Ω-cm and remnant polarization decreases with $y = 0$ to $y = 0.9$ mol%, respectively. The room temperature piezoelectric coefficient for undoped ~ 172 pC/N and increases up to ~ 179 pC/N for $y = 0.3$ mol% and decreases with further increase in Er content. The sensitivity of temperature dependent

light upconversion for $y = 0.5$ mol% as a function of temperature reaches its maximum values of ~ 0.00344 K^{-1} at ~ 450 K.

Acknowledgements One of the authors (Mohd Azaj Ansari) is thankful to University Grants Commission (UGC), India, for providing financial assistant vide letter No. Ref: FOS-1/114/PhD/4805, and Roll No 512429, for the Junior Research Fellowship (JRF:2015-2017) and Senior Research Fellowship (SRF:2017-2020). The authors wish to thank University of Delhi for the central instrument facility at the University Science Instrument Centre (USIC) and Guru Jambheshwar University of Science and Technology, Hisar (Haryana) for using the Novo-Control impedance analyzer under DST (FIST) sponsored programme for the measurement of dielectric properties.

References

1. W. Liu, J. Wang, X. Ke, S. Li, J. Alloy. Compd. **30**, 1397–1400 (2004)
2. Y. Yao, C. Zhao, D. Lv, D. Wang, H. Wu, Y. Yang, X. Ren, J. Epl. **98**, 27008 (2012)
3. A.K. Kalyani, H.K. Krishnan, A. Sen, A. Senyshyn, R. Ranjan, Phys. Rev. B **91**, 024101 (2015)
4. L. Xu, Z. Wang, B. Su, C. Wang, X. Yang, R. Su, X. Long, C. He, Crystals **10**, 434 (2020)
5. A.K. Kalyani, A. Senyshyn, R. Ranjan, J. Appl. Phys. **114**, 014102 (2013)
6. X. Zhang, L. Wu, S. Gao, J.Q. Liu, B. Xu, Y.D. Xia, J. Yin, Z.G. Liu, AIP Adv. **5**, 047134 (2015)
7. A. Movchikova, O. Malyshkina, G. Suchaneck, G. Gerlach, R. Steinhausen, H.T. Langhammer, C. Pientschke, H. Beige, J. Electroceram. **20**, 43–46 (2008)
8. W. Li, Z. Xu, R. Chu, Z. Wu, J. Hao, P. Fu, J. Du, Z. Yue, RSC Adv. **5**, 91903 (2015)
9. X. Wang, C.N. Xu, H. Yamada, K. Nishikubo, X.G. Zheng, Adv. Mater. **17**, 1254–1258 (2005)
10. M.K. Mahata, T. Koppe, K. Kumar, H. Hofsass, U. Vetter, Sci. Rep. **10**, 8775 (2020)
11. R. Bokolia, M. Mondal, V.K. Rai, K. Sreenivas, J. Appl. Phys. **121**, 084101 (2017)
12. E. Antonelli, M. Letonturier, J.-C. M' Peko, A.C. Hernandes, J. Eur. Ceram. Soc. **29**, 1449–1455 (2009)
13. L. Shao, L. Sheng, F. Wen, Z. Ying, H. Huang, P. Zheng, W. Wu, Ceram. Int. **44**, 99159922 (2018)
14. P. Du, L. Luo, W. Li, Y. Zhang, H. Chen, Mater. Sci. Eng. B **178**, 1219–1223 (2013)
15. F.A. Ismail, R.A.M. Osman, M.S. Idris, A.Z. Jamal, EPJ Conf. **162**, 01051 (2017)
16. W. Li, Z. Xu, R. Chu, P. Fu, G. Zang, J. Alloy. Compd. **583**, 305–308 (2014)
17. D. Peng, H. Zou, C. Xu, X. Wang, X. Yao, J. Lin, T. Sun, AIP Adv. **2**, 042187 (2012)
18. X. Wu, K.W. Kwok, J. Am. Ceram. Soc. **97**, 1504–1510 (2014)
19. D. Peng, X. Wang, C. Xu, X. Yao, J. Lin, T. Sun, J. Am. Ceram. Soc. **96**, 184–190 (2013)
20. T. Wei, Z. Dong, C.Z. Zhao, Y.J. Ma, T.B. Zhang, Y.F. Xie, Q.J. Zhou, Z.P. Li, Cerm. Int. **42**, 5537–5545 (2016)
21. Y. Zhao, Y. Ge, X. Zhang, Y. Zhao, H. Zhou, J. Li, H. Jin, J. Alloy. Compd. **683**, 171–177 (2016)
22. X.N. Zhu, W. Zang, X.M. Chen, AIP Adv. **3**, 082125 (2013)
23. P.V. Balachandran, D. Xue, T. Lookman, Phys. Rev. B **93**, 144111 (2016)
24. Y. Yu, H. Zou, Q.F. Cao, X.S. Wang, Y.X. Li, X. Yao, Ferroelectric **487**, 77–85 (2015)
25. R. Kumar, I. Singh, R. Meena, K. Asokan, B. Birajdal, S. Patnaik, Mat. Res. Bull. **123**, 110694 (2020)
26. V. Mueller, H. Beige, H.-P. Arbicht, C. Eisenschmidt, J. Mater. Res. **19**, 2834 (2004)
27. C. Liu, Q. Wang, X. Wu, B. Sa, H. Sun, L. Luo, C. Lin, X. Zheng, T. Lin, Z. Sun, ACS Omega **4**, 11004–11013 (2019)
28. M.A. Ansari, K. Sreenivas, Ceram. Int. **45**, 20738–20749 (2019)
29. V.V. Mitic, Z.S. Nikolic, V.B. Paunaic, M. Milijkovic, B. Jordovic, L. Zivkovic, J. Am. Ceram. Soc. **93**, 132–137 (2010)
30. R. Mahani, O. El-sayed, S.K. El-Mahy, I.K. Battisha, Acta Phys. Polo. A **137**, 410–415 (2020)
31. Y.K. Kshetri, T. Kamiyama, S. Torii, S.H. Jeong, T.H. Kim, H. Choi, J. Zhou, Y.P. Feng, S.W. Lee, Sci. Rep. **10**, 4952 (2020)
32. F. Sun, R. Wang, C. Aku-leh, H.X. Yang, R. He, J. Zhao, Sci. Rep. **4**, 6429 (2014)
33. G. Philippot, M. Albino, R. Epherre, G. Chevallier, Y. Beynet, C. Maniere, A. Weibel, A. Peigney, M. Deluca, C. Elissalde, M. Maglione, C. Aymonier, C. Estourners, Adv. Electron. Mater. **1**, 1500190–1500198 (2015)
34. A. Wei, L.T. Hui, W.C. Hui, D.C. Ling, L.N. Neng, L. Yong, Q.Z. Ming, S. Tao, W.Y. Yin, J. Huan, T.G. Shan, J.X. Ping, Acta Phys. Chim. Sin. **31**, 1059–1068 (2015)
35. C.H. Perry, D.B. Hall, Phys. Rev. Lett. **15**, 700–702 (1965)
36. V.V. Lemanov, Phys Colid State **39**, 318–322 (1997)
37. A.S. Deepa, S. Vidya, P.C. Manu, S. Solomon, A. John, J.K. Thomas, J. Alloys Compd. **509**, 1830–1835 (2011)
38. J.D. Freire, R.S. Katiyar, Phys. Rev. **37**, 2074–2085 (1988)
39. Y. Leyet, R. Pena, Y. Zulueta, F. Guerrero, J.A. Rivera, Y. Romagura, J. Perez de la Cruz, Mat. Sci. Eng. B **177**, 832–837 (2012)
40. L. Zhang, M. Zhang, L. Wang, C. Zhang, Z. Zhang, Y. Yao, L. Zhang, D. Xue, X. Lou, X. Ren, Appl. Phys. Lett. **105**(16), 2908 (2014)
41. F. Auzel, Chem. Rev. **104**, 139–173 (2004)
42. D. Peng, L. Laihui, L. Weiping, Y. Qingying, C. Hongbing, Appl. Phys. Lett. **104**, 152902 (2014)

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.

Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

End-to-End Convolutional Neural Network Feature Extraction for Remote Sensed Images Classification

Abebaw Alem & Shailender Kumar

To cite this article: Abebaw Alem & Shailender Kumar (2022) End-to-End Convolutional Neural Network Feature Extraction for Remote Sensed Images Classification, Applied Artificial Intelligence, 36:1, 2137650, DOI: 10.1080/08839514.2022.2137650

To link to this article: https://doi.org/10.1080/08839514.2022.2137650

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 30 Oct 2022.

Submit your article to this journal

Article views: 50

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uaai20

APPLIED ARTIFICIAL INTELLIGENCE
2022, VOL. 36, NO. 1, e2137650 (3328 pages)
https://doi.org/10.1080/08839514.2022.2137650

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

End-to-End Convolutional Neural Network Feature Extraction for Remote Sensed Images Classification

Abebaw Alem [a,b] and Shailender Kumar [a]

[a]Department of Computer Science and Engineering, Delhi Technological University, Delhi, India;
[b]Department of Computer Science and IT, Debre Tabor University, Debre Tabor, Ethiopia

ABSTRACT

Recently, land cover and land use (LCLU) classification in remote sensing imagery has attracted research interest. The LCLU contains dynamic remote sensed images due to sensor technology ability, seasonal changes, and distance for resolution. Therefore, the deep learning-based LCLU classification system needs more investigation using deep learning techniques. Deep learning approaches have gotten more attention for their powerful performance improvements. Most recent studies have been performed on deep convolutional neural networks (CNNs) that have been trained on pre-trained networks in remote sensing classification. However, designing CNNs from scratch has not yet been widely investigated in remote sensed images as they need ample training time and a powerful processor. Therefore, we used hyperparameters and early stopping techniques to apply an end-to-end CNN feature extractor (CNN-FE) model for LCLU classification in the UC-Merced dataset. We approved the model's applicability in the domain area by retraining it on another dataset called SIRI-WHU and building the VGG19 pre-trained feature extractor model built on the same hyperparameters. The CNN-FE has outperformed the state-of-the-art baseline studies' accuracy and the VGG19 pre-trained model. Moreover, a better CNN-FE performance was achieved when trained in the UC-Merced dataset than the model performance when trained in the SIRI-WHU dataset.

ARTICLE HISTORY
Received 11 April 2022
Revised 6 June 2022
Accepted 13 October 2022

Introduction

Land cover and land use (LCLU) classification contains dynamic remote sensed (RS) images that are inconsistent data due to the ability of sensor technologies, variations in annual seasons, and the distance for resolution. RS is the art and science of extracting information about an object/phenomenon without making physical contact using advanced sensing technologies. The RS technologies produce many RS images daily, and they could be collected from the earth's environment or space. Sensing technologies (Wang et al. 2021) are remote sensors used to collect large amounts of RS images from the observed earth environment.

CONTACT Abebaw Alem abebale@dtu.edu.et Department of Computer Science and Engineering, Delhi Technological University, Delhi 110042, India

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.
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.

Theland is one of the four pillars of sustainable development (social, human, economic, and environment or land). Therefore, managing, controlling, and planning the land could be critical for development. It could be better to support the tasks in machine-aided based LCLU classification systems. Thus, the LCLU classification problem is becoming the recent focal point research area in RS images (Du et al. 2019; Fan et al. 2020; Kang et al. 2022; Li et al. 2020; Sang et al. 2020; Radamanthys Stivaktakis, Tsagkatakis, and Tsakalides 2019).

Therefore, the LCLU classification problem in RS images could be solved by proposing the deep learning (DL) approach. DL is a robust recent machine learning (ML) approach that enables performance improvement for RS images (Kang et al. 2022; Sang et al. 2020; Scott et al. 2017; Shao, Yang, and Zhou 2018). Convolutional neural networks (CNNs) are prevalent DL techniques that consist of more than two layers (Chen et al., 2019) involving convolution filters. Convolution is the weighted sum of pixel values of the RS images. The purpose of using convolution is to reduce the size of the input image shape and the total number of parameters in the network (Senecal, Sheppard, and Shaw 2019). Therefore, we propose the convolutional neural network feature extractor (CNN-FE) DL model using the convolutional feature extractor technique and other DL hyperparameters for image feature extraction that will be described in section 3.1.

Nowadays, CNN methods get more civility in RS image classification problems for their powerful performance improvements. Previous studies, such as (Bahri et al. 2020; Dong et al. 2020; Liang, Deng, and Zeng 2020; Lin et al. 2019; Mateen et al. 2019; Qian et al. 2020; Zheng et al. 2022) have approved the performance improvements of the deep CNN methods in RS images. The CNNs DL approach consists of three main layers: convolutional, pooling, and fully connected (Cheng et al. 2017; Long et al. 2017). We used various optimization techniques in each of these layers. Thus, CNNs make different convolution processes from the input to the full CNNs. This process makes end-to-end predictions (Kang et al. 2022).

Deep CNNs are pillars and efficient end-to-end approaches for outstanding results in computer vision trends, specifically in RS image classifications. The CNN models have powerful feature extraction capability for classification improvement in RS images (Liu et al. 2019). Similarly, the end-to-end approaches can also improve the performance, as illustrated by Fan et al. (2020), Li et al. (2020), and Peng et al. (2019). The CNN-FE is an end-to-end learning approach that extracts the image features from the input to the output processes without using other feature extractor algorithms.

Deep CNN models could build RS images in three ways: creating from scratch, using pre-trained models, and retraining the pre-trained models. The pre-trained models are trained from earlier trained models on other large datasets such as “imagenet” images (Russakovsky et al. 2015). Training DL

models using a pre-trained network could have limitations for classifying RS images since the RS images have inconsistent properties compared to the "imagenet" images. Moreover, according to (Nogueira, Penatti, and Santos 2016), the pre-trained based CNN models could have limitations in extracting RS image features due to the properties of the natural images and the RS images. Therefore, building deep CNNs from scratch could resolve such constraints.

However, training deep CNNs from scratch has not been widely investigated in RS yet. This could be why building CNN models from scratch is difficult due to the lack of comprehensive training data and the significant amount of time needed for training (Bosco, Wang, and Hategekimana 2021; Yin et al. 2017). As we reviewed earlier studies, very few recent studies (Helber et al. 2019; Radamanthys Stivaktakis, Tsagkatakis, and Tsakalides 2019) have tried to develop CNNs from scratch for RS image analysis. Nevertheless, more studies are required to build CNN models from scratch. We took (Radamanthys Stivaktakis, Tsagkatakis, and Tsakalides 2019) and (Helber et al. 2019) earlier studies as our state-of-the-art baseline. We have created an end-to-end CNN-FE DL model by deploying new regularization and early stopping techniques, as shown in Table 1.

The motivation of this paper is to apply DL approaches, especially CNNs, by deploying regularization and early stopping techniques. This paper is also motivated by the state-of-the-art baseline earlier studies that have been studied by Stivaktakis, Tsagkatakis, and Tsakalides (2019) and Helber et al. (2019), considering their limitations as stated in Table 1.

Therefore, the objective of this study is to apply the deep CNN-FE approach built from scratch for LCLU classification in RS images and improve the state-of-the-art baseline study performances. Moreover, approving the developed model in the cases of developing the DL pre-trained network and retraining the developed model on the other RS dataset is another objective of this paper. To achieve these objectives, we preprocessed the datasets and trained the model with 17 layers (four Conv2D, four-batch normalization, four pooling,

Table 1. Hyperparameter settings compared with earlier comparative studies.

DL Hyperparameters	Chosen values for each DL hyperparameters in both earlier and our studies		
	(Helber et al. 2019)	CNN (Radamanthys Stivaktakis, Tsagkatakis, and Tsakalides 2019)	CNN-FE (Ours)
Optimizers	Stochastic gradient descent (SGD)	Adagrad	Adam
Batch size	16	10	64
Learning rate	le-3	-	0.0001
Iteration Epochs	120	300	100
Loss function	categorical_cross_entropy	binary_cross_entropy	categorical_cross_entropy
Activation functions	Relu	Relu, sigmoid	Relu, softmax
Dropout	-	0,0.25, 0.50, 0.75	0.5
Early stopping	-	-	Automatically stopping

one dropout, one flatten, three dense including the output (softmax) at the top) as figured in Figure 1. Finally, we evaluated the model with test set samples and compared the results from the state-of-the-art baseline studies and the VGG19 pre-trained model, as described in section 4.

Our significant contributions to this paper are:

(1) The significance of applying the DL method, CNN-FE, for the LCLU classification problem using the UC-Merced RS dataset from scratch development using the convolutional feature extractor, early stopping technique, and hyperparameters regularizations;
(2) Evaluating and comparing the performance of the CNN-FE model using evaluating metrics and improving the state-of-the-art baseline studies performance in the target UC-Merced dataset; and
(3) Approving the CNN-FE model by comparing its performance in building the VGG19 pre-trained model and retraining the CNN-FE model in another dataset called SIRI-WHU, which has different properties from the target UC-Merced dataset.

Related Work

LCLU contains dynamic sources of information on the observed earth surfaces because the nature of RS images is very complex. The LCLU classification in RS images has been a recent comprehensive study area (Dong et al. 2020). The deep CNNs could be applied to various domains in RS imagery data. RS image

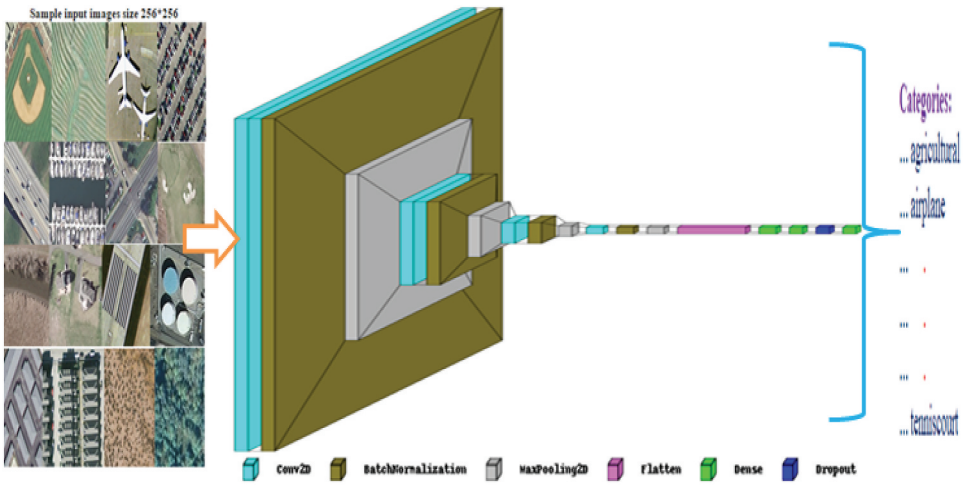


Figure 1. Structure of CNN-FE model.

classifications (Chen, Hu, and Duan 2019; Helber et al. 2019; Huang, Wang, and Li 2019; Liang, Deng, and Zeng 2020; Stivaktakis, Tsagkatakis, and Tsakalides 2019), and object detections (Hou et al. 2019; Hu et al. 2019; Long et al. 2017; Pang et al. 2019) are some of such applications. RS image classification is one of the application domains in computer vision (Shabbir et al. 2021). This is a challenging problem in large datasets with many classes and different conditions (Alhichri et al. 2021). Moreover, DL models are also powerful for validating feature extraction capabilities in computer vision (Gong et al. 2022). Thus, we propose DL methods for this challenging problem, which needs more investigation.

Despite the prominent results of deep CNNs, there are some problems to be solved regarding selecting fit hyperparameters and dataset sampling. Recent studies have shown that varying the hyperparameters affects the model's performance, such as (Long et al. 2017; Zheng et al. 2022). These hyperparameters, the kernel size (Chen, Hu, and Duan 2019; Peng et al. 2019), dropout (Li, Zhang, and Zhu 2019; Stivaktakis, Tsagkatakis, and Tsakalides 2019), and learning rate (Li, Zhang, and Zhu 2019; Long et al. 2017; Zhang et al. 2020), could affect the performance and produce different performance results. In addition to the DL hyperparameters, studies such as (Bosco, Wang, and Hategekimana 2021; Helber et al. 2019; Liang, Deng, and Zeng 2020; Zhang et al. 2020) have also tried to investigate the influences of the train-test dataset splitting percentages on the CNN models by differentiating train-test sampling sizes. In general, using various hyperparameters influences the DL model performances. Thus, we incorporated such reliable hyperparameters with their valuable values in this paper.

As we explained in the introduction section earlier, many researchers have tried to investigate the DL methods using pre-trained feature extraction architecture. Nevertheless, very few studies, such as those (Helber et al. 2019; Stivaktakis, Tsagkatakis, and Tsakalides 2019), have tried to create CNN models in RS image classification from scratch.

The study (Stivaktakis, Tsagkatakis, and Tsakalides 2019) has studied the CNN model by focusing on data augmentation and dropout hyperparameters and training the model in 1600 training sets, and testing the performance of the model in 500 test sets. However, this study's sigmoid activation function and binary-cross-entropy loss function could influence the performance since these hyperparameters are recommended for binary classification rather than multiclass classifications, as the UC-Merced dataset is a multiclass classification problem. As we observed from the reviewed literature, the DL hyperparameters significantly influence the model's performance.

The study (Helber et al. 2019) has also investigated the Bag-of-Visual-Words (BoVW), CNN (with three layers), ResNet-50, and GoogleNet models by varying the training-test split set ratio starting from 10/90 to 90/10 on EuroSAT, UC-Merced, AID, Sat-6 and BCS dataset. Still, the hyperparameter

optimization technique is required to increase the CNN model performance that is built from scratch. We took only the CNN model accuracy result trained on the UC-Merced dataset for performance comparison in our paper, as shown in Table 8.

Thus, considering these state-of-the-art studies' limitations, we built the CNN-FE model from scratch to LCLU classification in RS images and compared its performance with state-of-the-art studies. To fight over-fitting and increase the CNN-FE performance, we deployed new optimal regularizations and early stopping techniques implemented in this paper, as shown in Table 1.

Materials and Methods

This paper proposed to build the CNN-based feature extraction (CNN-FE) method from scratch for the LCLU classification problem in the inconsistent RS images using various DL layers, as shown in Figure 1. To compare the performance and approve the applicability of the CNN-FE model, we built the VGG19 pre-trained model. We have used two datasets, namely, the UC-Merced dataset, which is our target dataset for building the model, and the SIRI-WHU dataset used for model approval purposes. The DL Keras and Tensorflow open-source software were also the essential materials we used for experimental executions.

Convolutional Neural Networks

In this paper, we used the most prominent DL approach CNNs in the form of Conv2D, which took the image shape (height, width, channel), i.e. (256, 256, 3). The CNNs are multi-layer neural networks used to extract image features or pixels. The CNNs consist of convolution, pooling, and fully connected layers integrated with other DL hyperparameters. We described each sequential process for the end-to-end DL approach in the following and depicted in Figure 1.

The Input Layer and Convolutional (Conv2d) Layers

The input layer is the entire input image layer with height*width*channel pixel shapes. It is introduced into the convolutional layer to be processed. Convolutional layers then receive the input layers and image pixels. The convolutional layers compute the perceptron with a given filter (f, f), pooling, stride, and padding to transform the input image volume into a new output volume.

The CNNs can successfully capture deep spatial feature representations for RS scene classification (Zeng, Chen, and Song 2021) with convolution. The CNN convolution could operate the mathematical operation of matrix multiplications in the given layers, and every image is represented in the form of an

array of values or pixels. In convolutional operations, the arrays are multiplied pixel-wise, and the product is summed to create a new array or feature map representing **height_new * width_new** based on Equation (1) computations. The CNNs are different from other conventional ML approaches in input data types and weight calculations (Kim et al. 2018) with the convolution method. The overall process of the convolutional layers makes the model's feature map.

The feature map of the model can be transformed into other resolution feature maps using the downsampling and upsampling techniques. The downsample is a convolution operation with strides to reduce the input image size and double the number of filters. In contrast, upsampling is a bilinear interpolation operation to double the input image size and reduce the number of filter sizes (Wang et al. 2021).

The convolutional layers consist of convolution filters or kernels with learnable parameters (Maggiori et al. 2016; Peng et al. 2019). Convolution could be performed with valid convolution (no padding), same convolution (with padding), and stride (slide or shift) convolution. The mathematical computation of the output volume of the image in each layer could be calculated using the input volume (height*width), stride(s), and padding (p) parameters. Stride (s) of the filter ($f \times f$) is the interval of the filter jumps/shifts s number of transitions from the first elements in a pixel or each spatial dimension. At the same time, padding (p) is the number of pixels added at the outer edges of the input image volumes (height × width). A filter is usually odd, and smaller in size is 1×1, 3×3, 5×5, and 7×7 with 0, 1, 2, and 3 paddings, respectively. In the keras DL tool, there is no padding for image border (0) to valid convolution and padding for image border to same convolution. Thus, the output volume ($height_{new}$*$width_{new}$) of a layer could be computed using (1), and the number of padding for same convolution could be calculated using (2). The default values of p and s are 0 and 1, respectively.

In this paper, we use the same convolution with the filter size (3,3) and three paddings. The Conv2D layers are used to extract the input image features by sliding a convolution filter size of (3, 3) to produce a new output hierarchical feature map. There are four convolutional block layers in our sequential model training, including 64, 128, and 256 convolution kernels with a filter size of three each. The convolution activates image features using convolutional filters or kernels that denote the weight matrix. Therefore, convolution is used as our feature extraction method for RS images.

$$\text{height_new} = \text{width_new} = [(\text{height_new} - \text{f} + 2\text{p})/\text{s}] + 1 \tag{1}$$

$$p = (f - 1)/2 \tag{2}$$

In the convolution process, the number of parameters (params) could be calculated with (3) and (4). The total parameter numbers of the model are the summations of the calculated results from Conv2D and Dense layers. We

design the model with four Conv2D layers that calculate the number of parameters for those layers in the same norm by (3) and two dense layers (4). However, the calculation formula of dense parameters differs from Conv2D as equated in (4). Number 1 refers to the bias associated with each filter for learning.

$$\text{Conv2Dparam\#} = \#\text{outputchannel} * (\#\text{inputchannel} * \text{filterheight} * \text{filterwidth} + 1) \tag{3}$$

$$\text{Denseparam\#} = \#\text{outputchannel} * (\#\text{inputchannel} + 1) \tag{4}$$

Pooling Layer

The pooling layer is used to resize and downsample the spatial representations, followed by convolutional operations. We use a common pooling technique called max pooling. Max pooling could pick the most activated feature and could be used to reduce overfitting and reduce the number of parameters.

Therefore, the pooling layers in CNNs are essential for downsampling processes used to reduce the size of the input RS images. In addition, the block layers involve various max-pooling with two, the stride with two, and the padding with "same."

Fully Connected Layers (FCNs)

FCNs are feature classifiers in the last couple of layers of the network. They include flatten layers, dense layers, and an output layer at the end. Perceptrons in an FCN are fully connected to all activations of the previous layer.

The CNNs also have various activation functions, which should be nonlinear as linear functions have a constant derivative, as described in our previous work (Alem and Kumar 2021). These are softmax, rectified linear unit (Relu), hyperbolic tangent (tanh), and sigmoid or logistic functions. We use the result at the entire convolutional layers to activate the weights in each convolution process and the softmax at the output layer since it is reliable for our multiclass classification problem. The softmax function is a feature classifier, and it introduces a probability score for each class. The highest probability score among each class is predicted as our predicted class. This probability score will be used for performance evaluations in section 4.

Dataset Descriptions

The RS dataset has been collected through advanced sensor technologies, and then they could label manually for research or other commercial purposes. We used the common UC-Merced dataset and the rarely investigated SIRI-WHU dataset to check the possible applicability of the built model on the target UC-

Merced. We divided both datasets into 60%, 20%, and 20% for training, validation, and tasting samples for each labeled class.

The UC-Merced dataset is an LCLU data set collected from the earth, labeled manually, and introduced by (Yang and Newsam 2010). It contains 21 classes with 100 images each, measuring 256×256 pixels with a spatial resolution of about 30 cm per pixel. However, the UC-Merced dataset is inconsistent as about 44 images have different pixel shapes. The variety of properties of the dataset could affect the performance results. Sample images in each class are depicted in Figure 2. This dataset is available at http://weegee.vision.ucmerced.edu/datasets/landuse.html.

The SIRI-WHU dataset was collected from Google Earth that covered urban areas in China, and it was introduced by (Zhao et al. 2016). The dataset contains 12 categories and 200 images per category with 200×200 pixels in a spatial resolution of 200 cm per pixel. Sample images in each category are depicted in Figure 3. The dataset is publicly available for research purposes at https://figshare.com/articles/dataset/SIRI_WHU_Dataset/8796980.

Experimental Results and Discussions

Our experiment was executed with an Intel Core i3-4000 M CPU 2.40 GHz RAM = 4GB laptop personal computer integrated with the Collaboratory on its Tesla T4 GPU. Keras and Tensorflow open-source DL software packages were used for this experiment.

Experimental Setting

The dataset and the DL hyperparameters could be considered for their appropriate settings to build our model. As we described in section 3.2, there are

Figure 2. Sample images in each classes of the UCM dataset.

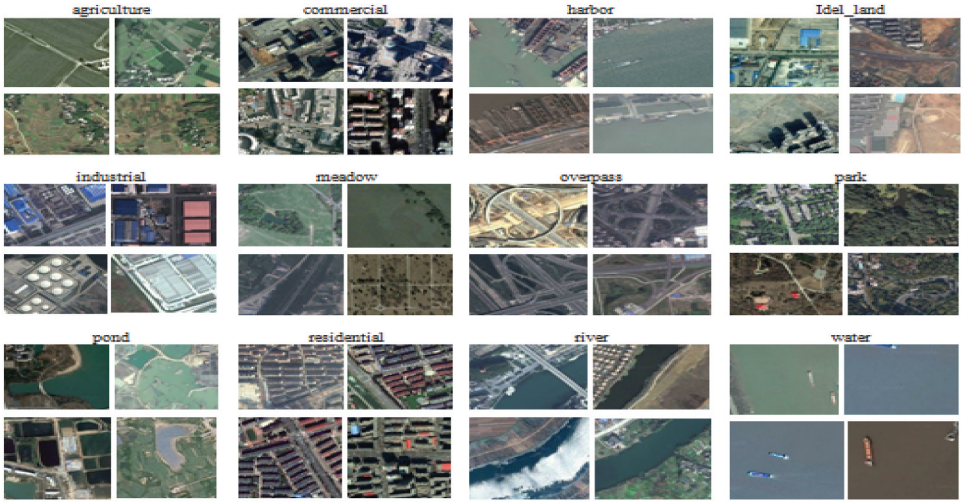


Figure 3. Sample images in each class of the SIRI-WHU dataset used for CNN-FE model checking.

2100 images in the UC-Merced dataset and 2400 images in the SIRI-WHU dataset. Therefore, to reduce the overfitting of the model, we split both the UC-Merced and SIRI-WHU datasets into three sets: training set, validation set, and test set in 60%, 20%, and 20% of the dataset, respectively. Then after splitting, the total sample images in the training set, validation set, and test set becomes 1260, 420, and 420 for UC-Merced and 1440, 480, and 480 for SIRI-WHU, respectively. Each dataset is loaded into the experiment and preprocessed. First, we built the model on the UC-Merced dataset as follows: then, we rebuilt the model on the SIRI-WHU dataset for its applicability approval in the same manner.

The training set is a collection of 1260 images that have been used to fit and train our model with a batch size of 64 and hundreds of epochs, as shown in Table 1 (right column). In each epoch, the same training images are fed to CNN-FE architecture recurrently, and the model could learn and continue to learn from the hidden image features. In general, the model has been trained on a training set in four CNNs sequential layers and its performance has been evaluated with the validation set during the training and with test set after the training.

The validation set is a collection of 420 images separate from the training set that was used to validate our model performance during the training. Splitting the dataset into a validation set is critical in reducing the overfitting of training and evaluating the model during its development.

On the other hand, the test set is a set of 420 images used to evaluate the performance of our model after completing the training. The test set is the support as shown in the last column of Tables 2–5. It is used to analyze the

Table 2. Summarization of the classification performance of CNN-FE for each class with performance measurement metrics in the UC-Merced dataset.

Class name	Precision	Recall	F1-score	Support
Agricultural	0.91	1.00	0.95	20
Airplane	0.94	0.80	0.86	20
Baseballdiamond	0.85	0.85	0.85	20
Beach	0.95	1.00	0.98	20
Buildings	0.79	0.75	0.77	20
Chaparral	1.00	0.95	0.97	20
Denseresidential	0.57	0.80	0.67	20
Forest	0.86	0.95	0.90	20
Freeway	0.95	0.90	0.92	20
Golfcourse	0.80	0.60	0.69	20
Harbor	0.91	1.00	0.95	20
Intersection	0.88	0.75	0.81	20
Mediumresidential	0.58	0.90	0.71	20
Mobilehomepark	0.92	0.55	0.69	20
Overpass	0.81	0.85	0.83	20
Parkinglot	1.00	0.85	0.92	20
River	0.78	0.90	0.84	20
Runway	0.95	0.90	0.92	20
Sparseresidential	0.75	0.90	0.82	20
Storagetanks	1.00	0.90	0.95	20
Tenniscourt	1.00	0.70	0.82	20

performance evaluation metrics, including accuracy, loss, precision, recall, F1-score, and confusion matrix, as we described in section 4.2.

In addition to setting the dataset splitting, we have chosen the DL hyperparameters to build, compile, and fit our model on the UC-Merced dataset and evaluate the model's performance as shown in Table 1. Both dropout and early stopping hyperparameters are used for reducing over-fitting. Early stopping is a technique that could automatically stop the training when either validation

Table 3. Summarizations of the classification performance of VGG19 for each class in performance measurement metrics in the UC-Merced dataset.

Class name	Precision	Recall	F-score	Support
Agricultural	1.00	1.00	1.00	20
Airplane	0.95	0.90	0.92	20
Baseballdiamond	1.00	0.90	0.95	20
Beach	1.00	0.95	0.97	20
Buildings	0.82	0.70	0.76	20
Chaparral	1.00	1.00	1.00	20
Denseresidential	0.50	0.55	0.52	20
Forest	0.82	0.90	0.86	20
Freeway	1.00	0.85	0.92	20
Golfcourse	0.86	0.60	0.71	20
Harbor	1.00	1.00	1.00	20
Intersection	0.81	0.85	0.83	20
Mediumresidential	0.69	0.90	0.78	20
Mobilehomepark	0.61	0.55	0.58	20
Overpass	0.84	0.80	0.82	20
Parkinglot	0.95	0.95	0.95	20
River	0.76	0.95	0.84	20
Runway	0.87	1.00	0.93	20
Sparseresidential	0.90	0.90	0.90	20
Storagetanks	0.95	0.95	0.95	20
Tenniscourt	0.85	0.85	0.85	20

Table 4. Summarizations the classification performance of CNN-FE for each individual class with performance measurement metrics in SIRI-WHU dataset.

Class name	Precision	Recall	F1-score	Support
Agriculture	0.81	0.65	0.72	40
Commercial	0.94	0.82	0.88	40
Harbor	0.88	0.75	0.81	40
Idle_land	0.62	0.78	0.69	40
Industrial	0.90	0.90	0.90	40
Meadow	0.55	0.60	0.57	40
Overpass	0.92	0.90	0.91	40
Park	0.79	0.68	0.73	40
Pond	0.72	0.82	0.77	40
Residential	0.83	0.95	0.88	40
River	0.81	0.75	0.78	40
Water	0.93	1.00	0.96	40

Table 5. Summarizations of the classification performance of VGG19 for each class with performance measurement metrics in the SIRI-WHU dataset.

Class name	Precision	Recall	F1-score	Support
Agriculture	0.91	0.50	0.65	40
Commercial	0.72	0.95	0.82	40
Harbor	0.88	0.93	0.90	40
Idle_land	0.86	0.62	0.72	40
Industrial	0.85	0.88	0.86	40
Meadow	0.63	0.60	0.62	40
Overpass	0.80	0.97	0.88	40
Park	0.50	0.60	0.55	40
Pond	0.67	0.75	0.71	40
Residential	0.92	0.88	0.90	40
River	0.85	0.72	0.78	40
Water	0.98	1.00	0.99	40

loss has stopped decreasing or validation accuracy has stopped increasing. In addition to these techniques, the convolutional techniques were applied to preprocess and extract feature maps by reducing the image shape (256, 256, 3) into other reduced feature maps.

Performance Evaluation Metrics and Experimental Results

After validating the model using the validation set during training, we retrained it by combining the training set and validation set with an early stopping technique. Hereafter, the training set sample images become 80% of the dataset for validating the model's performances with 20% of the test set sample images.

After building the model, we evaluated its performance using the evaluation measurement metrics of accuracy, precision, recall, F1-score, and confusion or error matrix (CM). In addition to these evaluation metrics, we used the loss function, i.e., the categorical_cross_entropy, to evaluate the training and validation errors. The training losses are calculated during each epoch, whereas validation losses are computed after each training epoch for the

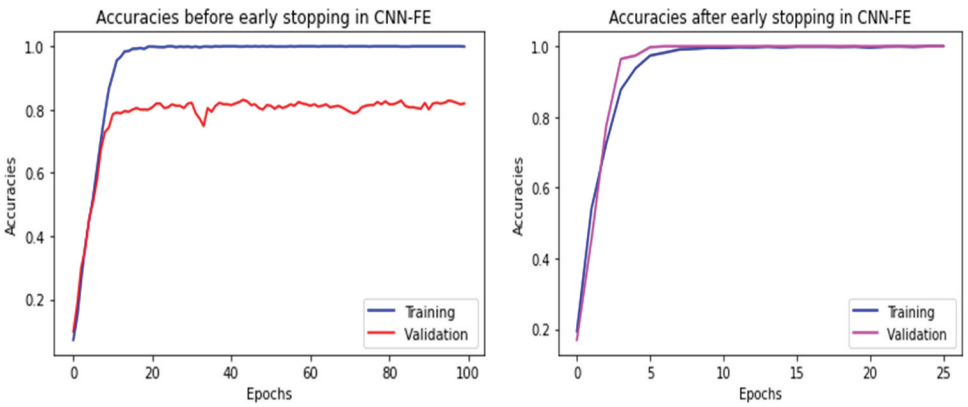
errors. At most, when the number of epochs increases, the losses decrease, and the accuracies increase.

The model's accuracy was evaluated in two ways, i.e., with and without using the early stopping technique. The early stopping has been stopped at a random iteration epoch from out of 100 epochs when either the validation accuracy has been stopped increasing (as depicted in Figures 4b, 5b, 6b, and 7b) or the validation loss stopped decreasing (as depicted in Figures 8b, 9b, 10b, and 11b) while evaluating the models with test set sample images. Therefore, from the experiments with and without early stopping, we observed that the accuracy result increased using the early stopping technique in each model of CNN-FE and VGG19 trained on both datasets, as shown in Table 7.

In addition to evaluating the overall accuracy of both models, we assessed each class with 20 sample images per class using precision, recall, and F1-score performance measurement merits as stated in Tables 2–5. Furthermore, the CM metric was also used to identify the predicted classes based on the higher normalized probability values in each class intersection. CM considers the normalized probability values for each class category in rows (True labeled class) and columns (predicted labeled class), as shown in Figures 12–15. CM measures the performance of the DL model, whether each class is correctly classified or incorrectly classified. Therefore, according to Figures 12–15, the score in the diagonal intersection showed the correct classified classes with higher normalized probability. In contrast, the results in other rows-columns wise are predicted in misclassified classes with lower normalized probability.

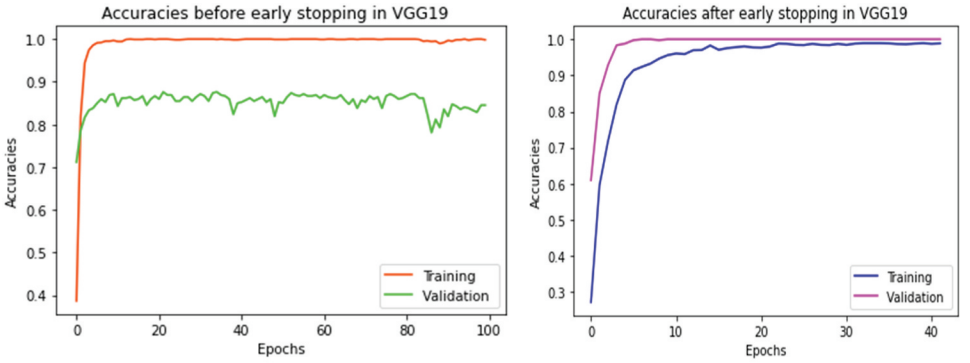
Model Validations with VGG19 Pre-Trained Network and SIRI-WHU Dataset

After building and evaluating the CNN-FE model, we assured its possible applicability to the LCLU classification in RS images by comparing its



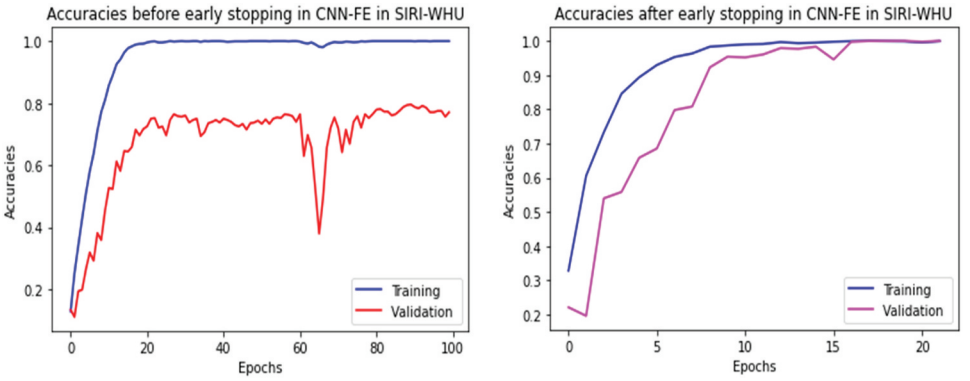
a) Before applying early stopping b) After applying early stopping

Figure 4. Training and validation accuracies with and without applying early stopping technique. (a) Before applying early stopping. (b) After applying early stopping.



a) Before applying early stopping b) After applying early stopping

Figure 5. Training and validation accuracies in VGG19 with and without applying early in stopping technique in UC-Merced dataset. (a) Before applying early stopping. (b) After applying early stopping.

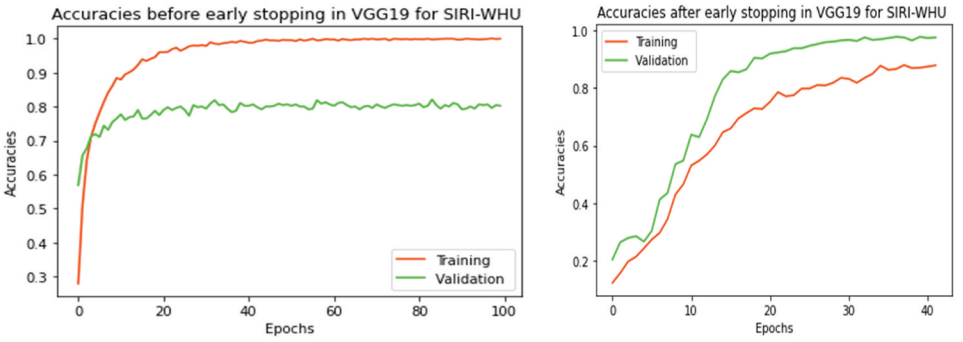


a) Before applying early stopping b) After applying early stopping

Figure 6. Training and validation accuracies of CNN-FE model in SIRI-WHU dataset with and without applying early stopping technique. (a) Before applying early stopping. (b) After applying early stopping.

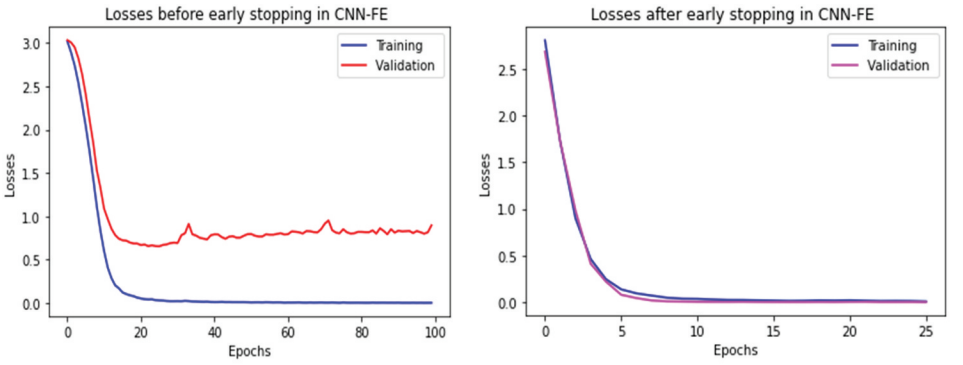
performance with the VGG19 feature extractor network and retrained on another dataset called SIRI-WHU.

The VGG19 pre-trained feature extractor was trained on the pre-trained network, which was trained on the large dataset "imagenet" in the same hyperparameters to check the applicability of CNN-FE for LCLU classification in RS images. The VGG19 was designed by (Simonyan and Zisserman 2015) to analyze the neural network depth effect on the accuracy of image recognition. Therefore, we created the VGG19 pre-trained model to compare its performance with CNN-FE trained on UC-Merced and SIRI-WHU. While comparing the accuracy performances of both DL models, CNN-FE outperformed VGG19 as compared in Table 7. Using the early stopping technique improved



a) Before applying early stopping b) After applying early stopping

Figure 7. Training and validation accuracies of VGG19 in the SIRI-WHU dataset with and without applying the early stopping technique. (a) Before applying early stopping. b) After applying early stopping.

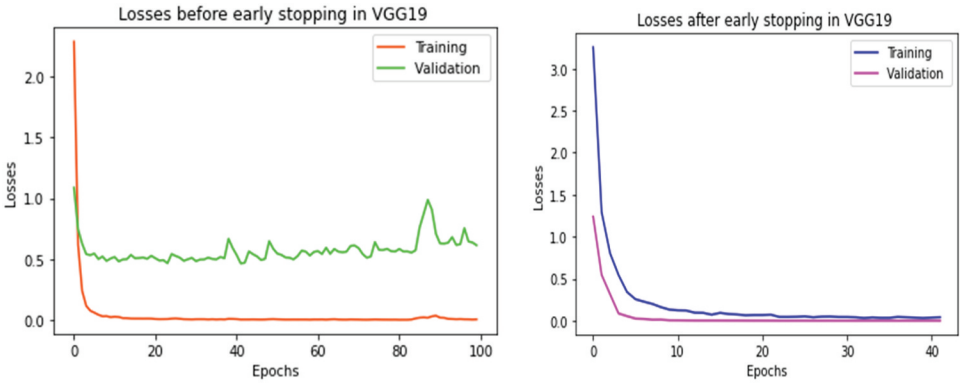


a) Losses before applying early stopping b) Losses after applying early stopping

Figure 8. Training and validation losses with and without applying early stopping technique. (a) Losses before applying early stopping. (b) Losses after applying early stopping.

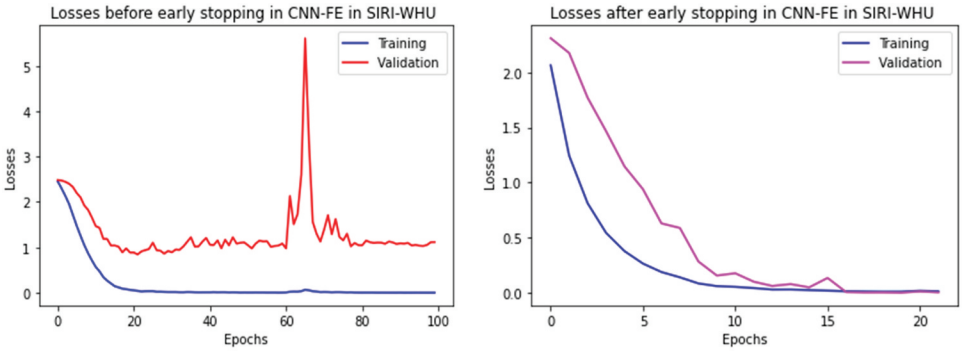
the accuracy performance of VGG19 in both datasets as CNN-FE, as shown in Table 7.

In addition to checking its applicability on the other DL pre-trained model, we retrained the CNN-FE model on the SIRI-WHU dataset. As we stated earlier, the properties of the dataset could influence the performance of DL models. We used the SIRI-WHU dataset with different properties from the target dataset UC-Merced to see this effect. After training the CNN-FE model on the SIRI-WHU dataset, the validation accuracy and loss fluctuated, especially between epochs 60 and 80 than the validation accuracy and loss trained in UC-Merced, as shown in Figures 6a and 10a.



a) Before applying early stopping b) After applying early stopping

Figure 9. Training and validation losses in VGG19 with and without applying the early stopping technique in the UC-Merced dataset. (a) Before applying early stopping. b) After applying early stopping.



a) Before applying early stopping b) After applying early stopping

Figure 10. Training and validation losses of CNN-FE model in SIRI-WHU dataset with and without applying early stopping technique. (a) Before applying early stopping. (b) After applying early stopping.

Discussions

This study investigated the application of an end-to-end DL approach called CNN-FE for LCLU classification using RS images. We showed the possibility of designing CNNs from scratch to develop LCLU classification in complex RS images using two different datasets. We also developed a comparative VGG19 pre-trained network using the same hyperparameters. In addition to validating this DL pre-trained model, we retrained the CNN-FE on the SIRI-WHU dataset and assured its applicability in the domain area. Therefore, as far as our knowledge, CNN-FE is significant in this study.

a) Before applying early stopping b) After applying early stopping

Figure 11. Training and validation losses of VGG19 in SIRI-WHU dataset with and without applying early stopping technique. (a) Before applying early stopping. (b) After applying early stopping.

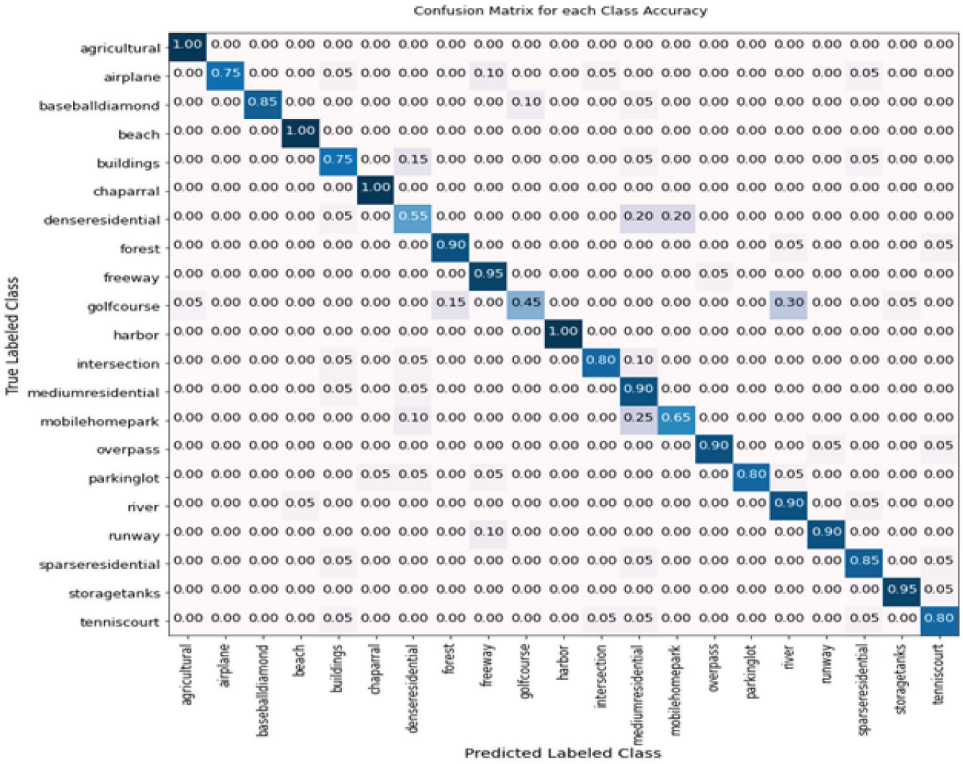


Figure 12. CM performance results in the CNN-FE model for each labeled class.

Discussions on Results

The CNN-FE model performance result indicates the possibility of building the DL models from scratch. It is comparable to those trained from pre-trained models in the UC-Merced dataset. Significant results were reported when

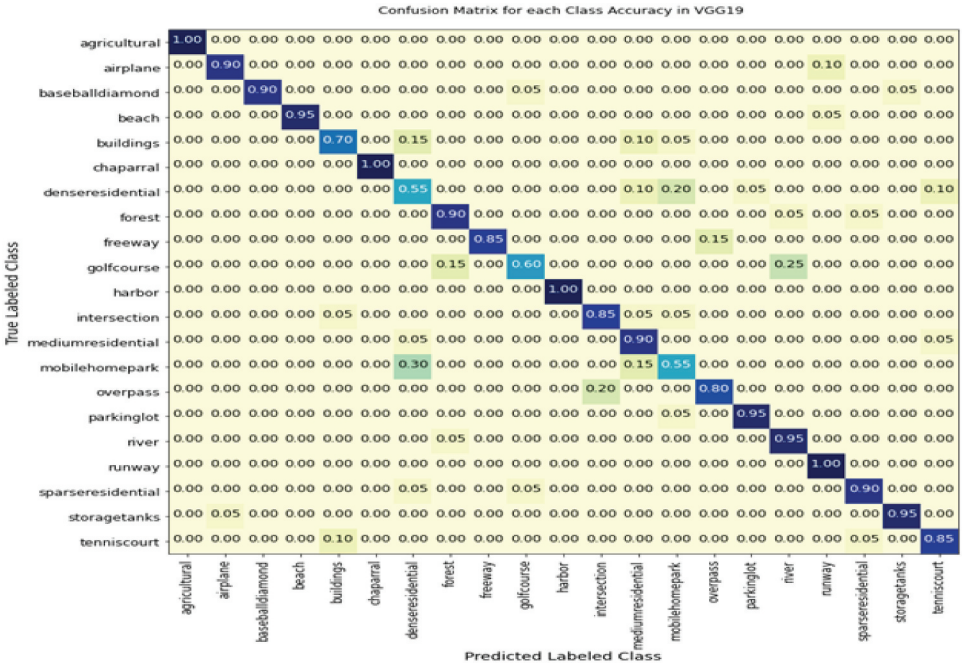


Figure 13. CM performance results of VGG19 pre-trained for each labeled class.

Figure 14. CM performance results of CNN-FE for each class classification in SIRI-WHU.

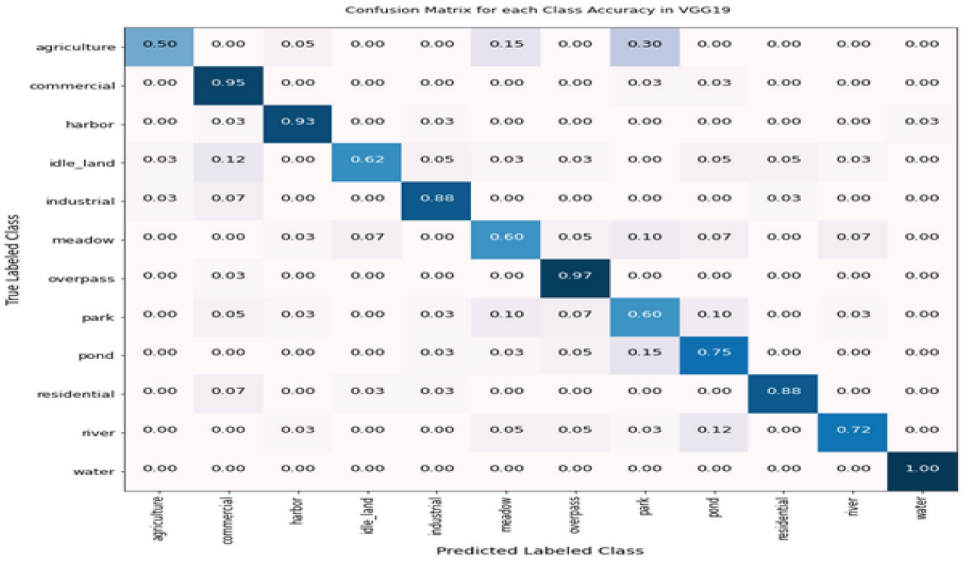


Figure 15. CM performance results of VGG19 for each class classification in SIRI-WHU.

compared with VGG19 pre-trained architecture. In addition, the CNN-FE was retrained on the SIRI-WHU dataset, and a considerable accuracy performance was achieved in the UC-Merced, as shown in Table 7. To sum up, the performance of the CNN-FE model resulting from various measurement metrics showed that it is possible to prove its applicability to the classification problem in RS images.

Each class classification performance was evaluated with precision, recall, and F1-score. Therefore, according to Table 2, the classes such as chaparral, parking lot, storage tanks and tennis court have the best precision performed, which means that these classes were precisely predicted. However, the lower result precisions were reported in dense residential (i.e., 0.57), which means that it has inflexible properties to predict precisely. The classes such as agricultural beach and harbor were classified in best recall performance, whereas mobile homepark class scored the lower recall performance. Classes with the perfect or lower performance in both precision and recall are also the perfect or lower results in F1-score. Thus, there were no same classes with perfect or lower performance in both metrics, and there were no perfect classes in F1-score. However, a lower F1-score was recorded in dense residential (0.67) class. Perfect performance means 100% accurately and precisely classified when measured in given metrics.

To sum up, the individual class performance of the two models in the two datasets is summarized in Table 6. The dense residential class has lower performance in both methods than other classes in the UC-Merced dataset, while the meadow and park classes have lower performance in CNN-FE and VGG19, respectively, in the SIRI-WHU dataset. In the case of CM metrics,

better result performance for each class has been observed in both methods in the UC-Merced dataset than in the SIRI-WHU dataset, as compared and shown in Figures 12–15.

The classes, such as agricultural, harbor, overpass, and river, are common in both datasets. However, most of these classes have different performance values, as shown in Tables 2–6. This could be why the two datasets have inconsistent properties, which were collected from different locations with different resolutions and pixel values.

In addition to evaluating the individual classes, we also evaluated the two methods within the two datasets. Thus, while comparing the CNN-FE from the pre-trained VGG19 network outperformed results in CNN-FE have been achieved in both datasets, as shown in Table 7.

Discussions on Similar Studies

In this paper, we aimed to improve the performance of the DL model from the existing state-of-the-art studies studied by Stivaktakis, Tsagkatakis, and Tsakalides (2019) and Helber et al. (2019) by considering their limitations of the DL hyperparameters. The DL hyperparameters influence the DL model performances. Therefore, to see the effect, we used various hyperparameters, such as dropout, learning rate, batch size, epochs, and early stopping, with their valuable values as stated in section 4.1. The study (Stivaktakis, Tsagkatakis, and Tsakalides 2019) has analyzed the dropout hyperparameter effects on the CNN performance with different values (null, 0.25, **0.50**, and 0.75), which generates the accuracy of 81.2, 81.3, **81.4**, and 79.7 with augmentation and 68.0, 73.7, **75.7**, and 77.7 without data augmentation technique, respectively. Among these provided accuracies and dropout values, we listed and compared the last two accuracy performances with unaugmented data with corresponding dropout values of 0.**5** and 0.75, respectively, shown in Table 8. The CNN-FE model has achieved 89.76% and 80% accuracy in the UC-Merced and the SIRI-WHU dataset, respectively. The VGG19 pre-trained model has also achieved 85.95% and 78.33% accuracy, as shown in Table 7. Moreover, the CNN-FE model outperformed the state-of-the-art studies and the pre-trained network, as shown in Table 8.

Conclusion

In this paper, we have applied the CNN-FE model built from scratch to address the challenge of LCLU classification in RS images. Although CNNs are powerful DL approaches to analyzing RS images for LCLU classification systems, designing CNN models from scratch has not been solved yet. It is crucial to build CNN models from scratch for RS images since RS images are

Table 6. Class comparisons in precision, recall, and F1-score (%) on the two models and datasets.

		Precision performance		Recall performance		F1-score performance	
Dataset	Method	Perfect (1) classes	Lower (-) classes	Perfect (1) classes	Lower (-) classes	Perfect (1) classes	Lower (-) classes
UC-Merced	CNN-FE	Chaparral, parkinglot, storagetanks, and tenniscour	Denseresidential (0.57)	Agricultural, beach and harbor	Mobilehomepark (0.55) and golfcourse (0.60)	None	Denseresidential (0.67)
	VGG19	Agricultural, baseballdiamond, beach, charparral, freeway, and harbor	Denseresidential (0.50)	Agricultural, charparral, harbor and runway	Denseresidential (0.55)	Agricultural, charparral, and harbor	Denseresidential (0.52)
SIRI-WHU	CNN-FE	None	Meadow (0.55)	Water	Meadow (0.60)	None	Meadow (0.57)
	VGG19	None	Park (0.50)	Water	Agriculture	None	Park (0.55)

Table 7. Results of accuracy (%) performances at random early stopping technique.

Dataset	Methods	Stopped at epoch # out of 100	Accuracy performance results	
			Before early stopping	After early stopping
UC-Merced	CNN-FE	26	85.95	**89.76**
	VGG19	42	85.00	85.95
SIRI-WHU	CNN-FE	22	78.67	**80.00**
	VGG19	41	76.88	78.33

Table 8. General comparison of the accuracy (%) with the state-of-the-arts in the UC-Merced target dataset.

Method	Dropout	Precision	Recall	F1-score	Accuracy
CNNs (Radamanthys Stivaktakis, Tsagkatakis, and Tsakalides 2019)	0.50	85.4	83.3	84.3	75.7
	0.75	85.5	85.8	85.7	77.7
CNN (Helber et al. 2019)	-	-	-	-	82.38
CNN-FE (Ours)	0.50	90.00	88.00	88.99	89.76
VGG19 (Ours)	0.50	88.00	86.00	86.99	85.95

inconsistent, and modeling them from pre-trained networks could affect the result.

Therefore, we applied an end-to-end CNN-FEDL model to extract the inconsistent UC-Merced RS image features for LCLU classification in RS images. We retrained this model on the other SIRI-WHU dataset to analyze whether the dataset influences the model performance. We also built a VGG19 pre-trained DL model on both datasets and evaluated their performances to validate the CNN-FE possible applicability in the domain area. We compare its result with the state-of-the-art earlier studies and the VGG19 pre-trained model, trained in the same hyperparameters. The CNN-FE outperformed the accuracy performance from state-of-the-art earlier studies and the VGG19 Pre-trained model. Therefore, we proved that the developed CNN-FE model is possibly applicable to the domain area and improved performance. However, the model needs more improvements, validations, and comparisons with other traditional machine learning approach in large datasets. Therefore, building DL and traditional ML approaches in other large RS datasets is our future task to compare their performances.

Acknowledgments

The Ethiopian Ministry of Education (MoE), cooperatively with Debre Tabor University, has provided a monthly stipend for this paper. Reviewing a manuscript is a challenging task in which the editors and reviewers share their deep and experienced knowledge with different authors from various disciplines. Therefore, we authors would like to thank the anonymous editors and reviewers of the Applied Artificial Intelligence International Journal (E-ISSN: 1087-6545) for their constructive reviews and comments by devoting their precious time to improving this quality paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Abebaw Alem http://orcid.org/0000-0001-8154-4648

References

Alem, A., and S. Kumar. 2021. Transfer learning models for land cover and land use classification in remote sensing image. *Applied Artificial Intelligence* 36 :2014192. doi:10.1080/08839514.2021.2014192.

Alhichri, H., A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan. 2021. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. *IEEE Access* 9 ():14078–94. doi:10.1109/ACCESS.2021.3051085.

Bahri, A., S. G. Majelan, S. Mohammadi, M. Noori, and K. Mohammadi. 2020. Remote Sensing Image Classification via Improved Cross-Entropy Loss and Transfer Learning Strategy Based on Deep Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters* 17 (6):1087. doi:10.1109/LGRS.2019.2937872.

Bosco, M. J., G. Wang, and Y. Hategekimana. 2021. Learning multi-granularity neural network encoding image classification using DCNNs for easter Africa community countries. *IEEE Access* 9 ():146703–18. doi:10.1109/ACCESS.2021.3122569.

Cheng, G., Z. Li, X. Yao, L. Guo, and Z. Wei. 2017. Remote sensing image scene classification using bag of convolutional features. *IEEE Geoscience and Remote Sensing Letters* 14 (10):1735–39. doi:10.1109/LGRS.2017.2731997.

Chen, D., P. Hu, and X. Duan. 2019. Complex scene classification of high resolution remote sensing images based on DCNN model. *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, 1–4. 10.1109/Multi-Temp.2019.8866895

Chen, Y., Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia. 2019. Deep learning ensemble for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (6):1882–97. doi:10.1109/JSTARS.2019.2915259.

Dong, S., Y. Zhuang, Z. Yang, L. Pang, H. Chen, and T. Long. 2020. Land cover classification from VHR optical remote sensing images by feature ensemble deep learning network. *IEEE Geoscience and Remote Sensing Letters* 17 (8):1396–400. doi:10.1109/LGRS.2019.2947022.

Du, P., E. Li, J. Xia, A. Samat, and X. Bai. 2019. Feature and model level fusion of pretrained CNN for remote sensing scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (8):2600–11. doi:10.1109/JSTARS.2018.2878037.

Fan, R., R. Feng, L. Wang, J. Yan, and X. Zhang. 2020. Semi-MCNN: A semisupervised multi-CNN ensemble learning method for urban land cover classification using submeter HRRS images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:4973–87. doi:10.1109/JSTARS.2020.3019410.

Gong, N., C. Zhang, H. Zhou, K. Zhang, Z. Wu, and X. Zhang. 2022. Classification of hyperspectral images via improved cycle-MLP. *Institutions of Engineering and Technology Computer Vision* 16 (5):468–78. doi:10.1049/cvi2.12104.

Helber, P., B. Bischke, A. Dengel, and D. Borth. 2019. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected*

Topics in Applied Earth Observations and Remote Sensing 12 (7):2217–26. doi:10.1109/JSTARS.2019.2918242.

Hou, B., J. Li, X. Zhang, S. Wang, and L. Jiao (2019). Object detection and tracking based on convolutional neural networks for high-resolution optical remote sensing video. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 5433–36. 10.1109/IGARSS.2019.8898173

Huang, W., Q. Wang, and X. Li (2019). Feature sparsity in convolutional neural networks for scene classification of remote sensing image school of computer science and center for optical imagery analysis and learning (OPTIMAL). *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 3017–20. 10.1109/IGARSS.2019.8898875

Hu, Y., X. Li, N. Zhou, L. Yang, L. Peng, and S. Xiao. 2019. A sample update-based convolutional neural network framework for object detection in large-area remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 16 (6):947–51. doi:10.1109/LGRS.2018.2889247.

Kang, W., Y. Xiang, F. Wang, and H. You. 2022. DO-net : Dual-output network for land cover classification from optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 19 (Art no. 8021205):1–5. doi:10.1109/LGRS.2021.3114305.

Kim, M., J. Lee, D. Han, M. Shin, J. Im, J. Lee . . ., and L. J. Quackenbush, Z. Gu. 2018. Convolutional neural network-based land cover classification using 2-D spectral reflectance curve graphs with multitemporal satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11 (12):4604–17. doi:https://doi.org/10.1109/JSTARS.2018.2880783.

Liang, J., Y. Deng, and D. Zeng. 2020. A deep neural network combined CNN and GCN for remote sensing scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:4325–38. doi:10.1109/JSTARS.2020.3011333.

Lin, P., M. Sun, C. Kung, and T. Chiueh. 2019. FloatSD : A new weight representation and associated update method for efficient convolutional neural network training. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 9 (2):267–79. doi:10.1109/JETCAS.2019.2911999.

Li, B., W. Su, H. Wu, R. Li, W. Zhang, W. Qin, S. Zhang, and J. Wei. 2020. Further Exploring Convolutional Neural Networks' Potential for Land-Use Scene Classification. *IEEE Geoscience and Remote Sensing Letters* 17 (10):1687–91. doi:https://doi.org/10.1109/LGRS.2019.2952660.

Liu, X., Y. Zhou, J. Zhao, R. Yao, B. Liu, and Y. Zheng. 2019. Siamese convolutional neural networks for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters* 16 (8):1200–04. doi:10.1109/LGRS.2019.2894399.

Li, Y., Y. Zhang, and Z. Zhu (2019). Learning deep networks under noisy labels for remote sensing image scene classification. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 3025–28. 10.1109/IGARSS.2019.8900497

Long, Y., Y. Gong, Z. Xiao, and Q. Liu. 2017. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 55 (5):2486–98. doi:10.1109/TGRS.2016.2645610.

Maggiori, E., Y. Tarabalka, G. Charpiat, and P. Alliez (2016). Fully convolutional neural networks for remote sensing image classification. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5071–74. 10.1109/IGARSS.2016.7730322

Mateen, M., J. Wen, Nasrullah, S. Song, and Z. Huang. 2019. Fundus image classification using VGG-19 architecture with PCA and SVD. *Symmetry* 11 (1):1. doi:10.3390/sym11010001.

Nogueira, K., O. A. B. Penatti, and J. A. D. Santos. 2016. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition* 61:539–56. doi:10.1016/j.patcog.2016.07.001.

Pang, J., C. Li, J. Shi, Z. Xu, and H. Feng. 2019. R2-CNN: Fast tiny object detection in large-scale remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 57 (8):5512–24. doi:10.1109/TGRS.2019.2899955.

Peng, C., Y. Li, L. Jiao, Y. Chen, and R. Shang. 2019. Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (8):2612–26. doi:10.1109/JSTARS.2019.2906387.

Qian, Y., W. Zhou, W. Yu, L. Han, W. Li, and W. Zhao. 2020. Integrating backdating and transfer learning in an object-based framework for high resolution image classification and change analysis. *Remote Sensing* 12 (24):4094. doi:10.3390/rs12244094.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and A. C. Berg. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* 115 (3):211–52. doi:https://doi.org/10.1007/s11263-015-0816-y.

Sang, Q., Y. Zhuang, S. Dong, G. Wang, and H. Chen. 2020. FRF-Net : Land cover classification from large-scale VHR optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 17 (6):1057–61. doi:10.1109/LGRS.2019.2938555.

Scott, G. J., M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis. 2017. Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters* 14 (4):549–53. doi:10.1109/LGRS.2017.2657778.

Senecal, J. J., J. W. Sheppard, and J. A. Shaw (2019). Efficient convolutional neural networks for multi-spectral image classification. *2019 International Joint Conference on Neural Networks (IJCNN)*, 14-19 July 2019, 1–8. 10.1109/IJCNN.2019.8851840

Shabbir, A., N. Ali, J. Ahmed, B. Zafar, A. Rasheed, M. Sajid, A. Ahmed, and S. H. Dar. 2021. Satellite and Scene image classification based on transfer learning and fine tuning of ResNet50. *Mathematical Problems in Engineering* 2021:1–18. doi:https://doi.org/10.1155/2021/5843816.

Shao, Z., K. Yang, and W. Zhou. 2018. Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset. *Remote Sensing* 10 (6):964. doi:10.3390/rs10060964.

Simonyan, K., and A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition ArXiv:1409.1556v6 doi:10.48550/arXiv.1409.1556.

Stivaktakis, R., G. Tsagkatakis, and P. Tsakalides. 2019. Deep learning for multilabel land cover scene categorization using data augmentation. *IEEE Geoscience and Remote Sensing Letters* 16 (7):1031–1035. doi:10.1109/LGRS.2019.2893306.

Wang, J., Y. Zhong, Z. Zheng, A. Ma, and L. Zhang. 2021. Rsnet: The search for remote sensing deep neural networks in recognition tasks. *IEEE Transactions on Geoscience and Remote Sensing* 59 (3):2520–34. doi:10.1109/TGRS.2020.3001401.

Yang, Y., and S. Newsam (2010). Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 270–79. 10.1145/1869790.1869829

Yin, X., W. Chen, X. Wu, and H. Yue (2017). Fine-tuning and visualization of convolutional neural networks. *12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 1310–15. 10.1109/ICIEA.2017.8283041

Zeng, Z., X. Chen, and Z. Song. 2021. MGFN: A multi-granularity fusion convolutional neural network for remote sensing scene classification. *IEEE Access* 9:76038–46. doi:10.1109/ACCESS.2021.3081922.

Zhang, J., C. Lu, J. Wang, X.-G. Yue, S.-J. Lim, Z. Al-Makhadmeh, and A. Tolba. 2020. Training convolutional neural networks with multi - size images and triplet loss for remote sensing scene classification. *Sensors* 20 (4):1188. doi:10.3390/s20041188.

Zhang, Y., X. Zheng, Y. Yuan, and X. Lu. 2020. Attribute-cooperated convolutional neural network for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 58 (12):8358–71. doi:10.1109/TGRS.2020.2987338.

Zhao, B., Y. Zhong, G. S. Xia, and L. Zhang. 2016. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing* 54 (4):2108–23. doi:10.1109/TGRS.2015.2496185.

Zheng, X., T. Gong, X. Li, and X. Lu. 2022. Generalized scene classification from small-scale datasets with multitask learning. *IEEE Transactions on Geoscience and Remote Sensing* 60:5609311. doi:10.1109/TGRS.2021.3116147.

Journal of Renewable and Sustainable Energy

Enhancing wind power forecasting from meteorological parameters using machine learning models

Cite as: J. Renewable Sustainable Energy **14**, 063302 (2022); https://doi.org/10.1063/5.0117662
Submitted: 01 August 2022 • Accepted: 12 October 2022 • Published Online: 08 November 2022

Upma Singh and M. Rizwan

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Can lidars assess wind plant blockage in simple terrain? A WRF-LES study
Journal of Renewable and Sustainable Energy **14**, 063303 (2022); https://doi.org/10.1063/5.0103668

The effects of submergence depth on energy harvesting from the VIV of a four-cylinder oscillator with rigid connection
Journal of Renewable and Sustainable Energy **14**, 064501 (2022); https://doi.org/10.1063/5.0109454

An adaptive wind power forecasting method based on wind speed-power trend enhancement and ensemble learning strategy
Journal of Renewable and Sustainable Energy **14**, 063301 (2022); https://doi.org/10.1063/5.0107049

APL Energy
Bridging basic research and innovative technology that will impact the future
Now Open for Submissions

AIP Publishing

© 2022 Author(s).

Journal of Renewable and Sustainable Energy ARTICLE scitation.org/journal/rse

Enhancing wind power forecasting from meteorological parameters using machine learning models

Cite as: J. Renewable Sustainable Energy **14**, 063302 (2022); doi: 10.1063/5.0117662
Submitted: 1 August 2022 · Accepted: 12 October 2022 ·
Published Online: 8 November 2022

View Online Export Citation CrossMark

Upma Singh[1,2,a)] and M. Rizwan[1,b)]

AFFILIATIONS

[1]Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India
[2]Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, Delhi 110058, India

[a)]**Author to whom correspondence should be addressed:** upma@msit.in
[b)]**Email:** rizwan@dce.ac.in

ABSTRACT

Short-term wind power forecasting is crucial for power system stability, dispatching, and cost control. Wind energy has the potential to be a viable source of renewable energy. Wind power generation forecasting is vital for resolving the supply and demand challenges of the smart grid. Moreover, one of the most problematic aspects of wind power is its high fluctuation and intermittent nature, which makes forecasting difficult. The goal of this research is to create machine learning models that can properly estimate wind power production. Significantly, the major contributions of this work are highlighted in the following significant elements. First, a data analysis framework for visualizing the gathered dataset from the SCADA system is presented. Second, for forecasting wind power time-series dataset values, we examine the predicting performance of various machine learning models using various statistical indices. The experimental findings demonstrate that with a minor reconstruction error, the proposed forecast approaches can minimize the complexity of the forecasting. Furthermore, in terms of forecast accuracy, a gradient boosting regression model outperforms other benchmark models. According to the analysis, our methodology might be applied in real-world circumstances to assist the management group in regulating the power provided by wind turbines.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0117662

I. INTRODUCTION

It is imperative to accelerate the development of renewable energy due to the increasing use of fossil fuels and their nonrenewable properties. Wind power is recognized as one of the most essential strategic alternatives for the long-term growth of electric power on account of its low emissions and lack of pollution (El Bourakadi *et al.*, 2022). Because of the unpredictability and volatility of wind energy output, vast-scale integration into the electricity grid will pose significant problems to the power system's dispatch and safe operation (Jiang *et al.*, 2019). As a result, a very precise wind power forecast is crucial for its real-time deployment in the power system. Wind power forecast approaches now accessible include physical models, statistical models, and machine learning models. To generate mathematical models, the physical model relies heavily on prior numerical weather predictions' (NWPs) meteorological datasets (Han *et al.*, 2022). However, the physical modeling method is complicated, and the computation cost is high (Dumas *et al.*, 2022). Statistical models can identify the relationship between previous weather conditions and wind power and then forecast future wind power based on that relationship (Toubeau *et al.*, 2021). In related domains of wind power forecasting, statistical methodologies, such as autoregressive (AR), autoregressive moving average (ARMA) (Wu *et al.*, 2014), and variants of ARMA, have been used (Hodge *et al.*, 2011). The non-stationary and nonlinear data, however, limit the prediction ability of these techniques. Some artificial intelligence models have been effectively implemented for enhancing wind power forecasting in recent years due to their theoretical suitability for mining complicated nonlinear correlations in recent data (Nazir *et al.*, 2020). For wind power forecasting one hour in advance, Nazaré *et al.* suggested artificial neural networks (ANNs) integrated with an optimization technique. Researchers discovered that by using ANNs, the training outcomes can quickly converge to the local minimum (Nazaré *et al.*, 2020). Acikgoz *et al.* (2020) proposed extreme learning machine (ELM) based approach to predict wind power.

In another paper, Yin and Liu (2022) evolved a forecasting model based on residual correction, ensemble reinforcement learning, and

Published under an exclusive license by AIP Publishing

outlier correction. Initially, outliers in the original data are corrected using the Hampel identification (HI) and to predict the rectified wind power data, three basic learners are chosen: extreme learning machine, echo state network, and group technique of data handling. To define the dependency structure and uncertainty of wind turbine outputs in the wind farm, Ye *et al.* (2017) developed a short-term wind energy forecast model based on the geographical correlation and physical methodology. Tasnim *et al.* (2017) created a two-stage auto-encoder (a special kind of deep learning) for predicting wind power. Similarly, Zheng *et al.* (2017) proposed a two-stage hierarchical adaptive neuro-fuzzy inference system [double-stage hybrid adaptive neuro-fuzzy inference system (ANFIS)] for short-term wind energy forecasting of a micro-grid wind farm in Beijing, China. As neural network theories and technology have progressed, more researchers have used deep neural networks (DNNs) to forecast the wind power of wind farms, such as RNN, convolutional neural networks (CNNs), and deep belief networks (DBNs). The hybrid strategy integrates two or more types of approaches. In recent years, researchers have begun to employ intelligent algorithms, deep learning, machine learning, and other algorithms to develop wind power forecasting models. For instance, sparrow search algorithms (SSAs), support vector machines (SVMs), linear regression, artificial neural networks (ANNs), and other algorithms. Each of these algorithms has its own set of benefits and drawbacks (Nazir *et al.*, 2020; Jafarian-Namin *et al.*, 2019).

Various research works are striving to construct reliable wind power prediction models. Since then, most of these algorithms have failed to produce appropriate results for diverse wind farm regions when the forecast is done in erratic and unstable wind patterns. Because of these considerations, there are far more input parameters required. Nowadays, linear regression, support vector regression, auto-regression, and other ML-based regression prediction approaches are frequently employed. Electric load forecasting, solar irradiance prediction for photovoltaic systems, grid management, wind energy, and power generation and consumption all use these technologies. Because the methodologies and algorithms are incapable of producing adequate results with high precision, an accurate and precise forecasting approach is required for the production of wind power. Also, the formerly built models lack the presentation of time-series data logs, which might aid the managing group in analyzing and understanding formerly undetected problems. Over the recent decade, machine learning approaches have established themselves as a significant area of research within a data-driven framework by dealing with different types of challenging problems in the electricity market (Donadio *et al.*, 2021; Yatiyana *et al.*, 2017), the gas market (Hu *et al.* 2021), and other real-world applications (Jiao *et al.*, 2018). Chen *et al.* (2021) performed a systematic comparison analysis on multivariate and univariate wind power forecasts for five Arctic wind farms. Multiple machine learning methods for forecasting wind power are used in this work. It analyses the performance of models based on whether they consider weather parameters in the Arctic. Distinct kinds of machine-learning algorithms are used to forecast multiple univariate time series, and their performance is compared. Similarly, Malakouti *et al.* (2022) presented a novel algorithmic method for predicting wind turbine power using machine learning approaches. Ensemble approaches, light gradient boosting machines, extremely randomized trees, and the CNN-LSTM approach have been used. According to the outcomes, the convolutional neural network-long short-term memory (CNN-LSTM) approach has the lowest mean square error value, indicating that it is more accurate. Several machine learning algorithms have been proposed in the literature to improve wind power prediction. However, the performance under short-term wind power prediction is still not satisfactory.

In this paper, we scrutinize the predicting performance of various machine learning models, such as random forest (RF) regression, K-nearest neighbor (KNN) regression, extreme gradient boost (xGBoost) regression, and gradient boosting (GB) regression, using various statistical indices. The analytic outcomes demonstrate that with a minor reconstruction error, the proposed forecast approaches can minimize the intricacy of the forecasting. Furthermore, in terms of forecast accuracy, a gradient boosting (GB) regression model outperforms other benchmark models. According to the analysis, our methodology might be applied in real-world circumstances to assist the management group in regulating the wind turbine power.

The following are the main contributions and novelties in this work:

- The purpose of the research is to develop a convenient and simple machine learning method for forecasting wind power output.
- For the wind power forecast, four machine learning algorithms are examined. The results show that the gradient boosting regression (GBR) has better performance than the other models.
- An exploratory data analysis and forecasting aspects of wind power production have been the focus of our research work. The exploratory data analysis is carried out in order to better understand wind and generated electricity. The exploratory data analysis provides a full insight into the wind turbine data logs by exploring accessible factors such as the wind direction, wind speed, theoretical power, and active power. As a result, the overall complexity of data logs would be decreased, making them more comprehensible.
- The scatter curves displaying the relationships between wind speed and power generated are drawn for all methodologies utilized in this paper. Then, using the plotted error curves, the forecasted average wind power and the actual average power from a turbine are compared.
- To compare the performance of the machine learning (ML) models, performance assessment metrics, namely, mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and coefficient of determination (R^2) are used to evaluate and examine the prediction capabilities of the proposed ML models.
- The presented framework gives short-term dataset visualization (i.e., the time stamp of one year) as well as the prediction of wind power generation, which might aid smart grid management.

The following is the planned organizational structure of this paper: Section II examines the collection of SCADA datasets. Section III describes the mechanism adopted for wind power forecasting. Section IV provides a brief overview of the machine learning techniques used in this study. Section V contains numerical outcomes and discussion, and Sec. VI presents the conclusions.

II. DESCRIPTION OF SCADA DATASET

The target onshore Yalova wind farm is situated in west Turkey. It has been operating since 2016. It is equipped with 36 wind turbines

Published under an exclusive license by AIP Publishing

with a combined nominal output of 54 000 kW and two turbine variants (SL 1500/90, SL 1500/82) with a capacity of 1.5 MW. The wind turbine has set 3.5 m/s cut-in speed, 10.9 m/s rated speed, and 22 m/s cutout speed. A SCADA system was used to control and monitor the target turbine, which can produce power outputs without additional expenses by default. SCADA records were retrieved for wind power forecasting in this study. The SCADA datasets under analysis were captured at a sampling rate of 10 min. The SCADA system captured the wind speed and direction as well as the produced and theoretical power, based on the power curves of the turbine. The employed data for model construction were for the duration of January 2018 to December 2018 (Erisen, 2018). In this investigation, a train-test split percentage of 80%–20% was chosen. The 47 033-point dataset is divided into two sections: a training set (37 567 points) and a test set (9466 points).

A. SCADA dataset pre-processing

Though the SCADA dataset can be utilized to forecast wind turbine output, it is still difficult to find the best approach owing to the possibility of erroneous data points in the dataset. During wind turbine operations, these inaccurate data points are mostly caused by system processing faults, sensor degradation/malfunction, and maintenance all of which are unfavorable to the forecast model (Bonanno *et al.*, 2015; Lu *et al.*, 2020).

As a result, it is anticipated that SCADA data be pre-processed before being used to develop a model.

B. Elimination of outliers

In the raw SCADA records, there were some timestamp where dataset values were negative. These negative values in the raw dataset indicate noticeable outliers. Although these negative values are technically feasible, they have no practical significance with respect to wind energy output. As a result, these evident outliers, as well as the related variables under the same time stamps, would be deleted.

C. Anomalies detection

The wind turbine power curve indicates the relation between the wind speed and produced wind power, which is an essential indicator of wind turbine productivity. Ideally, the shape of the power curve should be in the sigmoid function ("S") shape. The power curve of the target onshore wind turbine still exhibits few anomalies which depart from typical data after evident anomalies elimination, when compared to a regular S shape. Figure 1 depicts a graph of the wind speed power scatter curve with the ideal power production curve generally matching the actual power production curve. A detailed examination of this curve shows some anomalies.

These anomalies are created during operation and may be classified into three categories, as seen in Fig. 1 (Aly, 2020). Type 1: This anomaly is primarily produced by turbine downtime, when the wind speed exceeds the cut-in speed (3.5 m/s) when the wind power is near zero. Type 2: This type of anomalies is mostly produced by wind-curtailment, in which the produced power is intentionally reduced by the operator owing to a variety of issues such as problems with large-capacity power storage or grid supply constraints. On the contrary, sensor malfunction/degradation is the primary source of type 3 anomalies. (Phan *et al.*, 2021). Figure 2 presents visualization of distribution of the dataset.

III. MECHANISM FOR WIND POWER FORECAST

The goal of this research is to see if machine learning approaches can be used to estimate or foretell wind power. To predict wind power dataset, four machine learning algorithms are analyzed and tested against each other. Figure 3 illustrates the entire approach for wind power estimation used in this research. Prior to developing the machine learning model, the SCADA dataset is analyzed and pre-processed. To enhance the dataset quality, we eliminate outliers and impute missing data values. Outliers are often detected and removed to enhance the forecasting accuracy of the machine learning models. Otherwise, the generated models might be inaccurate. To replace the outliers, the median of the training dataset is utilized. Missing wind speed measurements can be caused by a variety of factors, including inaccurate data recording, thunderstorms, deterioration, and other anemometer faults.

In this investigation, a train-test split percentage of 80%–20% was chosen. The normalized data were then separated into train and test sets. The model parameters are calculated after the models have been trained using the training set. To identify the best parameters for the machine learning models, we used the trial-and-error method. Finally, five statistical indicators are employed to assess the quality of each model: R^2, MAE, MSE, RMSE, and MAPE.

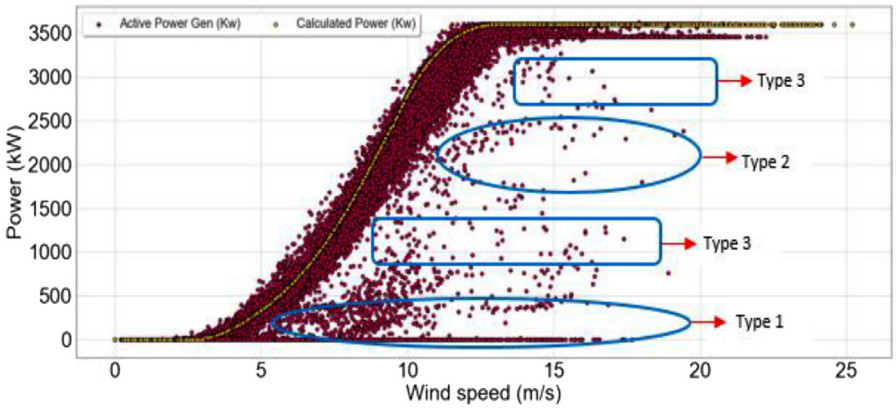


FIG. 1. Wind power vs wind speed curve with the original wind turbine dataset.

Published under an exclusive license by AIP Publishing

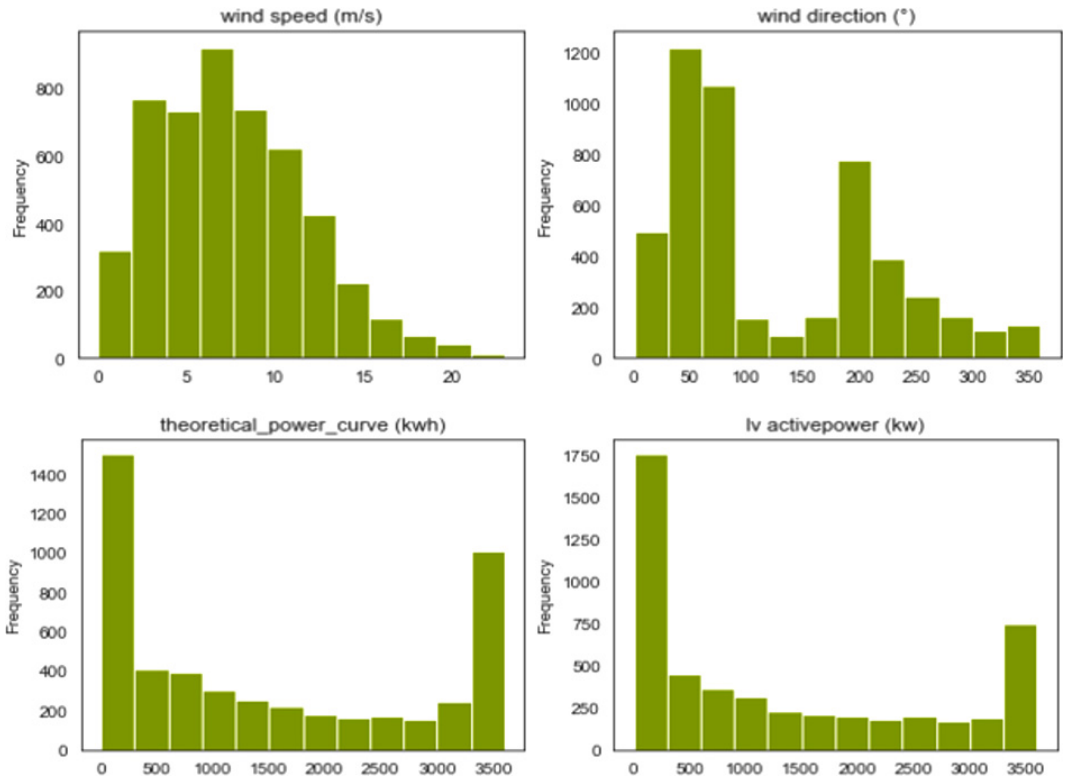


FIG. 2. Dataset distribution visualization.

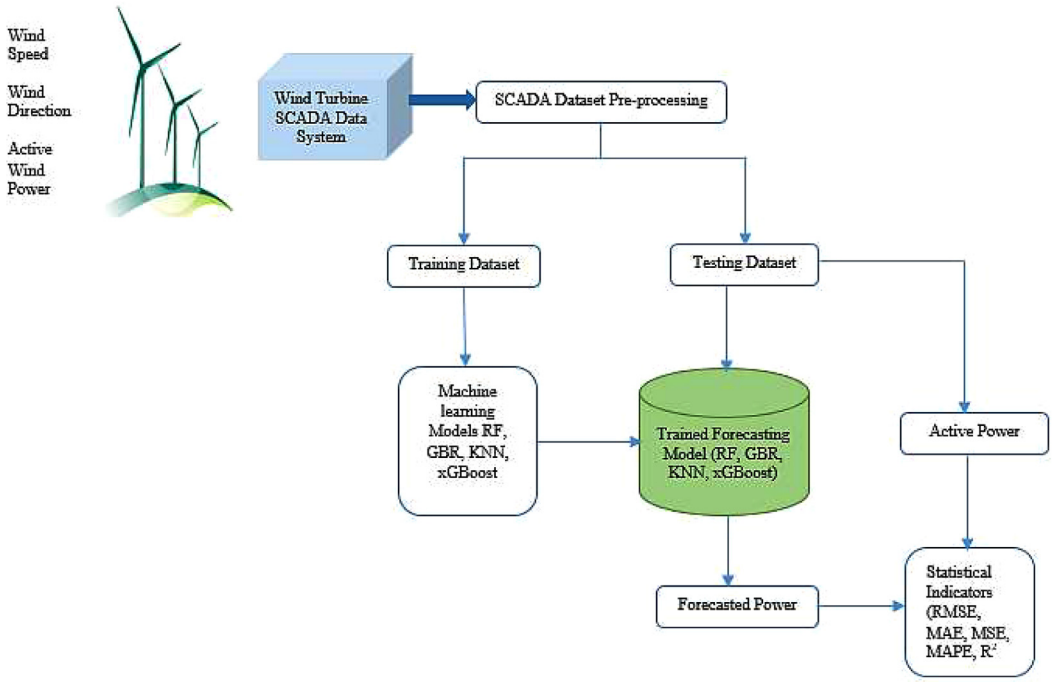


FIG. 3. The proposed model for analyzing and predicting wind turbine data.

Published under an exclusive license by AIP Publishing

IV. METHODOLOGY

It is difficult to say earlier whether a particular ML algorithm will work for a particular problem, even if one already knows that a particular approach works on a relevant topic. As a result, an appropriate approach is selected with the help of the trial and error method. In addition to the techniques used, most of the methods have a number of so-called hyper-parameters. These are the parameters that must be set before the training begins.

A. Random forest regression

Random forest regression is an ensemble supervised learning approach, which involves fitting numerous regression trees on random selections of the training dataset. Initially, the algorithm randomly divides the dataset into numerous subparts and then generates various decision trees for each subpart. All trees are utilized for prediction, and their forecasts are averaged to produce a more reliable and precise forecast. Also, it is capable of effectively handling huge datasets. This approach also manages variables quickly, which makes it suitable for complex tasks. The random forest cannot forecast values that are not in the training dataset target range (Lahouar and Slama, 2017; Shi *et al.*, 2018). We utilized scikit-learn for random forest regression implementation.

In RF, a random vector of k is generated, which is a subset of the dataset's feature space, and each tree is built utilizing k and the training data. In a random forest, the margin function and the generalization error are stated as

$$PE^{*} = P_{X,Y}(mg(X,Y) < 0)$$

where

$$mg(X,Y) = av_k I(h_k(X) = Y) - max_{j \neq Y} av_k I\left(h_k(X) = j\right), \quad (1)$$

where h_k are the classifiers, $I(\cdot)$ is the indicator function, mg is the margin function, which governs the average votes at random vectors for the right output as compared to any other output, PE^{*} is the generalization error, and X, Y are random vectors.

In the RF algorithm, the random state is set to 40, and the number of trees is set to 100, because increasing the number of trees to more than 100 did not substantially increase the forecast outcome. In order to improve runtime and forecast performance, an adequate number of trees must be determined. Figure 4(a) depicts the scatterplot employing RF regression showing the relationship between the speed of wind (m/s) and turbine power generation (kW) and Fig. 4(b) compares the forecast average power with the actual average power (kW) produced from the turbine.

B. K-nearest neighbor regression

The K-nearest neighbors (KNNs) technique is a non-parametric approximation approach that may be used to solve problems like classification and regression. The assumption behind KNN is that an item belongs to the same class as its nearest neighbors. The approach demands the input of a positive integer k at initially. For every sample, the method finds the k points on the dataset that have a similar pattern to that of the sample. The distance between all of the samples in the dataset and the newly analyzed sample is needed for this selection procedure (Silva, 2014; Kusiak *et al.*, 2009). It has an algorithmic framework that is simple to understand and implement, and it does not involve any model fitting or function estimation. The KNN regressor is widely considered as one of the most prominent data mining algorithms in the area of research due to its sophisticated capabilities. The KNN approach, when applied in forecasting applications, determines the neighbors: components from the training set that meet the reference conditions based on certain preset criteria. The recent data on wind power, wind direction, and speed are the significant features of this study. These data are organized in a matrix X_{ij}, where each row represents a feature vector for a certain estimate period. The nearest neighbor for a new data point at time t, as defined by the feature vector y_j, is compared to all of the rows in X_{ij}, and the result is recorded in the Euclidean distances vector d_i

$$d_i = \sqrt{\sum_i (X_{ij} - y_j)^2}. \quad (2)$$

The first k matches are obtained after sorting the distance values in an increasing order. The numerical value for y_j is the average of all the "K" nearest neighbors variable numerical values.

In this algorithm, the Euclidean distance was used as the distance metric, and the number of neighbors for the kNN algorithm was

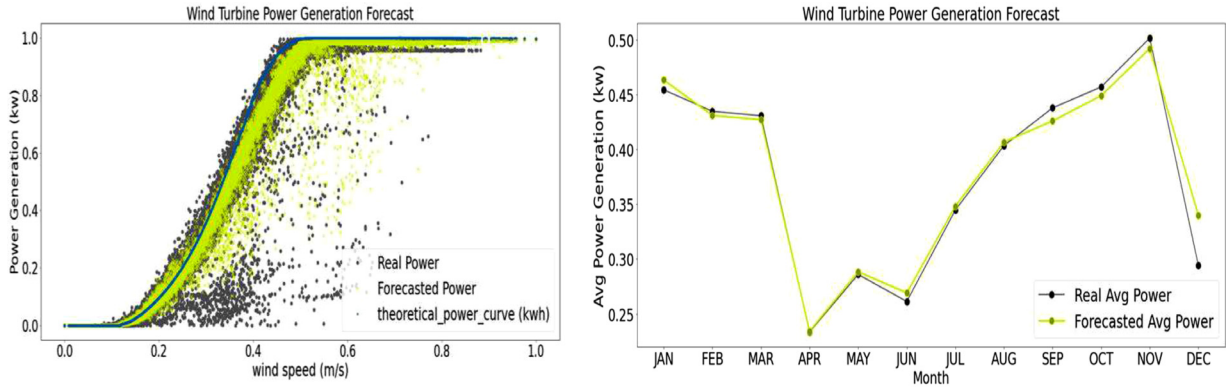
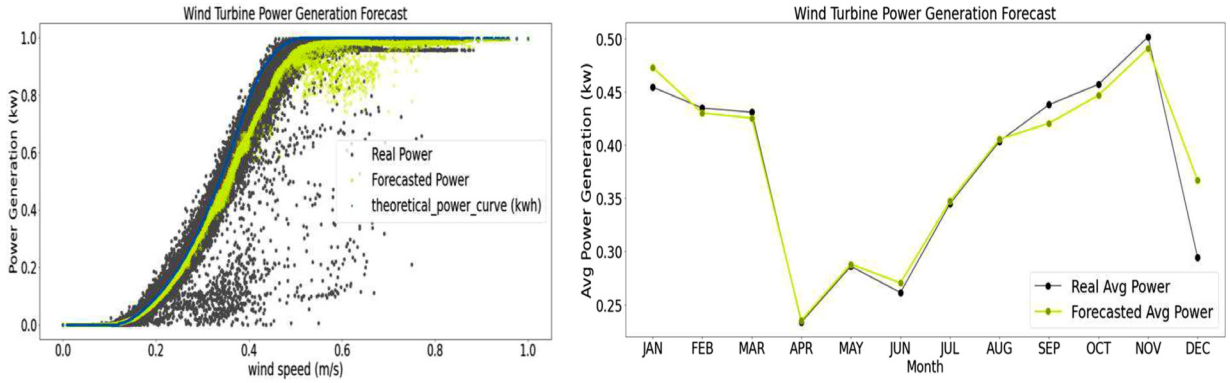


FIG. 4. (a) Scatterplot employing RF regression showing the relationship between the wind speed (m/s) and turbine power production (kW); (b) forecast average power compared to actual average power (kW) employing RF regression.

Published under an exclusive license by AIP Publishing



(a) (b)

FIG. 5. (a) Scatterplot employing KNN regression depicts the relationship between the wind speed (m/s) and turbine power production (kW); (b) forecast average power compared to actual average power (kW) employing KNN regression.

determined to be 7. Figure 5(a) depicts a scatterplot employing KNN regression, which shows the relationship between the speed of wind (m/s) and turbine power production (kW), whereas Fig. 5(b) depicts the forecast average power compared to the actual average power (kW) employing KNN regression.

C. Extreme gradient boost regression

The extreme gradient boosting (xGBoost) technique is a more efficient and improved variant of the gradient boosting machine (GBM) algorithm. As a result, the xGBoost method is both quick and scalable. The goal of the method is to design better decision trees in order to minimize the objective function. The xGBoost technique may be used to do regression analysis as well. It can identify optimal solutions faster, because it is fast and more efficient than other boosting options (Cai *et al.*, 2020; Chatterjee and Dethlefs, 2020). xGBoost uses the decision tree as its base learner. By adding more new base learners, the error between the prediction value and the target value is lowered. The final predictive values are obtained by adding all of the base learners. The objective function $F_{obj}(\theta)$ is given as

$$F_{obj}(\theta) = L(\theta) + \Omega(\theta),$$

where

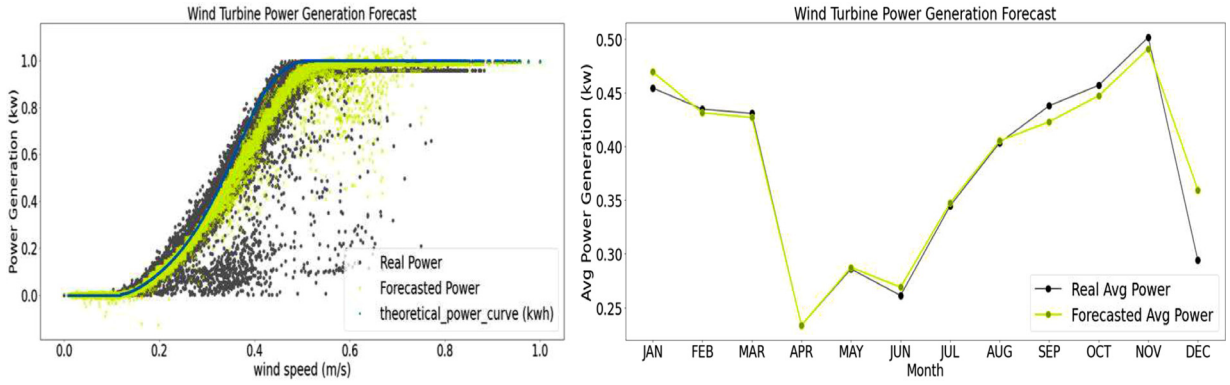
$$L(\theta) = l(\hat{y}_i, y_i) \quad \text{and} \quad \Omega(\theta) = \gamma T + \frac{1}{2}\lambda \|w\|^2. \tag{3}$$

Here, w is the leaves weight, λ is the regularization parameter, the number of leaves on the tree is denoted by T, γ is the learning rate, $\Omega(\theta)$ is the regularization term, and $L(\theta)$ is the loss function between real value y_i and prediction $\hat{y}_i$.

In this algorithm, the number of trees was set to 500, the depth was set to 7, and a 0.1 learning rate is chosen. Figure 6(a) depicts the scatterplot employing xGBoost regression shows the relationship between the speed of wind (m/s) and turbine power production (kW), whereas Fig. 6(b) depicts the forecasted average power compared to actual average power from a turbine (kW) employing xGBoost regression.

D. Gradient boosting regression

Gradient boosting (GB) is a machine learning approach for solving classification and regression tasks. GB creates optimum estimates



(a) (b)

FIG. 6. (a) Scatterplot employing xGBoost regression shows the relationship between the wind speed (m/s) and turbine power production (kW); (b) forecast average power compared to actual average power (kW) employing xGBoost regression.

Published under an exclusive license by AIP Publishing

by minimizing cost functions by merging weak prediction models such as decision trees. The number of weak prediction models may be adjusted; for example, in our models, the number of trees is set to 100. GBR is a weak learner strategy that runs a number of models periodically so as to create a subsequent model by reducing the error of the prior models. Thus, by integrating weak models, this algorithm generates a high-precision forecasting model. This algorithm has the advantage of being able to handle lost data. An additional benefit of GBR is that it provides various hyper-parameter tuning possibilities, which boosts the algorithm's flexibility. It is one of the ensemble models (Bochenek *et al.*, 2021). It creates a forecast model that identifies a final linear combination $\hat{F}(x)$, which is a weighted sum of functions $h_i(x)$ from various classes $'H$,

$$\hat{F}(x) = \sum_{i=1}^{M} \gamma_i h_i(x) + const, \tag{4}$$

where *const* is the number of one's first guess that remains constant, $h_i(x)$'s are base learners from $'H$, and M is the number of iterations of GB (Barque *et al.*, 2018). This approach decreases the bias of a training set approximation $\hat{F}(x)$ according to the empirical risk reduction principle. It creates a model using a constant function $F_0(x)$ in a greedy pattern to minimize the average value of the loss function

$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma), \tag{5}$$

$$F_m(x) = F_{m-1}(x) + \underset{hm \in 'H}{argmin} \sum_{i=1}^{n} L\big(y_i, F_{m-1}(x_i) + h_m(x_i)\big), \tag{6}$$

where $h_m \in 'H$ is a base learner function.

The following are the brief steps of the GB technique in pseudo-code:

Step 1: Initialize the variable with a constant value as shown below:

$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma). \tag{7}$$

Step 2: Determine the pseudo-residuals for $1 \leq m \leq M$

$$r_{im} = -\left[\frac{\partial L\big(y_i, F(x_i)\big)}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad for \quad i = 1, 2, \ldots, n. \tag{8}$$

Step 3: Solve the one-dimensional optimization problem

$$\gamma_m = \underset{\gamma}{argmin} \sum_{i=1}^{n} L\big(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)\big). \tag{9}$$

Step 4: Make changes to the model depending on

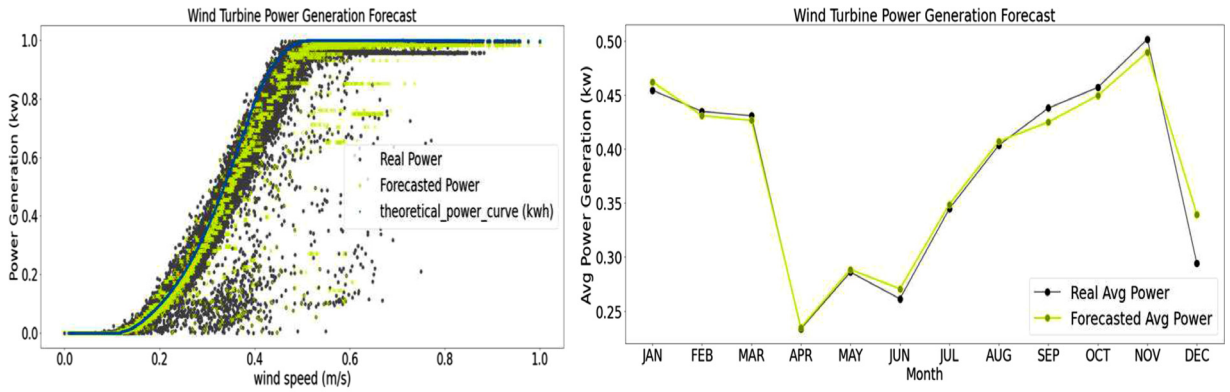
$$F_m(x) = v F_{m-1}(x) + \gamma_m h_m(x). \tag{10}$$

Finally, stop the procedure until m reaches M iterations of gradient boosting.

In this investigation, the estimate value of 100 and 0.1 learning rate is chosen, as low learning rates makes it smoother to halt before over-fitting. Figure 7(a) shows the scatterplot employing GBR shows the relationship between the speed of (m/s) and turbine power generation (kW), and Fig. 7(b) shows the forecasted average power compared to actual average power from the turbine (kW) employing GBR.

V. NUMERICAL OUTCOMES AND DISCUSSIONS

This segment discusses the findings and significant observations obtained from the final outcomes of wind power forecast models. All of the algorithms presented and detailed above were trained and tested on a Jupiter notebook (Python 3.9.5 version) equipped with 12 GB of 16 MHz DDR3 RAM and a 1.6 GHz Intel Core i3 processor. All the models are trained by utilizing the selected input parameters, e.g., wind speed, wind direction, and the actual output power produced from the wind turbine following that, the model is given a testing/validation dataset depending on which it provides production forecasts (active power). The forecast accuracy of all machine learning models is evaluated by utilizing the statistical indicators R^2, MAE, MSE, RMSE, and MAPE. In general, errors in the train set demonstrate the applicability of the generated model, whereas errors in the test set demonstrate the model's prediction performance. Algorithms with minimal



(a) (b)

FIG. 7. (a) Scatterplot employing GBR shows the relationship between the wind speed (m/s) and turbine power production (kW); (b) forecast average power compared to actual average power (kW) employing GBR.

Published under an exclusive license by AIP Publishing

TABLE I. Selection of optimum parameters for each algorithm.

Algorithms	Optimal parameters details
RF	Random state selected as 40, number of trees in the random forest N = 100.
KNN	Nearest neighbor search algorithm with Euclidean distance and number of nearest neighbors k = 7
xGBoost	Number of trees N = 500, learning rate for tree booster = 0.1, depth = 7, booster parameter = gb-tree
GBR	Learning rate = 0.1, estimator value = 100, depth = 4

errors represent the most accurate and acceptable method. The statistical indicator, R^2 indicates how efficiently the model reproduces the measured outputs. The MSE represents the average square of the errors. The RMSE is the standard deviation of the forecast errors; the lower the RMSE, the superior the model is regarded to be. Moreover, a model is deemed good and free of over-fitting if the RMSE values of the testing and training sets fall within a restricted range. The MAE is calculated by adding the absolute differences between the real and forecast values. The MAPE determines accuracy by comparing the real and forecast data points. Among the five evaluation criteria assessed here, RMSE should be seen as the key focus with errors squared before being averaged and a significant weight assigned to significant errors. As a result, the smallest value of the RMSE deduced the actual error rate. The fact that the values of the root mean square are comparable to the mean absolute error implies that there is no considerable variance in the magnitudes of error, indicating the model's efficacy and generality. The forecast model parameters have been evaluated utilizing numbers of runs for the separate algorithms based on the random state, value of k, distance measure, the number of trees, learning rate, and booster parameter among other things in order to optimize the model performance and accuracy. Table I shows the selection of optimum parameters for each algorithm. The performance of all forecast models can be assessed utilizing the overlapping scatterplots, which indicate the correlations between the power developed by the turbine and wind speed as well as the plot between the forecast average power values in comparison to the actual average power generated by the turbine that pictorially reveals the performance of regression models, as can be seen in Figs. 4–7. The RF regression model's outcomes are shown in Fig. 4. As seen in Figs. 4(a) and 4(b), this approach was unable to anticipate accurate power values. The RF method, out of the four regression algorithms, performed poorly and had a low accuracy rate and large predicting error, as demonstrated by the performance measures presented in Table II. Additionally, the R^2 value of this algorithm was lower than that of the other models with longer execution times.

TABLE III. Training and testing time for each algorithm with optimal parameters.

Algorithm	Training (fit) time with optimal parameters (s)	Forecast time with optimal parameters (ms)
RF	11.1	1.12
KNN	4.14	1.26
xGBoost	4.01	0.009
GBR	5.77	0.07

The KNN regression results are shown in Fig. 5. The KNN method could forecast power values favorably, as shown in the graph; nonetheless, its performance was better than the RF regression model, despite the fact that this algorithm could not make proper forecasts at high wind speeds as shown in Fig. 5(a). The majority of the forecast values are found overlapping or near to the actual average power levels, as illustrated in Fig. 5(b), and also the model has a stronger R^2 value than that of the RF regression model. As a result, the overall performance of the KNN regression model was found to be very good. The results of xGBoost are displayed in Fig. 6, and the outcomes of the GBR regression model are shown in Fig. 7. In contrast to the other regression models, the GBR method offered the most accurate outcomes for predicting both low and strong values of speed of the wind and is able to manage high wind speeds, as shown in Fig. 7(a). Furthermore, as seen in Fig. 7(b), the forecast curve closely follows or totally covers the real average power curve. As a result, when compared with the other algorithms, the GBR model performance is perceived to be the best with the highest R^2 value. From Table II, it is clearly evident that the outcomes of GBR and xGBoost are comparable. On the contrary, in terms of execution time, the xGBoost algorithm outperforms all other algorithms.

Table III illustrates how long each algorithm takes to execute after being given the query/input. Clearly, all algorithms can respond to the query/input in less than a fraction of a second, which is far under the time constraint of a few seconds. It is worth noting that KNN has a longer response time, because it is a lazy learning algorithm, which means it does not learn until a query is received. Note

TABLE II. Dataset statistical indicators; better performance is indicated by italics.

	Performance evaluation on training data					Performance evaluation on testing data				
Regression model	MSE	RMSE	MAPE	MAE	R^2	MSE	RMSE	MAPE	MAE	R^2
RF	0.0041	0.0584	0.3264	0.0286	0.956	0.005	0.0671	0.323	0.029	0.955
KNN	0.0033	0.057	0.277	0.023	0.976	0.004	0.068	0.322	0.028	0.964
xGBoost	0.0031	0.039	0.2227	0.021	0.981	0.004	0.066	0.301	0.021	0.965
GBR	***0.0030***	***0.036***	***0.2118***	***0.0155***	***0.985***	***0.003***	***0.064***	***0.282***	***0.026***	***0.969***

Published under an exclusive license by AIP Publishing

that KNN examines the whole training data set for each query, thus as the data expands, the response time will grow as well.

VI. CONCLUSION

For a safe and long-term balancing of the electricity grid, a correct short-term wind power forecast is crucial. Wind energy is currently being used extensively as an alternative energy source on a huge scale. Because of the volatile nature of wind, prediction of turbine power is difficult, and integration into the primary power grid is a significant problem. Because predicting can never be deemed error-free, we must develop improved models to eliminate these imperfections. A comparison of several machine learning approaches for forecasting wind power depending on wind direction, wind speed, and actual power (generated from the turbine) dataset is presented in this work. To simulate the wind power forecasting model, this study utilized and examined multiple machine learning algorithms specifically RF, KNN, xGBoost, and GBR regression algorithms. The SCADA dataset from the Yalova wind farm for the period of one year at a sampling frequency of 10 min is employed to assess the effectiveness of the machine learning models. Various statistical criteria were employed to assess the forecast accuracy of the ML models. The outcomes reveal that the ML models contain important information from historical observations of wind power series enhanced forecast outcomes. For all studies, we found that no single strategy outperformed the others, although on average, the GBR model achieved an adequate forecast performance with the R^2 of about 0.969, MSE of 0.003, RMSE of 0.064, MAPE of 0.282, and MAE of 0.026. To enhance the performance of a ML model, gradient boosting employs an ensemble of weak learners. Decision trees are commonly used as poor learners. When their outputs are combined, they provide better models. Additionally, because decision trees are structurally weak and not stable, minor differences in the training sample might cause major changes in the topologies of the trees and differing forecasts for the same validation samples. When using regression, the final outcome is calculated by taking the average of all weak learners. In GBR, weak learners function in a sequential fashion, with each model attempting to improve on the prior model's error. The created tree-based ensemble models impart robust and precise hourly predictions and may be employed for sustainable power grid integration and management. As previously stated, providing forecasts for wind power that can be produced in a day based on input parameters (wind direction and wind speed) is incredibly valuable, and our machine learning models have been found to be fairly efficient for such applications. Despite the positive findings obtained using dynamic machine learning approaches for wind power prediction, future research will also focus on strengthening the resilience of ML methods to noisy wind energy data by creating wavelet-based dynamic machine learning approaches. In addition, we anticipate that fine-tuning of all the model parameters will enhance the outcomes much further.

ACKNOWLEDGMENTS

This research has not received a specific grant from any funding agencies/organizations.

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 own work, and it does not contain material that was taken from other sources. All data measures are real, 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 in the research presented in this paper.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Upma Singh: Conceptualization (equal); Methodology (equal); Software (equal); Validation (equal); Writing – original draft (equal). **M. Rizwan:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in the SCADA system of the Yalova wind farm situated in western Turkey.

REFERENCES

Acikgoz, H., Yildiz, C., and Sekkeli, M., "An extreme learning machine based very short-term wind power forecasting method for complex terrain," Energy Sources, Part A **42**(22), 2715–2730 (2020).

Aly, H. H., "A novel deep learning intelligent clustered hybrid models for wind speed and power forecasting," Energy **213**, 118773 (2020).

Barque, M., Martin, S., Vianin, J. E. N., Genoud, D., and Wannier, D., "Improving wind power prediction with retraining machine learning algorithms," in *2018 International Workshop on Big Data and Information Security (IWBIS)* (IEEE, 2018), pp. 43–48.

Bochenek, B., Jurasz, J., Jaczewski, A., Stachura, G., Sekuła, P., Strzyżewski, T., Wdowikowski, M., and Figurski, M., "Day-ahead wind power forecasting in Poland based on numerical weather prediction," Energies **14**(8), 2164 (2021).

Bonanno, F., Capizzi, G., Sciuto, G. L., and Napoli, C., "Wavelet recurrent neural network with semi-parametric input data preprocessing for micro-wind power forecasting in integrated generation systems," in *2015 International Conference on Clean Electrical Power (ICCEP)* (IEEE, 2015), pp. 602–609.

Cai, R., Xie, S., Wang, B., Yang, R., Xu, D., and He, Y., "Wind speed forecasting based on extreme gradient boosting," IEEE Access **8**, 175063–175069 (2020).

Chatterjee, J. and Dethlefs, N., "Deep learning with knowledge transfer for explainable anomaly prediction in wind turbines," Wind Energy **23**(8), 1693–1710 (2020).

Chen, H., Birkelund, Y., Anfinsen, S. N., and Yuan, F., "Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region," J. Renewable Sustainable Energy **13**(2), 023314 (2021).

Donadio, L., Fang, J., and Porté-Agel, F., "Numerical weather prediction and artificial neural network coupling for wind energy forecast," Energies **14**(2), 338 (2021).

Dumas, J., Wehenkel, A., Lanaspeze, D., Cornélusse, B., and Sutera, A., "A deep generative model for probabilistic energy forecasting in power systems: Normalizing flows," Appl. Energy **305**, 117871 (2022).

El Bourakadi, D., Yahyaouy, A., and Boumhidi, J., "Improved extreme learning machine with AutoEncoder and particle swarm optimization for short-term wind power prediction," Neural Comput. Appl. **34**(6), 4643–4659 (2022).

Erisen, B., see https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset for "Wind Turbine SCADA Dataset, 2018" (last accessed May 18, 2020).

Jafarian-Namin, S., Goli, A., Qolipour, M., Mostafaeipour, A., and Golmohammadi, A. M., "Forecasting the wind power generation using Box–Jenkins and hybrid artificial intelligence: A case study," Int. J. Energy Sector Manage. **13**, 1038 (2019).

Published under an exclusive license by AIP Publishing

Jiang, P., Li, R., and Li, H., "Multi-objective algorithm for the design of prediction intervals for wind power forecasting model," Appl. Math. Modell. **67**, 101–122 (2019).
Jiao, R., Huang, X., Ma, X., Han, L., and Tian, W., "A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting," IEEE Access **6**, 17851–17858 (2018).
Han, Y., Mi, L., Shen, L., Cai, C. S., Liu, Y., Li, K., and Xu, G., "A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting," Appl. Energy **312**, 118777 (2022).
Hodge, B. M., Zeiler, A., Brooks, D., Blau, G., Pekny, J., and Reklatis, G., "Improved wind power forecasting with ARIMA models," in *Computer Aided Chemical Engineering* (Elsevier, 2011), Vol. 29, pp. 1789–1793.
Hu, S., Xiang, Y., Huo, D., Jawad, S., and Liu, J., "An improved deep belief network based hybrid forecasting method for wind power," Energy **224**, 120185 (2021).
Kusiak, A., Zheng, H., and Song, Z., "Models for monitoring wind farm power," Renewable Energy **34**(3), 583–590 (2009).
Lahouar, A. and Slama, J. B. H., "Hour-ahead wind power forecast based on random forests," Renewable Energy **109**, 529–541 (2017).
Lu, P., Ye, L., Zhong, W., Qu, Y., Zhai, B., Tang, Y., and Zhao, Y., "A novel spatio-temporal wind power forecasting framework based on multi-output support vector machine and optimization strategy," J. Cleaner Prod. **254**, 119993 (2020).
Malakouti, S. M., Ghiasi, A. R., Ghavifekr, A. A., and Emami, P., "Predicting wind power generation using machine learning and CNN-LSTM approaches," Wind Eng. **46**, 1853–1869 (2022).
Nazaré, G., Castro, R., and Gabriel Filho, L. R., "Wind power forecast using neural networks: Tuning with optimization techniques and error analysis," Wind Energy **23**(3), 810–824 (2020).
Nazir, M. S., Alturise, F., Alshmrany, S., Nazir, H., Bilal, M., Abdalla, A. N., Sanjeevikumar, P., and Ali, Z. M., "Wind generation forecasting methods and proliferation of artificial neural network: A review of five years research trend," Sustainability **12**(9), 3778 (2020).
Phan, Q. T., Wu, Y. K., and Phan, Q. D., "An overview of data preprocessing for short-term wind power forecasting," in *2021 7th International Conference on Applied System Innovation (ICASI)* (IEEE, 2021), pp. 121–125.
Shi, K., Qiao, Y., Zhao, W., Wang, Q., Liu, M., and Lu, Z., "An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness," Wind Energy **21**(12), 1383–1394 (2018).
Silva, L., "A feature engineering approach to wind power forecasting: GEFCom 2012," Int. J. Forecast. **30**(2), 395–401 (2014).
Tasnim, S., Rahman, A., Oo, A. M. T., and Haque, M. E., "Autoencoder for wind power prediction," Renewables **4**(1), 6 (2017).
Toubeau, J. F., Dapoz, P. D., Bottieau, J., Wautier, A., De Greve, Z., and Vallée, F., "Recalibration of recurrent neural networks for short-term wind power forecasting," Electr. Power Syst. Res. **190**, 106639 (2021).
Wu, B., Song, M., Chen, K., He, Z., and Zhang, X., "Wind power prediction system for wind farm based on auto regressive statistical model and physical model," J. Renewable Sustainable Energy **6**(1), 013101 (2014).
Yatiyana, E., Rajakaruna, S., and Ghosh, A., "Wind speed and direction forecasting for wind power generation using ARIMA model," in *2017 Australasian Universities Power Engineering Conference (AUPEC)* (IEEE, 2017), pp. 1–6.
Ye, L., Zhao, Y., Zeng, C., and Zhang, C., "Short-term wind power prediction based on spatial model," Renewable Energy **101**, 1067–1074 (2017).
Yin, S. and Liu, H., "Wind power prediction based on outlier correction, ensemble reinforcement learning, and residual correction," Energy **250**, 123857 (2022).
Zheng, D., Eseye, A. T., Zhang, J., and Li, H., "Short-term wind power forecasting using a double-stage hierarchical ANFIS approach for energy management in microgrids," Prot. Control Mod. Power Syst. **2**(1), 13 (2017).

Published under an exclusive license by AIP Publishing

2022 International Conference on Industry 4.0 Technology (I4Tech)
Vishwakarma Institute of Technology, Pune, India. Sep 23-24, 2022

Explicate Toxicity By eXplainable Artificial Intelligence

Nakul Tanwar
Department of Biotechnology
Delhi Technological University
Delhi, India
tbret41@gmail.com

Jaishree Meena
Department of Biotechnology
Delhi Technological University
Delhi, India
jaishreedtu@gmail.com

Yasha Hasija*
Department of Biotechnology
Delhi Technological University
Delhi, India
yashahasija6oct@gmail.com
*Corresponding Author

***Abstract*— Chemical exposure can cause formative neurotoxicity, which requires fast and exact testing techniques. Current methods, notably in vivo research on animal models and assessments of primary cell cultures derived from animal and humans, have difficulties in terms of time, expense, and application to human physiology. For example, *in vivo* animal studies may take years to complete. In this study, eXplainable Artificial Intelligence (XAI) was combined with XGBoost machine learning (ML) models that were prepared to utilize binary classification as a potent mix of datasets to identify genes that may be associated with neurotoxicity. Significant genes were found and connected to the progression of neurotoxicity after SHAP values were effectively integrated into the ML models.**

Keywords—Neurotoxicity, XAI, XGBoost ML classifier, SHAP values

I. INTRODUCTION

The Toxic Substances Control Act (TSCA) is in charge of about 84,000 mixtures, almost all of which have not been checked for their effects on developing neurons. The growing human brain is particularly vulnerable to hazardous exposures, and it is believed that the consequences of formative neurotoxicity are significant. In this way, we need right-away methods for evaluating formative neurotoxicity that is fast, don't cost much, and are accurate [1]. Current methods take into account *in vivo* studies and evaluations of important cell and tissue culture practices on both animals and humans. These strategies are limited by time, money, the lack of important cells and tissues, and the fact that they have nothing to do with how the human body functions. Because of these problems, drug approval rates have gone down, even though the number of people working on new ideas has grown. Human pluripotent stem cells can help solve these problems because they are a flexible source of clinically useful human cells at a reasonable price [1]. The National Institutes of Health (NIH), the Defense Advanced Research Projects Agency (DARPA), and the U.S. Food and Drug Administration (FDA) started the Micro physiological Systems Program in 2012 with the goal of making human tissue chips with bioengineered tissue models that replicate human functioning. This program aims to use these tissue chips to envisage how safe and also effective the expected 30 medicines will be. Schwartz et al. used the high-quality information from these signs to make 3D models of budding human brain tissue from different sorts of cells and AI models to predict early neurotoxicity. The 3D model is ready to do exactly what it needs to do, which is to find dangerous substances and show the ability to summarize important natural frameworks, showing the need for more information. However, building this 3D model is not an easy task, and it requires not only a large amount of expertise but also a lot of manual labor in order to utilize it successfully [1].

ML and deep learning (DL) are two of the most progressive algorithms that are accessible today. They offer excellent results and frequently eliminate the need for humans to handle, store, and process data [2]. The main purpose of these models is to improve accuracy, speed up the processes, and do dynamic tasks, which increases the dependency of humans on such models. Still, these models need to be more straightforward, objective, and clear, which makes them irrelevant to real problems [3]. XAI has an objective of producing the insides of the ML layers useful, which makes the models more reliable and easier to understand. XAI has made plans so that it gives even the smallest bit of real evidence on the side of the result and includes features that might change the final conclusion. Since this framework is still fairly juvenile, it should be looked into to see how it could be used in forecasting, medical services, and industry. The words 'explainable' or 'explain-ability' refers to how well a model can defend its results. Also, explainability makes sure that the model works within the standard protocol and guarantees the precision of the models [4]. It can also mean that a model gives a complete and accurate picture of the end result. Explainability can be called either "local" (for an example of choices that doesn't come with anything else) or "global" (for understanding the decision-making process behind the model). The growing importance of logic puts more emphasis on the idea of tools and technologies that help people understand how black-box models work. XAI is the most common term used to describe these technologies [5].

The age of huge amounts of data is brought on by the acceptance of cutting-edge innovations. Artificial intelligence (AI) and its technologies (ML & DL) manage and utilize this data, which makes people to rely on these models [6]. Because of this trust, people are in a great position to figure out the real goal behind any result produced by these models. The responsibility of XAI is to ensure that the user comprehends the reasoning behind the output, resulting in enhanced performance and more dependability for particular models. The XAI is embodied in the accompanying characters:

A. Clarification

The essential target of XAI is to give far-reaching substantiation to support the result. In artificial intelligence, there are five kinds of explanations: explanations for users, explanations to gain public confidence, explanations to

2022 International Conference on Industry 4.0 Technology (I4Tech) | 978-1-6654-7196-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/I4TECH55392.2022.9952865

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 30,2022 at 04:09:42 UTC from IEEE Xplore. Restrictions apply.

satisfy institutional and consistency demands, explanations to construct AI predictive models with AI calculations, and explanations for system administrators [7].

B. Significant

These guidelines suggest that XAI must be valuable for the administrator or organization as far as appreciating a clarification in an assortment of ways and managing genuine outcomes related to different levels of examination [7].

C. Accuracy

Accuracy is basic for XAI; it makes sense of the justification behind the result as well as gives a moment of knowledge into the dynamic interaction for a genuine issue. This will likewise build the end-trust clients in such models [7].

D. Information Sufficiency

Every innovation has its own downsides; by following this rule, XAI should make use of knowledge boundaries with respect to training data, as this will assist in the prevention of the development of inappropriate insights [7].

II. AI'S HIERARCHY:

AI algorithms can categorize, recognize, and interpret [8]. Using computer-based intelligence in medical services/biomedical facilities provides individuals with additional treatment alternatives. Still, knowing how far AI can go raises questions about its utility [9]. DARPA established the "three rushes of AI" to evaluate AI's strengths and downsides [10]. AI came in three waves:

A. Inaugural Wave : Aborginial Knowledge

Indigenous Knowledge calls AI a "master's framework" built up of human computations and programming based on input. They provide projects with logical, breakable concepts. AI relies on uniform standards and a framework that examines fundamental limits to determine the optimal way to tackle a problem. These frameworks can't grasp novel circumstances since human specialists establish their bounds [11].

B. The subsequent Wave: Quantifiable Education

This wave involves the least intervention of humans. Models based on facts may educate themselves to deliver correct results [12]. In many respects, these models can anticipate the future better than humans can. These prediction models are substructures of ANN (artificial neural networks) and feature several hidden layers that may learn autonomously using a training data. Each layer analyses the data to generate reliable findings [13]. The second wave framework is superior to the first wave. These models work well in many disciplines. These outcomes stimulate better direction and eliminate errors, saving time, money, and effort. The data defines how these models appear; therefore, they are "data subservient." These models fail to convey to the client how the layer operates inside [14].

C. The Ternary Wave: The Transformation Process

The third generation of AI constructs program-explaining models by uncovering logical rules that solidify their decision-making. These models will be built by the AI systems themselves. Third-wave structures employ real-world models to grasp situations and the dynamic cycle. The client may comprehend inner layer functions, which overcomes the AI/DL transparency issue [15].

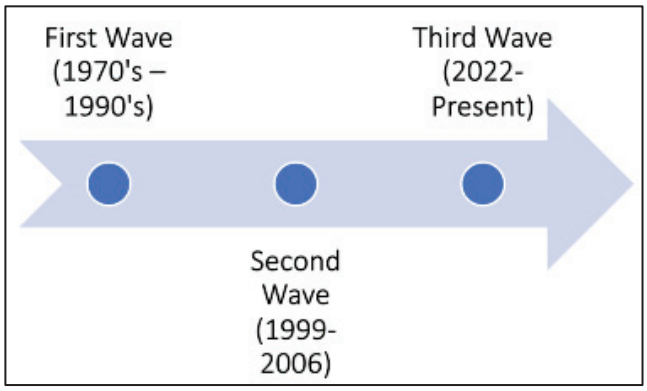


Fig.1 Three Waves of Artificial Intelligence.

A fourth wave is autonomous. This futuristic trend comprises devices with the ability to perceive and interact with the environment. This futuristic wave will permit AI to move and act financially, despite the fact that it is yet theoretical and models must be built [16].

III. AI'S FRAMEWORK

XAI is another human-made AI line. People must trust models to make progress in many areas. The quantitative, hard-to-figure models are less rational and more difficult. AI is the finest technique to persuade others. XAI structures are algorithms that analyze a model's behavior. SHAP, LIME, ELI5, What-if-device, AIX360, and Skaters are the most prevalent XAI frameworks. Imagine a Google-made gizmo that transforms how AI models are seen. AIX360 and Skaters are open-source model-understanding extensions. This helps construct a comprehensible AI framework, which is essential for real-world application cases [17].

A. SHAP

Shapley Additive Explanations also known as SHAP can give insight on a range of models, like logistic regression, tree-based models, and much more complex machine learning models like deep learning. SHAP is predicated on the Shapley value from game theory, which reveals how various factors can change the outcome. It also uses Shapley's value and their associated developments to combine credit allocations with local simplification [18].

B. LIME

LIME is similar to SHAP, except it can perform calculations at a considerably quicker rate and can employ many classes to interpret the results of an impermeable black-box model. The output of LIME is the contribution of each feature to the prediction of a data set. The ML classifier must be able to take in unprocessed text and calculate a probability for each class. Local justification for black-box models can be determined by observing how models respond to information that must be explained in a particular area. LIME can then determine, depending on what the model predicted, which contradictions in the model's predictions were the most persuasive.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 30,2022 at 04:09:42 UTC from IEEE Xplore. Restrictions apply.

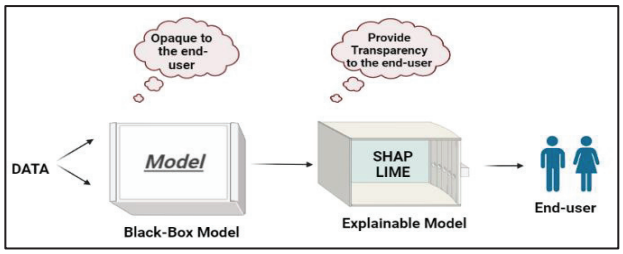


Fig.2 Illustrating SHAP and LIME to provide transparency to end-user.

IV. Methodology

Data collection: The data came from GitHub, and DevTox2D provided data from a correlation that compared 3D human tissue models and 2D human tissue models to predict neurotoxicity in developing brains. There are 180 rows and 11 columns in the data. These are the gene expression profiles that were taken from 3D and 2D models of human neural tissue that were exposed to different substances. The datasets are described in terms of the number of records per million (TPM), where each line represents a single example and each section embodies a gene. The names of the samples are listed in the first column. These names provide the chemical and the duration of the exposure, with the naming formats for the 2D and 3D datasets being distinct from one another.

Expression_2D:

The example name D39-C15 in the 2D tissue model dataset compares to the non-toxic substance c15 following 39 days of exposure. D27Hi-4 is the example name for the toxic synthetic t4 after 27 days of exposure.

Expression_3d (Used for our Model):

The sample names in the 3D tissue model collection all adhere to a consistent pattern. The 3D dataset, as opposed to the 2D dataset, comprises two biological replicates of each sample. This is in contrast to the 2D dataset. For instance, the name of the sample, "d2c15b," correlates with the non-toxic component "c15" after two days of exposure in the second biological copy. The name of the sample, "d7t4a," matches the hazardous chemical "t4" after seven days of exposure in the first biological duplicate.

Feeding dataset to ML classifiers: Randomly, the datasets were split 80:20 into sets for training and sets for testing. Because AI techniques such as SVMs, KNNs, and DL are rising in public prominence in disciplines such as omics data processing, data processing of structure and sequence, medical bioimaging, and signal data processing, we have decided to pioneer AI on our datasets. Three XGBoost models were created by running the ML classifier on training dataset. Extreme Gradient Boosting (XGBoost) is an AI method that uses decision trees to improve performance through a process called boosting. XGBoost systems are available for a number of scripting languages, but Python is by far the most popular. They work well with the well-known Scikit-learn AI package that Python data scientists use frequently. After configuring the XGBoost machine learning classifier with the Scikit-learn program on the training sets, the testing datasets were used to evaluate the models' confusion matrix and accuracy [20].

Applying XAI on trained XGBoost ML model: The SHAP library in Python was used to do an XAI investigation on the ready-made XGBoost models. Using an XAI investigation to look into model choices can help find things that make the model more confident in its predictions. For each data feature, the global importance was found by taking the average or mean value of all the elements. Then, this information was utilized to create the bar plot, which was used to highlight the most important and significant genes having a role in causing neurotoxicity.

Confusion Matrix: The NxN matrix is what tells us how well the model works for classification. The parts of the confusion are used to keep an eye on three important limits: precision, sensitivity, and specificity. Finding the best limit between classes will help you figure out what the classes should expect from the information in a grouping problem. One can find the best limit by weighing the benefits of precision, sensitivity, and specificity [20].

V. Results

Lastly, our model is now 86 percent accurate, which helps us to figure out which genes are most significant and how well our XGBoost prediction model works. According to SHAP summary charts, the most important genes in all datasets are ID4, WWC1, SRXN1, BMP7, GRIN3A, NCF2, DDN, F13A1, RNF207, RNASE1, SRPX, KDSR, CELF4, HIST1H3H, MMP12, SRSF6, PRKCSH, TAP1, DALRD3, and their effects on model predictions are both huge and good. With the accuracy of the dataset, we can test how our predicted genes affects prediction model and figure out what that means for our AI model to conclude about the causal genes of neurotoxicity. You can see the confusion matrix in the box below (Fig.3).

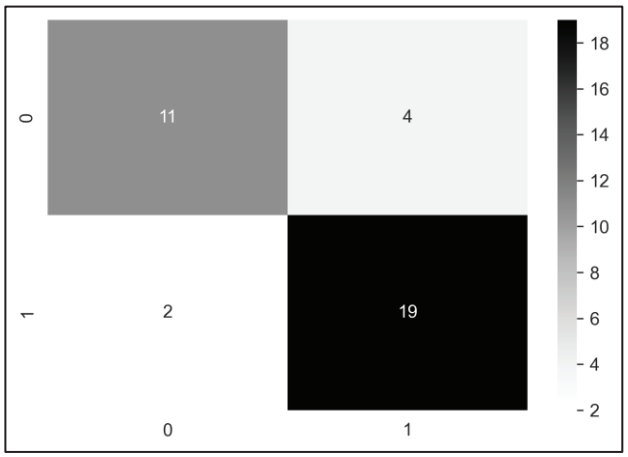


Fig.3. True positive (TP) cases are shown by the grey squares in the grid. False-positive (FP) and false-negative (FN) models are shown by the dark squares in the grid, and true negative (TN) cases are shown in a similar way by the white squares.

The positive and negative connections between the predictors and the target gene were visualized using SHAP summary graphs (Fig. 4). The genes are given preference in descending order of their attribute values; the horizontal arrangement indicates whether the presence of a gene is associated with the extent of toxicity prediction, indicating its impact on the inclusive model's prediction output; and the color indicates the consequence of a specific gene is substantial (red shade) or marginal (blue shade) in prediction of toxicity. For instance, ID4 gene is found to be most

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 30,2022 at 04:09:42 UTC from IEEE Xplore. Restrictions apply.

relevant as per the SHAP summary graph demonstrating its high influence on prediction model.

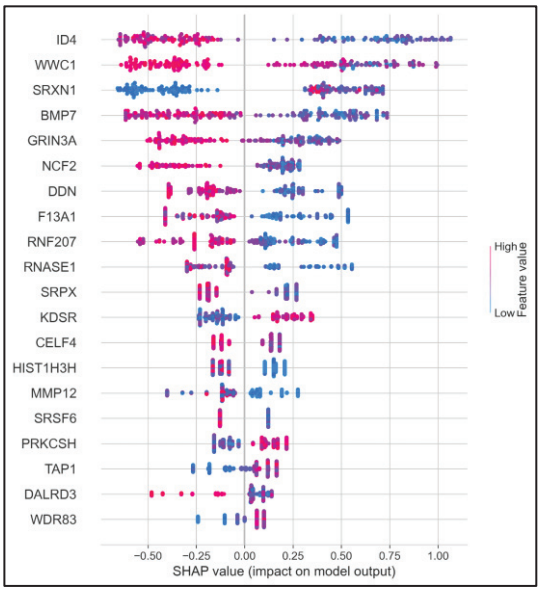


Fig.4 The plot of the SHAP summary shows the most relevant genes and their effect in causing neurotoxicity. The y-axis is utilized to rank the significance of every element on the y-hub. Shown on the x-axis are the consequences of a genes impact on the model's results. Red demonstrates a feature's impact is genuinely critical, while blue shows no impact by any means.

The prescient force of the ML model is helped by genes towards the highest point of the tree, as opposed to those at the base. Bar plots shown in (Fig.5) displays the main and most relevant genes at the top and the genes that are not impacting the prediction model at the base.

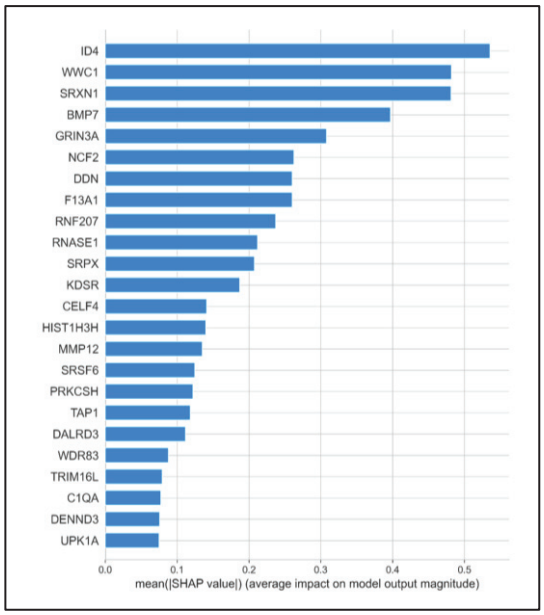


Fig.5 From the given bar plot, genes of most noteworthy significance namely ID4, WWC1, SRXN1, BMP7, GRIN3A, NCF2, DDN, F13A1, RNF207, RNASE1, SRPX, KDSR, CELF4, HIST1H3H, MMP12, SRSF6, PRKCSH, TAP1, DALRD3 were driven by carrying out SHAP values on the prepared models.

VI. TESTING CONTRIBUTION OF SUBSTANTIAL GENES ON PREDICTION MODEL

The precision of our model has been used to determine the genes that have had the greatest influence on it. The accuracy has increased significantly after these genes have been retrieved and then provided as an input once again in our XGBoost machine learning classifier. As a result, this demonstrates that our model was able to correctly identify the involvement of critical genes that influenced the accuracy and performance of the model. The confusion matrix (Fig.6) for those genes that were repossessed using SHAP values earlier showing an accuracy of 91.67 percent.

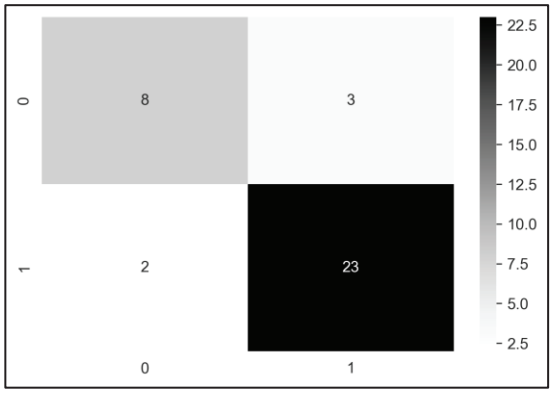


Fig.6 Confusion matrix shows the improved accuracy of the model containing all genes after applying SHAP values on previously determined genes of relevance.

In this SHAP summary plot (Fig.7) we can see the genes of highest relevance at top. This plot indicates the strong and positive impact of ID4 gene while SRXN1 gene shown to inversely correlated with the XGBoost prediction model.

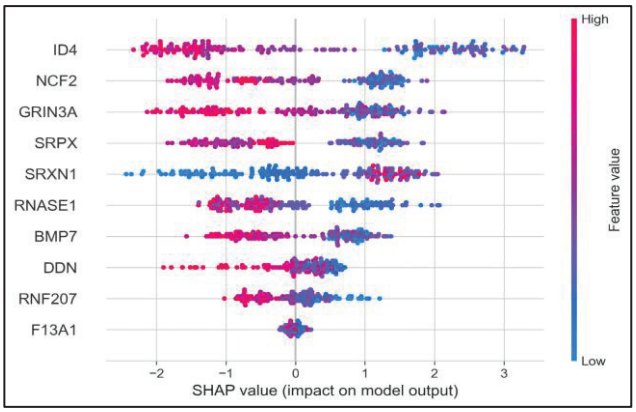


Fig.7 SHAP summary values of the validation model indicating ID4 gene to be of highest relevance in neurotoxicity.

Once more, SHAP values were utilized to produce a global feature significance plot. This bar plot (Fig.8) of feature significance assists us in determining, in descending order, which genes are the most important. Figure 8, depicts bar plots with the most significant genes on top and the least significant genes on the bottom. ID4 has the best predictive value and the greatest contribution to our ML prediction model, whereas F13A1 has the lowest contribution.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 30,2022 at 04:09:42 UTC from IEEE Xplore. Restrictions apply.

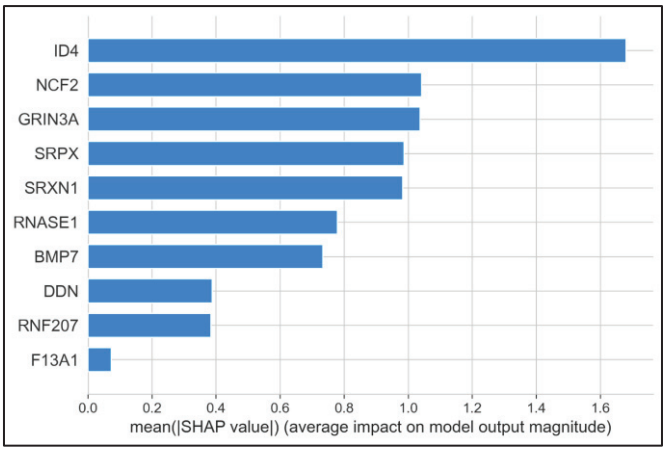


Fig. 8 showing SHAP Bar plot demonstrating the most significant genes for neurotoxicity

Top identified genes related to Neurotoxicity:

ID4: ID4 encodes the protein "Inhibitor of DNA Binding 4, HLH". This gene instructs the body to create an Inhibitor of DNA binding (ID) protein to disrupt protein attachment to DNA and help to control cell functions such as tumor development and disease propagation. ID proteins are a collection of four helix-loop-helix proteins in mammalian cells, govern cell growth, separation, senescence, cell cycle movement, angiogenesis, vasculogenesis, etc. Alzheimer's disease (AD) causes mental deterioration, unusual behavior, and loss of daily duties. Alzheimer's is the most prevalent neurodegenerative adult illness. The clumping of amyloid-beta-peptides, which make up Alzheimer's disease plaques, accelerates the disease's development. New research indicates that abnormal cell cycle re - entry may play a pivotal role in Amyloid-beta peptide induced neuronal degeneration. Numerous signaling intermediates, notably ID1, HIF-1, CDK5, and Shh, may aid to Amyloid-beta peptides instigated cell cycle re-entry in post-mitotic neurons; moreover, ID1 and CDK5/p25 inhibit the activity of each other. ID proteins may therefore have potential clinical implications in AD. [21].

SRXN1: Protein oxidation, which is caused by oxidative stress, is thought to play a role in mental illnesses like Parkinson, Alzheimer, and stroke. Sulfiredoxin-1 (SRXN1) is a protein that acts as a natural cancer-preventive agent and protects nerve cells. It has been shown that SRXN1 can protect against oxidative damage and could be used as an aim for neuro-protective mediation in studies of oxidative trauma [22].

BMP7: This gene expresses a secreted TGF-beta protein superfamily ligand. TGF-beta proteins are involved in cell growth and separation. The ligands of this family bind to a variety of TGF-beta receptors, which then recruits and activates SMAD family transcription factors, which are responsible for controlling gene expression. The proteolytic processing of the encoded preproprotein sets up each part of the disulfide-connected homo-dimer, which helps the bone, kidney, and brown fat tissue grow and develop. This protein has also been shown to speed up the growth of extra bones in animal models and help people heal from broken bones [23].

NCF2: Oxidative stress is caused by disequilibrium between the synthesis of reactive oxygen/nitrogen species and the organism's ability to counteract their harmful effects using antioxidative mechanisms. High quantities of reactive oxygen species (ROS) promote oxidative stress and disease propagation, whereas low levels are unhealthy. Changes in NADPH oxidase gene architectures reduce ROS generation. Hereditary mutations in phagocyte Nox2 oxidase (CYBB, CYBA, NCF1, NCF2, NCF4) cause chronic granulomatous disease (CGD), an inherited immunological condition [24]. Nox2 catalyzes superoxide generation by transferring electrons from cytosolic NAPDH to oxygen. Nox2 resides mostly on phagosomes or cellular membranes, where it is positioned to generate superoxide within phagosomal vesicles or in the extracellular space. Increased Nox2 activity has been linked to a number of neurodegenerative diseases, and it results in increased inflammatory response and neuronal cell death due to excessive oxidative stress [25].

VII. CONCLUSIONS & FUTURE SCOPE

The currently available methods of testing for toxicity are too time-consuming and pricey to be used for the purpose of conducting an exhaustive investigation of any newly discovered or poorly known chemical compounds. In spite of the fact that high-throughput screening methods come with a substantial assurance, it is possible that these models will continue to be used for a considerable amount of time to come. When taken as a whole, the consequences of AI-set yield are often proportionate to the need for XAI. When working with XAI, it is possible to place an emphasis on various points of view that have the potential to alter the final result. This is possible because XAI provides the administrator with valuable information regarding models that can be used to determine how to make use of those models in the future. XAI is constructed on a variety of existing framework. The ones that are utilized the most are SHAP and LIME, although there are a few others as well. Despite this, the progression of any new concept is directly correlated to the number of challenges it must overcome.

In this work, we looked at how to uncover competitor genes using XAI. More specifically, we looked at how to locate ID4, WWC1, SRXN1, BMP7, GRIN3A, NCF2, DDN, F13A1, RNF207, RNASE1, SRPX, KDSR, CELF4, HIST1H3H, MMP12, SRSF6, PRKCSH, TAP1, and DALRD3. The ability of this model to accurately predict neurotoxicity is significantly impacted by each and every one of these genes. We have once again used the result as the starting point for our analysis into the effectiveness of our model so that we may more thoroughly verify its performance. As a direct consequence of this, the accuracy of our model has been effectively raised from 86.67 percent to 91.67 percent.

People's capabilities are constrained by technology, yet on the whole, it benefits them. At the moment, AI models employ implicit mastery, which makes it difficult for humans to comprehend what they are trying to do. Because of this, technology has given rise to XAI, which bases its decisions on the explicit knowledge of people. The environment of artificial intelligence is made up of both explicit and implicit forms of information. Analysts are now bringing these two items together in a manner that might assist in finding solutions to future significant issues. It also demonstrates how eager tech enthusiasts are to make the most of XAI, which opens up a variety of research avenues and possibilities for the future.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 30,2022 at 04:09:42 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] F. Kuusisto, V. Santos Costa, Z. Hou, J. Thomson, D. Page, and R. Stewart, "Machine learning to predict developmental neurotoxicity with high-throughput data from 2D bio-engineered tissues," *Proc Int Conf Mach Learn Appl*, vol. 2019, p. 293, Dec. 2019, doi: 10.1109/ICMLA.2019.00055.

[2] K. Alikhademi, B. Richardson, E. Drobina, and J. E. Gilbert, "Can Explainable AI Explain Unfairness? A Framework for Evaluating Explainable AI," Jun. 2021, doi: 10.48550/arxiv.2106.07483.

[3] R. K. E. Bellamy *et al.*, "AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias," Oct. 2018, Accessed: Dec. 28, 2021. [Online]. Available: http://arxiv.org/abs/1810.01943

[4] J. Dodge, Q. V. Liao, Y. Zhang, R. K. E. Bellamy, and C. Dugan, "Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment," *International Conference on Intelligent User Interfaces, Proceedings IUI*, vol. Part F147615, pp. 275–285, Jan. 2019, doi: 10.1145/3301275.3302310.

[5] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models," Aug. 2017, doi: 10.48550/arxiv.1708.08296.

[6] N. Haefner, J. Wincent, V. Parida, and O. Gassmann, "Artificial intelligence and innovation management: A review, framework, and research agenda☆," *Technological Forecasting and Social Change*, vol. 162, p. 120392, Jan. 2021, doi: 10.1016/J.TECHFORE.2020.120392.

[7] Shane T. Mueller and Elizabeth S. Veinott, ", 'Principles of Explanation in Human-AI Systems,' ," 2021.

[8] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature 2015 521:7553*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[9] Mervis Jeff, "'When will we get there?,'" *Science V*, 2017.

[10] O. C. Jenkins, "Next Wave Artificial Intelligence: Robust, Explainable, Adaptable, Ethical, and Accountable A Computing Community Consortium (CCC) Quadrennial Paper".

[11] K. Kersting, "Rethinking Computer Science Through AI," *KI - Künstliche Intelligenz 2020 34:4*, vol. 34, no. 4, pp. 435–437, Dec. 2020, doi: 10.1007/S13218-020-00692-5.

[12] "The Four Waves of A.I. | Fortune." https://fortune.com/2018/10/22/artificial-intelligence-ai-deep-learning-kai-fu-lee/ (accessed Jul. 04, 2022).

[13] Y. Lecun, Y. Bengio, and G. Hinton, "Artificial Neural Networks as Models of Neural Information Processing," 2015.

[14] FeiyuXu and Hans Uszkoreit, "Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges," Sep. 2019.

[15] HolgerHoos and KristianKersting, "'The third wave of artificial intelligence,'" Dec. 2020.

[16] "Artificial Intelligence Tech Will Arrive in Three Waves." https://futurism.com/artificial-intelligence-tech-will-arrive-in-three-waves (accessed Jul. 04, 2022).

[17] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, "Designing theory-driven user-centric explainable AI," *Conference on Human Factors in Computing Systems - Proceedings*, May 2019, doi: 10.1145/3290605.3300831.

[18] M. Chromik, "reSHAPe: A Framework for Interactive Explanations in XAI Based on SHAP," 2020, doi: 10.18420/ecscw2020_p06.

[19] J. ¨ Urgen Dieber and S. Kirrane, "Why model why? Assessing the strengths and limitations of LIME," Nov. 2020, doi: 10.48550/arxiv.2012.00093.

[20] J. Meena and Y. Hasija, "Application of explainable artificial intelligence in the identification of Squamous Cell Carcinoma biomarkers," *Computers in Biology and Medicine*, vol. 146, 2022.

[21] S. der Chen, J. L. Yang, Y. C. Lin, A. C. Chao, and D. I. Yang, "Emerging Roles of Inhibitor of Differentiation-1 in Alzheimer's Disease: Cell Cycle Reentry and Beyond," *Cells 2020, Vol. 9, Page 1746*, vol. 9, no. 7, p. 1746, Jul. 2020, doi: 10.3390/CELLS9071746.

[22] Q. Li, S. Yu, J. Wu, Y. Zou, and Y. Zhao, "Sulfiredoxin-1 protects PC12 cells against oxidative stress induced by hydrogen peroxide," *Journal of Neuroscience Research*, vol. 91, no. 6, pp. 861–870, Jun. 2013, doi: 10.1002/JNR.23218.

[23] "BMP7 bone morphogenetic protein 7 [Homo sapiens (human)] - Gene - NCBI." https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=655 (accessed Jul. 04, 2022).

[24] "NCF2 gene: MedlinePlus Genetics." https://medlineplus.gov/genetics/gene/ncf2/ (accessed Jul. 04, 2022).

[25] Bido, S., Muggeo, S., Massimino, L. *et al.* Microglia-specific overexpression of α-synuclein leads to severe dopaminergic neurodegeneration by phagocytic exhaustion and oxidative toxicity. *Nat Commun* **12,** 6237 (2021). https://doi.org/10.1038/s41467-021-26519-x

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 30,2022 at 04:09:42 UTC from IEEE Xplore. Restrictions apply.

biosensors

MDPI

Article

Fabrication of a Molecularly Imprinted Nano-Interface-Based Electrochemical Biosensor for the Detection of CagA Virulence Factors of *H. pylori*

Kirti Saxena [1], Bayu Tri Murti [2,3], Po-Kang Yang [3], Bansi Dhar Malhotra [4], Nidhi Chauhan [1,*] and Utkarsh Jain [1,*]

[1] Amity Institute of Nanotechnology, Amity University Uttar Pradesh (AUUP), Sector-125, Noida 201313, India
[2] Graduate Institute of Biomedical Materials and Tissue Engineering, Taipei Medical University, Taipei 11031, Taiwan
[3] Department of Biomedical Sciences and Engineering, National Central University, Chung-li 32001, Taiwan
[4] Nanobioelectronics Laboratory, Department of Biotechnology, Delhi Technological University, Delhi 110042, India
* Correspondence: nidhichauhan2007@rediffmail.com (N.C.); utkarsh.jain@ddn.upes.ac.in (U.J.)

Abstract: *H. pylori* is responsible for several stomach-related diseases including gastric cancer. The main virulence factor responsible for its establishment in human gastric cells is known as CagA. Therefore, in this study, we have fabricated a highly sensitive MIP-based electrochemical biosensor for the detection of CagA. For this, an rGO and gold-coated, screen-printed electrode sensing platform was designed to provide a surface for the immobilization of a CagA-specific, molecularly imprinted polymer; then it was characterized electrochemically. Interestingly, molecular dynamics simulations were studied to optimize the MIP prepolymerization system, resulting in a well-matched, optimized molar ratio within the experiment. A low binding energy upon template removal indicates the capability of MIP to recognize the CagA antigen through a strong binding affinity. Under the optimized electrochemical experimental conditions, the fabricated CagA-MIP/Au/rGO@SPE sensor exhibited high sensitivity (0.275 μA ng^{-1} mL^{-1}) and a very low limit of detection (0.05 ng mL^{-1}) in a linear range of 0.05–50 ng mL^{-1}. The influence of other possible interferents in analytical response has also been observed with the successful determination of the CagA antigen.

Keywords: *H. pylori*; CagA; biosensors; molecularly imprinted polymers; molecular dynamics simulations; electrochemical

check for updates

Citation: Saxena, K.; Murti, B.T.; Yang, P.-K.; Malhotra, B.D.; Chauhan, N.; Jain, U. Fabrication of a Molecularly Imprinted Nano-Interface-Based Electrochemical Biosensor for the Detection of CagA Virulence Factors of *H. pylori*. *Biosensors* **2022**, *12*, 1066. https://doi.org/10.3390/bios12121066

Received: 21 October 2022
Accepted: 18 November 2022
Published: 23 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

Gastric cancer is the fifth most common and prevalent type of cancer that affects older males especially [1,2]. The incidence of gastric cancer mostly depends upon food habits and *Helicobacter pylori* (*H. pylori*) [3]. *H. pylori* infection is the main factor in 70% of cases of gastric cancer. It is a spiral-shaped, microaerophilic gram-negative bacteria that is responsible for various stomach-related diseases such as gastritis, peptic ulcers, and gastric cancer in advanced cases [4–6]. Many virulence factors of *H. pylori* are responsible for the establishment of infection in gastric cells or tissues [7–9]. CagA or cytotoxin-associated gene A is the most studied toxin of *H. pylori*, which is 120–145 KDa. CagA toxin is known to disrupt the normal differentiation of gastric epithelial cells that include cell polarity, cell adhesion, and cell migration inhibition [10–12]. Due to its role in the development of gastric cancer, it is classified as an oncogenic protein of *H. pylori* bacteria. CagA is also utilized as a biomarker in several studies for the successful detection of *H. pylori* [13]. Therefore, as a considerable biomarker, CagA is utilized herein for the development of a biosensing surface for the detection of *H. pylori*. However, conventional methods for *H. pylori* detection are available but they have several disadvantages such as low sensitivity, high cost, a requirement for skilled professionals, and painfully invasive methods (i.e., endoscopy) [14]. In order to overcome these limitations, biosensors show remarkable advantages over conventional

methods as they are quick, easy to handle, highly sensitive, and low-cost [15,16]. Biosensors of different types are characterized based on their transducer surface and biorecognition element. Natural receptors such as antigens/antibodies, enzymes, cells, DNA/RNA, aptamers, and peptides are used as biorecognition elements [17–19]. They exhibit a very high affinity for their target analyte but also have some challenges in laboratory applications such as poor durability, high-temperature sensitivity, and low stability in high/low pH solutions. In this context, imprinted polymers have been reported as robust artificial receptors/biomimetic materials. They have great potential to overcome the above-mentioned limitations of natural receptors for the recognition of desired analytes. The molecular imprinting technique used the stereochemistry of the targeted analyte for specific molecular recognition in a simple and highly effective manner. Molecularly imprinted polymers, or MIPs, are chemically and physically stable and able to form specific cavities in polymer matrices corresponding to binding analytes. The synthesis process involved the printing of a macromolecular structure onto the polymeric material's surface using a template molecule to create similar sites to the analyte. The basic synthesis principle of MIP is based on a process in which the copolymerization of functional monomers and cross-linked polymers takes place in the presence of a target analyte. Furthermore, the removal of the template from the polymer matrix after the polymerization process creates specific imprinted cavities (recognition sites) complementary in shape, size, and functionality to the template molecules. Thus, MIPs are attractive to researchers for the development of biosensors for the detection of several infectious diseases [20]. MIPs can be categorized into non-covalent and covalent imprinting on the basis of imprinting strategy. However, the fabrication of MIPs involved various methods; among them, the electropolymerization technique provides easy adherence of the polymeric film on the electrode surface, thus enabling it to control the film thickness of any size and shape. The electropolymerization technique is also a simple and easy-to-handle method [21].

Nanomaterials have also played an important role in the fabrication of electrochemical-based sensing platforms [22,23]. The best-known nanomaterials, reduced graphene oxide (rGO) and gold (Au), were employed in this study to amplify the signals on the electrode surface. rGO and Au nanomaterials provide high conductivity, large surface area for immobilization of biomolecules, and facilitate electron transfer on an electrode [24,25]. Therefore, in the present work, a screen-printed carbon electrode coated with rGO and gold nanoparticles was utilized as a matrix for the fabrication of a CagA detection platform [26,27]. For this, CagA-specific MIP was immobilized on an Au/rGO/SPE surface, which provides a biocompatible and electron-conductive platform. The molecular dynamics (MD) simulations were carried out to predict the optimized molar ratio of the MIP prepolymerization along with their possible modes of interaction. Density functional theory (DFT) calculations were also incorporated to assess the electronic properties and the active sites of the MIP component. Upon MD simulations, the binding energies were calculated, emphasizing the excellent functionality of the MIP-based electrodes in the as-prepared electrochemical biosensor.

2. Materials and Methods

2.1. Chemical Reagents Used

CagA antigen and antibodies were procured from the Max von Pettenkofer Institute of Hygiene and Medical Microbiology, Ludwig Maximilian University, Munich, Germany and stored at −20 °C till further use. The monomer, cross-linker, initiator, and methanol were purchased from Sigma-Aldrich, India, and were used without any further purification. Other chemicals, including $K_3[Fe(CN)_6]$, $K_4[Fe(CN)_6]$, and KCl, were purchased from Sisco Research Laboratory (SRL), Mumbai, India. All other chemicals were of analytical reagent (AR) grade and all aqueous solutions were prepared in double distilled water throughout the experiments. Cyclic voltammetric (CV) and electrochemical impedance studies (EIS) were performed using Bio-Logic Science Instruments SP-200 with EC-Lab® software with an impedance analysis module. The working electrode was the screen-printed carbon

electrode (SPE). Scanning electron microscopy (SEM) and Brunauer–Emmett–Teller (BET) were used to study the surface morphology and surface area of synthesized MIP with and without template CagA. The Material Studio version 2018r1 software package (Accelrys Inc., San Diego, CA, USA) [28] was used to conduct computational simulations.

2.2. Molecular Dynamics Simulations

All the preparation and geometry optimization of the bulk molecules during MIP prepolymerization were conducted using Material Studio version 2018r1 software package (Accelrys Inc., San Diego, CA, USA). To accurately mimic the experimental condition, the primary virulence determinant of *H. pylori* (i.e., CagA (PDB ID: 4IRV) [29]) was used as the template for MIP, while methyl methacrylate (MMA) served as the functional monomer of the polymerization process (Figure 1a,b). Ethylene glycol dimethacrylate (EGDMA) and acetonitrile (ACN) were employed as the cross-linker and the porogen, respectively (Figure 1c,d). For the small organic molecules (i.e., MMA, EGDMA, ACN), the molecular geometries were optimized using spin-restricted DFT described in the DMol3 code [30,31] with the B3LYP hybrid functional [32,33]. To evaluate the electronic properties of the studied molecules, a double numerical atomic orbital (DNP) basis set with a 3.5 basis file was used along with 3.7 Å of global orbital cut-off. The optimum energy, gradient, and displacement convergences were set to be 1×10^{-5} Ha, 2×10^{-3} Ha/Å, and 5×10^{-3} Å, respectively. Instead, the density functional-based tight binding (DFTB+) [34] method was considered to optimize the MIP template due to the macromolecular structure of the CagA protein which contains thousands of atoms. Non-self-consistent charge (SCC) DFTB was used herein to achieve the convergence during optimization along with the 3rd order parametrization for organic and biological systems (3ob) of the Slater−Koster library and 0.005 Ha smearing parameter. The energy tolerance was recorded to be 1×10^{-1} kcal/mol throughout geometry optimization.

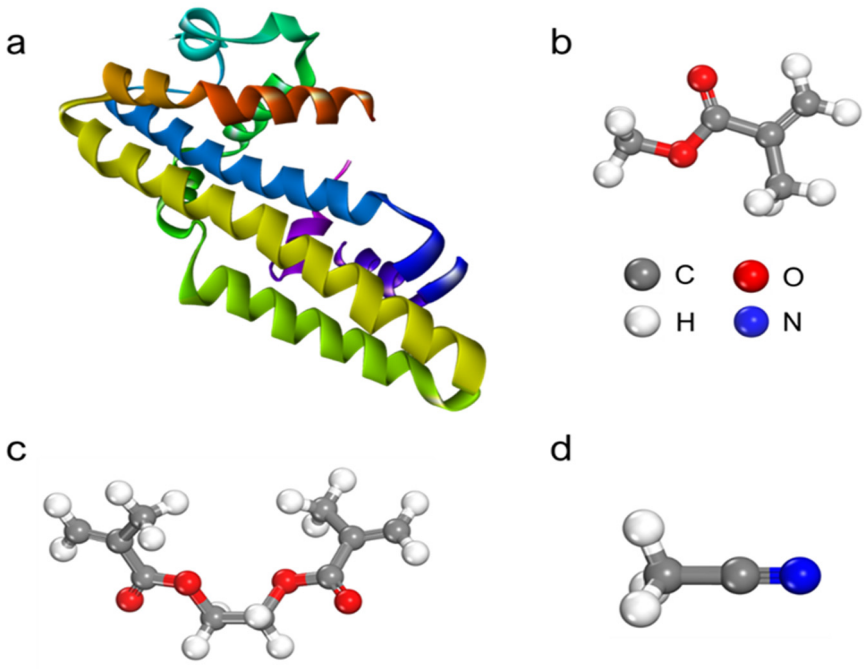


Figure 1. Chemical structures of MIP optimization components including (**a**) CagA, (**b**) MMA, (**c**) EGDMA, and (**d**) ACN. CagA was represented in the ribbon diagram, while other molecules were displayed in the ball and stick model.

Before the MD simulation of MIP prepolymerization systems, all studied molecules were energetically minimized using a Condensed-phase Optimized Molecular Potential for Atomistic Simulation Studies (COMPASS) force field [35,36] with 5000 steepest descent

and 5000 conjugate gradients to eliminate any bad contacts [37]. To provide a suitable environment for the molecular simulation, five periodic cells were then constructed with the 0.1 g/cm^3 density and 298 K bias temperature containing all polymerization components for MIP synthesis. As shown in Table 1, the periodic cells were loaded with different molar ratios of MIP components in accordance with the experiment. The all-periodic systems were then optimized under the Smart algorithm and the COMPASS force field to reach their minimum energy configuration.

Table 1. Prepolymerization mixtures for the molecular simulation of MIP.

Polymer	CagA (mmol)	MMA (mmol)	EGDMA (mmol)	ACN (mmol)	Molar Ratio
MIP 1	1	2	2	80	1:2:2
MIP 2	1	4	4	80	1:4:4
MIP 3	1	6	6	80	1:6:6
MIP 4	1	8	8	80	1:8:8
MIP 5	1	10	10	80	1:10:10

The MD simulations were conducted in several stages with the use of the COMPASS force field to achieve the accurate and final energy-optimized systems. First, the polymer systems were equilibrated under canonical ensemble (i.e., constant number of particles, volume, and temperature ensemble; NVT) at room temperature (298 K) with a time step of 1.0 fs and a total simulation time of 200 ps. Canonical ensemble allows control over both the volume and temperature during simulation. Afterwards, the MD was carried out by heating up the system to 333 K for 200 ps with canonical ensemble, followed by cooling the systems back to normal temperature (298 K) with the similar dynamics parameter. This step was implemented to allow the prepolymerization system to cross over the energy barriers as well as to reproduce the experimental temperature during polymerization. The Nosé−Hoover thermostat was used throughout the equilibration stage to control the temperature of the NVT simulation. After the optimization and equilibration steps, a simulated annealing process was introduced over all periodic systems with the mid-cycle temperature of 698 K. This step was followed by gradual cooling at room temperature in 10,000 steps under the COMPASS force field and a constant number of particles, volume, and temperature (NPT) ensemble. The Nosé−Hoover thermostat was applied to control the temperature, while the Andersen barostat was used to control the pressure during the simulation. Lastly, the systems were re-equilibrated at 298 K for 200 ps with an NPT ensemble. The Andersen algorithm was applied to control both the temperature and pressure of the system. Radial distribution functions (RDFs), denoted as g(r), were analyzed upon MD simulations to evaluate the structural characteristics of the prepolymerization system and the probabilities of reactive atoms of the functional monomer at various separation distances from the MIP template (i.e., CagA) [38].

2.3. Binding Energy Calculation

The binding energy (E_b) of the prepolymerization systems were calculated as per Equation (1):

$$E_b = E_{complex} - E_{component} \quad (1)$$

where $E_{complex}$ is the total energy of the MIP complex and $E_{component}$ is the total energy of the MIP component including the monomer, template, cross-linker, and porogen. Total energy is defined as the sum of valence energy and non-bond energy upon geometry optimization with the Smart algorithm and COMPASS force field. The convergence tolerances for energy, force, stress, and displacement were recorded to be 2×10^{-5} kcal/mol, 1×10^{-3} kcal/mol/Å, 1×10^{-3} GPa, and 1×10^{-5} Å.

2.4. Synthesis of Nanomaterials

The gold nanoparticles (AuNPs) and rGO were synthesized by using earlier reported protocols. According to the protocol, hydrogen tetrachloroaurate was dissolved in distilled water and kept at 100 °C with continuous stirring. Five ml of 1% trisodium citrate was

heated separately and quickly added into a boiling solution of hydrogen tetrachloroaurate. The yellowish color of the solution became transparent and, finally, a wine red color was obtained. The solution was further heated up for 15 more minutes and allowed to cool at room temperature [39]. Furthermore, rGO was synthesized by using a modified Hummer's method. Two gm of graphite powder was measured and dissolved slowly in 50 mL sulphuric acid (H_2SO_4). The flask was kept in an ice bath continuously and left undisturbed for 30 min. After that, 6 gm of $KMnO_4$ was slowly added to the above solution. The prepared suspension was stirred continuously for about 2 h and then sonicated. Next, 150 mL of distilled water was slowly added to dilute the sulphuric acid and a further 8 mL H_2O_2 was added to stop the reaction. Immediately, the yellow color of the solution was observed. The prepared solution was centrifuged, and the obtained black powder was washed with diluted HCl and distilled water several times. The obtained product was vacuum-dried for 24 h and the final product was dried and stored until further used. Furthermore, to prepare rGO, 0.1 gm graphene oxide powder was dissolved in 150 mL dist. water and sonicated for 2 h. The obtained brown solution continuously stirred in a paraffin oil bath and 1 mL of hydrazine monohydrate was added dropwise very slowly. Again, the solution was kept on the stirrer for 18 h at 90 °C. The color of the mixture changed to black and when the stirring stopped, the hydrophobic particles seemed to settle down at the bottom [40,41].

2.5. Synthesis of CagA-Specific MIP

The MIP was synthesized by using a bulk polymerization and electrochemical polymerization method with an optimized ratio of methyl methacrylate (MMA) (monomer), EGDMA (cross-linker), and dilution of CagA analyte. Briefly, monomer, cross-linker, and CagA antigen were mixed together in acetonitrile (ACN) with an initiator Azobisisobutyronitrile (AIBN). This prepared solution was continuously stirred for 12–16 h at 70 °C for bulk polymerization. After the polymerization process, the obtained product was crushed and dried in a vacuum oven overnight to completely remove the moisture. Again, the CagA antigen removal was performed by using methanol and acetic acid in a 4:1 ratio twice and pure methanol once. The end-product was further dried and kept in vacuum-dried condition till further use. The prepared MIP with antigen and without antigen (after washing) were characterized through SEM analysis of the polymer matrix.

2.6. Fabrication of Biosensing Surface

2.6.1. Electrodeposition of Nanomaterials on an Electrode Surface

The screen-printed electrodes (SPE) were used for the fabrication of the detection platform. rGO and gold nanomaterials were deposited on SPE using the cyclic voltammetry method. First, the rGO was deposited on the SPE by using a potential range of 0.5 to −1.5 for 6 cycles at a 50 mV/sec scan rate. Again, the gold nanomaterial was deposited through electrodeposition at a 50 mV/s scan rate with a 0.3 to −0.5 potential range for 5 cycles [42,43].

2.6.2. Electropolymerization of CagA-MIP

The rGO and gold nanomaterial-decorated SPE (Au/rGO@SPE) was further modified with CagA-specific MIP. The modification step was performed by electropolymerization (CV) in a potential range of −0.2 to 0.6 V with a scan rate of 50 mV/s for 15 cycles. To remove the CagA template from MIP, the SPE surface was washed with methanol and acetic acid mixture 2 times and only methanol 1 time each for 15 min and dried at room temperature [44].

2.7. Electrochemical Response Study

The fabricated sensing platform was studied electrochemically after each modification step to analyze the reaction response on the SPE surface. All the measurements were taken in the presence of electrolytes containing 0.1 M $K_3[Fe(CN)_6]$ and $K_4[Fe(CN)_6]$ by using a three-electrode cell configuration. The study involved electrochemical measurement

through cyclic voltammetry (CV), electrochemical impedance spectroscopy (EIS), and differential pulse voltammetry (DPV) [45].

2.8. Study of the Analytical Performance of Fabricated Biosensor

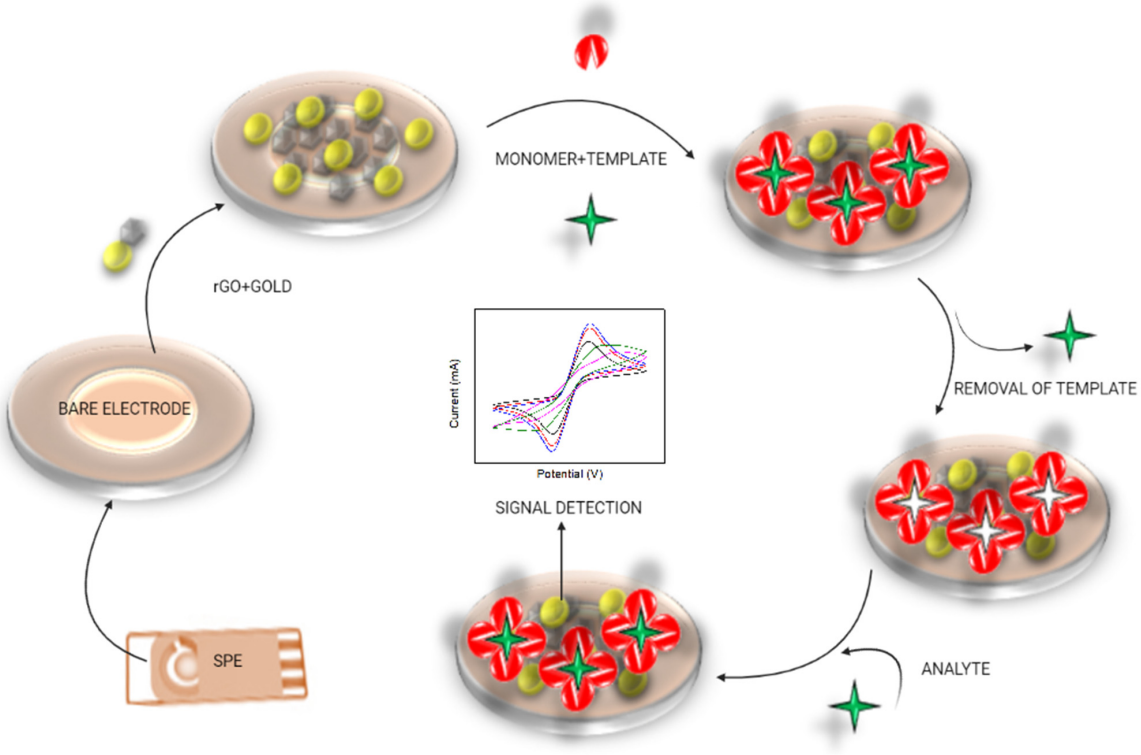
The analytical performance of the fabricated sensing platform was examined with different concentrations of the CagA template. The electrode was also optimized for different pH and temperatures. The effect of different pH (ranging from 4–10) was observed in the presence of 0.1 PBS buffer prepared by the standard method. To optimize temperature for better functionality of the fabricated sensing electrode, various temperatures were applied such as 10, 20, 30, 40 50, 60, and 70 °C. Other than these basic parameters, the selectivity and stability of the developed electrochemical biosensor were assessed by observing the current response in the presence of various interferents like acetylcholine (Ach), ascorbic acid, glucose, uric acid, and BabA and VacA (*H. pylori*-specific other antigens). Furthermore, the stability of the fabricated electrode was studied for up to 60 days by performing DPV once per week under optimum conditions.

2.9. Real Sample Analysis

The efficacy of the fabricated sensing platform was observed by electrochemically measuring the analytical performance in the presence of spiked blood samples. For this, blood samples were spiked with known concentrations of the analyte CagA antigen. The prepared spiked blood samples were applied to the working electrode and incubated at room temperature for 10 min. Then, after 10 min, the current response was measured using the CV technique [46].

3. Results and Discussion

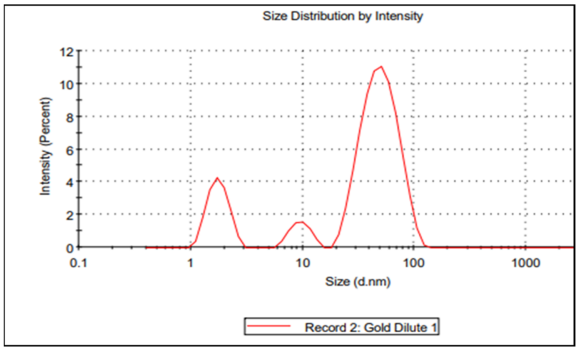
A novel MIP-based electrochemical biosensor was constructed for the detection of the CagA antigen of *H. pylori* bacteria as illustrated in Scheme 1. The deposition of nanoparticles on the electrode surface facilitates the deposition of the biorecognition element by providing a high surface area and a remarkable electrochemical response toward the target analyte.



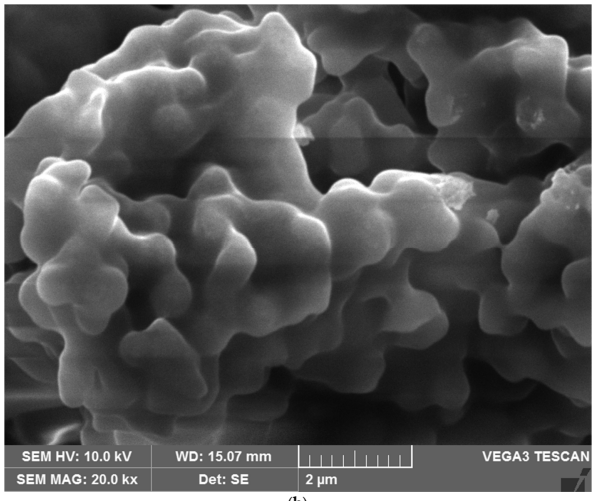
Scheme 1. Graphical representation of the fabrication steps involved in the development of the MIP-based biosensor for the detection of the CagA virulence factor of *H. pylori*.

3.1. Morphological Study of Synthesized CagA-MIP by SEM

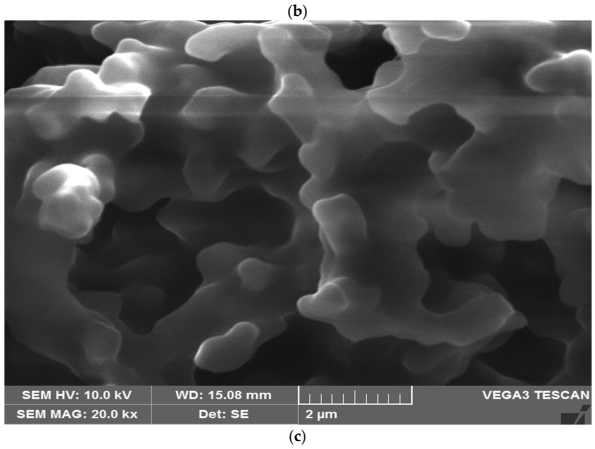
The size of synthesized gold nanomaterials was observed via the dynamic light scattering (DLS) method. The particles are able to scatter light depending on their diameter to the sixth power. The size of particles measured by DLS is the hydrodynamic diameter of the spheres and is influenced by all the adsorbed substances on the surface of the nanoparticles. Figure 1a indicates that the monodisperse nanoparticles with sizes 60 ± 20 nm are present in colloidal solution. The intensity is highest for larger particles, as can be seen in Figure 2a, whereas the middle peak shows the lowest intensity which may be due to the coagulation of nanoparticles in the sample or the presence of any impurities on the particle's surface. A possible reason behind this could be the improper sonication of samples to obtain a homogenous solution before measurement.



(a)



(b)



(c)

Figure 2. *Cont.*

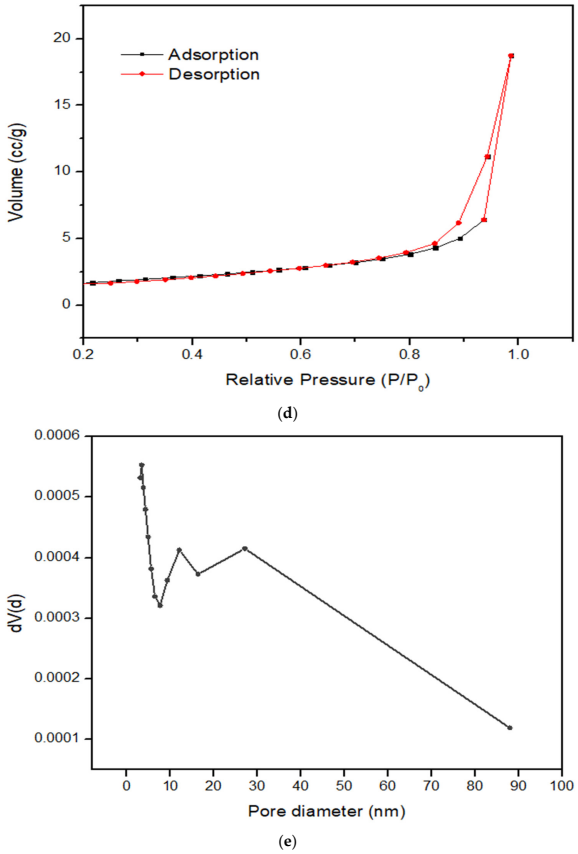


Figure 2. (**a**) Determination of the size of gold nanoparticles via the DLS technique. (**b**) Surface characterization of synthesized MIP with CagA template. (**c**) Surface characterization of synthesized MIP after removal of the CagA template. (**d**) BET isotherm; adsorption and desorption curve. (**e**) Analysis of pore size distribution for MIP by BET.

The synthesized MIP was characterized by SEM to study the surface morphology of the polymer matrix [47,48]. SEM images of unwashed MIP and washed MIP (after removal of CagA antigen) are shown in Figure 2b,c. Figure 2b shows the smooth surface of CagA-MIP before the removal of antigen, which is obviously different from Figure 2c where the rough surface with a large number of holes represents the removal of the template from the polymer matrix. The presence of cavities as shown in Figure 2c indicates the successful imprinting of CagA antigen in the polymer matrix. Pore size distribution and nitrogen adsorption−desorption isotherm of synthesized MIP were observed through BET analysis after removing the template. Figure 2d,e described the adsorption and desorption curve and pore size distribution, respectively. The BET analysis shows the microporous structure of synthesized MIP. Figure 2d depicts the absorption and desorption isotherm of MIP that showed monolayer adsorption. The presence of micropores provides large surface area to access electrode materials very quickly for active species. According to this characterization, the BET surface area of MIP was observed as 5.074 m^2/g. The pore diameter for most of the MIP's pores are centered at approximately 10 nm and 28 nm, which showed a bimodal size distribution (Figure 2e). The pore volume and pore diameter obtained for MIP was 0.029 cc/g and 3.374 nm, respectively, as shown in Table 2. From this BET isotherm and pore size distribution analysis, MIP is expected to facilitate electrochemical applications [49,50].

Table 2. Parameters for synthesized MIP.

Surface Area	**5.074 m^2/g**
Pore Volume	0.029 cc/g
Pore Diameter Dv(d)	3.374 nm

3.2. Geometry Optimizations

Computational simulation has been effectively used to elucidate the nano−bio interaction in biosensor platforms [51,52]. In this study, the polymerization process of MIP was optimized through molecular modelling in order to determine the energetic contributions towards the binding event of the MIP template. To realistically simulate the MIP component, the overall molecular structures of MIP were constructed and energetically optimized. In the first step, the electrostatic potential (ESPs) and the highest occupied molecular orbital (HOMO)−lowest unoccupied molecular orbital (LUMO) isosurfaces were analyzed via B3LYP-based DFT optimizations to observe the electron localization and reactive sites of the functional monomer. The ESPs display the region of the electric field where yellow- and blue-colored areas represent negative and positive potentials, respectively [53]. As seen in Figure 3a, the electron cloud is mostly accumulated around the O atoms of the MMA due to their high electronegativities as compared to the C-H backbone, which tends to be the electropositive area. ESPs enunciate the reactive site of the MMA, which is predominantly driven by electronegative O atoms. These two O atoms in the monomer could form recognition imprinting sites toward the template molecule during MIP synthesis. Frontier orbital analysis was further studied to assess the chemical reactivity of the MMA (Figure 2b,c). It is also well-known that HOMO−LUMO plots depict the electronic density and, therefore, demonstrate the accumulation of reactive sites in molecules [54]. As a result, HOMO−LUMO lobes were mainly delocalized at the O atoms and neighboring C backbone of the monomer, with a small contribution coming from the atoms which are away from the electronegative oxygens. This result emphasizes the chemical reactivity of the MMA functional monomer, which is predominantly derived from the O atoms, in agreement with the ESP analysis.

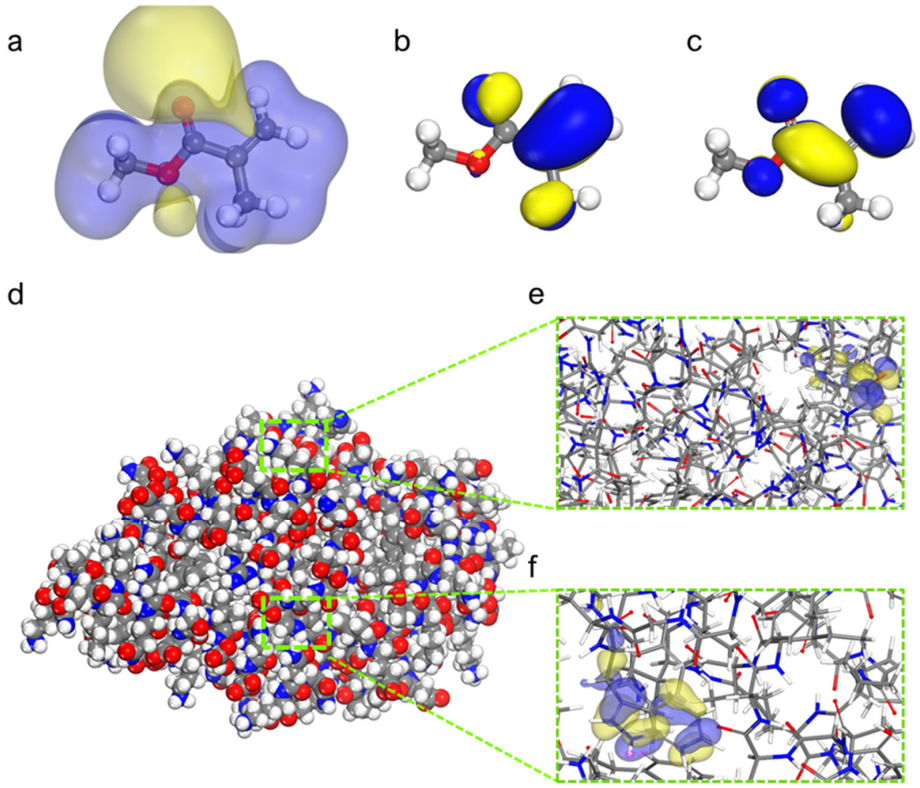


Figure 3. (**a**) Electrostatic potential, (**b**) HOMO, and (**c**) LUMO isosurfaces of MMA after geometry optimization using the B3LYP hybrid functional. (**d**) Optimized structure, (**e**) LUMO, and (**f**) HOMO isosurfaces of CagA after geometry optimization with DFTB+ with 3ob Slater−Koster library.

Furthermore, the optimized structure and frontier orbital isosurfaces of the CagA template is depicted in Figure 3d–f. After the DFTB+-based geometry optimization, the

HOMO isosurface was primarily distributed in tryptophan (Trp) amino acid, while the LUMO contour was delocalized in phenylalanine (Phe). The results show that the main reactivity of the CagA was driven by two aromatic amino acids in the protein building block due to the resonance delocalization of π-electrons in the benzene ring of Phe and indole ring of Trp [55,56]. However, due to the lack of hydrogen donor in the Phe side chain, only Trp was included in the RDF observation in subsequent analyses.

3.3. Molecular Dynamics Simulations and Binding Energy Calculation

The proximity of the CagA to MMA degrees are discussed in this section from the RDF standpoint after the MD treatments. Figure 4 shows the labeled atoms at the MMA monomer and Trp of CagA that could form recognition imprinting sites on the basis of ESPs and frontier orbital analysis from DFT calculations [37,57,58]. In MMA, two oxygen atoms (i.e., MO1 and MO2) were assigned as the possible hydrogen bonding acceptors (Figure 4a), while the TH at Trp was assigned as the proton donor for the RDF analysis (Figure 4b). In general, RDF is essential to the statistical mechanics and demonstrates the probability of finding a particle in an infinitesimal shell at a distance r from another particle in the liquid state [59]. RDF is useful to predict the conditions of the MMA molecules around the CagA template [37].

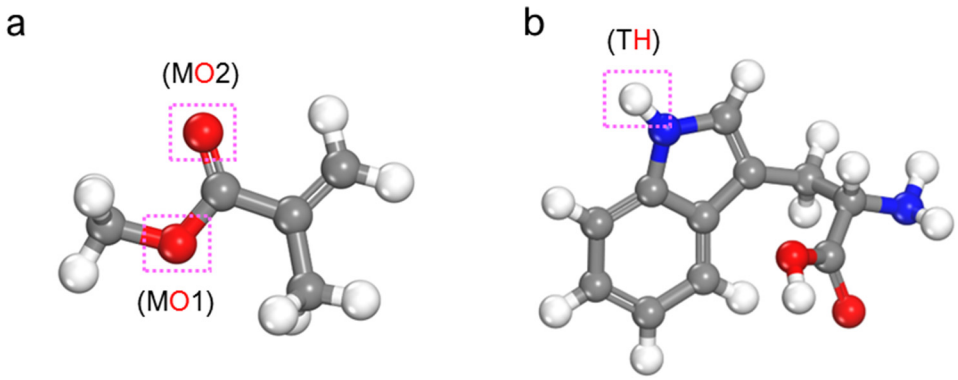


Figure 4. Chemical structures of (**a**) MMA and (**b**) tryptophan of CagA along with their labelled atoms for RDF analysis.

As a result, TH (atom H labeled in Trp)−MO1 (atom O1 labeled in MMA) atomic contact appeared (Figure 5a). Among other MIP optimization mixtures, MIP 2 shows the lowest radius (distance) to TH, indicating a stronger molecular interaction between functional monomers and CagA. Due to the 5–6 Å radius recorded to TH, the TH, the TH−MO1 in the MIP 2 interaction may be dominated by weak hydrogen bonds. The highest RDF was observed in MIP 2 at 13.7 of g(r) peak with 5.62 Å of r, which demonstrates a greater possibility of finding MMA at a closer atomic distance to the CagA template. Meanwhile, the RDF of the TH−MO2 is depicted in Figure 5b. Herein, both MIP 2 and MIP 5 showed very close radii to TH with a higher g(r) peak observed in MIP 2 than in MIP 5 at a closer distance to Trp. In terms of MO1 and MO2 interactions in MIP 2, MO1 is predicted to be the main interaction site of monomer with CagA due to its comparatively stronger action site than MO2. These interactions may belong to weak hydrogen bonding rather than strong hydrogen bonding at a short distance (~2.7–3.0 Å) [60,61]. Thus, from the RDF point of view, the adsorption capacity of MIP 2 could be greater than other studied MIP mixtures shown in Table 1. This finding is in good agreement with the experimental molar ratio during MIP synthesis.

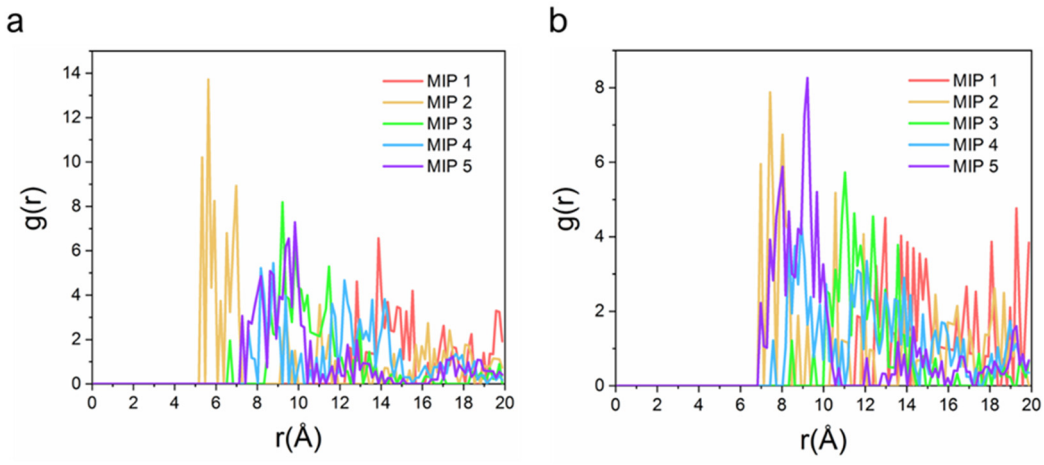


Figure 5. RDFs of (**a**) TH−MO1 and (**b**) TH−MO2. g(r) shows the RDF values, while r(Å) represents the distance from a solute atom and the average bulk atom.

Binding energy was further characterized to elucidate the interaction energy of the MIP complex, particularly before and after template removal. It is well-known that template removal is a critical step in the synthesis to form an imprinted site which is responsible for the recognition properties of the MIP [62]. The E_b was calculated according to Equation (1) from the MIP 2 as the optimized mixture of MIP preparation (Table 3). As a result, E_b before the template removal was significantly lower than after template removal, indicating a high interaction strength between the receptor (MIP) and target (CagA). This could lead to the high sensitivity of the biosensor for the recognition of the specific target. After the formation of the imprinted site, it is well expected that E_b remains higher due to the absence of stable CagA protein in the complex system. As seen in Table 3, CagA possesses a remarkably low total energy than other molecules (i.e., −6878.95 kcal/mol). Indeed, a low E_b upon template removal indicates the capability of MIP to recognize the biosensor target (CagA) through a robust binding mode and affinity.

Table 3. Total energy and binding energy of the MIP prepolymerization system.

Complex/Molecules	Total Energy (kcal/mol)	Binding Energy (kcal/mol)
Before template removal (CagA-MMA-EGDMA-ACN)	−9049.77	−886.48
After template removal (MMA-EGDMA-ACN)	−1317.29	−32.94
CagA	−6878.95	
MMA	−53.32	
EGDMA	−153.70	
ACN	−1077.32	

3.4. Electrochemical Experiment Results

3.4.1. Validation of Fabrication Steps through Electrochemical Study

The electrochemical behavior of the modified electrode was investigated at each step by cyclic voltammetry and electrochemical impedance study (EIS) curves. Performance was determined through a comparison of CV signals after each deposition step. Figure 6a shows the CV curves in a potential range of −0.8 to 0.8 V, belonging to the bare electrode, rGO and gold NP-coated electrodes, MIP-modified electrode, and the electrode modified by MIP after the removal of the template in 0.1 M ferro/ferri solution. After modification with rGO and gold NPs, the current intensity was increased in comparison to bare SPE due to the high conductivity and extended surface area of rGO and gold. Further modification of SPE by MIP decreased the current intensity, which indicates the large transfer resistance of polymer-modified SPE. This could be attributed to the large electron transfer resistance

of MIPs. After the removal of the template, the current increased significantly due to the formation of CagA-specific imprinting cavities that resulted in the enhancement of electrolyte diffusion and the acceleration of electron transfer on the SPE surface. The same steps were validated through EIS and also showed the resistance developed at each step in Figure 6b [63].

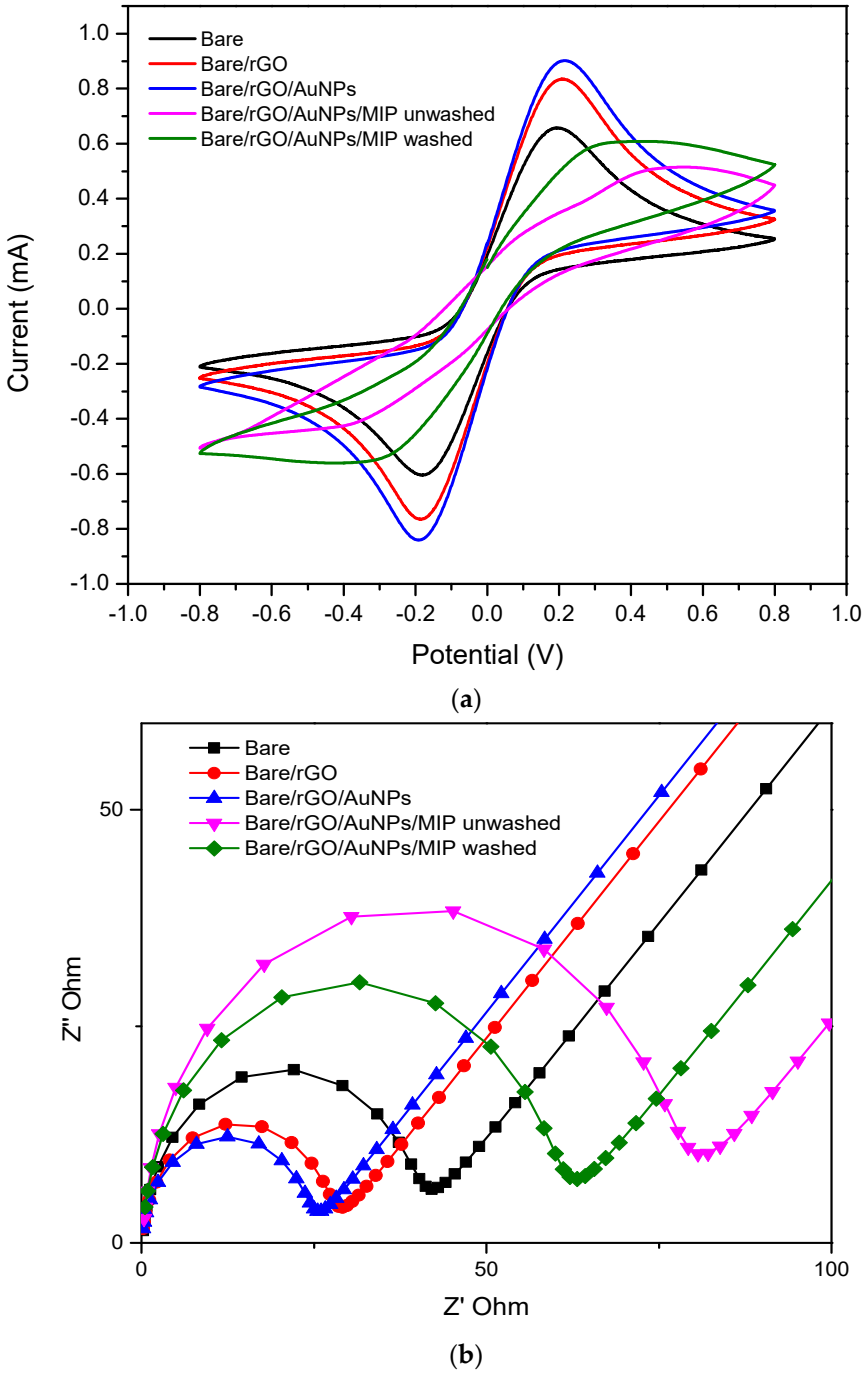


Figure 6. (**a**) Electrochemical response studies via the CV technique for Bare SPE (black), rGO@SPE (red), AuNPs/rGO@SPE (blue), MIP unwashed/AuNPs/rGO@SPE (green), and MIP washed/AuNPs/rGO@SPE (magenta) obtained in the presence of $[Fe(CN)_6]^{3-/4-}$ electrolyte at 50 mV/s. (**b**) The EIS study of Bare SPE (black), rGO@SPE (red), AuNPs/rGO@SPE (blue), MIP unwashed/AuNPs/rGO@SPE (magenta), and MIP washed/AuNPs/rGO@SPE (green) at the frequency range of 10–2000 Hz.

3.4.2. Optimization of Fabricated Electrodes for Various Sensing Parameters

To achieve the optimum sensing conditions like pH and temperature, the fabricated electrode was examined at various ranges. The influence on sensing performance due to the effect of various pH values is illustrated in Figure 7. The increase in the pH value from 4 to 10 reflected the improvement in electrode performance. However, after 7, the current intensity decreased to the pH range of 8 to 10, which indicates the binding of MIPs was prevented beyond 7 pH. Thus, the optimum pH was observed as 7 based on which MIPs showed remarkable recognition performance. The fabricated sensing platform was also examined with varying temperatures in the range of 10–70 °C. The bar graph was plotted against the different DPV currents at various temperatures (Figure 8). The DPV current was increased from 10- 40 °C and then dramatically decreased beyond 40 °C. Thus, we selected 40 °C as the optimum temperature for this fabricated electrode.

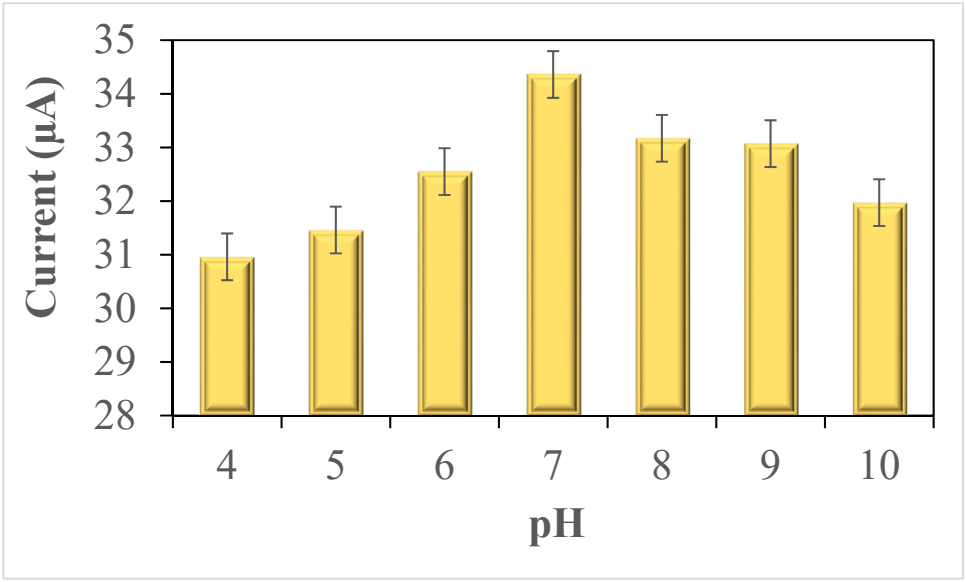


Figure 7. Effect of various pH on the analytical performance of the MIP-based electrochemical biosensor.

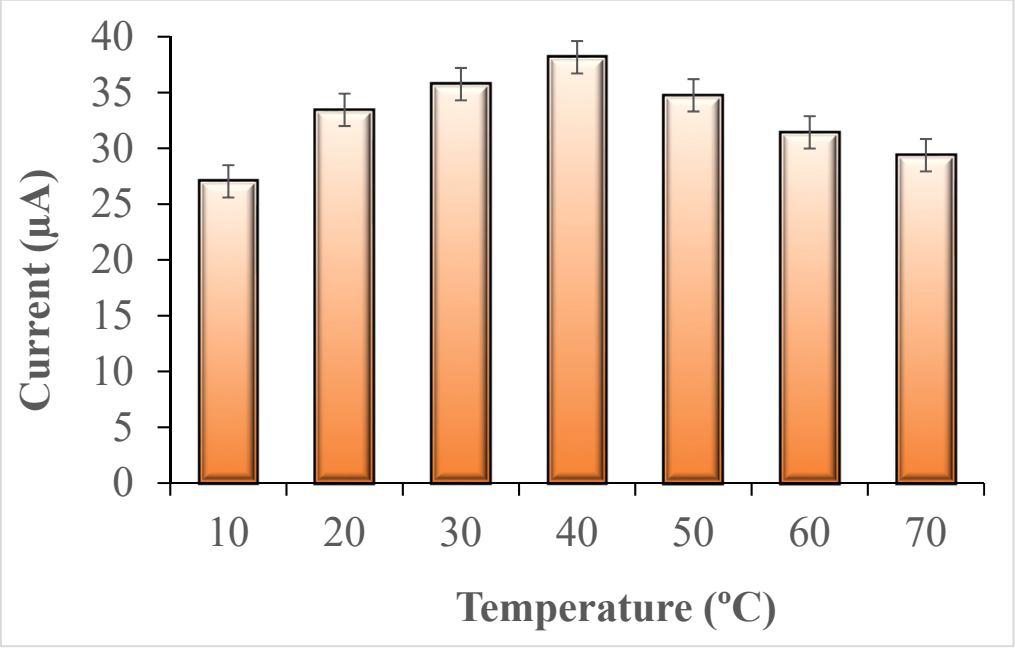


Figure 8. Determination of optimum temperature for the fabricated sensing platform.

3.4.3. Sensor's Analytical Performance

The electrochemical response of the fabricated electrode CagA-MIPs/Au/rGO@SPE was further investigated by using various concentrations of CagA. The differential pulse voltammetry (DPV) was used to record current responses of various concentrations (Figure 9a). The concentration study was performed within a concentration range of 0.05 ng/mL to 50 ng mL^{-1} (0.05 ng mL^{-1}, 0.1 ng mL^{-1}, 0.5 ng mL^{-1}, 1 ng mL^{-1}, 5 ng mL^{-1}, 10 ng mL^{-1}, 50 ng mL^{-1}). The concentration study was carried out in ferro/ferri buffer solution [64]. The DPV graphs show that the magnitude of peak current response decreases significantly with an increase in CagA concentration. The acquired result depicts that at higher concentrations of CagA, active sites on the surface of the electrode are blocked due to greater occupancy of cavities, which is attributed to the blocking of electron transfer via buffer solution. Thus, the low current is generated at the highest provided (50 ng mL^{-1}) concentration. The calibration graph is also plotted between the current response and different concentrations, as shown in Figure 9b. The graph shows a linear curve with high sensitivity as 0.325 µA ng^{-1} mL^{-1}. The limit of detection was reported as 0.05 ng mL^{-1}.

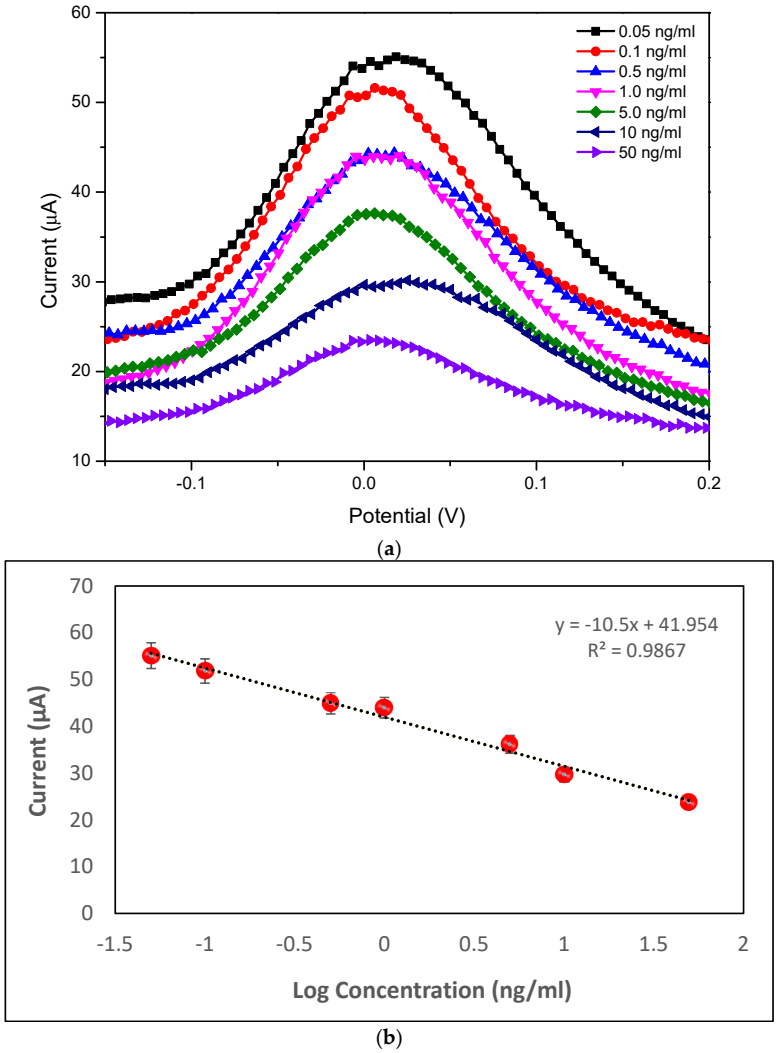


Figure 9. (**a**) DPV response study of MIP-CagA/Au/rGO@SPE with different concentrations of CagA antigen in the presence of $[Fe(CN)_6]^{3-/4-}$ electrolytes. (**b**) Representation of the standard calibration curve of MIP-CagA/Au/rGO@SPE with different concentrations of target analytes with reference to DPV results.

3.4.4. Scan Rates Study to Observe Sensing Response

In Figure 10a, the CV study demonstrates the biosensing response of the modified electrode at different scan rates, i.e., from 20–80 mV/s. The voltage for scan rate study was applied from +0.8 to −0.8 V. This study stated a 'quasi-reversible' process to observe electron transfer by diffusion reaction. The CV plot shows the symmetrical reduction and oxidation peaks that depict the relatively equivalent potential response. With the increase in the scan rate range, the peak currents also increased, which shows the steady electron transfer kinetics on the modified electrode surface. In Figure 10b, the square root of the scan rates has been demonstrated which is observed from 20–80 mV/s. The scan rate is equivalent to the redox peak current. The relationship between the current response and the square root of the scan rate has been plotted as follows:

$$y = 0.0054x + 0.4507; R^2 = 0.9935 \quad (2)$$

$$y = -0.0052x - 0.3739; R^2 = 0.9888 \quad (3)$$

The oxidation, as well as the reduction current peaks, are linearly relative to the relevant scan rate's square root.

3.4.5. Analysis of Selectivity and Stability

The interference of various biomolecules like glucose, cholesterol, ascorbic acid, *H. pylori*-specific BabA, and VacA was studied for modified electrode. For clinical use, a sensor must show high specificity for the target analyte in comparison to other interfering biomarkers that are present in the blood serum. As shown in Figure 11, the magnitude of the DPV current was observed as 34.68 µA for CagA-MIP/Au/rGO@SPE. After that, the DPV's current response was taken in the presence of the various abovementioned interferents. Although there were slight variations in the current response, there was no major alteration observed in the current's response to other interferents as compared to blank. However, after the addition of CagA analyte, a significant increase in the current response was observed. The interference study shows that the fabricated electrode is highly specific for the CagA virulence factor of *H. pylori*. The retained activity of the fabricated biosensor MIP-CagA/Au/rGO@SPE was analyzed in the presence of 0.05 ng/mL of CagA antigen to observe current response incubating in a mixture at regular intervals (every week). The fabricated electrode was stored at 4 °C and observed for up to 60 days for any reduction in its performance. The fabricated sensing platform retained its original activity by approximately 80% over the period of 60 days (Figure 12).

3.4.6. Real Sample Analysis

Further, the fabricated MIP-based sensing platform was observed for clinical applications by investigating it with spiked blood samples. Results showed that the developed biosensor is very selective for the diagnosis of *H. pylori* and easy to handle. CagA antigen was analyzed in the spiked sample and RSD (relative standard deviation) was calculated to determine the precision of the biosensor. Additionally, the recovery percentage was also calculated by using the standard method in the presence of 0.05 ng/mL CagA antigen. The entire experiment was repeated five times to obtain the average recovery in similar conditions. The results showed approximately 96% recovery, which shows the efficacy of the fabricated biosensor for its clinical use in the future.

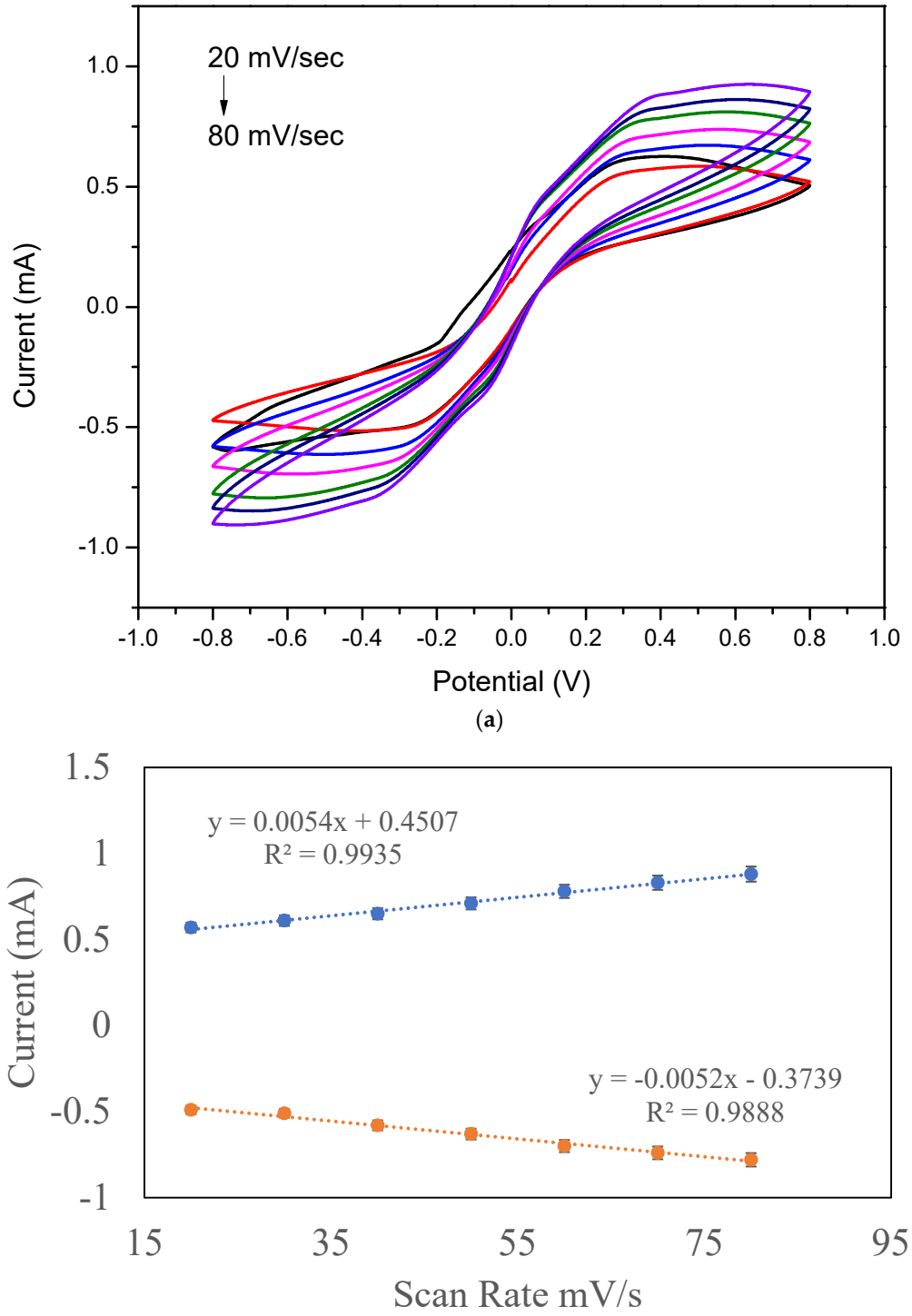


Figure 10. (**a**) CV recorded for MIP-CagA/Au/rGO@SPE at different scan rates in the presence of $[Fe(CN)_6]^{3-/4-}$ electrolytes. (**b**) Representation of variations in the peak oxidation current (Ipa) and reduction peak current (Ipc) at different scan rates than the linear graph.

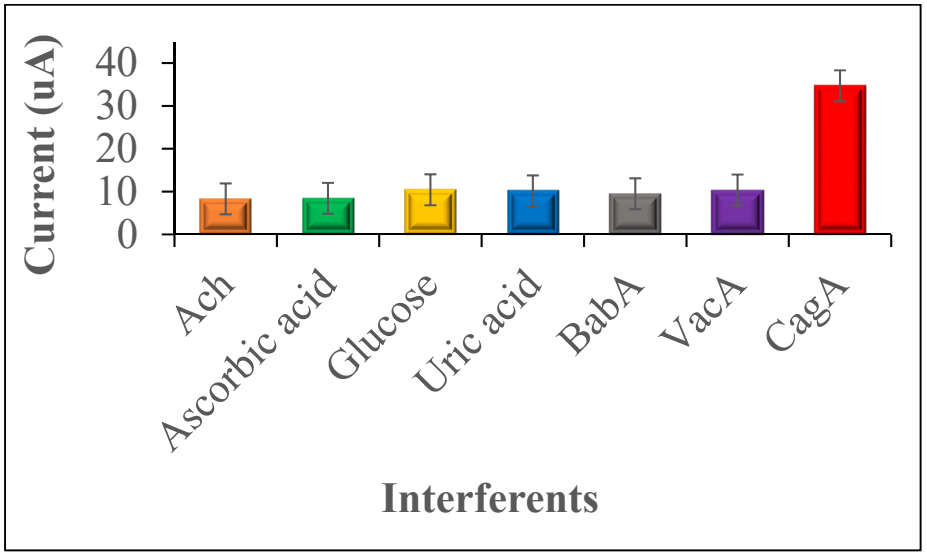


Figure 11. Plot of interferent studies of MIP-CagA/Au/rGO@SPE with 0.05 ng/mL CagA antigens.

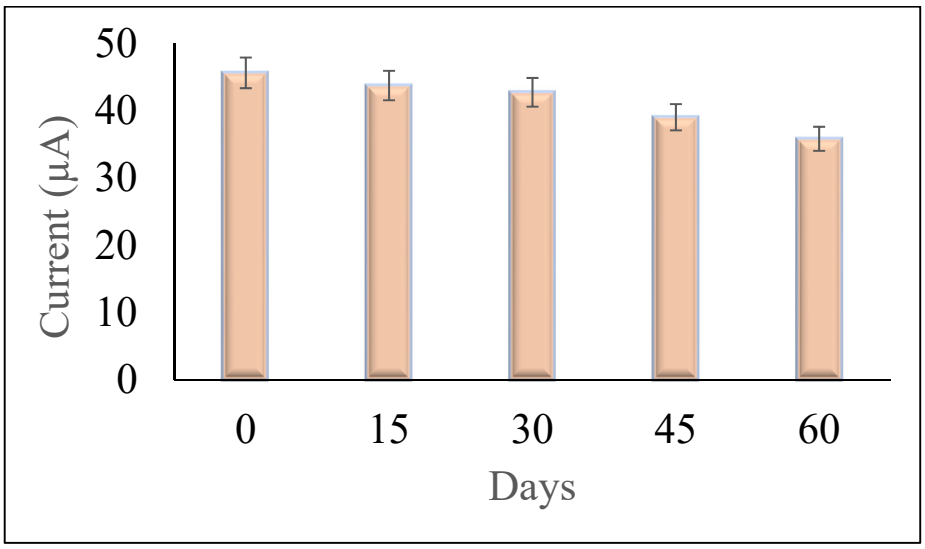


Figure 12. Demonstration of the fabricated sensing electrode's stability over the period of 60 days.

4. Conclusions

Herein, we have developed a highly sensitive rGO and gold NP-based MIP sensor for the detection of the CagA antigen of *H. pylori*. An electrodeposition technique (CV) was used for the deposition of nanomaterials, followed by the electropolymerization of MIP on a screen-printed electrode. The reactive site of functional monomers and the CagA template were elucidated by ESP and frontier orbital analysis, resulting in the potential atomic contacts that are responsible for the formation of imprinted sites. From the RDF standpoint, the adsorption capacity of MIP 2 with CagA as a template, MMA as a functional monomer, and EGDMA as a cross-linker at a ratio of 1:4:4 was greater than other simulated MIP mixtures which matched well to the experimental molar ratio during MIP synthesis. Further, a low E_b upon template removal suggests the ability of as-synthesized MIP to recognize the biosensor target through a robust binding affinity. The sensitivity of the fabricated electrode was calculated as 0.275 µA ng^{-1} mL^{-1} with a limit of detection of 0.05 ng/mL. The CagA antigen is the major virulence factor responsible for the establishment of *H. pylori* infection in the human stomach. Thus, the proposed MIP-based

biosensor can be developed as a point-of-care device in the future for the detection of gastric cancer in its early stage.

Author Contributions: Conceptualization, K.S., U.J. and N.C.; methodology, K.S.; validation, P.-K.Y., B.D.M. and N.C.; formal analysis, B.T.M.; investigation, K.S. and B.T.M.; resources, U.J. and N.C. and P.-K.Y.; writing—original draft preparation, K.S.; writing—review and editing, P.-K.Y. and B.D.M.; visualization, U.J.; supervision, U.J. and N.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Department of Science and Technology (DST), New Delhi (File No. DST/WOS-B/HN-12/2021, TDP/BDTD/33/2019). One of us (Bansi Dhar Malhotra) thanks the Science and Engineering Board (DST-SERB, Govt. of India) for granting the Distinguished Fellowship award (SB/DF/011/2019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We would like to acknowledge Wolfgang Fischer and Evelyn Weiss of the Max von Petten kofer Institute of Hygiene and Medical Microbiology, Ludwig Maximilians University, Munich, Germany for providing CagA antigen for the present analysis. We are also thankful to the Department of Science and Technology (DST), New Delhi (File No. DST/WOS-B/HN-12/2021, TDP/BDTD/33/2019). The authors thank The National Center for High-Performance Computing, Hsinchu City, Taiwan for providing the essential computational resources. One of us (Bansi Dhar Malhotra) thanks the Science and Engineering Board (DST-SERB, Govt. of India) for granting the Distinguished Fellowship award (SB/DF/011/2019).

Conflicts 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

1. Claros, J.; Tocino, R.V.; Fonseca, E.; Cigarral, B.; Barrios, B.; Casado, D.; Escalera, E.; Olivares, A.; Figuero, L.; Terán, E.; et al. Gastric cancer. *Medicine* **2021**, *13*, 1328–1334. [CrossRef]
2. Sitarz, R.; Skierucha, M.; Mielko, J.; Offerhaus, J.; Maciejewski, R.; Polkowski, W. Gastric cancer: Epidemiology, prevention, classification, and treatment. *Cancer Manag. Res.* **2018**, *10*, 239–248. [CrossRef] [PubMed]
3. Molina-Castro, S.; Pereira-Marques, J.; Figueiredo, C.; Machado, J.C.; Varon, C. Gastric cancer: Basic aspects. *Helicobacter* **2017**, *22*, e12412. [CrossRef] [PubMed]
4. Kienesberger, S.; Zechner, E.L. Helicobacter pylori. In *Reference Module in Biomedical Sciences*, 4th ed.; Academic Press: Cambridge, MA, USA, 2017; pp. 518–524. [CrossRef]
5. Garza-González, E.; Perez-Perez, G.I.; Maldonado-Garza, H.J.; Bosques-Padilla, F.J. A review of *Helicobacter pylori* diagnosis, treatment, and methods to detect eradication. *World J. Gastroenterol.* **2014**, *20*, 1438–1449. [CrossRef] [PubMed]
6. Kao, C.-Y.; Sheu, B.-S.; Wu, J.-J. Helicobacter pylori infection: An overview of bacterial virulence factors and pathogenesis. *Biomed. J.* **2016**, *39*, 14–23. [CrossRef] [PubMed]
7. Wang, Y.-K.; Kuo, F.-C.; Liu, C.-J.; Wu, M.-C.; Shih, H.-Y.; Wang, S.S.; Wu, J.-Y.; Kuo, C.-H.; Huang, Y.-K.; Wu, D.-C. Diagnosis of *Helicobacter pylori* infection: Current options and developments. *World J. Gastroenterol.* **2015**, *21*, 11221–11235. [CrossRef]
8. Saxena, K.; Chauhan, N.; Jain, U. Advances in diagnosis of Helicobacter pylori through biosensors: Point of care devices. *Anal. Biochem.* **2021**, *630*, 114325. [CrossRef]
9. Yamaoka, Y.; Ojo, O.; Fujimoto, S.; Odenbreit, S.; Haas, R.; Gutierrez, O.; El-Zimaity, H.M.T.; Reddy, R.; Arnqvist, A.; Graham, D.Y. Helicobacter pylori outer membrane proteins and gastroduodenal disease. *Gut* **2006**, *55*, 775–781. [CrossRef]
10. Zhang, J.; Wang, M.; Shi, Y.; Wang, Q.; Zhao, W. Rapid detection of *Helicobacter pylori* using cytotoxin-associated gene A based on loop-mediated isothermal amplification assay and magnetic nanoparticles. *Mater. Express* **2020**, *10*, 283–289. [CrossRef]
11. Chang, W.-L.; Yeh, Y.-C.; Sheu, B.-S. The impacts of H. pylori virulence factors on the development of gastroduodenal diseases. *J. Biomed. Sci.* **2018**, *25*, 1–9. [CrossRef]
12. Magalhaes, A.; Reis, C. Helicobacter pylori adhesion to gastric epithelial cells is mediated by glycan receptors. *Braz. J. Med Biol. Res.* **2010**, *43*, 611–618. [CrossRef] [PubMed]
13. Gupta, S.; Tiwari, A.; Jain, U.; Chauhan, N. Synergistic effect of 2D material coated Pt nanoparticles with PEDOT polymer on electrode surface interface for a sensitive label free Helicobacter pylori CagA(Ag-Ab) immunosensing. *Mater. Sci. Eng. C Mater. Biol. Appl.* **2019**, *103*, 109733. [CrossRef] [PubMed]
14. Granstrom, M.; Lehours, P.; Bengtsson, C.; Mégraud, F. Diagnosis of *Helicobacter pylori*. *Helicobacter* **2008**, *13*, 7–12. [CrossRef]
15. Chauhan, N.; Saxena, K.; Tikadar, M.; Jain, U. Recent advances in the design of biosensors based on novel nanomaterials: An insight. *Nanotechnol. Precis. Eng.* **2021**, *4*, 045003. [CrossRef]

16. Blum, L.J.; Coulet, P.R. What Is Biosensor, M. Dekker. 1991. Available online: https://www.routledge.com/Biosensor-Principles-and-Applications/lum/p/book/9780824785468 (accessed on 2 September 2022).
17. Dede, S.; Altay, F. Biosensors from the First Generation to Nano-biosensors. *Int. Adv. Res. Eng. J.* **2018**, *2*, 200–207.
18. Kizek, R.; Krejcova, L.; Michalek, P.; Rodrigo, M.M.; Heger, Z.; Krizkova, S.; Vaculovicova, M.; Hynek, D.; Adam, V. Nanoscale virus biosensors: State of the art. *Nanobiosens. Dis. Diagn.* **2015**, *4*, 47–66. [CrossRef]
19. Zarei, M. Infectious pathogens meet point-of-care diagnostics. *Biosens. Bioelectron.* **2018**, *106*, 193–203. [CrossRef]
20. Gui, R.; Jin, H.; Guo, H.; Wang, Z. Recent advances and future prospects in molecularly imprinted polymers-based electrochemical biosensors. *Biosens. Bioelectron.* **2018**, *100*, 56–70. [CrossRef]
21. Roy, A.C.; Nisha, V.; Dhand, C.; Ali, A.; Malhotra, B. Molecularly imprinted polyaniline-polyvinyl sulphonic acid composite based sensor for para-nitrophenol detection. *Anal. Chim. Acta* **2013**, *777*, 63–71. [CrossRef]
22. Hashwan, S.S.B.; Khir, M.H.B.M.; Al-Douri, Y.; Ahmed, A.Y. Recent Progress in the Development of Biosensors for Chemicals and Pesticides Detection. *IEEE Access* **2020**, *8*, 82514–82527. [CrossRef]
23. Subramani, K.; Elhissi, A.; Subbiah, U.; Ahmed, W. Introduction to nanotechnology. *Nanobiomater. Clin. Dent.* **2019**, 3–18. [CrossRef]
24. Xia, M.-Y.; Xie, Y.; Yu, C.-H.; Chen, G.-Y.; Li, Y.-H.; Zhang, T.; Peng, Q. Graphene-based nanomaterials: The promising active agents for antibiotics-independent antibacterial applications. *J. Control. Release* **2019**, *307*, 16–31. [CrossRef] [PubMed]
25. Baptista, P.; Fernandes, A.; Figueiredo, S.; Vinhas, R.; Cordeiro, M.; Carlos, F.F.; Mendo, S. Gold nanoparticle-based theranostics: Disease diagnostics and treatment using a single nanomaterial. *Nanobiosens. Dis. Diagn.* **2015**, *4*, 11–23. [CrossRef]
26. Pourmadadi, M.; Shayeh, J.S.; Arjmand, S.; Omidi, M.; Fatemi, F. An electrochemical sandwich immunosensor of vascular endothelial growth factor based on reduced graphene oxide/gold nanoparticle composites. *Microchem. J.* **2020**, *159*, 105476. [CrossRef]
27. Krishnan, S.K.; Singh, E.; Singh, P.; Meyyappan, M.; Singh Nalwa, H. A review on graphene-based nanocomposites for electrochemical and fluorescent biosensors. *RSC Adv.* **2019**, *9*, 8778–8881. [CrossRef]
28. Biovia, D.S. *Material Studio Modelling*; Dassault Systemes: San Diego, CA, USA, 2018.
29. Nešić, D.; Buti, L.; Lu, X.; Stebbins, C.E. Structure of the *Helicobacter pylori* CagA oncoprotein bound to the human tumor suppressor ASPP2. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1562–1567. [CrossRef] [PubMed]
30. Delley, B. An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys.* **1998**, *92*, 508–517. [CrossRef]
31. Delley, B. From molecules to solids with the DMol3 approach. *J. Chem. Phys.* **2000**, *113*, 7756–7764. [CrossRef]
32. Stephens, P.J.; Devlin, F.J.; Chabalowski, C.F.; Frisch, M.J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627. [CrossRef]
33. Tirado-Rives, J.; Jorgensen, W.L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *J. Chem. Theory Comput.* **2008**, *4*, 297–306. [CrossRef]
34. Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684. [CrossRef] [PubMed]
35. Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102*, 7338–7364. [CrossRef]
36. Sun, H.; Ren, P.; Fried, J. The COMPASS force field: Parameterization and validation for phosphazenes. *Comput. Theor. Polym. Sci.* **1998**, *8*, 229–246. [CrossRef]
37. Kong, Y.; Wang, N.; Ni, X.; Yu, Q.; Liu, H.; Huang, W.; Xu, W. Molecular dynamics simulations of molecularly imprinted polymer approaches to the preparation of selective materials to remove norfloxacin. *J. Appl. Polym. Sci.* **2016**, *133*, 42817. [CrossRef]
38. Patel, R.R.; Mohanraj, R.; Pittman, C.U., Jr. Properties of polystyrene and polymethyl methacrylate copolymers of polyhedral oligomeric silsesquioxanes: A molecular dynamics study. *J. Polym. Sci. Part B Polym. Phys.* **2005**, *44*, 234–248. [CrossRef]
39. Synthesis of Gold Nanoparticles via Chemical Reduction Method. Available online: https://www.researchgate.net/publication/306360987_Synthesis_of_Gold_Nanoparticles_via_Chemical_Reduction_Method (accessed on 20 October 2022).
40. Jo, J.; Lee, S.; Gim, J.; Song, J.; Kim, S.; Mathew, V.; Alfaruqi, M.H.; Kim, S.; Lim, J.; Kim, J. Facile synthesis of reduced graphene oxide by modified Hummer's method as anode material for Li-, Na- and K-ion secondary batteries. *R. Soc. Open Sci.* **2019**, *6*, 181978. [CrossRef]
41. Alibeyli, R.; Ata, A.; Topaç, E. Reduced Graphene Oxide Synthesis via Improved Hummers' Method, (n.d.). In Proceedings of the Conference: ISITES., Pune, India, 17 August 2014.
42. Jain, U.; Gupta, S.; Chauhan, N. Detection of glycated hemoglobin with voltammetric sensing amplified by 3D-structured nanocomposites. *Int. J. Biol. Macromol.* **2017**, *101*, 896–903. [CrossRef]
43. Pareek, S.; Jain, U.; Balayan, S.; Chauhan, N. Ultra-sensitive nano-molecular imprinting polymer-based electrochemical sensor for Follicle-Stimulating Hormone (FSH) detection. *Biochem. Eng. J.* **2022**, *180*, 108329. [CrossRef]
44. Balayan, S.; Chauhan, N.; Chandra, R.; Jain, U. Molecular imprinting based electrochemical biosensor for identification of serum amyloid A (SAA), a neonatal sepsis biomarker. *Int. J. Biol. Macromol.* **2022**, *195*, 589–597. [CrossRef]
45. Saxena, K.; Kumar, A.; Chauhan, N.; Khanuja, M.; Malhotra, B.D.; Jain, U. Electrochemical Immunosensor for Detection of H. pylori Secretory Protein VacA on g-C3N4/ZnO Nanocomposite-Modified Au Electrode. *ACS Omega* **2022**, *7*, 32292–32301. [CrossRef]

46. Gupta, S.; Jain, U.; Murti, B.T.; Putri, A.D.; Tiwari, A.; Chauhan, N. Nanohybrid-based immunosensor prepared for Helicobacter pylori BabA antigen detection through immobilized antibody assembly with @ Pdnano/rGO/PEDOT sensing platform. *Sci. Rep.* **2020**, *10*, 1–14. [CrossRef] [PubMed]
47. Stoian, I.-A.; Iacob, B.-C.; Dudaș, C.-L.; Barbu-Tudoran, L.; Bogdan, D.; Marian, I.O.; Bodoki, E.; Oprean, R. Biomimetic electrochemical sensor for the highly selective detection of azithromycin in biological samples. *Biosens. Bioelectron.* **2020**, *155*, 112098. [CrossRef] [PubMed]
48. Li, Y.; Zhang, L.; Dang, Y.; Chen, Z.; Zhang, R.; Li, Y.; Ye, B.-C. A robust electrochemical sensing of molecularly imprinted polymer prepared by using bifunctional monomer and its application in detection of cypermethrin. *Biosens. Bioelectron.* **2018**, *127*, 207–214. [CrossRef] [PubMed]
49. Electrochemical Based C-Reactive Protein (CRP) Sensing Through Molecularly Imprinted Polymer (MIP) Pore Structure Coupled with Bi-Metallic Tuned Screen-Printed Electrode. *Biointerface Res. Appl. Chem.* **2021**, *12*, 7697–7714. [CrossRef]
50. Jin, M.; Zhang, G.; Yu, F.; Li, W.; Lu, W.; Huang, H. Sponge-like $Ni(OH)_2$–NiF_2 composite film with excellent electrochemical performance. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1601–1605. [CrossRef]
51. Bathinapatla, A.; Kanchi, S.; Singh, P.; Sabela, M.I.; Bisetty, K. An ultrasensitive performance enhanced novel cytochrome c biosensor for the detection of rebaudioside A. *Biosens. Bioelectron.* **2016**, *77*, 116–123. [CrossRef] [PubMed]
52. Ashaduzzaman, M.; Deshpande, S.R.; Murugan, N.A.; Mishra, Y.K.; Turner, A.P.F.; Tiwari, A. On/off-switchable LSPR nano-immunoassay for troponin-T. *Sci. Rep.* **2017**, *7*, srep44027. [CrossRef]
53. Basiuk, E.V.; Martínez-Herrera, M.; Álvarez-Zauco, E.; Henao-Holguín, L.V.; Puente-Lee, I.; Basiuk, V.A. Noncovalent functionalization of graphene with a Ni(II) tetraaza[14]annulene complex. *Dalton Trans.* **2014**, *43*, 7413–7428. [CrossRef]
54. Murti, B.T.; Putri, A.D.; Kanchi, S.; Sabela, M.I.; Bisetty, K.; Inamuddin; Asiri, A.M. Light induced DNA-functionalized TiO2 nanocrystalline interface: Theoretical and experimental insights towards DNA damage detection. *J. Photochem. Photobiol. B Biol.* **2018**, *188*, 159–176. [CrossRef]
55. Hu, J.; He, P.; Li, Y. Chemical modifications of tryptophan residues in peptides and proteins. *J. Pept. Sci.* **2021**, *27*, e3286. [CrossRef]
56. King, T.A.; Kandemir, J.M.; Walsh, S.J.; Spring, D.R. Photocatalytic methods for amino acid modification. *Chem. Soc. Rev.* **2021**, *50*, 39–57. [CrossRef] [PubMed]
57. Sun, H.-W.; Qiao, F.-X. Recognition mechanism of water-compatible molecularly imprinted solid-phase extraction and determination of nine quinolones in urine by high performance liquid chromatography. *J. Chromatogr. A* **2008**, *1212*, 1–9. [CrossRef] [PubMed]
58. Tokonami, S.; Shiigi, H.; Nagaoka, T. Review: Micro- and nanosized molecularly imprinted polymers for high-throughput analytical applications. *Anal. Chim. Acta* **2009**, *641*, 7–13. [CrossRef] [PubMed]
59. Thorneywork, A.L.; Roth, R.; Aarts, D.G.A.L.; Dullens, R.P.A. Communication: Radial distribution functions in a two-dimensional binary colloidal hard sphere system. *J. Chem. Phys.* **2014**, *140*, 161106. [CrossRef]
60. Harris, T.K.; Mildvan, A.S. High-Precision Measurement of Hydrogen Bond Lengths in Proteins by Nu-clear Magnetic Resonance Methods. *Proteins* **1999**, *35*, 275–282. [CrossRef]
61. Sarkhel, S.; Desiraju, G.R. N-H . . . O, O-H . . . O, and C-H . . . O hydrogen bonds in protein-ligand complexes: Strong and weak interactions in molecular recognition. *Proteins* **2004**, *54*, 247–259. [CrossRef]
62. Caro, E.; Marce, R.; Borrull, F.; Cormack, P.; Sherrington, D. Application of molecularly imprinted polymers to solid-phase extraction of compounds from environmental and biological samples. *TrAC Trends Anal. Chem.* **2006**, *25*, 143–154. [CrossRef]
63. Soni, S.; Jain, U.; Burke, D.H.; Chauhan, N. A label free, signal off electrochemical aptasensor for amphetamine detection. *Surf. Interfaces* **2022**, *31*, 102023. [CrossRef]
64. Jain, U.; Soni, S.; Balhara, Y.P.S.; Khanuja, M.; Chauhan, N. Dual-Layered Nanomaterial-Based Molecular Pattering on Polymer Surface Biomimetic Impedimetric Sensing of a Bliss Molecule, Anandamide Neurotransmitter. *ACS Omega* **2020**, *5*, 10750–10758. [CrossRef]

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/345483581

Fake News Detection Using Passive-Aggressive Classifier

Chapter · September 2020

DOI: 10.1007/978-981-15-7345-3_13

CITATIONS
7

READS
153

2 authors:

Saloni Gupta
Indira Gandhi Institute of Technology
17 PUBLICATIONS **108** CITATIONS

SEE PROFILE

Priyanka Meel
Delhi Technological University
36 PUBLICATIONS **327** CITATIONS

SEE PROFILE

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

Fake News and Rumour Detection View project

M.Tech Dissertation View project

All content following this page was uploaded by Priyanka Meel on 14 January 2022.

The user has requested enhancement of the downloaded file.

Springer Link

Fake News Detection Using Passive-Aggressive Classifier

Inventive Communication and Computational Technologies pp 155-164 | Cite as

- Saloni Gupta (1) Email author (salonigupta140398@gmail.com)
- Priyanka Meel (1)

1. Delhi Technological University, , Delhi, India

Conference paper
First Online: 25 September 2020

- 506 Downloads

Part of the Lecture Notes in Networks and Systems book series (LNNS, volume 145)

Abstract

People can get infected with fake news very quickly with misleading words and images and post them without any fact-checking. The social media life has been used to distribute counterfeit data, which has a significant negative influence on individual consumers and on a wider community. The fake news problem is tackled using a machine learning algorithm. Different classifiers are used for the purpose of identifying fake news. In this paper, Passive-Aggressive Classifier is implemented for this purpose. The approach is implemented on two datasets of fake and real news. After performing the experiment, it is observed that Passive-Aggressive Classifier provides an accuracy of 97.5%. The performance of the proposed model is compared with the existing methods. The Passive-Aggressive Classifier provides the best result compared to others.

Keywords

Fake news detection Machine learning model Classifier
Passive-Aggressive Classifier TfIdfVectorizer

This is a preview of subscription content, log in to check access.

References

1. Metz C (2016) The bittersweet sweepstakes to build an AI that destroys fake news, Dec 2016 (Online). Available https://www.wired.com/2016/12/bittersweet-sweepstakes-build-ai-destroys-fake-news/ (https://www.wired.com/2016/12/bittersweet-sweepstakes-build-ai-destroys-fake-news/)

2. Granik M, Mesyura V (2017) Fake news detection using naive Bayes classifier. In: 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON), Kiev, Ukraine
Google Scholar (https://scholar.google.com/scholar?q=Granik%20M%2C%20Mesyura%20V%20%282017%29%20Fake%20news%20detection%20using%20naive%20Bayes%20classifier.%20In%3A%202017%20IEEE%20first%20Ukraine%20conference%20on%20electrical%20and%20computer%20engineering%20%28UKRCON%29%2C%20Kiev%2C%20Ukraine)

3. Bhowmik D, Zargari S, Ajao O (2018) Fake news identification on twitter with hybrid CNN and RNN models. In: Proceedings of the 9th international conference on social media and society
Google Scholar (https://scholar.google.com/scholar?q=Bhowmik%20D%2C%20Zargari%20S%2C%20Ajao%20O%20%282018%29%20Fake%20news%20identification%20on%20twitter%20with%20hybrid%20CNN%20and%20RNN%20models.%20In%3A%20Proceedings%20of%20the%209th%20international%20conference%20on%20social%20media%20and%20society)

4. Zheng L, Zhang J, Cui Q, Li Z, Yang PS, Yang Y (2018) TI-CNN: convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749 (http://arxiv.org/abs/1806.00749)

5. Lakshmanarao A, Swathi Y, Kiran TSR (2019) An efficient fake news detection system using machine learning. Int J Innov Technol Exploring Eng (IJITEE) 8(10)
Google Scholar (https://scholar.google.com/scholar?q=Lakshmanarao%20A%2C%20Swathi%20Y%2C%20Kiran%20TSR%20%282019%29%20An%20efficient%20fake%20news%20detection%20system%20using%20machine%20learning.%20Int%20J%20Innov%20Technol%20Exploring%20Eng%20%28IJITEE%29%208%2810%29)

6. Ahmed H, Traore I, Saad S (2017) Detection of online fake news using n-gram analysis and machine learning techniques. In: International conference on intelligent, secure, and dependable systems in distributed and cloud environments ISDDC 2017
Google Scholar (https://scholar.google.com/scholar?q=Ahmed%20H%2C%20Traore%20I%2C%20Saad%20S%20%282017%29%20Detection%20of%20online%20fake%20news%20using%20n-gram%20analysis%20and%20machine%20learning%20techniques.%20In%3A%20International%20conference%20on%20intelligent%2C%20secure%2C%20and%20dependable%20systems%20in%20distributed%20and%20cloud%20environments%20ISDDC%202017)

7. Khattar D, Goud JS, Gupta M, Varma V (2019) MVAE: multimodal variational autoencoder for fake news detection. In: The web conference-2019, San Francisco
Google Scholar (https://scholar.google.com/scholar?q=Khattar%20D%2C%20Goud%20JS%2C%20Gupta%20M%2C%20Varma%20V%20%282019%29%20MVAE%3A%20multimodal%20variational%20autoencoder%20for%20fake%20news%20detection.%20In%3A%20The%20web%20conference-2019%2C%20San%20Francisco)

8. Wang Y, Ma F, Jin Z, Yuan Y, Xun G, Jha K, Su L, Gao J (2019) EANN: event adversarial neural networks for multi-modal fake news detection. In: 24th ACM SIGKDD international conference on knowledge discovery & data mining, London

Google Scholar (https://scholar.google.com/scholar?q=Wang%20Y%2C%20Ma%20F%2C%20Jin%20Z%2C%20Yuan%20Y%2C%20Xun%20G%2C%20Jha%20K%2C%20Su%20L%2C%20Gao%20J%20%282019%29%20EANN%3A%20event%20adversarial%20neural%20networks%20for%20multi-modal%20fake%20news%20detection.%20In%3A%2024th%20ACM%20SIGKDD%20international%20conference%20on%20knowledge%20discovery%20%26%20data%20mining%2C%20London)

9. Markines B, Cattuto C, Menczer F (2009) Social spam detection. In: 5th international workshop on adversarial information retrieval on the web, 2009
Google Scholar (https://scholar.google.com/scholar?q=Markines%20B%2C%20Cattuto%20C%2C%20Menczer%20F%20%282009%29%20Social%20spam%20detection.%20In%3A%205th%20international%20workshop%20on%20adversarial%20information%20retrieval%20on%20the%20web%2C%202009)

10. Lu J, Zhao P, Hoi SCH (2016) Online passive-aggressive active learning.
https://doi.org/10.1007/s10994-016-5555-y
(https://doi.org/10.1007/s10994-016-5555-y)

11. Available
https://drive.google.com/file/d/0B3e3qZpPtccsMFo5bk9Ib3VCc2c/view
(https://drive.google.com/file/d/0B3e3qZpPtccsMFo5bk9Ib3VCc2c/view)
(Online)

12. Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y (2006) Online passive-aggressive algorithms. J Mach Learn Res 7:551–585
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=2274378)
zbMATH (http://www.emis.de/MATH-item?1222.68177)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Online%20passive-aggressive%20algorithms&author=K.%20Crammer&author=O.%20Dekel&author=J.%20Keshet&author=S.%20Shalev-Shwartz&author=Y.%20Singer&journal=J%20Mach%20Learn%20Res&volume=7&pages=551-585&publication_year=2006)

Copyright information

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

About this paper

Cite this paper as:

Gupta S., Meel P. (2021) Fake News Detection Using Passive-Aggressive Classifier. In: Ranganathan G., Chen J., Rocha Á. (eds) Inventive Communication and Computational Technologies. Lecture Notes in Networks and Systems, vol 145. Springer, Singapore. https://doi.org/10.1007/978-981-15-7345-3_13

- First Online 25 September 2020
- DOI https://doi.org/10.1007/978-981-15-7345-3_13
- Publisher Name Springer, Singapore
- Print ISBN 978-981-15-7344-6
- Online ISBN 978-981-15-7345-3

- eBook Packages Engineering Engineering (RO)

- Buy this book on publisher's site
- Reprints and Permissions

Personalised recommendations

SPRINGER NATURE

© 2020 Springer Nature Switzerland AG. Part of Springer Nature.

Not logged in Delhi Technological University (2000607895) - Convener, UGC-Infonet Digital Library Consortium (3000132959) - INDEST-AICTE-Level III (3000168247) - Information and Library Network (INFLIBNET) Centre (3994475188) 14.139.251.100

Neural Computing and Applications
https://doi.org/10.1007/s00521-022-07938-3

ORIGINAL ARTICLE

Framework for detection of probable clues to predict misleading information proliferated during COVID-19 outbreak

Deepika Varshney[1] · Dinesh Kumar Vishwakarma[1]

Received: 11 August 2021 / Accepted: 11 October 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract
Spreading of misleading information on social web platforms has fuelled huge panic and confusion among the public regarding the Corona disease, the detection of which is of paramount importance. To identify the credibility of the posted claim, we have analyzed possible evidence from the news articles in the google search results. This paper proposes an intelligent and expert strategy to gather important clues from the top 10 google search results related to the claim. The N-gram, Levenshtein Distance, and Word-Similarity-based features are used to identify the clues from the news article that can automatically warn users against spreading false news if no significant supportive clues are identified concerning that claim. The complete process is done in four steps, wherein the first step we build a query from the posted claim received in the form of text or text additive images which further goes as an input to the search query phase, where the top 10 google results are processed. In the third step, the important clues are extracted from titles of the top 10 news articles. Lastly, useful pieces of evidence are extracted from the content of each news article. All the useful clues with respect to N-gram, Levenshtein Distance, and Word Similarity are finally fed into the machine learning model for classification and to evaluate its performances. It has been observed that our proposed intelligent strategy gives promising experimental results and is quite effective in predicting misleading information. The proposed work provides practical implications for the policymakers and health practitioners that could be useful in protecting the world from misleading information proliferation during this pandemic.

Keywords COVID-19 · Information pollution · Fake news detection

1 Introduction

In the recent scenario, a new coronavirus disease spread around the world. The disease emerges as a respiratory infection with significant concern for global public health hazards. Initially, it is suspected that the disease is transmitted from animal to humans, later the paradigm is shifted that the infection is transmitted toward human to human via droplets, close contacts creating huge panic with approximate 6,359,182 confirmed cases and 380,663 deaths[1] have been encountered till now and growth rate is still high which has alarmed the global authorities including world health organization (WHO) [1]. The COVID-19 pandemic affects worldwide badly. However, there is no shortage of people who are taking this crisis as an opportunity for malicious activities/gaining profit. Many health-related misleading information some of the fake cures suggested for COVID-19 have been posted by malicious users, creating confusion and misconceptions about the disease. During this pandemic, people have their eye on any new announcement from the government official or some news that can help to get rid of COVID-19. As the disease is deadly, the people are also desperate to know some cure and rush to find a new coronavirus disease. Some of the fake cures posted on social media are harmful and give bad health advice. The recent examples of fake cures are shown

✉ Dinesh Kumar Vishwakarma
dinesh@dtu.ac.in

Deepika Varshney
deepikavarshney06@gmail.com

[1] Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Delhi 110042, India

[1] https://in.search.yahoo.com/search?fr=mcafee&type=E211IN885G0&p=coronavirus.

Published online: 13 November 2022

in Fig. 1. where Fig. 1a shows the image with a false claim is gone viral that drinking water a lot and gargling with warm water and salt or vinegar eliminates the coronavirus, however, there is no significant evidence has been found concerning to this claim. Another cure in Fig. 1b reported that the silver solution can kill coronavirus within 12 h. The proliferation of these misleading information creates many misconceptions in the mind-set of people related to coronavirus disease. Some of the users are spreading it without verification and fuelled panic among people regarding the COVID-19. According to [2], misleading or fake can be defined as any post that shares content that does not faithfully represent the event that it refers to. We followed this definition in our work and define *"misleading information as the content that does not faithfully represent the event that it refers to and having no significant evidence of proof to validate the claim."* Recent research shows that numerous misleading content is circulating about the coronavirus and it is becoming difficult to differentiate fake news from the real one [3]. The propagation of misleading content on the virus could also be deleterious to mankind. This has led to the dire need for a system that can differentiate fake from real. Earlier, many of the previous research has reported methods of detecting fake news in online social media considering various applications. Most of the previous research has counter fake news problems mainly in the following types: Image-based and Text-based algorithms. Earlier, many researchers have worked on fake news detection by applying a text-based approach. Text-based approaches mainly use text patterns and match them with already existing patterns of fake news. They are sometimes referred to as the linguistic approach. Along with this lots of researchers have shifted their interest in the credibility detection of posts/tweets using text-based features [4]. Like a text-based approach, research has also been done by employing an image-based approach. From the study, it has been seen that researchers have explored images based algorithms for the analysis of fake images or images attached with false claims in mainly following ways, Text additive images, and Manipulated images. *The manipulated images are termed as images whose piece/-part or certain region is manipulated with respect to visual context.* Various image-based features have been explored for the classification of images. The authors of [5] propose 5 visual features and 7 statistical features for the verification of news events. Along with the manipulated images, some researchers have also considered text additive images to analyze misleading content. *The text additive images termed as images embedded with false claims instead of having any manipulation from visual context.* The authors of [6] have incorporated text additive images, where they have applied a rule-based algorithm for the prediction of fake news. From recent research, it has been observed that none of the works have shown and reported fake news prediction analysis propagated during one of the major pandemic *CORONAVIRUS*. Many people are sharing fake cures to get rid of coronavirus disease without any verification and create lots of misconceptions. Government and officials have also urged peoples to check the authenticity of the post before sharing [3]. This also motivates us to build an intelligent system for the prediction of fake news spreading during this pandemic. We, therefore, developed a generalized model of detecting misleading content on social media platforms, where we have considered COVID-19 as a special issue which is a huge pandemic and taken as one of the application case studies in this work. However, our model is generalized and works for other applications as well. COVID-19 is an emerging issue and very few research have been reported yet in this context that leads to motivates us to build an efficient framework to predict misleading content spreading during the COVID outbreak. One of the major lacks in this case that there is no dataset concerning corona fake news that is publically available to test our proposed algorithm. One of the novelty and contribution of the work is we have built a dataset with 335 claim samples having 168 fake and 167 real news related to COVID-19, also working to expand the dataset. Along with this to validate the generalizability of our model, it has also been tested over one of the standard dataset "LIAR" that further validates the effectiveness of the model. Three sets of novel features (N-gram, Levenshtein Distance, and Word Similarity) are proposed in this work that is found to be efficient in predicting misleading content and results reveal that the model outperforms the other state-of-the-art method, discussed in detail in the later sections. To the best of our knowledge, none of the proposed work has been employed these features concerning news article headlines and its content. The major key contributions of the work are highlighted in the following points.[2]

- The proposed work gives a significant contribution in providing a novel generalized framework for the collection and annotation of misleading content in an online social network where considering COVID-19 (fake news spreading during Corona outbreak) as one of the special case studies from the application perspective, where considered post in the form of images and text related to unverified cures to get rid out from the disease.
- For the analysis perspective, the data are collected related to Covid-19 and builds a self-generated dataset "*Covidfakenews 2019*" of misleading post spreading during corona. As the problem is new/emerging, no

[2] https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-fake-news-disinformation-rumors-hoaxes.

(a) (b)

Fig. 1 Examples of misinformation

datasets are publically available for analysis related to this area of research. This leads to giving a significant contribution.

- To the best of our knowledge, the proposed work is first to provide three sets of novel features (N-gram, Levenshtein Distance, and Word Similarity) concerning to news article headlines and its content from the google search results, that further be used as an efficient clue for the prediction of misleading content.
- As the COVID-19 is one of the emerging issues, none of the work has been reported yet to predict the fake news propagating during this phase and gives a major contribution by providing the analysis, which greatly helps researchers for further study.
- We investigate the model performances with different classifiers and comparative analysis reveals that our proposed method outperforms other states-of-the-art on the same dataset.

The remaining of this paper is organized as follows. In Sect. 2, we are going to discuss the previous work that has been done related to this field, where Sect. 3, gives a detailed description of how the data collection and annotation have been employed as well as the strategy/method that we have employed for the misleading information detection, which followed by a discussion of experimental results in Sect. 4. Lastly, the paper is concluded with some suggested future work aspects.

2 Related work

In the current era spreading false information is one of the crucial problems nowadays, where it is quite difficult for the online user to discriminate fake news from the real one and that's why the development of an intelligent system is required. Most of the methods proposed in earlier states to detect misleading information considered it a task of classification problem intending to associate label as true or false with a particular claim/post. From the survey analysis, it has been observed that the classification approaches are turn divided into approaches based on machine learning and deep learning. A detailed description is given below.

2.1 Machine learning

From the study, it has been proven that machine learning algorithms are extremely useful in countering numerous problems in the information engineering field. In particular, many machine learning approaches implemented for misleading/fake information detection are applied as a supervised learning strategy. In machine learning classification algorithms support vector machines(SVMs) are one of the widely used methods for classifications. The authors of [7] have proposed a method to employ a graph-kernel-based SVM classifier to detect rumors using propagation structure and content features with an accuracy of 0.91 on the Sina-Weibo dataset. Whereas, in [8] the author reported a set of features to distinguish among fake news, real news, and satire. The SVMs are also employed for clickbait detection in [9]. The random forest has also been exploited in

numerous works for fake news and rumor detection like SVM. Most of the studies have reported random forests as a strong performer among other machine learning algorithms [10–14]. In [11], the author has proposed a set of temporal, structural, and linguistic features for the classification of rumors in a tweet graph by employing a random forest with an accuracy of 0.90. The random forest has also been used for stance detection in [10, 15]. The comparative studies of a different approach in the context of rumor and fake news have shown competitive performance for logistic regression [12, 16–18]. The authors of [19], employed logistic regression for stance classification of news articles or headlines and claims. A decision tree is another widely studied algorithm proposed particularly for misleading content detection [20]. The effectiveness of decision tree algorithms like j48 with respect to other machine learning paradigms, including SVMs has been reported in [4, 12]. The authors of [4], have used the content and context-based features to perform credibility evaluation of tweets and the model is performing well with an accuracy of 0.86. In [21], to evaluate the trustworthiness of users in social media via decision tree, the author has proposed a series of user trust metrics and reported an accuracy of 0.75.

2.2 Deep learning

Deep learning is one of the prominent and widely explored research topics in machine learning. The main advantage of deep learning over traditional machine learning approaches is that they are not based on manually crafted features and reduce feature extraction time. Along with this, the deep learning framework can learn hidden representations from simpler inputs both in context and content variation [22]. The two prominent and widely used paradigms in Morden artificial neural network are RNN and CNN. In [22], authors have proposed novel RNN architectures, namely tanh-RNN, LSTM, and Gated Recurrent Unit(GRU), to detect rumors. The results show that GRU has obtained the best results in both the datasets considered with 0.88 and 0.91 accuracies, respectively. Whereas, in [23], the author has proposed a multi-task learning approach and designed a multi-task learning framework with an LSTM layer shared among all tasks to counter the problem of rumor classification. Like RNN, CNNs have also been explored and widely studied for image recognition and many other fields of computer vision. However, it is now gaining popularity in the NLP field [24]. The authors of [15], have explored a technique using CNN with single and multi-word embedding to counter problems concerning both stance and veracity classification of tweets. The author has reported an accuracy of 0.70 for the stance classification problem and 0.53 for the problem of veracity classification. Whereas, Paragraph embedding is explored to learn the representation of a small group of posts in a specific event and used them as input for their CNN model in [25] and achieved an accuracy of 0.93 for Sina Weibo and 0.77 for Twitter. From the study, it has been observed that most of the recent work has explored the combination of RNN and CNN in their model [26–28]. The authors of [28] proposed an architecture applied on the LIAR dataset that encodes text information via a CNN and metadata about the author of the text using an LSTM layer as well as it has been also found that the hybrid model has proved to outperform all other baselines along with a bi-LSTM architecture with an accuracy of 0.27 on the testing dataset. Whereas, in [26] author has proposed an approach based on repost sequence patterns for the detection of false rumors.

All the different approaches discussed above have considered different machine and deep learning methods for the prediction of misleading content. The previous study shows that very few works have reported the important clues concerning news article headlines and news article content to predict misleading information. The study has observed that crucial pieces of evidence can be identified from News article headlines and its content to predict false information. None of the previous work has been included these three sets of novel clues (N-gram, Levenshtein Distance, and Word Similarity) concerning news articles that were found to be efficient. That gives novelty to our proposed work.

3 Methodology

In this section, we elaborate on each phase of the proposed methodology. We first detail the data creation step adopted for evaluating our model. The self-generated dataset is the collection of claims in the form of images and text posted by users related to coronavirus. Here, images we mean images embedded with text. In the next phase, the given claim has been processed with the removal of stop words, punctuation, and unnecessary words for query building. The query is then given as an input to google search, where the top 10 search results are considered for analysis. To maintain the claim's relevance and retrieve related results, the top 10 links have been used for processing. In the next phase, important clues are identified from both news article headlines as well as news article content concerning Word Similarity, N-gram, Levenshtein Distance Ratio, and N-gram, Levenshtein Distance Ratio measures, respectively. Lastly, these important clues are used for binary class classification real or misleading.

3.1 Problem definition

In this paper, we have considered a binary class classification problem. We assume that the posted source claim $c = \{c1, c2, c3...ck\}$ can be divided into two classes Class $= \{M, R\}$.: (1) Real (R), namely the posted claim faithfully represents the event that it refers to, (2) Misleading (M), namely the posted claim does not faithfully represent the event that it refers to. The claims that have been included here are related to coronavirus COVID-19. Take an example of the post related to coronavirus. The post that "chlorine and alcohol products cannot kill viruses within the body" is true information and belongs to the real class since many authoritative and authentic news media have reported relevant news and has also acknowledged by WHO.[3] While one of the posts says that "coronavirus is caused by 5G technology" is absolutely a false rumor, it not only lacks the factual sport but also deviates from the scientific principles. So, the goal is to learn a classifier from the labeled feature set, that is $f : X^k \mapsto Y^k$, where Y^k takes one of the two fine-grained classes: {R, M}. Given the input feature set X^k, the classifier f can output the classification result for the posted claim C^k.

3.2 Dataset creation and query searching

In the coronavirus outbreak, lots of misleading information was posted by users, and most of the misleading information reading fake cures, lockdown, and others have gone viral during the period. The posts have gone viral on different social media platforms like Twitter, Facebook, and others. The dataset is created during the covid-pandemic 2020. The dataset is the collection of post available in the form of text embedded images and text-only form. Here, text embedded images we mean that the claim is inbuilt in an image. Sometimes to represent certain event users post an image embedded with text. The example of text embedded images we can also find from Fig. 1. For building the dataset, the prominent claims are collected related to corona posted during the period from different web sources (News Media, WHO, Buzzfednews) in the form of images and text. Some of the sources of misleading news are Buzzfeed news,[4] Buzzfeednews provides many coronavirus hoaxes that are posted over various social web platform in different forms(images, audio, text), along with it they also provide veracity, mentioning whether the given claim is fake or not, another source that we have incorporated in this work is WHO (world health organization), the website is providing fake and real claims, the another web sources that are reporting myths regarding coronavirus.[5] The empirical analysis has been done on these web sources for the collection of fake samples (Fig. 2). The sources provide the running list of the latest hoaxes spreading about the coronavirus. Whereas most of the real claims are retrieved from the WHO website for more reliable collection. In all total of 335 queries are collected with having 168 misleading and 167 real claims (*Covid-fakenews 2019*). As there is no source providing clear-cut veracity until and unless we read the complete text of it. That's why after empirical analysis we prepare our ground truth dataset consisting of 335 samples, including fake and real as a target class. The world cloud representation of fake and real claims in the proposed dataset is shown in Fig. 3a and b, respectively.

There are 9837 fake samples and 8634 real samples are considered with respect to 335 claims. The text additive images are processed further using the OCR technique, where textual data has been extracted to build the query for further processing shown in Fig. 2. In Fig. 2, the input is taken in the form of either plain text or the form of text additive image. If it is a simple text, then it is directly passed to the query building phase; otherwise, if it is in the form of text additive image, the OCR technique has been applied to extract the text from an image, and further, it goes further to the query building phase. The query is searched on google, where top 10 google results $L_i = L_1, L_2...L_{10}$ are retrieved and used for further processing as shown in Fig. 2. Each link is passed to the processing phase, where important clues are identified for prediction.

3.3 Clues identification for misleading information prediction

In this phase, important clues are identified from the retrieved top 10 URLs that further be used to predict the misleading information. The pieces of evidence/clues are identified with respect to news article headlines/titles and news article content/text. In this phase, the clues are identified in two steps. Firstly, the titles are analyzed and important features are extracted with respect to Word Similarity Ratio, N-grams, and Levenshtein Distance Ratio. The LDR[6] and N-gram[7] are the earlier techniques that helps in reteriving efficient features in our work. LDR (Levenstein distance ratio) is one of the good method of reteriving similarity among two sentences. Similarly, N-Gram is also useful in finding important phrases in the

[3] https://www.indiatoday.in/world/story/drinking-alcohol-will-not-protect-you-from-covid-19-says-who-1653555-2020-03-08.

[4] https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-fake-news-disinformation-rumors-hoaxes.

[5] https://www.medicalnewstoday.com/articles/coronavirus-myths-explored.

[6] Levenshtein distance—Wikipedia.

[7] What are N-Grams?—Kavita Ganesan, PhD (kavita-ganesan.com).

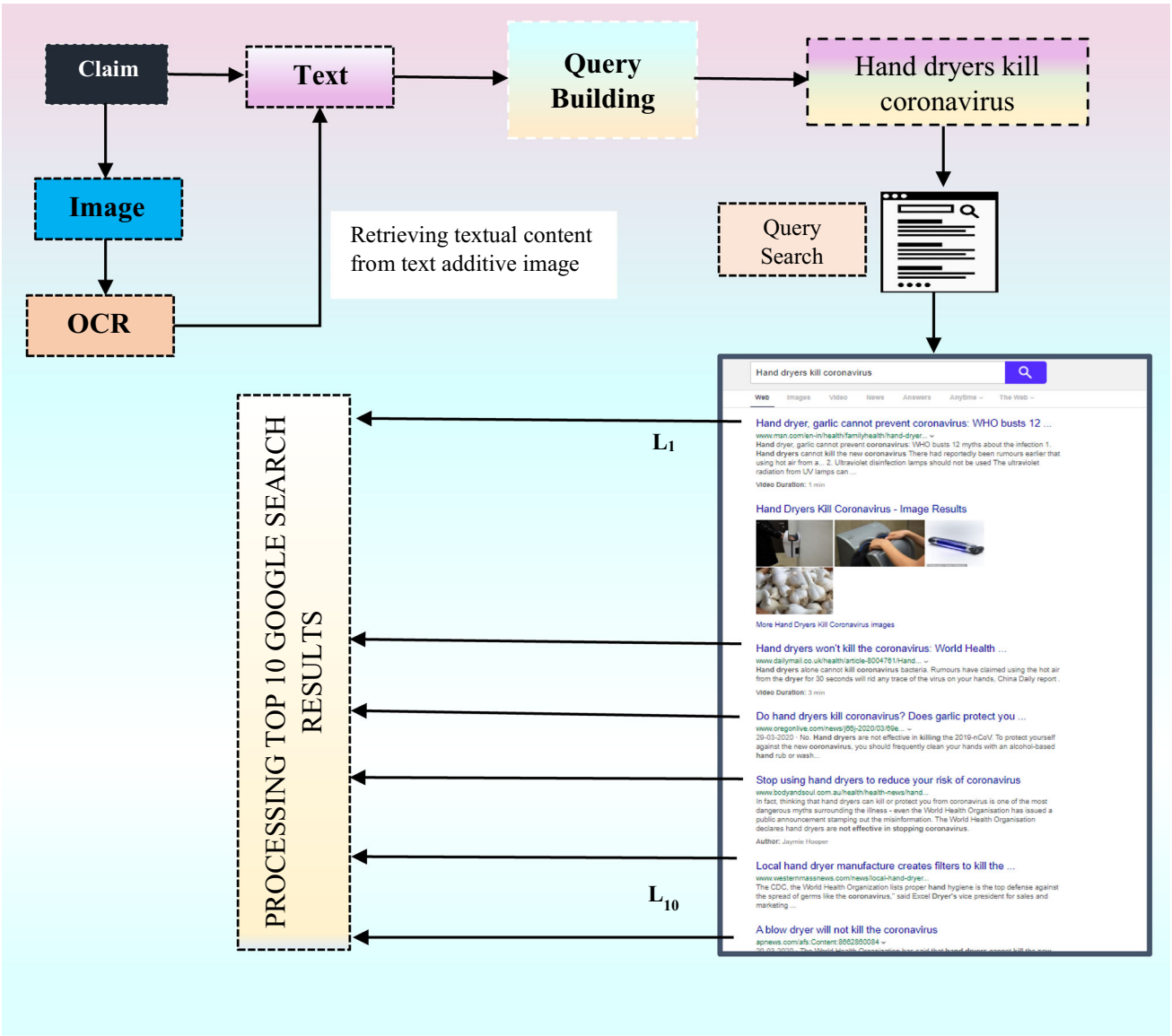


Fig. 3 Example of building query and search Process

sentences concerning false content. Some of the earlier studies have also mentioned that the N-gram method contributes to finding important clues from the news article title [2, 3]. We have created a corpus of most prominently used words to represent false information in our work. The N-Gram technique has been used to extract features (unigram, bigram, and trigram) concerning to each paragraphs. In the same way LDR method is one of the text-similarity measures between words, this gives a ratio indicating how similar the two sentences. Earlier studies have employed many different methods for text similarity (Cosine, LDR, Jaccard, etc.). The LDR is a good method for finding out or measuring the distance between sequence of characters.

The detailed description of the complete process is shown in the following steps.

3.3.1 Clues identification from news article headlines/titles

The effective clues are retrieved from news article title/headlines based on three proposed measures (Word Similarity Ratio, Levenshtein Distance Ratio, and N-Grams). A detailed description of each of these measures have shown below:

3.3.1.1 Word similarity ratio (WSR) This is the first measure considered for retrieving some important clues from the news headline. This is the ratio of the number of keywords (K_i) that are found to be similar in *news title* and the *query* to the total number of words in the query (t_i). The Ratio helps in identifying useful or relevant titles. The

Fig. 2 Word cloud representation for fake and real claims in the proposed dataset

more the word similarity ratio, the more relevant the title to the query.

$$\text{WSR} = \frac{K_i}{t_i} \tag{1}$$

3.3.1.2 Levenshtein distance ratio (LDR) Levenshtein Distance Ratio has been used to find the similarity between the query and the news article headline/title. To calculate the ratio, Levenshtein distance needs to be calculated. Levenshtein distance is a metric to measure how apart are two sequences of words or in another way can say it is a measure of the minimum number of edits that you need to do to change a one-word sequence into the other.

The formal definition of the Levenshtein distance between two string a and b can be seen as follows:

$$\text{lev}_{(a,b)}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0 \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j)+1 \\ \text{lev}_{a,b}(i,j-1)+1 \\ \text{lev}_{a,b}(i-1,j-1)+1_{(ai \neq bj)} \end{cases} & \text{otherwise} \end{cases} \tag{2}$$

where, $1_{(ai \neq bj)}$ is an indicator function equals to 0 when $a_i = b_i$ and 1 otherwise. The Levenshtein distance ratio can be retrieved based on the Levenshtein distance and can be calculated using the following formula:

$$\text{LDR} = \frac{(|a| + |b|) - \text{lev}_{(a,b)}(i,j)}{|a| + |b|} \tag{3}$$

where, $|a|$ and $|b|$ are the length of sequence a and sequence b, respectively. Here a is query and b is news article headline.

The main purpose of including Levenshtein distance is for string matching, and it is helpful to find matches for short strings in many longer texts.[8] In our case, it will be helpful, as the string may not be of the same length. The length of the title and the length of the query may not be the same and quite effective than other distance measures.

3.3.1.3 N-grams N-grams is an ordered n-tuple of characters. N-gram also contributes to finding important clues from the news article title [29]. Unigram takes a sentence and gives all the words in that we fence. Whereas, Bigram takes a sentence and gives us a set of two consecutive words in the sentences. A trigram gives sets of three consecutive words in a sentence. We manually analyzed titles used to represent and find a prominent keyword used in false query search responses and build a *false phrases dictionary* with 20 keywords. Some examples are (*"misleading," "misinformation," "is it," "is it true," "not known," "no proof," "no known," "no scientific evidence," "no evidence," "not verified," "hoax," "clickbait," "not proven," "denied," "deny," "unverified," "false," "fake," "fake news," "falsely," 'myth," "ridiculous," "rumour"*).

The word cloud representation of the respective keyword is shown in Fig. 4. The thing to be noted here, the keywords considered here are mainly focused on the news reported related to coronavirus. To extract all possible combinations, we have considered three types of n-gram (Unigram, Bigram, Trigram) to reduce the chance of missing any possible case. The count of unigram, bigram, and trigram is used as a clue, which reveals the number of titles reporting false keyword phrases. The detailed

[8] https://en.wikipedia.org/wiki/Levenshtein_distance.

Fig. 4 Word cloud representing key phrases used for false information prediction

description of features belonging to each category concerning to News article headline/ title and News article content is shown in Table 1.

3.3.2 Clues identification from news article content

As we have identified the clues in the previous section related to the top 10 news headlines/titles respective to each query. In the same way, the analysis has also been applied over new article content. New article content is one of the important parts that reveals what the article trying to convey, and whether the given content reporting the same thing as it claiming for. Analyzing news article content is needful to extract possible clues to predict whether the query is misleading or not. The article is quite long, analyzing it at once, and summarization is also not quite effective in getting an overview. To resolve this problem instead of analyzing the whole article at once or summarizing it, we extracted each article's first 8 paragraph (< p >) tags. The reason of considering first 8 < p > tags as from the analysis it has been found that most of the good articles covering the information within 8 paragraphs, as well as from the manual inspection it has also been observed that the 8 paragraphs are sufficient in getting the overview about an article. As discussed previously, each paragraph is inspected concerning two clue measures Levenshtein Distance Ratio and N-Grams. Both measures are applied in the same way but the context is changed, instead of applying it into news headline/title, news articles paragraphs are considered for fetching the possible clues. For each title, 8 paragraphs are employed to examine with respect to two clue measures as shown below:

3.3.2.1 Levenshtein distance ratio As we have previously discussed the Levenshtein distance ratio, the same is applied here. The LDR is calculated between the query and each paragraph text of the news article to inspect the similarity among them. This can be defined as for each query Q there are n paragraphs where the range is from $i = 1$ to $i = 8$ and the Levenshtein distance ratio is calculated between Q and n_i.

3.3.2.2 N- grams The N-gram measures are used in the same way as applied earlier. The n-grams are employed over each paragraph text The count of false keyword phrases is identified considering all three types 1-gram, 2-gram, and 3 grams that further be used to predict misleading information. For each n_i (n paragraphs), where the range is from $i = 1$ *to* $i = 8$, and for each paragraph, the count of keywords belonging to *false phrases dictionary* are calculated.

The proposed algorithm for the prediction of misleading content is given below in Algorithm 1: The proposed algorithm firstly checks for the user input whether it is an image or text. If it is an image embedded with text, then it first processed through the OCR algorithm, and the textual part is extracted which further be pre-processed to build the query for search. In the pre-processing part, stop words and punctuations have been removed. On the other hand, if the given input is text, it will go directly through the processing phase and the query building phase. The important clues for each query have been found out with respect to three feature sets based on word similarity, Levenshtein distance, and N-Gram. Finally, all the features are fed into the classification model for binary class classification and get the status (Misleading or real). Along with this, the

Table 1 Description of a set of features employed for misleading content detection

Feature category	Feature	Feature description	Clue identification type
Word similarity based	Word matching count (WMC)	The total number of words in the title/heading that are found to be similar and matching the words in the query	News Article Headline
Word similarity based	Query length (QL)	The word count of a query	News Article Headline
Word similarity based	Word matching ratio (WMR)	It is the ratio of word matching count and Query length	News Article Headline
Levenshtein Distance based	Title-query ratio	It is the Levenshtein distance ratio with respect to the title and the query	News Article Headline
N-gram based	1 Gram	Count of words(unigrams) found in the title/headline that belong to the false phrase dictionary	News Article Headline
N-gram based	2 Gram	Count of two consecutive words(Bigrams) found in the title/headline that belongs to the false phrase dictionary	News Article Headline
N-gram based	3 Gram	Count of three words(trigrams) found in the title/headline that belong to the false phrase dictionary	News Article Headline
Levenshtein distance based	Text-Query Ratio	It is the Levenshtein distance ratio with respect to the news article paragraph and the query	News Article Content
N-gram based	1 Gram(article paragraph)	Count of words(unigrams) found in each n article paragraph that belong to the false phrase dictionary	News Article Content
N-gram based	2 Gram(article paragraph)	Count of two consecutive words(Bigrams) found in each n article paragraph that belongs to the false phrase dictionary	News Article Content
N-gram based	3 Gram(article paragraph)	Count of three consecutive words(trigrams) found in each n article paragraph that belong to the false phrase dictionary	News Article Content

detailed flow diagram of the proposed approach as discussed above is shown in Fig. 5.

the self-generated dataset *Covid-fakenews2019* dataset and publically available dataset (Liar Dataset). Currently, no fake news dataset is available for COVID-19, the experiment is conducted on the proposed dataset, and the model's performance is evaluated on the same. To validate the model's generalizability, the testing has also been performed on the publicly available dataset "Liar." The metric used in this paper is Accuracy(Acc), Precision(Pre),

Algorithm 1.(Misleading information Detection)

Input(Text/Text additive image) and Output(Status)

```
def. main ():
  if(image)
       claim= OCR(image)
  else:
  claim= raw input(claim)
  query= pre-processing(claim)
  Feature1= Word_Similarity_based_feature (query);
  Feature2= Levenshtein_Distance_based feature (query);
  Feature3= N-Gram based Feature (query);
  Status= Classification_model (Feature1, Feature 2, Feature 3)
  Output(Status) //Misleading or real
```

4 Experimental results

This section describes the experiments conducted to respond to the questions we draw up in the introduction part (Sect. 1). The performance experiment is made with

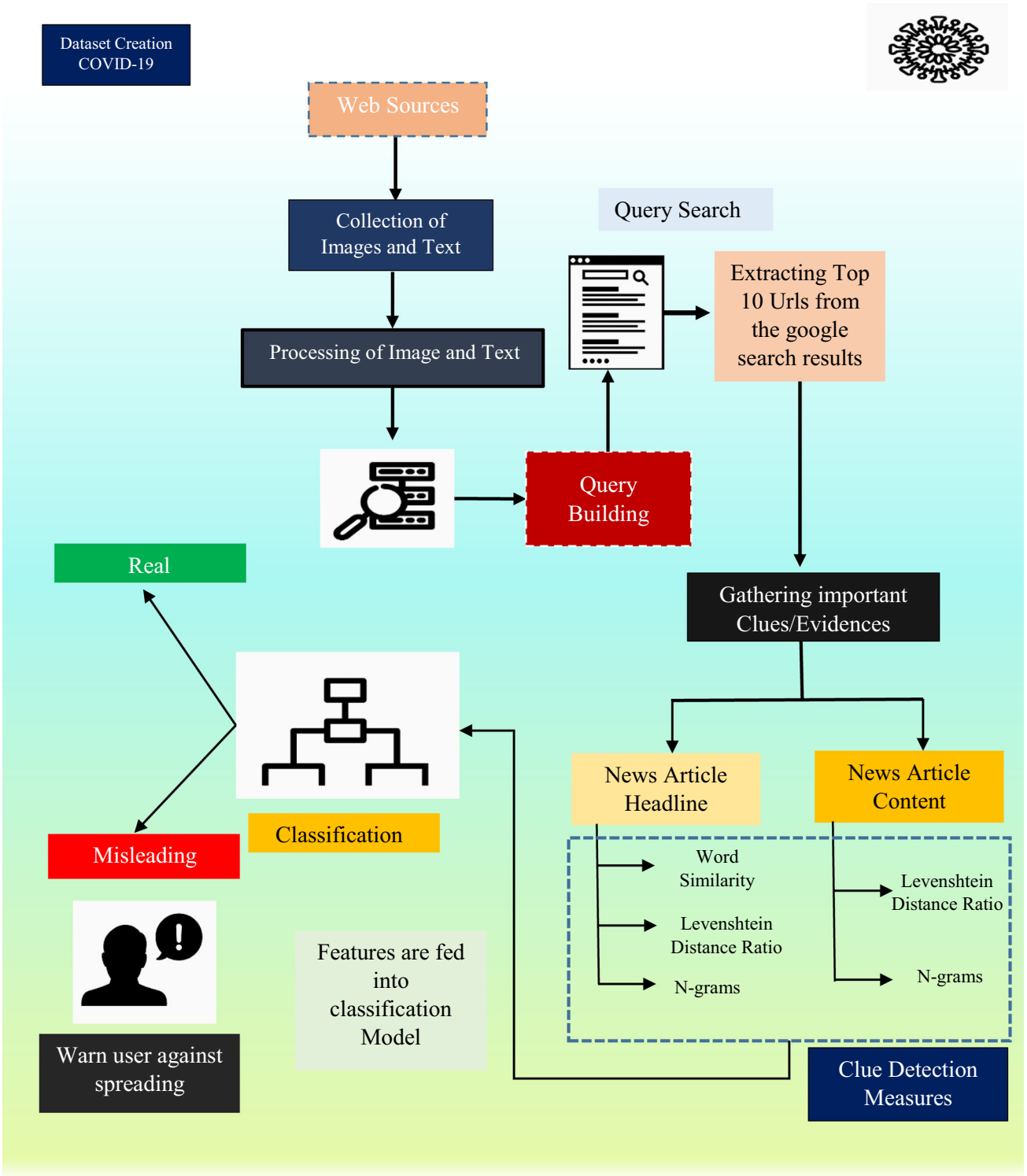


Fig. 5 The figure shows the flow diagram of the proposed method

Recall(Rec), and F1-Score(F1). The experiment is conducted to analyze the performance of the proposed model. The evaluation results analysis on the self-generated and Liar datasets is discussed in Sects. 4.1 and 4.2.

4.1 Evaluation on self-generated dataset (*Covid-fakenews2019)*

To evaluate the proposed algorithm on the Self-Generated Dataset, the percentage split technique has been

Table 2 Effectiveness of proposed model on the Covid-fakenews2019 dataset with respect to New article headline

Classifier	Word similarity				Levenshtein distance ratio				N-Grams				All- features			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Logistic Regression	**99.98**	0.99	0.99	0.99	**99.98**	0.99	0.99	0.99	99.93	0.99	0.99	0.99	99.85	0.99	0.99	0.99
Naïve Bayes	99.02	0.99	0.99	0.99	99.02	0.99	0.99	0.99	99.03	0.99	0.99	0.99	98.62	0.98	0.98	0.98
Random forest	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99
Adaboost	67.46	0.67	0.67	0.67	59.01	0.59	0.59	0.56	56.74	0.66	0.56	0.51	70.16	0.70	0.70	0.69
K-Nearest	**99.98**	0.99	0.99	0.99	**99.98**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	99.96	0.99	0.99	0.99
SGD	**99.98**	0.99	0.99	0.99	**99.98**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99

Table 3 Effectiveness of proposed model on the Covid-fakenews 2019 dataset with respect to New article content

Classifier	Levenshtein distance ratio				N-grams				All- features			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Logistic regression	99.94	0.99	0.99	0.99	99.96	0.99	0.99	0.99	**99.94**	0.99	0.99	0.99
Naïve bayes	99.94	0.99	0.99	0.99	99.87	0.99	0.99	0.99	99.80	0.99	0.99	0.99
Random forest	**99.99**	0.99	0.99	0.99	**99.99**	0.99	0.99	0.99	99.89	0.99	0.99	0.99
Adaboost	59.48	0.60	0.59	0.57	56.25	0.67	0.56	0.50	60.54	0.62	0.60	0.57
K-nearest	**99.99**	0.99	0.99	0.99	99.92	0.99	0.99	0.99	99.89	0.99	0.99	0.99
SGD	**99.99**	0.99	0.99	0.99	99.98	0.99	0.99	0.99	99.89	0.99	0.99	0.99

Table 4 Effectiveness of proposed model on the Covid-fakenews2019 dataset with respect to both new article headline and content

Classifier	10-fold cross-validation				20-fold cross-validation				Percentage split(70:30)			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
Logistic regression	99.95	0.99	0.99	0.99	99.96	0.99	0.99	0.99	99.92	0.99	0.99	0.99
Naïve bayes	98.41	0.98	0.98	0.98	98.41	0.98	0.98	0.98	98.33	0.98	0.98	0.98
Random forest	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**
Adaboost	70.42	0.71	0.70	0.69	69.98	0.70	0.70	0.69	69.98	0.70	0.70	0.69
K-nearest	99.96	0.99	0.99	0.99	99.97	0.99	0.99	0.99	99.96	0.99	0.99	0.99
SGD	99.98	0.99	0.99	0.99	99.98	0.99	0.99	0.99	99.92	0.99	0.99	0.99
Linear SVM	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**	**99.99**	**0.99**	**0.99**	**0.99**

employed with 70% samples are used to train the model and 30% to test the data. The performance has been analyzed based on three different ways. Firstly, the effectiveness of the proposed model is evaluated with respect to news article title/headline using percentage split technique (70:30) as shown in Table 2. Secondly, performance concerning to the news article content is evaluated and lastly, the effectiveness is evaluated by employing percentage split as well as cross-fold validation technique (tenfold and 20 fold) by combining the clues identified from both news article headline and its content as shown in Tables 3 and 4, respectively. The effectiveness of the proposed model has been evaluated with respect to the features employed for the news article headline which shows that the SVM classifier outperforms all other classifiers when considering all sets of features. Form the observation analysis shown in Tables 2, 3, and 4, most of the classifiers perform best on our proposed method by employing different set of categories, which clearly shows the effectiveness of the proposed model. The effectiveness of the proposed model by incorporating all set of features in terms of accuracy measure is shown in Fig. 6. From the figure, it can be clearly seen that most of the classifiers perform well with respect to different testing schemes (tenfold, 20 fold, 70:30 split). The highest value achieved by the different models are shown in Bold text in Table 2, 3, 4, and 5.

4.2 Evaluation on Liar dataset

To validates the generalizability of the proposed model and to do a comparative analysis with the state-of-the-art methods to show the effectiveness of the proposed

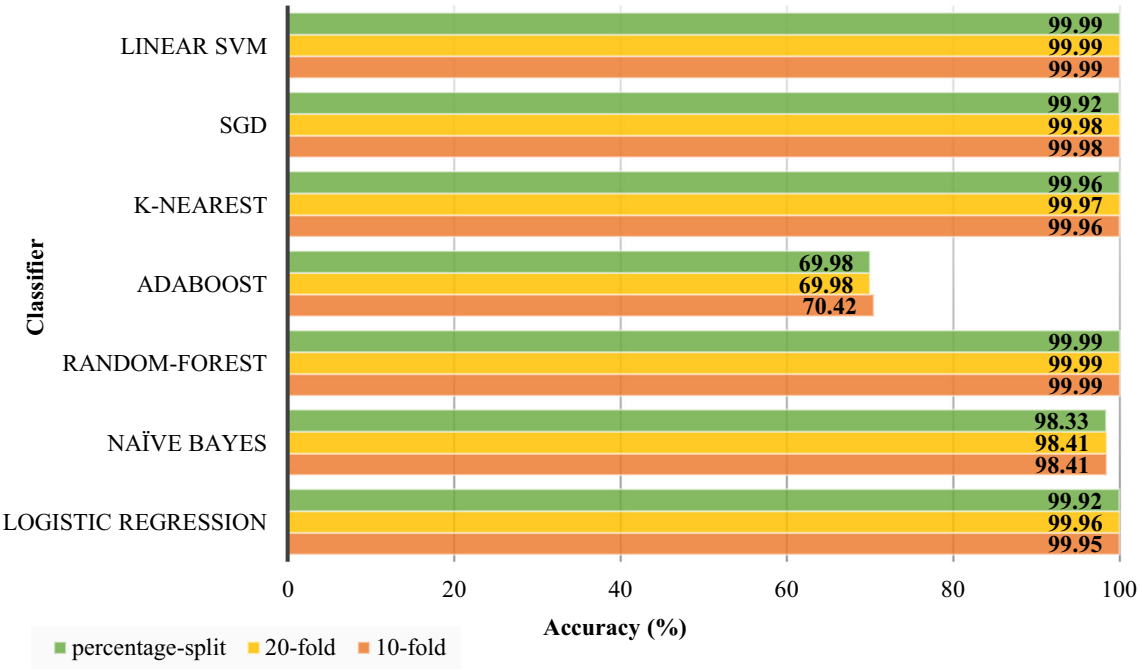


Fig. 6 Effectiveness of the proposed model on Covid-fakenews 2019 dataset

algorithm for predicting the misleading post on the social media platform, we have employed another standard dataset "Liar Dataset" to evaluate the performance of the proposed algorithm. Liar dataset was created by [28], which is the collection of 12,836 short-statements from Politifact platform[9](Fact checking website) which verify and label any text claim into six categories "true," "mostly-true," "half-true," "mostly-false," "false" and "pants-fire." As per the probabilistic analysis, the Politifact considered the true and mostly true sample in the credible category and labeled as true, whereas half true, mostly false, false and pants-on-fire samples are considered in the false category.[10] This categorization is based on the Politifact description given for the true and false label, which classified true and mostly true as accurate statements and others as inaccurate. Accordingly, we have considered these six label categories into true and false samples for the binary classification. The authors of [1], also applied the same way for the binary classification. The proposed solution has been applied on the given dataset, and the detailed evaluation measure like Accuracy, Precision, Recall, F-Measure, and analysis results is shown in Table 5. From Table 5, it can be observed that the proposed technique outperforms the existing algorithm by employing different classification algorithms with respect to Accuracy, Precision, Recall, and F1score. Most of the classifiers perform best, whereas SVM outperforms all other with an F1 score of 0.99 on both percentages split and cross-validation scheme. The existing work analysis also reports SVM as an effective classifier for the prediction of fake news [29, 30].

In the field of misleading information detection, very few standard datasets are available and one of the popular datasets that have been widely used for analysis is

Table 5 Comparative study of the proposed method with the existing state-of-the-art methods

References	Model	Acc	Pre	Rec	F1
[29]	SVM	0.56	0.57	0.56	0.48
	LR	0.56	0.56	0.56	0.51
	Decision tree	0.51	0.51	0.51	0.51
	Ad boost	0.56	0.56	0.56	0.54
	Naïve-bayes	0.60	0.59	0.60	0.59
	K-NN	0.53	0.53	0.53	0.53
[28]	SVM	0.25	–	–	–
	LR	0.24	–	–	–
Our method	Random forest	0.98	0.98	0.98	0.98
	LR	0.97	0.98	0.97	0.96
	Naïve bayes	0.97	0.98	0.97	0.95
	Decision tree	0.97	0.98	0.97	0.96
	k-NN	0.97	0.98	0.97	0.95
	SVM	**0.99**	**0.99**	**0.99**	**0.99**

[9] https://www.politifact.com/.

[10] https://towardsdatascience.com/identifying-fake-news-the-liar-dataset-713eca8af6ac.

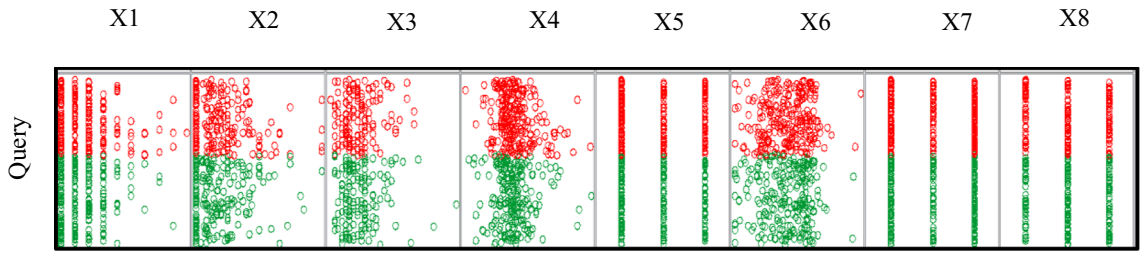


Fig. 7 Scatter plot of each features corresponding to the query

Table 6 List of features and their descriptions

Feature no	Feature description
X1	Word_matching_title_query
X2	Word_match_ratio
X3	Query_length
X4	Title_query_ratio
X5	Title_sentiment
X6	Query_text_ratio
X7	Text_sentiment
X8	Query_sentiment

the LIAR dataset. The comparative analysis has been performed with the existing algorithm on the same dataset. The authors of [29], have employed lexical and sentiment features as well as also included n-grams features for fake news detection. Lexical and sentiment features have included average word length, article length, and exclamation mark count as lexical. Every article's positive, negative, and neutral sentiment have been measured. Unigrams, bigrams, and TF-IDF vectorizer functions have been incorporated to generate TF-IDF n-gram features. The performance analysis over traditional machine learning models reveals that Naïve Bayes with n-gram features perform best in comparison with all other classifiers with an accuracy of 0.60. In [28], the author has exploited text-based features by employing SVM and Logistic regression classifier and achieving an accuracy of 0.25 and 0.24, respectively. The proposed algorithm outperforms the state-of-the-art methods that have been discussed above and SVM found to be the best with an F1 score of 0.99. From the proposed feature analysis, it has been found that Levenshtein distance and N-grams based features are quit efficient and significant for the prediction of misleading information [29, 29]. It has also been observed that news article title and paragraph wise content analysis gives efficient clues for the prediction and is useful in improving the model's performance. None of the previous work has included these paradigms.

It is seen from Table 5, that approximately all given classifiers are performing well on our data. To conclude, the method suggested in the study can identify misinformation at 99% confidence level. The procedure we have applied to show statistical analysis is as follows. Firstly, we identify and depict the nature of our data samples. Next we establish a relation between the data samples concerning to query w.r.t each derived feature $X1, X2, X3 \ldots X8$. To show the clear pattern of data samples concerning fake and real class, a scatter plot is shown in Fig. 7. As in the diagram, it can be clearly visualize that the data samples are linearly separable. The decision boundary can distinguish the real and fake samples correctly because the data does not contain noise and less overlapping. The classification can be easily done as shown in Fig. 7. Here, green samples are considered as true and the red samples are considered as fake. The feature $X1, X2, X3 \ldots X8$ are described briefly in Table 6.

A statistical analysis of the result is performed through a hypothesis test, where we have considered the level of confidence is 99%. We considered a Null Hypothesis (H0) and Alternative Hypothesis (H1). The alternative hypothesis is the initial hypothesis that predicts a relationship between variables. Whereas the null hypothesis is a prediction of no relationship between the variables. The hypothesis for our case is as follows:

H0: None of the given features are effective in predicting misinformation.

H1: The given features can predict misinformation effectively.

The confidence interval method has been applied here, the test statistic between or outside the confidence interval. We performed the Z-Test,[11] were two samples(query_text_ratio, title_query_ratio) for mean on 99% confidence interval, and it can be clearly seen from Table 7 that the test is rejecting the null hypothesis because p (significance value) < 0.01, this implies it accept the alternative hypothesis (The given features can predict misinformation

[11] https://www.geeksforgeeks.org/z-test/.

Table 7 Statistical analysis results on Z-Test: Two Sample for Means

Parameters	Query_text_ratio	Title_query_ratio
Mean	0.282686586	0.402105109
Known variance	1.38E−02	1.10E−02
Observations	22,366	22,366
Hypothesized mean difference	0	
Z	− 113.366	
$p(Z <= z)$ one-tail	0.000	
Z Critical one-tail	2.326	
$p(Z <= z)$ two-tail	0.000	
Z Critical two-tail	2.576	

effectively) with the confidence level of 99%. Hence, the results are statistically significant and accept H1 (The given features can predict misinformation effectively).

5 Conclusion

In this paper, we developed an intelligent generalized strategy for identifying possible clues to predict misleading information where fake news proliferated during the COVID-19 outbreak is considered as a special case study and detailed analysis has been discussed. To extract important clues for the verification of news content, the scheme leverages three sets of novel features proposed in this paper that are based on Word Similarity, Levenshtein Distance, and N-Grams. These features are extracted from the news article headlines/titles and their content. The set of features is given as an input to the machine learning model. Performance has been analyzed by considering different variants of feature category by employing a different set of classifier applied on self-generated dataset *Covid-fakenews2019*. and the *Liar dataset*. It has been observed that most of the classifiers are performing well by employing both the dataset, where SVM outperforms all other classifiers with an f1-score of 0.99. In future, we are planning to build the real-time application of the proposed work and also extend the proposed dataset by incorporating a large number of samples with different categories of news that proliferated during the COVID-19 outbreak, like in this work, we are more focusing into the fake news propagated with respect to cures/treatment of the disease. In future, one can also include other types of news like false information regarding lockdowns in the cities, fake claims that are pretended to be posted by some government officials, etc.

Data availability Data will be made available upon request.

Declarations

Conflict of interest It is declare that we do not have any conflict of interest.

References

1. Sahu KK, Mishra AK, Lal A (2020) Comprehensive update on current outbreak of novel coronavirus infection (2019-nCoV). Ann Transl Med 8(6)
2. Boididou C, Papadopoulos S, Zampoglou M, Apostolidis L, Papadopoulou O, Kompatsiaris Y (2018) Detection and visualization of misleading content on Twitter. Int J Multimed Inf Retr 7(1):71–86
3. Zhou W, Wang A, Xia F, Xiao Y, Tang S (2020) Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak. Math Biosci Eng 17(3):2693–2707
4. Castillo C, Mendoza M, Poblete B (2011) Information credibility on Twitter. In: Proceedings of the 20th international conference on world wide web. pp 675–684
5. Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2017) Novel visual and statistical image features for microblogs news verification. Trans Multi 19(3):598–608
6. Vishwakarma DK, Varshney D, Yadav A (2019) Detection and veracity analysis of fake news via scrapping and authenticating the web search. Cogn Syst Res 58:217–229
7. Wu K, Yang S, Zhu KQ (2015) False rumors detection on Sina Weibo by propagation structures. In: 2015 IEEE 31st Int. Conf. Data Eng. pp 651–662
8. Horne BD, Adali S (2017) This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Eleventh international AAAI conference on web and social media
9. Chakraborty A, Paranjape B, Kakarla S, Ganguly N (2016) Stop clickbait: detecting and preventing clickbaits in online news media. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). pp 9–16
10. Aker A, Derczynski L, Bontcheva K (2017) Simple open stance classification for rumour analysis. arXiv Prepr arXiv:1708.05286
11. Briscoe EJ, Appling DS, Hayes H (2014) Cues to deception in social media communications. In: 2014 47th Hawaii international conference on system sciences. pp 1435–1443
12. Giasemidis G et al. (2016) Determining the veracity of rumours on Twitter. CoRR, vol. abs/1611.0
13. Swon S, Cha M, Jung K, Chen W, Wang Y (2013) Prominent features of rumor propagation in online social media. In: 2013 IEEE 13th Int. Conf. Data Min. pp 1103–1108
14. Zeng L, Starbird K, Spiro ES (2016) "# unconfirmed: classifying rumor stance in crisis-related social media messages. In: Tenth international AAAI conference on web and social media
15. Derczynski L, Bontcheva K, Liakata M, Procter R, Hoi GWS, Zubiaga A (2017) SemEval-2017 task 8: RumourEval: determining rumour veracity and support for rumours. CoRR, vol. abs/1704.0
16. Tacchini E, Ballarin G, Della Vedova ML, Moret S, de Alfaro L (2017) Some like it hoax: automated fake news detection in social networks. arXiv Prepr arXiv:1704.07506
17. Zhou L, Twitchell DP, Qin T, Burgoon JK, Nunamaker JF (2003) An exploratory study into deception detection in text-based computer-mediated communication. In: 36th annual Hawaii international conference on system sciences. pp 10

18. Ferreira W, Vlachos A (2016) Emergent: a novel data-set for stance classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. pp 1163–1168
19. Breiman L (2017) Classification and regression trees. Routledge, London
20. Loh W-Y (2011) Classification and regression trees. Wiley Interdiscip Rev Data Min Knowl Discov 1(1):14–23
21. Bodnar T, Tucker C, Hopkinson K, Bilén SG (2014) Increasing the veracity of event detection on social media networks through user trust modelling. In 2014 IEEE international conference on Big Data (Big Data). pp 636–643
22. Yu F et al (2016) Detecting rumors from microblogs with recurrent neural networks. CoRR 8(1):1435–1443
23. Kochkina E, Liakata M, Zubiaga A (2018) All-in-one: multi-task learning for rumour verification. In: Proceedings of the 27th international conference on computational linguistics. pp 3402–3413
24. Jacovi A, Shalom OS, Goldberg Y (2018) Understanding convolutional neural networks for text classification. arXiv Prepr arXiv:1809.08037
25. Yu F, Liu Q, Wu S, Wang L, Tan T (2019) Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. Comput Secur 83:106–121
26. Song C, Yang C, Chen H, Tu C, Liu Z, Sun M (2019) CED: Credible early detection of social media rumors. IEEE Trans Knowl Data Eng 33(8):3035–3047
27. Ajao O, Bhowmik D, Zargari S (2018) Fake news identification on twitter with hybrid cnn and rnn models. In: Proceedings of the 9th international conference on social media and society. pp 226–230
28. Wang W (2017) Liar, Liar Pants on Fire': a new benchmark dataset for fake news detection
29. Khan JY, Khondaker M, Islam T, Iqbal A, Afroz S (2019) A benchmark study on machine learning methods for fake news detection. arXiv Prepr arXiv:1905.04749
30. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Inf Sci 497:38–55
31. Ahmed H, Traore I, Saad S (2017) Detection of online fake news using N-gram analysis and machine learning techniques. In: International conference on intelligent, secure, and dependable systems in distributed and cloud environments. pp 127–138

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.

Noname manuscript No.
(will be inserted by the editor)

Geometric Properties of Functions Containing Derivatives of Bessel Functions

Sercan Kazımoğlu · Kamaljeet Gangania*

arXiv:2211.12697v1 [math.CV] 23 Nov 2022

Received: date / Accepted: date

Abstract In this paper our aim is to find the radii of γ-Spirallike of order α and convex γ-Spirallike of order α for three different kinds of normalizations of the function $N_\nu(z) = az^2 J_\nu''(z) + bzJ_\nu'(z) + cJ_\nu(z)$, where $J_\nu(z)$ is the Bessel function of the first kind of order ν. Moreover, the $\mathcal{S}^*(\varphi)-$radii and $\mathcal{C}(\varphi)-$radii of these normalized functions are investigated. The tables are created and visual verification with graphs are made by giving special values to the real numbers a, b and c in the obtained results.

Keywords γ-Spirallike functions · Radii of starlikeness and convexity · Bessel functions · Zeros of Bessel function derivatives · Radius.

Mathematics Subject Classification (2010) 30C45 · 30C80 · 30C15

1 Introduction

Let $\mathcal{A}$ be the class of analytic functions normalized by the condition $f(0) = 0 = f'(0) - 1$ in the unit disk $\mathbb{D} := \mathbb{D}_1$, where $\mathbb{D}_r := \{z \in \mathbb{C} : |z| < r\}$. We say that a function $f \in \mathcal{A}$ is γ-Spirallike of order α if and only if

$$\operatorname{Re}\left(e^{-i\gamma}\frac{zf'(z)}{f(z)}\right) > \alpha\cos\gamma,$$

where $\gamma \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $0 \leq \alpha < 1$. The class of such functions, we denote by $\mathcal{S}_p^\gamma(\alpha)$. In view of well known Alexender's relation, let $\mathcal{CS}_p^\gamma(\alpha)$ be the class of

Sercan Kazımoğlu
E-mail: srcnkzmglu@gmail.com
Department of Mathematics, Faculty of Science and Literature, Kafkas University, 36100, Kars-Turkey
Kamaljeet Gangania
E-mail: gangania.m1991@gmail.com
Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
*Corresponding author

convex γ-spirallike functions of order α, which is defined below

$$\operatorname{Re}\left(e^{-i\gamma}\left(1+\frac{zf''(z)}{f'(z)}\right)\right) > \alpha\cos\gamma.$$

Spacek [30] introduced and studied the class $\mathcal{S}_p^\gamma(0)$. Each function in $\mathcal{CS}_p^\gamma(\alpha)$ is univalent in $\mathbb{D}$, but they do not necessarily be starlike. Further, it is worth to mention that for general values of $\gamma(|\gamma| < \pi/2)$, a function in $\mathcal{CS}_p^\gamma(0)$ need not be univalent in $\mathbb{D}$. For example: $f(z) = i(1-z)^i - i \in \mathcal{CS}_p^{\pi/4}(0)$, but not univalent. Indeed, $f \in \mathcal{CS}_p^\gamma(0)$ is univalent if $0 < \cos\gamma < 1/2$, see Robertson [28] and Pfaltzgraff [27]. Note that for $\gamma = 0$, the classes $\mathcal{S}_p^\gamma(\alpha)$ and $\mathcal{CS}_p^\gamma(\alpha)$ reduce to the classes of starlike and convex functions of order α, given by

$$\operatorname{Re}\left(\frac{zf'(z)}{f(z)}\right) > \alpha \quad \text{and} \quad \operatorname{Re}\left(1+\frac{zf''(z)}{f'(z)}\right) > \alpha,$$

which we denote by $\mathcal{S}^*(\alpha)$ and $\mathcal{C}(\alpha)$, respectively.

In the recent past, connections between the special functions and their geometrical properties have been established in terms of radius problems [1, 2, 3, 5, 8, 9, 10, 15]. In particular, the $\mathcal{S}^*(\alpha)$-radius of a normalized function f is to find

$$\sup\{r \in \mathbb{R}^+ : \operatorname{Re}\left(\frac{zf'(z)}{f(z)}\right) > \alpha, z \in \mathbb{D}_r\}$$

and similarly, we can define $\mathcal{C}(\alpha)$-radius. In this direction, for $\mathcal{S}^*(\alpha)$-radius and more, we refer to Bessel functions [1, 3, 5] (see Watson's treatise [33] for more on Bessel function), Struve functions [1, 2], Wright functions [9], Lommel functions [1, 2] and Ramanujan type entire functions [15]. For their generalization to Ma-Minda classes [24] of starlike and convex functions, we refer to see [16, 21].

However, with the best of our knowledge, $\mathcal{S}_p^\gamma(\alpha)$-radius and $\mathcal{CS}_p^\gamma(\alpha)$-radius for special functions are not handled till date. Therefore, we define $\mathcal{S}_p^\gamma(\alpha)$-radius and its convex analog here below:

Definition 1 Let f in $\mathcal{A}$ be a special function. Then the radius of γ-Spirallike of order α is to find

$$r_{sp}^*(\alpha,\gamma;f) = \sup\left\{r \in \mathbb{R}^+ : \operatorname{Re}\left(e^{-i\gamma}\frac{zf'(z)}{f(z)}\right) > \alpha\cos\gamma, z \in \mathbb{D}_r\right\}$$

and the radius of convex γ-Spirallike of order α is given by

$$r_{sp}^c(\alpha,\gamma;f) = \sup\left\{r \in \mathbb{R}^+ : \operatorname{Re}\left(e^{-i\gamma}\left(1+\frac{zf''(z)}{f'(z)}\right)\right) > \alpha\cos\gamma, z \in \mathbb{D}_r\right\}.$$

In the present investigation, we consider our special normalized functions to be the derivatives of Bessel functions. For this to proceed further, recall that the Bessel function of the first kind of order ν is defined by [26, p.217]

$$J_\nu(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+\nu+1)}\left(\frac{z}{2}\right)^{2n+\nu} \quad (z \in \mathbb{C}). \tag{1.1}$$

We know that it has all its zeros real for $\nu > -1$. Here now we consider mainly the general function

$$N_\nu(z) = az^2 J_\nu''(z) + bzJ_\nu'(z) + cJ_\nu(z)$$

studied by Mercer [25]. Here, as in [25], $q = b - a$ and

$$(c = 0 \text{ and } q \neq 0) \text{ or } (c > 0 \text{ and } q > 0).$$

From (1.1), we have the power series representation

$$N_\nu(z) = \sum_{n=0}^{\infty} \frac{Q(2n+\nu)(-1)^n}{n!\Gamma(n+\nu+1)} \left(\frac{z}{2}\right)^{2n+\nu} \qquad (z \in \mathbb{C}) \tag{1.2}$$

where $Q(\nu) = a\nu(\nu-1)+b\nu+c$ $(a, b, c \in \mathbb{R})$. There are three important works on the function N_ν. Firstly, Mercer's paper [25] in which it has been proved that the k^{th} positive zero of N_ν increases with ν in $\nu > 0$. Secondly, Ismail and Muldoon [17] showed that under the conditions $a, b, c \in \mathbb{R}$ such that $c = 0$ and $b \neq a$ or $c > 0$ and $b > a$;

(i) For $\nu > 0$, the zeros of $N_\nu(z)$ are either real or purely imaginary.
(ii) For $\nu \geq \max\{0, \nu_0\}$, where ν_0 is the largest real root of the quadratic $Q(\nu) = a\nu(\nu-1) + b\nu + c$, the the zeros of $N_\nu(z)$ are real.
(iii) If $\nu > 0$, $\left(a\nu^2 + (b-a)\nu + c\right) / (b-a) > 0$ and $a/(b-a) < 0$, the zeros of $N_\nu(z)$ are all real except for a single pair which are conjugate purely imaginary.

Baricz et al. [6] obtained sufficient and necessary conditions for the starlikeness of a normalized form of N_ν by using results of Mercer [25], Ismail and Muldoon [17] and Shah and Trimble [29].

Note that N_ν do not belong to $\mathcal{A}$. To prove our main results, we consider the following normalizations of the function N_ν given by:

$$f_\nu(z) = \left[\frac{2^\nu \Gamma(\nu+1)}{Q(\nu)} N_\nu(z)\right]^{\frac{1}{\nu}}, \tag{1.3}$$

$$g_\nu(z) = \frac{2^\nu \Gamma(\nu+1) z^{1-\nu}}{Q(\nu)} N_\nu(z), \tag{1.4}$$

$$h_\nu(z) = \frac{2^\nu \Gamma(\nu+1) z^{1-\frac{\nu}{2}}}{Q(\nu)} N_\nu(\sqrt{z}). \tag{1.5}$$

In the rest of this paper, for the quadratic $Q(\nu) = a\nu(\nu-1) + b\nu + c$, we will always assume that $a, b, c \in \mathbb{R}$ $(c = 0 \text{ and } a \neq b)$ or $(c > 0 \text{ and } a < b)$. Moreover, ν_0 is the largest real root of the quadratic $Q(\nu)$ defined according to the above conditions.

One can see that the functions $N_\nu(z)$ and $N_\nu'(z)$ are entire functions of order zero, then they have infinitely many zeros. According to Hadamard factorization theorem [22]

$$N_\nu(z) = \frac{Q(\nu) z^\nu}{2^\nu \Gamma(\nu+1)} \prod_{n \geq 1} \left(1 - \frac{z^2}{\lambda_{\nu,n}^2}\right) \tag{1.6}$$

and

$$N'_\nu(z) = \frac{Q(\nu)z^{\nu-1}\nu}{2^\nu \Gamma(\nu+1)} \prod_{n\geq 1}\left(1 - \frac{z^2}{\lambda'^2_{\nu,n}}\right) \tag{1.7}$$

where $\lambda_{\nu,n}$ and $\lambda'_{\nu,n}$ denote the nth positive zero of $N_\nu(z)$ and $N'_\nu(z)$, respectively. Recently, for studies on the Geometric properties of Bessel functions, see [3,4,5,6,7,11,12,13,14,20,23,31].

Recall that Ma and Minda [24] subclasses of starlike and convex functions, respectively, are given by:

$$\mathcal{S}^*(\varphi) = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \varphi(z) \right\}$$

and

$$\mathcal{C}(\varphi) = \left\{ f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec \varphi(z) \right\},$$

where φ, a Ma-Minda function, is analytic and univalent with $\operatorname{Re}\varphi(z) > 0$, $\varphi'(0) > 0$ and $\varphi(\mathbb{D})$ is starlike with respect to $\varphi(0) = 1$ and symmetric about real axis. Note that $\varphi \in \mathcal{P}$, the class of normalized Carathéodory functions. We also recall that

Definition 2 Let $f \in \mathcal{A}$ be a special function. Then $\mathcal{S}^*(\varphi)$ −radius and $\mathcal{C}(\varphi)$ −radius of f are defined as follows:

$$r^*_\varphi(f) = \sup\left\{ r \in \mathbb{R}^+ : \frac{zf'(z)}{f(z)} \in \varphi(\mathbb{D}),\ z \in \mathbb{D}_r \right\}$$

and

$$r^c_\varphi(f) = \sup\left\{ r \in \mathbb{R}^+ : 1 + \frac{zf''(z)}{f'(z)} \in \varphi(\mathbb{D}),\ z \in \mathbb{D}_r \right\},$$

respectively.

In this paper, we deal with the radius of γ-Spirallike of order α and the radius of convex γ-Spirallike of order α for the functions $f_\nu(z)$, $g_\nu(z)$ and $h_\nu(z)$ defined by (1.3), (1.4) and (1.5) in the case when $\nu \geq \max\{0, \nu_0\}$. Also, we determine the $\mathcal{S}^*(\varphi)$ −radii and $\mathcal{C}(\varphi)$ −radii of these functions. The key tools in their proofs are some new Mittag-Leffler expansions for quotients of the function N_ν, special properties of the zeros of the function N_ν and their derivatives.

2 Zeros of hyperbolic polynomials and the Laguerre–Pólya class of entire functions

In this section, we recall some necessary information about polynomials and entire functions with real zeros. An algebraic polynomial is called hyperbolic if all its zeros are real. We formulate the following specific statement that we shall need, see [14] for more details.

By definition, a real entire function ψ belongs to the Laguerre–Pólya class $\mathcal{LP}$ if it can be represented in the form

$$\psi(x) = cx^m e^{-ax^2+\beta x} \prod_{k\geq 1} \left(1 + \frac{x}{x_k}\right) e^{-\frac{x}{x_k}},$$

with $c, \beta, x_k \in \mathbb{R}$, $a \geq 0$, $m \in \mathbb{N} \cup \{0\}$ and $\sum x_k^{-2} < \infty$. Similarly, ϕ is said to be of type $\mathcal{I}$ in the Laguerre-Pólya class, written $\varphi \in \mathcal{LPI}$, if $\phi(x)$ or $\phi(-x)$ can be represented as

$$\phi(x) = cx^m e^{\sigma x} \prod_{k\geq 1} \left(1 + \frac{x}{x_k}\right),$$

with $c \in \mathbb{R}$, $\sigma \geq 0$, $m \in \mathbb{N} \cup \{0\}$, $x_k > 0$ and $\sum x_k^{-1} < \infty$. The class $\mathcal{LP}$ is the complement of the space of hyperbolic polynomials in the topology induced by the uniform convergence on the compact sets of the complex plane while $\mathcal{LPI}$ is the complement of the hyperbolic polynomials whose zeros possess a preassigned constant sign. Given an entire function φ with the Maclaurin expansion

$$\varphi(x) = \sum_{k\geq 0} \mu_k \frac{x^k}{k!},$$

its Jensen polynomials are defined by

$$P_m(\varphi; x) = P_m(x) = \sum_{k=0}^{m} \binom{m}{k} \mu_k x^k.$$

The next result of Jensen [18] is a well-known characterization of functions belonging to $\mathcal{LP}$.

Lemma 1 *The function φ belongs to $\mathcal{LP}$ ($\mathcal{LPI}$, respectively) if and only if all the polynomials $P_m(\varphi; x)$, $m = 1, 2, ...$, are hyperbolic (hyperbolic with zeros of equal sign). Moreover, the sequence $P_m(\varphi; z/n)$ converges locally uniformly to $\varphi(z)$.*

The following result is a key tool in the proof of main results.

Lemma 2 *[19] If $\nu \geq \max\{0, \nu_0\}$ then the functions $z \longmapsto \Psi_\nu(z) = \frac{2^\nu \Gamma(\nu+1)}{Q(\nu) z^\nu} N_\nu(z)$ has infinitely many zeros and all of them are positive. Denoting by $\lambda_{\nu,n}$ the nth positive zero of $\Psi_\nu(z)$, under the same conditions the Weierstrassian decomposition*

$$\Psi_\nu(z) = \prod_{n\geq 1} \left(1 - \frac{z^2}{\lambda_{\nu,n}^2}\right)$$

is valid, and this product is uniformly convergent on compact subsets of the complex plane. Moreover, if we denote by $\lambda'_{\nu,n}$ the nth positive zero of $\Phi'_\nu(z)$, where $\Phi_\nu(z) = z^\nu \Psi_\nu(z)$, then the positive zeros of $\Psi_\nu(z)$ are interlaced with those of $\Phi'_\nu(z)$. In the other words, the zeros satisfy the chain of inequalities

$$\lambda'_{\nu,1} < \lambda_{\nu,1} < \lambda'_{\nu,2} < \lambda_{\nu,2} < \lambda'_{\nu,3} < \lambda_{\nu,3} < \cdots.$$

3 Main Results

3.1 Radii of $\gamma-$Spirallike and convex $\gamma-$Spirallike of The Functions f_ν, g_ν and h_ν

The first principal result we established concerns the radii of $\gamma-$Spirallike of order α and reads as follows.

Theorem 1 *Let $\alpha \in [0,1)$ and $\gamma \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. The following statements hold:*

a) If $\nu \geq \max\{0, \nu_0\}$, $\nu \neq 0$ then the radius of $\gamma-$Spirallike of order α of the function f_ν is the smallest positive root of the equation

$$\frac{1}{\nu}\frac{ar^3 J_\nu'''(r) + (2a+b)\, r^2 J_\nu''(r) + (b+c)\, r J_\nu'(r)}{ar^2 J_\nu''(r) + br J_\nu'(r) + c J_\nu(r)} = 1 - (1-\alpha)\cos\gamma.$$

b) If $\nu \geq \max\{0, \nu_0\}$, then the radius of $\gamma-$Spirallike of order α of the function g_ν is the smallest positive root of the equation

$$\frac{ar^3 J_\nu'''(r) + (2a+b)\, r^2 J_\nu''(r) + (b+c)\, r J_\nu'(r)}{ar^2 J_\nu''(r) + br J_\nu'(r) + c J_\nu(r)} = \nu - (1-\alpha)\cos\gamma.$$

c) If $\nu \geq \max\{0, \nu_0\}$, then the radius of $\gamma-$Spirallike of order α of the function h_ν is the smallest positive root of the equation

$$\frac{ar\sqrt{r} J_\nu'''(\sqrt{r}) + (2a+b)\, r J_\nu''(\sqrt{r}) + (b+c)\, \sqrt{r} J_\nu'(\sqrt{r})}{ar J_\nu''(\sqrt{r}) + b\sqrt{r} J_\nu'(\sqrt{r}) + c J_\nu(\sqrt{r})} = \nu - 2\,(1-\alpha)\cos\gamma.$$

Proof Firstly, we prove part **a** for $\nu \geq \max\{0, \nu_0\}$, $\nu \neq 0$ and **b** and **c** for $\nu \geq \max\{0, \nu_0\}$. We need to show that the following inequalities for $\alpha \in [0,1)$ and $\gamma \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$,

$$\operatorname{Re}\left(e^{-i\gamma}\frac{z f_\nu'(z)}{f_\nu(z)}\right) > \alpha\cos\gamma, \quad \operatorname{Re}\left(e^{-i\gamma}\frac{z g_\nu'(z)}{g_\nu(z)}\right) > \alpha\cos\gamma \text{ and } \operatorname{Re}\left(e^{-i\gamma}\frac{z h_\nu'(z)}{h_\nu(z)}\right) > \alpha\cos\gamma \tag{3.1}$$

are valid for $z \in \mathbb{D}_{r_{sp}^*(\alpha,\gamma;f_\nu)}$, $z \in \mathbb{D}_{r_{sp}^*(\alpha,\gamma;g_\nu)}$ and $z \in \mathbb{D}_{r_{sp}^*(\alpha,\gamma;h_\nu)}$ respectively, and each of the above inequalities does not hold in larger disks.
a): When we write the equation (1.3) in definition of the function $f_\nu(z)$, we get by using logarithmic derivation

$$\frac{z f_\nu'(z)}{f_\nu(z)} = \frac{1}{\nu}\frac{z N_\nu'(z)}{N_\nu(z)} = 1 - \frac{1}{\nu}\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2}, \quad (\nu \geq \max\{0,\nu_0\},\ \nu \neq 0), \tag{3.2}$$

It is known [13] that if $z \in \mathbb{C}$ and $\lambda \in \mathbb{R}$ are such that $|z| \leq r < \lambda$, then

$$\operatorname{Re}\left(\frac{z}{\lambda - z}\right) \leq \left|\frac{z}{\lambda - z}\right| \leq \frac{|z|}{\lambda - |z|}. \tag{3.3}$$

Then the inequality

$$\operatorname{Re}\left(\frac{z^2}{\lambda_{\nu,n}^2 - z^2}\right) \leq \left|\frac{z^2}{\lambda_{\nu,n}^2 - z^2}\right| \leq \frac{|z|^2}{\lambda_{\nu,n}^2 - |z|^2}$$

holds for every $|z| < \lambda_{\nu,1}$. Therefore, from (3.2) and (3.3), we have

$$\begin{aligned}\operatorname{Re}\left(e^{-i\gamma}\frac{zf_\nu'(z)}{f_\nu(z)}\right) &= \operatorname{Re}\left(e^{-i\gamma}\right) - \frac{1}{\nu}\operatorname{Re}\left(e^{-i\gamma}\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2}\right)\\ &\geq \cos\gamma - \frac{1}{\nu}\left|e^{-i\gamma}\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2}\right| \geq \cos\gamma - \frac{1}{\nu}\sum_{n\geq 1}\frac{2\,|z|^2}{\lambda_{\nu,n}^2 - |z|^2}\\ &= \frac{|z|\,f_\nu'(|z|)}{f_\nu(|z|)} + \cos\gamma - 1,\end{aligned}$$

with equality when $z = |z| = r$. Thus, for $r \in (0, \lambda_{\nu,1})$ it follows that

$$\inf_{z\in\mathbb{D}_r}\left\{\operatorname{Re}\left(e^{-i\gamma}\frac{zf_\nu'(z)}{f_\nu(z)} - \alpha\cos\gamma\right)\right\} = \frac{|z|\,f_\nu'(|z|)}{f_\nu(|z|)} + (1-\alpha)\cos\gamma - 1.$$

Now, the mapping $\Theta_\nu : (0, \lambda_{\nu,1}) \longrightarrow \mathbb{R}$ defined by

$$\Theta_\nu(r) = \frac{rf_\nu'(r)}{f_\nu(r)} + (1-\alpha)\cos\gamma - 1 = (1-\alpha)\cos\gamma - \frac{1}{\nu}\sum_{n\geq 1}\left(\frac{2r^2}{\lambda_{\nu,n}^2 - r^2}\right).$$

is strictly decreasing since

$$\Theta_\nu'(r) = -\frac{1}{\nu}\sum_{n\geq 1}\left(\frac{4r\lambda_{\nu,n}^2}{\left(\lambda_{\nu,n}^2 - r^2\right)^2}\right) < 0$$

for all $\nu \geq \max\{0, \nu_0\}, \nu \neq 0$. On the other hand, since

$$\lim_{r\searrow 0}\Theta_\nu(r) = (1-\alpha)\cos\gamma > 0 \quad \text{and} \quad \lim_{r\nearrow\lambda_{\nu,1}}\Theta_\nu(r) = -\infty,$$

in view of the minimum principle for harmonic functions imply that the corresponding inequality in (3.1) for $\nu \geq \max\{0, \nu_0\}, \nu \neq 0$ holds if and only if $z \in \mathbb{D}_{r_f}$, where r_f is the smallest positive root of equation

$$\frac{rf_\nu'(r)}{f_\nu(r)} = 1 - (1-\alpha)\cos\gamma$$

which is equivalent to

$$\frac{rN_\nu'(r)}{\nu N_\nu(r)} = 1 - (1-\alpha)\cos\gamma,$$

situated in $(0, \lambda_{\nu,1})$.

b): Since,

$$\frac{zg_\nu'(z)}{g_\nu(z)} = (1-\nu) + \frac{zN_\nu'(z)}{N_\nu(z)} = 1 - \sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2}, \quad (\nu \geq \max\{0, \nu_0\}). \quad (3.4)$$

By using the inequality (3.3), for all $z \in \mathbb{D}_{\lambda_{\nu,1}}$ we get the inequality

$$\begin{aligned}\operatorname{Re}\left(e^{-i\gamma}\frac{zg'_\nu(z)}{g_\nu(z)}\right) &= \operatorname{Re}\left(e^{-i\gamma}\right) - \operatorname{Re}\left(e^{-i\gamma}\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2}\right)\\ &\geq \cos\gamma - \left|e^{-i\gamma}\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2}\right| \geq \cos\gamma - \sum_{n\geq 1}\frac{2\left|z\right|^2}{\lambda_{\nu,n}^2 - \left|z\right|^2}\\ &= \frac{rg'_\nu(r)}{g_\nu(r)} + \cos\gamma - 1,\end{aligned}$$

where $|z| = r$. Thus, for $r \in (0, \lambda_{\nu,1})$ we obtain

$$\inf_{z\in\mathbb{D}_r}\left\{\operatorname{Re}\left(e^{-i\gamma}\frac{zg'_\nu(z)}{g_\nu(z)} - \alpha\cos\gamma\right)\right\} = \frac{rg'_\nu(r)}{g_\nu(r)} + (1-\alpha)\cos\gamma - 1.$$

The function $\Phi_\nu : (0, \lambda_{\nu,1}) \longrightarrow \mathbb{R}$ defined by

$$\Phi_\nu(r) = \frac{rg'_\nu(r)}{g_\nu(r)} + (1-\alpha)\cos\gamma - 1 = (1-\alpha)\cos\gamma - \sum_{n\geq 1}\left(\frac{2r^2}{\lambda_{\nu,n}^2 - r^2}\right).$$

is strictly decreasing and $\lim_{r\searrow 0}\Phi_\nu(r) = (1-\alpha)\cos\gamma > 0$ and $\lim_{r\nearrow\lambda_{\nu,1}}\Phi_\nu(r) = -\infty$.

Consequently, in view of the minimum principle for harmonic functions for $z \in \mathbb{D}_{r_g}$, we have that

$$\operatorname{Re}\left(e^{-i\gamma}\frac{zg'_\nu(z)}{g_\nu(z)}\right) > \alpha\cos\gamma$$

if and only if r_g is the smallest positive root of equation

$$\frac{rg'_\nu(r)}{g_\nu(r)} = 1 - (1-\alpha)\cos\gamma \quad \text{or} \quad \frac{rN'_\nu(r)}{N_\nu(r)} = \nu - (1-\alpha)\cos\gamma$$

situated in $(0, \lambda_{\nu,1})$.

c): Observe that

$$\frac{zh'_\nu(z)}{h_\nu(z)} = (1 - \frac{\nu}{2}) + \frac{1}{2}\frac{\sqrt{z}N'_\nu(\sqrt{z})}{N_\nu(\sqrt{z})} = 1 - \sum_{n\geq 1}\frac{z}{\lambda_{\nu,n}^2 - z}, \quad (\nu \geq \max\{0, \nu_0\}), \tag{3.5}$$

By using the inequality (3.2), for all $z \in \mathbb{D}_{\lambda_{\nu,1}^2}$ we obtain

$$\begin{aligned}\operatorname{Re}\left(e^{-i\gamma}\frac{zh'_\nu(z)}{h_\nu(z)}\right) &= \operatorname{Re}\left(e^{-i\gamma}\right) - \operatorname{Re}\left(e^{-i\gamma}\sum_{n\geq 1}\frac{z}{\lambda_{\nu,n}^2 - z}\right)\\ &\geq \cos\gamma - \left|e^{-i\gamma}\sum_{n\geq 1}\frac{z}{\lambda_{\nu,n}^2 - z}\right| \geq \cos\gamma - \sum_{n\geq 1}\frac{2\left|z\right|}{\lambda_{\nu,n}^2 - \left|z\right|}\\ &= \frac{rh'_\nu(r)}{h_\nu(r)} + \cos\gamma - 1,\end{aligned}$$

Equality holds for $z = |z| = r$. As a result, for $r \in (0, \lambda_{\nu,1})$ it follows that

$$\inf_{z\in\mathbb{D}_r}\left\{\mathrm{Re}\left(e^{-i\gamma}\frac{zh'_\nu(z)}{h_\nu(z)} - \alpha\cos\gamma\right)\right\} = \frac{rh'_\nu(r)}{h_\nu(r)} + (1-\alpha)\cos\gamma - 1.$$

The fucntion $\Psi_\nu : \left(0, \lambda^2_{\nu,1}\right) \longrightarrow \mathbb{R}$ defined by

$$\Psi_\nu(r) = \frac{rh'_\nu(r)}{h_\nu(r)} + (1-\alpha)\cos\gamma - 1 = (1-\alpha)\cos\gamma - \sum_{n\geq 1}\left(\frac{2r}{\lambda^2_{\nu,n} - r}\right).$$

is strictly decreasing, since

$$\Psi'_\nu(r) = -\sum_{n\geq 1}\left(\frac{\lambda^2_{\nu,n}}{\left(\lambda^2_{\nu,n} - r\right)^2}\right) < 0$$

for all $\nu \geq \max\{0, \nu_0\}$. On the other hand, since

$$\lim_{r\searrow 0}\Psi_\nu(r) = (1-\alpha)\cos\gamma > 0 \quad \text{and} \quad \lim_{r\nearrow\lambda^2_{\nu,1}}\Psi_\nu(r) = -\infty$$

in view of the minimum principle for harmonic functions imply that the corresponding inequality in (3.1) holds if and only if $z \in \mathbb{D}_{r_h}$, where r_h is the smallest positive root of equation

$$\frac{rh'_\nu(r)}{h_\nu(r)} = 1 - (1-\alpha)\cos\gamma$$

which is equivalent to

$$\frac{\sqrt{r}N'_\nu(\sqrt{r})}{N_\nu(\sqrt{r})} = 1 - (1-\alpha)\cos\gamma$$

and situated in $\left(0, \lambda^2_{\nu,1}\right)$. This completes the proof of part **a** when $\nu \geq \max\{0, \nu_0\}, \nu \neq 0$, and parts **b** and **c** when $\nu \geq \max\{0, \nu_0\}$. □

Remark 1 In Theorem 1, the choice of $\gamma = 0$ yields the result for the class of starlike functions of order α.

Example 1 Indeed, we can write the function $N_\nu(z)$ in terms of elementary trigonometric functions for the value $\nu = 1/2$ as follows:

$$N_{1/2}(z) = \frac{4(b-a)z\cos z + \left[a\left(3 - 4z^2\right) - 2b + 4c\right]\sin z}{2\sqrt{2\pi}\sqrt{z}}.$$

Thus, we have

$$f_{1/2}(z) = \frac{\left[4(a-b)z\cos z + \left(4az^2 - 3a + 2b - 4c\right)\sin z\right]^2}{(a - 2b - 4c)^2 z},$$

$$g_{1/2}(z) = \frac{4(a-b)z\cos z + \left(4az^2 - 3a + 2b - 4c\right)\sin z}{a - 2b - 4c}$$

and

$$h_{1/2}(z) = \frac{4(a-b)z\cos\sqrt{z} + (4az - 3a + 2b - 4c)\sqrt{z}\sin\sqrt{z}}{a - 2b - 4c}.$$

	$b=3$ and $c=0$			$a=1$ and $c=0$			$a=1$ and $b=2$		
	$a=2$	$a=3$	$a=4$	$b=2$	$b=3$	$b=4$	$c=2$	$c=3$	$c=4$
$R\left(f_{\frac{1}{2}}\right)$	0.1539	0.1190	0.0886	0.1746	0.1990	0.2130	0.3038	0.3401	0.3680
$R\left(g_{\frac{1}{2}}\right)$	0.2121	0.1639	0.1220	0.2409	0.2747	0.2942	0.4217	0.4730	0.5126
$R\left(g_{\frac{1}{2}}\right)$	0.0817	0.0486	0.0268	0.1056	0.1377	0.1582	0.3309	0.4193	0.4953

Table 1 $R(--)$: Radii of $\frac{\pi}{3}-$Spirallike of order $1/2$ for $f_{1/2}$, $g_{1/2}$ and $h_{1/2}$.

Taking $\nu = 1/2$, $a = 1$, $b = 2$, $c = 0$, $\alpha = 1/2$ and $\gamma = \pi/3$ in Theorem 1, we have following results.

Corollary 1 *The following statements are true.*

a) *The radius of $\frac{\pi}{3}-$Spirallike of order $1/2$ of the function $f_{1/2}$ is the smallest positive root of the equation*

$$\frac{4x\left(8x+1\right)\cos x+\left(68x^2-7\right)\sin x}{4\left(-4x\cos x+\left(4x^2+1\right)\sin x\right)}=0$$

b) *The radius of $\frac{\pi}{3}-$Spirallike of order $1/2$ of the function $g_{1/2}$ is the smallest positive root of the equation*

$$\frac{16x^3\cos x+3\left(12x^2-1\right)\sin x}{4\left(-4x\cos x+\left(4x^2+1\right)\sin x\right)}=0$$

c) *The radius of $\frac{\pi}{3}-$Spirallike of order $1/2$ of the function $h_{1/2}$ is the smallest positive root of the equation*

$$\frac{2\sqrt{x}\left(4x-1\right)\cos\sqrt{x}+\left(20x-1\right)\sin\sqrt{x}}{2\left(-4\sqrt{x}\cos\sqrt{x}+\left(4x+1\right)\sin\sqrt{x}\right)}=0.$$

The figure 1 was created by taking the special values in the Corollary 1.

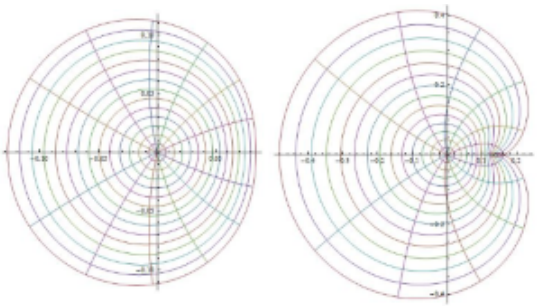


Fig. 1 Images of function $h_{1/2}(z)$ for $r = 0.1056$ and $r = 0.3$, respectively as per Corollary 1.

Remark 2 When $\nu = 0.5$, considering the special values of a, b and c real numbers, radii of $\gamma-$Spirallike of order α of the function f_ν is seen from the table above. In addition, according to the increasing values of a, b and c reel numbers, it is clear that radii of $\gamma-$Spirallike of order α of the function f_ν is monotone.

In the following, we deal with convex analogue of the class of γ-spiralllike functions of order α.

Theorem 2 *Let $\alpha \in [0,1)$ and $\gamma \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. The following statements hold:*

a) If $\nu \geq \max\{0, \nu_0\}$, $\nu \neq 0$ then the radius of the convex $\gamma-$Spirallike of order α of the function f_ν is the smallest positive root of the equation

$$\frac{rN_\nu''(r)}{N_\nu'(r)} + \left(\frac{1}{\nu} - 1\right)\frac{rN_\nu'(r)}{N_\nu(r)} = (\alpha - 1)\cos\gamma.$$

b) If $\nu \geq \max\{0, \nu_0\}$, then the radius of the convex $\gamma-$Spirallike of order α of the function g_ν is the smallest positive root of the equation

$$\frac{r^2N_\nu''(r) + (2-2\nu)\, rN_\nu'(r) + \left(\nu^2 - \nu\right)N_\nu(r)}{rN_\nu'(r) + (1-\nu)\, N_\nu(r)} = (\alpha - 1)\cos\gamma.$$

c) If $\nu \geq \max\{0, \nu_0\}$, then the radius of the convex $\gamma-$Spirallike of order α of the function h_ν is the smallest positive root of the equation

$$\frac{rN_\nu''(\sqrt{r}) + (3-2\nu)\sqrt{r}N_\nu'(\sqrt{r}) + \left(\nu^2 - 2\nu\right)N_\nu(\sqrt{r})}{2\sqrt{r}N_\nu'(\sqrt{r}) + 2\,(2-\nu)\, N_\nu(\sqrt{r})} = (\alpha - 1)\cos\gamma.$$

Proof We first prove the part (a). From (1.3), we have

$$1 + \frac{zf_\nu''(z)}{f_\nu'(z)} = 1 + \frac{zN_\nu''(z)}{N_\nu'(z)} + \left(\frac{1}{\nu} - 1\right)\frac{zN_\nu'(z)}{N_\nu(z)}$$

and by means of (1.6) and (1.7), we obtain

$$1 + \frac{zf_\nu''(z)}{f_\nu'(z)} = 1 - \left(\frac{1}{\nu} - 1\right)\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2} - \sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}'^2 - z^2}.$$

For $1 \geq \nu > \max\{0, \nu_0\}$ by using (3.3), we get

$$\begin{aligned}
&\mathrm{Re}\left(e^{-i\gamma}\left(1 + \frac{zf_\nu''(z)}{f_\nu'(z)}\right)\right) \\
&= \mathrm{Re}(e^{-i\gamma}) - \mathrm{Re}\left(e^{-i\gamma}\left(\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}'^2 - z^2} + \left(\frac{1}{\nu} - 1\right)\sum_{n\geq 1}\frac{2z^2}{\lambda_{\nu,n}^2 - z^2}\right)\right) \\
&\geq \cos\gamma - \left(\frac{1}{\nu} - 1\right)\sum_{n\geq 1}\frac{2r^2}{\lambda_{\nu,n}^2 - r^2} - \sum_{n\geq 1}\frac{2r^2}{\lambda_{\nu,n}'^2 - r^2} \\
&\geq \cos\gamma + \frac{rf_\nu''(r)}{f_\nu'(r)} \qquad (3.6)
\end{aligned}$$

	$b=3$ and $c=0$			$a=1$ and $c=0$			$a=1$ and $b=2$		
	$a=2$	$a=3$	$a=4$	$b=2$	$b=3$	$b=4$	$c=2$	$c=3$	$c=4$
$R^c\left(f_{\frac{1}{2}}\right)$	0.0875	0.0677	0.0504	0.0993	0.1131	0.1210	0.1722	0.1925	0.2082
$R^c\left(g_{\frac{1}{2}}\right)$	0.1221	0.0944	0.0703	0.1386	0.1579	0.1691	0.2411	0.2699	0.2921
$R^c\left(h_{\frac{1}{2}}\right)$	0.0406	0.0242	0.0134	0.0524	0.0682	0.0783	0.1622	0.2046	0.2408

Table 2 Radii of the convex $\frac{\pi}{3}-$Spirallike of order $1/2$ for $f_{1/2}$, $g_{1/2}$ and $h_{1/2}$

where $|z|=r$. Moreover, observe that if we use the inequality [4, Lemma 2.1]

$$\mu\,\mathrm{Re}\left(\frac{z}{a-z}\right)-\mathrm{Re}\left(\frac{z}{b-z}\right)\geq\mu\frac{|z|}{a-|z|}-\frac{|z|}{b-|z|}$$

where $a>b>0$, $\mu\in[0,1]$ and $z\in\mathbb{C}$ such that $|z|<b$, then we get that the inequality (3.6) is also valid when $\nu\geq 1$. Here we used that the zeros of N_ν and N_ν' are interlacing according to Lemma 2. The above inequality implies for $r\in(0,\lambda'_{\nu,1})$

$$\inf_{z\in\mathbb{D}_r}\left\{\mathrm{Re}\left(e^{-i\gamma}\left(1+\frac{zf_\nu''(z)}{f_\nu'(z)}\right)-\alpha\cos\gamma\right)\right\}=(1-\alpha)\cos\gamma+\frac{rf_\nu''(r)}{f_\nu'(r)}.$$

Now, the proof of part (a) follows on similar lines as of Theorem 1.

For the other parts, note that the functions g_ν and h_ν belong to the Laguerre-Pólya class $\mathcal{LP}$, which is closed under differentiation, their derivatives g_ν' and h_ν' also belong to $\mathcal{LP}$ and the zeros are real. Thus assuming $\delta_{\nu,n}$ and $\gamma_{\nu,n}$ are the positive zeros of g_ν' and h_ν', respectively, we have the following representations:

$$g_\nu'(z)=\prod_{n\geq1}\left(1-\frac{z^2}{\delta_{\nu,n}^2}\right)\quad\text{and}\quad h_\nu'(z)=\prod_{n\geq1}\left(1-\frac{z}{\gamma_{\nu,n}}\right),$$

which yield

$$1+\frac{zg_\nu''(z)}{g_\nu'(z)}=1-\sum_{n\geq1}\frac{2z^2}{\delta_{\nu,n}^2-z^2}\quad\text{and}\quad 1+\frac{zh_\nu''(z)}{h_\nu'(z)}=1-\sum_{n\geq1}\frac{z}{\gamma_{\nu,n}-z}.$$

Further, reasoning along the same lines as in Theorem 1, the result follows at once. □

Remark 3 In Theorem 2, the choice of $\gamma=0$ yields the result for the class of convex functions of order α.

3.2 Radii of Ma-Minda Starlikeness and Convexity of the Functions f_ν, g_ν and h_ν

In the following, now we establish the result for the complete family of Ma-Minda class of starlike functions.

Theorem 3 *Let $\nu \geq \max\{0, \nu_0\}$. The $\mathcal{S}^*(\varphi)$-radii of the functions f_ν, g_ν and h_ν are respectively, given by the smallest positive root of the equations*

a) $rN'_\nu(r) - (1-\beta)\nu N_\nu(r) = 0$ *with* $\nu \neq 0$;
b) $rN'_\nu(r) - (\nu-\beta)N_\nu(r) = 0$;
c) $\sqrt{r}N'_\nu(\sqrt{r}) - (\nu-2\beta)N_\nu(\sqrt{r}) = 0$

in $|z| < (0, \lambda_{\nu,1})$, $(0, \lambda_{\nu,1})$ *and* $(0, \lambda^2_{\nu,1})$, *where* $\varphi(-1) = 1-\beta$ *and* β *is the radius of the largest disk* $\{w : |w-1| < \beta\} \subseteq \varphi(\mathbb{D})$.

Proof We first prove part *a)* : from equation (3.2), we have

$$\frac{zf'_\nu(z)}{f_\nu(z)} = \frac{1}{\nu}\frac{zN'_\nu(z)}{N_\nu(z)} = 1 - \frac{1}{\nu}\sum_{n\geq 1}\frac{2z^2}{\lambda^2_{\nu,n} - z^2}, \quad (\nu \geq \max\{0,\nu_0\},\ \nu \neq 0)\,. \tag{3.7}$$

Let us consider the continuous function $T_{f_\nu} : (0, \lambda_{\nu,1}) \to \mathbb{R}$

$$T_{f_\nu}(r) = \frac{1}{\nu}\sum_{n\geq 1}\frac{2r^2}{\lambda^2_{\nu,n} - r^2} - \beta. \tag{3.8}$$

Then

$$T'_{f_\nu}(r) = \frac{1}{\nu}\sum_{n\geq 1}\left(\frac{4r\lambda^2_{\nu,n}}{(\lambda^2_{\nu,n} - r^2)^2}\right) > 0$$

for all $\nu \geq \max\{0,\nu_0\} \neq 0$ and for $r < \lambda_{\nu,1}$. Also, $T_{f_\nu}(0) = -\beta < 0$ and $\lim_{r\nearrow\lambda_{\nu,1}} T_{f_\nu}(r) = \infty$. This implies that there exist a unique positive root, say $R_\varphi(f_\nu)$ of the equation $T_{f_\nu}(r) = 0$ in $(0, \lambda_{\nu,1})$. Now let $\{w : |w-1| < \beta\} \subseteq \varphi(\mathbb{D})$ such that $\varphi(-1) = 1-\beta$. Therefore, using the Lemma [13] that if $z \in \mathbb{C}$ and $\lambda \in \mathbb{R}$ are such that $|z| \leq r < \lambda$, then

$$\mathrm{Re}\left(\frac{z}{\lambda - z}\right) \leq \left|\frac{z}{\lambda - z}\right| \leq \frac{|z|}{\lambda - |z|},$$

in (3.7), we get that

$$\left|\frac{zf'_\nu(z)}{f_\nu(z)} - 1\right| = \left|\frac{1}{\nu}\sum_{n\geq 1}\frac{2z^2}{\lambda^2_{\nu,n} - z^2}\right| \leq \frac{1}{\nu}\sum_{n\geq 1}\frac{2r^2}{\lambda^2_{\nu,n} - r^2} \leq \beta \tag{3.9}$$

implies that $f_\nu \in \mathcal{S}^*(\varphi)$ in $|z| < R_\varphi(f_\nu)$. Further, taking $z \in \mathbb{D}$ such that $z = r = -R_\varphi(f_\nu)$, from (3.8) and (3.9), it follows that

$$\left|\frac{zf'_\nu(z)}{f_\nu(z)} - 1\right| = \beta,$$

which implies that for $zf'_\nu(z)/f_\nu(z) \notin \varphi(\mathbb{D})$ for all $|z| = r \geq R_\varphi(f_\nu)$, which proves the sharpness part. Reasoning along the same lines, proofs of the parts *b)* and *c)* follow. □

	$b=3$ and $c=0$			$a=1$ and $c=0$			$a=1$ and $b=2$		
	$a=2$	$a=3$	$a=4$	$b=2$	$b=3$	$b=4$	$c=2$	$c=3$	$c=4$
$r^*_{e^z}\left(f_{\frac{1}{2}}\right)$	0.2354	0.1818	0.1352	0.2674	0.3050	0.3268	0.4696	0.5272	0.5718
$r^*_{e^z}\left(g_{\frac{1}{2}}\right)$	0.3140	0.2421	0.1798	0.3571	0.4078	0.4372	0.6350	0.7160	0.7792
$r^*_{e^z}\left(h_{\frac{1}{2}}\right)$	0.1611	0.0953	0.0523	0.2090	0.2736	0.3150	0.6845	0.8819	1.0559

Table 3 Radii of $\mathcal{S}^*(e^z)$ for the functions $f_{1/2}$, $g_{1/2}$ and $h_{1/2}$

The figure 2 was created by taking $a=1,\ b=2,\ c=0$ and $\varphi(z)=e^z$ in the Theorem 3.

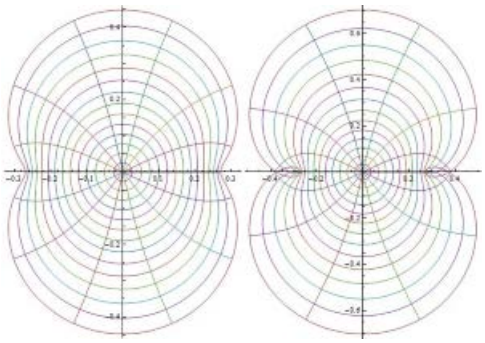


Fig. 2 Images of function $g_{1/2}(z)$ for $r=0.3571$ and $r=0.48$, respectively as per Theorem 3.

Remark 4 Theorem 3 with $\varphi(z)=(1+(1-2\beta)z)/(1-z)$ coincide with Theorem 1 with $\alpha=\beta$ and $\gamma=0$. In particular, choosing $\varphi(z)=(1+z)/(1-z)$ and $\beta=1$ in Theorem 3 reduces to Theorem 1 with $\alpha=0$ and $\gamma=0$.

As a consequence of Theorem 3, we get the sufficient conditions for the functions to belong in the class $\mathcal{S}^*(\varphi)$.

Corollary 2 *Let $\nu\geq\max\{0,\nu_0\}$. Further, $\varphi(-1)=1-\beta$ and β is the radius of the largest disk $\{w:|w-1|<\beta\}\subseteq\varphi(\mathbb{D})$. Then*

(i) if $1-\frac{1}{\nu}\frac{N'_\nu(1)}{N_\nu(1)}<\beta$ and $\nu\neq 0$, then the function f_ν belongs to $\mathcal{S}^*(\varphi)$.

(ii) if $\nu-\frac{N'_\nu(1)}{N_\nu(1)}<\beta$, then the function g_ν belongs to $\mathcal{S}^*(\varphi)$.

(iii) if $\frac{\nu}{2}-\frac{N'_\nu(1)}{2N_\nu(1)}<\beta$, then the function h_ν belongs to $\mathcal{S}^*(\varphi)$.

In the following, we prove the result for the Ma-Minda class of convex functions.

Theorem 4 *The $\mathcal{C}(\varphi)$-radii of the functions f_ν, g_ν and h_ν are respectively, given by the smallest positive root of the equations*

a) $rf''_\nu(r)+\beta\nu f'_\nu(r)=0$, *where* $\nu\geq\max\{0,\nu_0\}$, $\nu\neq 0$,
b) $rg''_\nu(r)+\beta g'_\nu(r)=0$, *where* $\nu\geq\max\{0,\nu_0\}$,
c) $\sqrt{r}h''_\nu(\sqrt{r})+\beta h'_\nu(\sqrt{r})=0$, *where* $\nu\geq\max\{0,\nu_0\}$

in $|z|<(0,\lambda_{\nu,1})$, $(0,\lambda_{\nu,1})$ *and* $(0,\lambda^2_{\nu,1})$, *where* $\varphi(-1)=1-\beta$ *and* β *is the radius of the largest disk* $\{w:|w-1|<\beta\}\subseteq\varphi(\mathbb{D})$.

Proof From (3.7), it follows that

$$\frac{zf''_\nu(z)}{f'_\nu(z)}=\frac{zN''_\nu(z)}{N'_\nu(z)}+\left(\frac{1}{\nu}-1\right)\frac{zN'_\nu(z)}{N_\nu(z)}. \tag{3.10}$$

Now using the Weierstrass decompositions of the functions N_ν and N'_ν, which are given below

$$N_\nu(z)=\frac{Q(\nu)z^\nu}{2^\nu\Gamma(\nu+1)}\prod_{n\geq 1}\left(1-\frac{z^2}{\lambda^2_{\nu,n}}\right)$$

and

$$N'_\nu(z)=\frac{Q(\nu)\nu z^{\nu-1}}{2^\nu\Gamma(\nu+1)}\prod_{n\geq 1}\left(1-\frac{z^2}{\tilde{\lambda}^2_{\nu,n}}\right),$$

where $\lambda_{\nu,n}$ and $\tilde{\lambda}_{\nu,n}$ are the n^{th} positive roots of N_ν and N'_ν, we have from (3.10) that

$$1+\frac{zf''_\nu(z)}{f'_\nu(z)}=1-\left(\frac{1}{\nu}-1\right)\sum_{n\geq 1}\frac{2z^2}{\lambda^2_{\nu,n}-z^2}-\sum_{n\geq 1}\frac{2z^2}{\tilde{\lambda}^2_{\nu,n}-z^2}. \tag{3.11}$$

Using the Lemma [13] in (3.11) that if $z\in\mathbb{C}$ and $\lambda\in\mathbb{R}$ are such that $|z|\leq r<\lambda$, then

$$\mathrm{Re}\left(\frac{z}{\lambda-z}\right)\leq\left|\frac{z}{\lambda-z}\right|\leq\frac{|z|}{\lambda-|z|},$$

it follows that the inequalities

$$\left|\frac{zf''_\nu(z)}{f'_\nu(z)}\right|\leq\left(\frac{1}{\nu}-1\right)\sum\frac{2r^2}{\lambda^2_{\nu,n}-r^2}+\sum_{n\geq 1}\frac{2r^2}{\tilde{\lambda}^2_{\nu,n}-r^2}=-\frac{rf''_\nu(r)}{f'_\nu(r)}$$

hold in $|z|=r<\tilde{\lambda}_{\nu,1}$ for $\nu\leq 1$. Observe that the inequality $|zf''_\nu(z)/f'_\nu(z)|\leq -rf''_\nu(r)/f'_\nu(r)$ also holds for $\nu>1$ in $|z|=r<\tilde{\lambda}_{\nu,1}$. Let $R^c_\varphi(f_\nu)$ be the smallest positive root of the equation

$$\frac{rf''_\nu(r)}{f'_\nu(r)}+\beta=0.$$

Now let $\{w:|w-1|<\beta\}\subseteq\varphi(\mathbb{D})$ such that $\varphi(-1)=1-\beta$. Then it follows that

$$\left|\frac{zf''_\nu(z)}{f'_\nu(z)}\right|\leq-\frac{rf''_\nu(r)}{f'_\nu(r)}\leq\beta$$

	$b=3$ and $c=0$			$a=1$ and $c=0$			$a=1$ and $b=2$		
	$a=2$	$a=3$	$a=4$	$b=2$	$b=3$	$b=4$	$c=2$	$c=3$	$c=4$
$r_\varphi^c(f_{\frac{1}{2}})$	0.1271	0.0983	0.0732	0.1443	0.1644	0.1761	0.2512	0.2812	0.3044
$r_\varphi^c(g_{\frac{1}{2}})$	0.1749	0.1352	0.1005	0.1987	0.2266	0.2427	0.3482	0.3906	0.4234
$r_\varphi^c(h_{\frac{1}{2}})$	0.0759	0.0451	0.0248	0.0983	0.1283	0.1474	0.3124	0.3979	0.4719

Table 4 Radii of $\mathcal{C}(\varphi)$ for the functions $f_{1/2}$, $g_{1/2}$ and $h_{1/2}$ with $\varphi(z)=z+\sqrt{1+z^2}$

hold in $|z|=r<R_\varphi^c(f_{f_\nu})$ for $\nu\geq\max\{0,\nu_0\}$, $\nu\neq 0$, which implies that the function $f_\nu\in\mathcal{C}(\varphi)$ in $|z|=r<R_\varphi^c(f_{f_\nu})$ for $\nu\geq\max\{0,\nu_0\}$, $\nu\neq 0$. For the sharpness, let us consider $z=-R_\varphi^c(f_\nu)$. Then

$$\left|\frac{zf_\nu''(z)}{f_\nu'(z)}\right|=-\frac{rf_\nu''(r)}{f_\nu'(r)}=\beta$$

such that $1+zf_\nu''(z)/f_\nu'(z)\notin\varphi(\mathbb{D})$ for all $|z|=r\geq R_\varphi^c(f_\nu)$, which proves the sharpness part. Reasoning along the same lines, proofs of the parts b) and c) follow. □

The following corollary yields sufficient conditions on parameters for functions to be in $\mathcal{C}(\varphi)$.

Corollary 3 *Let $\nu\geq\max\{0,\nu_0\}$, $\nu\neq 0$. Let $\varphi(-1)=1-\beta$ and β is the radius of the largest disk $\{w:|w-1|<\beta\}\subseteq\varphi(\mathbb{D})$. Then*

(i) The function f_ν belongs to $\mathcal{C}(\varphi)$, if

$$\left(1-\frac{1}{\nu}\right)\frac{N_\nu'(1)}{N_\nu(1)}-\frac{N_\nu''(1)}{N_\nu'(1)}<\beta\quad\text{and}\quad 1\geq\nu\geq\max\{0,\nu_0\},\ \nu\neq 0\quad\text{and}\quad\nu>1.$$

(ii) The function g_ν belongs to $\mathcal{C}(\varphi)$, if

$$\frac{N_\nu''(1)+2(1-\nu)N_\nu'(1)+(\nu^2+\nu)N_\nu(1)}{(1-\nu)N_\nu(1)-N_\nu'(1)}<\beta$$

(iii) The function h_ν belongs to $\mathcal{C}(\varphi)$, if

$$\frac{1}{2}\frac{N_\nu''(1)+(3-2\nu)N_\nu'(1)+(\nu^2-2\nu)N_\nu(1)}{(\nu-2)N_\nu(1)-N_\nu'(1)}<\beta.$$

The figure 3 was created by taking $\nu=1/2$, $a=2$, $b=3$, $c=0$ and $\varphi(z)=z+\sqrt{1+z^2}$ in Theorem 4.

Remark 5 Let β be the radius of the lagest disk $\{w:|w-1|<\beta\}$ inside $\varphi(\mathbb{D})$. Then Theorem 3 and Theorem 4 hold for each of the following cases and the radii are sharp, where

(i) $\beta=\min\left\{\left|1-\frac{1+A}{1+B}\right|,\left|1-\frac{1-A}{1-B}\right|\right\}=\frac{A-B}{1+|B|}$ when $\varphi(z)=\frac{1+Az}{1+Bz}$, where $-1\leq B<A\leq 1$;

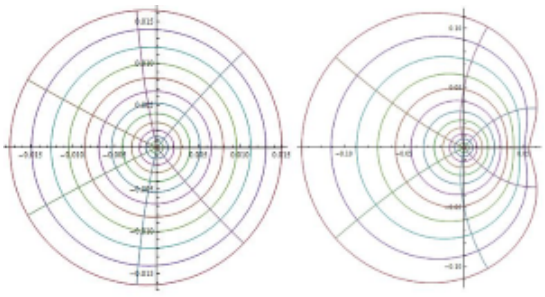


Fig. 3 Images of function $h_{1/2}(z)$ for $r = 0.1271$ and $r = 0.3$, respectively as per Theorem 4 and Remark 5.

(*ii*) $\beta = \sqrt{2 - 2\sqrt{2} + \sqrt{-2 + 2\sqrt{2}}}$ when $\varphi(z) = \sqrt{2} - (\sqrt{2} - 1)\sqrt{\frac{1-z}{1+2(\sqrt{2}-1)z}}$;
(*iii*) $\beta = \sqrt{2} - 1$ when $\varphi(z) = \sqrt{1+z}$;
(*iv*) $\beta = 1 - 1/e$ when $\varphi(z) = e^z$;
(*v*) $\beta = 2 - \sqrt{2}$ when $\varphi(z) = z + \sqrt{1+z^2}$;
(*vi*) $\beta = \frac{e-1}{e+1}$ when $\varphi(z) = \frac{2}{1+e^{-z}}$;
(*vii*) $\beta = \sin 1$ when $\varphi(z) = 1 + \sin z$;
(*viii*) $\beta = 1 - e^{e^{-1}-1}$ when $\varphi(z) = e^{e^z - 1}$;
(*ix*) for the domains bounded by the conic sections $\Omega_\kappa := \{w = u + iv : u^2 > \kappa^2(u-1)^2 + \kappa^2 v^2; \kappa \in [0, \infty)\}$, we have

$$\beta = \frac{1}{\kappa + 1},$$

where the boundary curve of Ω_κ for fixed κ is represented by the imaginary axis ($\kappa = 0$), the right branch of a hyperbola ($0 < \kappa < 1$), a parabola ($\kappa = 1$) and an ellipse ($\kappa > 1$). The univalent Carathéodory functions mapping $\mathbb{D}$ onto Ω_κ is given by

$$\varphi(z) := \varphi_\kappa(z) = \begin{cases} \frac{1+z}{1-z} & \text{for } \kappa = 0; \\ 1 + \frac{2}{1-\kappa^2}\sinh^2(A(\kappa)arctanh\sqrt{z}) & \text{for } \kappa \in (0,1); \\ 1 + \frac{2}{\pi^2}\log^2\frac{1+\sqrt{z}}{1-\sqrt{z}} & \text{for } \kappa = 1; \\ 1 + \frac{2}{\kappa^2-1}\sin^2\left(\frac{\pi}{2K(t)}F\left(\frac{\sqrt{z}}{\sqrt{t}}, t\right)\right) & \text{for } \kappa > 1, \end{cases}$$

where $A(\kappa) = (2/\pi)\arccos(\kappa)$, $F(w,t) = \int_0^w \frac{dx}{\sqrt{(1-x^2)(1-t^2x^2)}}$ is the Legender elliptic integral of the first kind, $K(t) = F(1,t)$ and $t \in (0,1)$ is choosen such that $\kappa = \cosh(\pi K'(t)/2K(t))$.

Statements and Declarations

– **Funding**: The work of Kamajeet Gangania is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No.:1051/(CSIR-UGC NET JUNE 2017).

– **Conflict of interest**: The authors declare that they have no conflict of interest
– **Availability of data and materials** : NA
– **Code availability**: NA
– **Authors' contributions** : All authors contributed Equally.

References

1. Aktaş, İ. Baricz, Á. and Orhan, H.: Bounds for radii of starlikeness and convexity of some special functions. Turkish J. Math. **42**, 211–226 (2018) doi: 10.3906/mat-1610-41.
2. Baricz, Á. Dimitrov, D.K., Orhan, H. and Yağmur, N.: Radii of starlikeness of some special functions. Proc. Amer. Math. Soc. **144**, 3355–3367 (2016). doi: 10.1090/proc/13120.
3. Á. Baricz, P. A. Kupán, R. Szász, The radius of starlikeness of normalized Bessel functions of the first kind. Proc. Amer. Math. Soc. 142(5) (2014), 2019-2025.
4. Á. Baricz, R. Szász, The radius of convexity of normalized Bessel functions of the first kind. Anal. Appl. 12(5) (2014), 485-509.
5. Á. Baricz, R. Szász, The radius of convexity of normalized Bessel functions. Anal. Math. 41(3) (2015), 141-151.
6. Á. Baricz, M. Çağlar, E. Deniz, Starlikeness of bessel functions and their derivatives. Math. Inequal. Appl. 19(2) (2016), 439-449.
7. Á. Baricz, H. Orhan, R. Szász, The radius of $\alpha-$convexity of normalized Bessel functions of the first kind. Comput. Method. Func. Theo. 16(1) (2016), 93-103.
8. Baricz, Á. and Prajapati, A.: Radii of starlikeness and convexity of generalized Mittag-Leffler functions. Math. Commun. **25**, 117–135 (2020).
9. Baricz, Á. Toklu, E. and Kadioğlu, E.: Radii of starlikeness and convexity of Wright functions. Math. Commun. **23**, 97–117 (2018).
10. Baricz, Á. and Yağmur, N.: Geometric properties of some Lommel and Struve functions. Ramanujan J. **42**, 325–346 (2017). doi: 10.1007/s11139-015-9724-6
11. R.K. Brown, Univalence of Bessel functions, Proc. Amer. Math. Soc. 11(2) (1960) 278–283.
12. M. Çağlar, E. Deniz, R. Szász, Radii of $\alpha-$convexity of some normalized Bessel functions of the first kind. Results Math. 72 (2017), 2023-2035.
13. E. Deniz, R. Szász, The radius of uniform convexity of Bessel functions. J. Math. Anal. Appl. 453(1) (2017) 572–588.
14. E. Deniz, S. Kazımoğlu, M. Çağlar, Radii of Starlikeness and Convexity of Bessel Function Derivatives. Ukr. math. J. (In Press).
15. Deniz, E.: Geometric and monotonic properties of Ramanujan type entire functions. Ramanujan J. **55**, 103–130 (2020). doi: 10.1007/s11139-020-00267-w.
16. Gangania, K. and Kumar, S.S.: $\mathcal{S}^*(\psi)$ and $\mathcal{C}(\psi)$-radii for some special functions. Iran. J. Sci. Technol. Trans. A Sci. **46**, 955–966 (2022).
17. M.E.H. Ismail, M.E. Muldoon, Bounds for the small real and purely imaginary zeros of Bessel and related functions, Meth. Appl. Anal. 2(1) (1995) 1–21.
18. J. L.W.V. Jensen, Recherches sur la théorie des équations. Acta Math. 36 (1913) 181–195.
19. S. Kazımoğlu, E. Deniz, Radius Problems for Functions Containing Derivatives of Bessel Functions. Computational Methods and Function Theory. (2022) 1-26.
20. E. Kreyszig, J. Todd, The radius of univalence of Bessel functions, Illinois J. Math. 4 (1960) 143–149.
21. Kumar, S.S. and Gangania, K.: Subordination and radius problems for certain starlike functions. arXiv:2007.07816 (2020).
22. B. Ya. Levin, Lectures on Entire Functions. Amer. Math. Soc.: Transl. Math. Monographs 150 (1996).
23. V. Madaan, A. Kumar and V. Ravichandran, Radii of Starlikeness and Convexity of Some Entire Functions, Bull. Malays. Math. Sci. Soc. 43 (2020) 4335-4359.

24. Ma, W.C. and Minda, D.: A unified treatment of some special classes of univalent functions. Proceedings of the Conference on Complex Analysis, Tianjin, Conf Proc Lecture Notes Anal, I Int Press. Cambridge, MA. 157–169 (1992).
25. A. McD. Mercer, The zeros of $az^2 J_\nu''(z)+bzJ_\nu'(z)+cJ_\nu(z)$ as functions of order, Internat. J. Math. Math. Sci 15 (1992) 319–322.
26. F. W. J. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark (Eds.), NIST Handbook of Mathematical Functions, Cambridge Univ. Press, Cambridge, 2010.
27. Pfaltzgraff, J.A.: Univalence of the integral of $f'(z)^\lambda$. Bull. London Math. Soc. **7**, no. 3, 254–256 (1975).
28. Robertson, M.S.: Univalent functions $f(z)$ for which $zf'(z)$ is spirallike. Michigan Math. J. **16**, 97–101 (1969).
29. S. M. Shah, S. Y. Trimble, Entire functions with univalent derivatives, J. Math. Anal. Appl. 33 (1971) 220–229.
30. Spacek, L.: Contribution á la thèorie des fonctions univalentes, Casop Pest. Mat.-Fys. **62**, 12-19 (1933).
31. R. Szász, About the radius of starlikeness of Bessel functions of the first kind, Monatshefte für Mathematik, 176 (2015), 323-330.
32. R. Zhang, Sums of zeros for certain special functions, Integral Transforms Spec. Funct., 21(5) (2010) 3511–365.
33. Watson, G.N. : A Treatise on the Theory of Bessel Functions, Cambridge University Press, 1944.

Journal of Electronic Materials
https://doi.org/10.1007/s11664-022-10035-9

BRIEF COMMUNICATION

Green Emission from Calcium Zirconium Silicate Doped with Trivalent Terbium Phosphor

Vijay Singh[1] · Ch. B. Annapurna Devi[2] · B. Rupa Venkateswara Rao[3] · A. S. Rao[4]

Received: 4 July 2022 / Accepted: 11 October 2022
© The Minerals, Metals & Materials Society 2022

Abstract

Trivalent terbium ion (Tb^{3+})-doped calcium zirconium silicate ($Ca_3ZrSi_2O_9$) phosphors ($Ca_3ZrSi_2O_9$:xTb^{3+}), with molar composition $0.005 \leq x \leq 0.13$, were effectively produced by the standard sol–gel method. The synthesized samples were characterized by x-ray diffraction (XRD), scanning electron microscopy (SEM), Fourier transform-infrared (FT-IR) spectroscopy, photoluminescence (PL) excitation and PL emission. XRD spectral profiles of the prepared samples demonstrate the formation of a monoclinic structure. SEM micrographs of the $Ca_3ZrSi_2O_9$:xTb^{3+} phosphor samples show chunky solids interlocked in a tile-like pattern. The FT-IR spectra reveal the structure and bond activities of the samples. The PL excitation spectra demonstrate the $4f^8 \rightarrow 4f^7 5d^1$ transitions of Tb^{3+} ions under 545 nm emission wavelength. The PL emission spectra of Tb^{3+} ions recorded under 243-nm excitation wavelength exhibit $^5D_3 \rightarrow {}^7F_J$ ($J = 6, 5, 4$) and $^5D_4 \rightarrow {}^7F_J$ ($J = 6, 5, 4, 3$) transitions over the wavelength range of 350–650 nm. Of all these emission transitions, the one pertaining to the $^5D_4 \rightarrow {}^7F_5$ transition around 545 nm is strong and attributed to green emission. The photometric analysis also confirms the green emission from the $Ca_3ZrSi_2O_9$:$0.09Tb^{3+}$ phosphor, which further supports its potential application in the fabrication of LEDs and in solid state display applications.

Keywords $Ca_3ZrSi_2O_9$ · Tb^{3+} ions · sol–gel · photoluminescence

Introduction

It is well documented that the fourth-generation illumination source of the twenty-first century is light-emitting diodes (LEDs). They are often employed in lighting and display applications because of their dominant traits, and they have become a part of daily life.[1] Because of their unique properties, scientists have recently focused on white LEDs (w-LEDs) for solid-state lighting (SSL) applications, which include chemical and physical stability, energy savings, high performance, and environmental friendliness.[2,3] Phosphors emit light of a specified wavelength when activated by photons, electron beams, and other energy technologies. Rare earths (RE) with desired characteristic emission and excitation spectra are utilized as activators, while inorganic materials such as phosphate, silicate, borate, and fluorides are employed as hosts because the electronic transitions of RE ions are not greatly impacted by host lattices.[4–8]

Calcium is a soft metal and with greater versatility as a phosphor host than many other compounds. Calcium-based phosphors are widely used in phosphor-converted w-LED applications, which have surpassed traditional light bulbs and fluorescent lamps in terms of energy efficiency, lifetime, and performance.[9,10] Silicate-based RE luminous materials have a strong chemical and thermal stability and are, therefore, commonly used as a luminous host material matrix due to the different silicate and metal cation configurations.[11–13] Researchers working on the problem of sensitized luminescence have taken an interest in the optical characteristics of calcium silicate phosphors doped with transition metals such as Zr, Pb, and Mn.[14] Zirconium (Zr) has unique properties such as exceptional

✉ Vijay Singh
vijayjiin2006@yahoo.com

1 Department of Chemical Engineering, Konkuk University, Seoul 05029, Republic of Korea

2 Department of Physics, Vignan's Lara Institute of Technology and Science, Vadlamudi, Guntur, Andhra Pradesh 522213, India

3 Department of Physics, V.R. Siddhartha Engineering College, Vijayawada 52007, India

4 Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110042, India

Published online: 02 November 2022

refractoriness, high chemical resistance, low coefficient of thermal expansion, low phonon capacity, greater refractive index, and high chemical stability. For these applications, zirconium is a desirable material from the perspective of both fundamental as well as applied sciences.[15]

There are many different types of phosphors, each emitting a distinct color, and the majority of them contain RE elements such as terbium, gadolinium, yttrium, europium, lanthanum and cerium.[16–21] Terbium (Tb^{3+}) ions have attracted the attention of many optical investigators due to their peculiar emission characteristics in two crucial spectrum areas: their emission bands are largely in the visible light area due to their transitions $^5D_4 \rightarrow {}^7F_J$ ($J = 3$, 4, 5, 6) following violet or near-ultraviolet illumination.[22] Terbium ions have been widely employed in synthesizing materials for use in many disciplines such as phosphors and lasers. In this study, the authors chose $Ca_3ZrSi_2O_9$ as the host lattice and Tb^{3+} ions as the activator. The host lattice of Tb^{3+}-doped $Ca_3ZrSi_2O_9$ has been widely researched for SSL lighting applications. The authors primarily explored the luminescence properties of $Ca_3ZrSi_2O_9$:Tb^{3+} phosphors and the structure of the samples using x-ray diffraction (XRD), scanning electron microscopy (SEM), Fourier transform-infrared (FT-IR) spectroscopy, photoluminescence (PL) excitation, PL emission and photometric analysis.

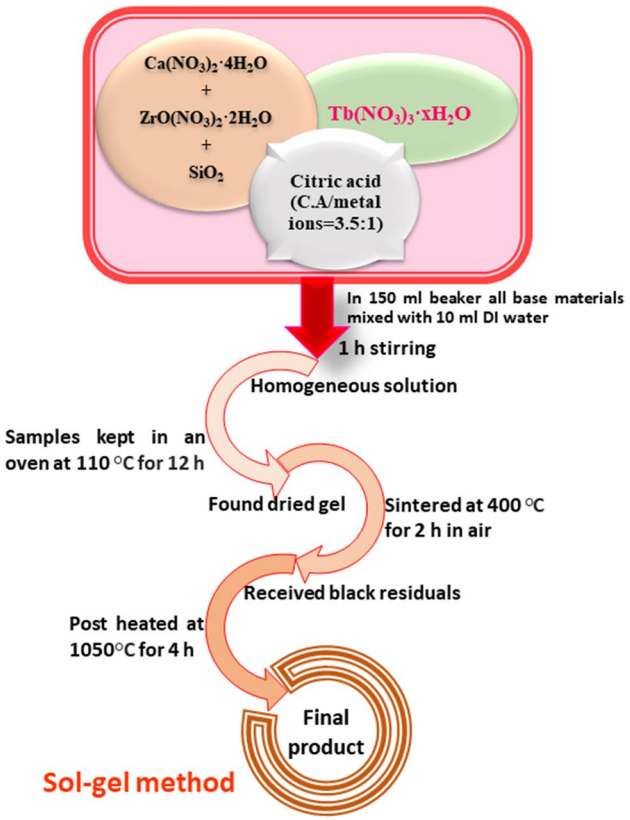


Fig. 1 Flow chart of synthesis procedure of $Ca_3ZrSi_2O_9$:xTb^{3+} phosphors.

Material Preparation and Analysis

$Ca_3ZrSi_2O_9$:xTb^{3+} samples were produced with a dopant concentration (x) range of $0.005 \leq x \leq 0.13$ using the sol–gel method. Stoichiometric quantities of $Ca = Ca(NO_3)_2{\cdot}4H_2O$ (host material), $Zr = ZrO(NO_3)_2{\cdot}2H_2O$ (host material), $Si = SiO_2$ (host material), $Tb = Tb(NO_3)_3{\cdot}xH_2O$ (dopant), and C.A = citric acid (fuel) were weighed in a 150-ml beaker and, after 10 ml distilled water was added, the solution was mixed for 1 h with a magnetic stirrer. The resultant solution was maintained in an oven until a dry gel developed. This gel was placed in a hot furnace at 400°C for 2 h in air. The obtained precursors were crushed and again fired at 1050°C for 4 h in a muffle furnace. The final sample was finely granulated for better characterization. This procedure is illustrated in a flow diagram in Fig. 1.

An x-ray diffractometer (Rigaku MiniFlex II, XRD) was used to identify the prepared samples. The XRD patterns were recorded over the 2θ range of 13–63° at a 4°/min scanning rate. A SEM instrument (S-3400, Hitachi, Japan) was used to examine the surface morphology. FT-IR spectra were collected from a spectrometer (Thermo Fisher Nicolet 6700). Photoluminescence characterizations were done on a spectrofluorophotometer (Shimadzu, RF-5301PC).

Results and Discussion

Crystallographic Structural Analysis

Figure 2a shows the standard orientation of the $Ca_3ZrSi_2O_9$ crystal structure, and Fig. 2b, c, and d shows its orientation along the (001), (100) and (010) planes, respectively. These figures reveal that calcium zirconium silicate is crystallized in an uncommon monoclinic structure with ZrO_6 octahedron pairs shared at the edges of its unique layer. Along with ZrO_6 and CaO_n layers, SiO_4 tetrahedrons are inserted along the c-axis to form a three-dimensional crystal structure. Thus, $Ca_3ZrSi_2O_9$ is a complicated silicate structure. For this monoclinic system ($\overline{a} \neq \overline{b} \neq \overline{c}$ and $\alpha = \gamma = 90° \neq \beta$) the lattice parameters are $\overline{a} = 7.3603(1)$ Å, $\overline{b} = 10.1766(3)$ Å, $\overline{c} = 10.4514(3)$ Å and $\beta = 90.875(2)°$.[23] $Ca_3ZrSi_2O_9$ related to cuspidine ($Ca_4Si_2O_7$)F_2 group. In its crystal structure, there exist disilicate groups and four columns of wide "octahedral" walls stretched along the [001] direction. In $Ca_3ZrSi_2O_9$, those octahedral walls contain three Ca-O polyhedra with 7–8 coordination, and a single Zr-O polyhedron exists with 6 coordination.[23,24] The doped cation replaces the cation with similar ionic radius (r) in

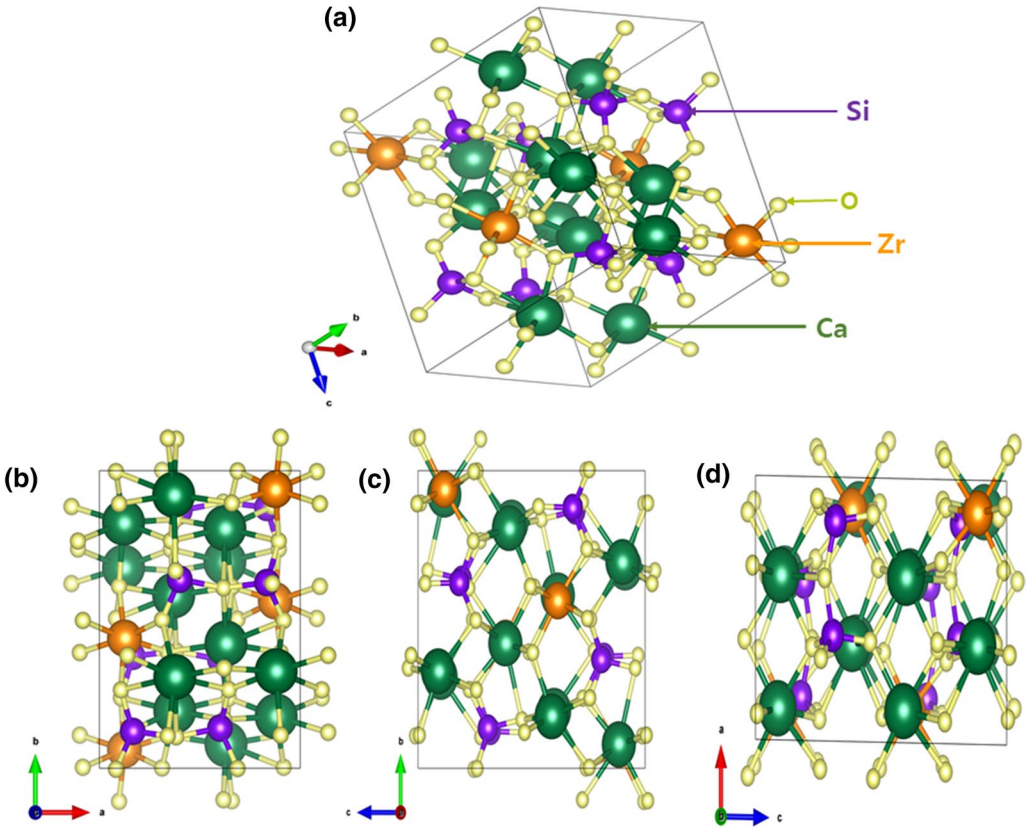


Fig. 2 (a) A schematic of the crystal structure of the $Ca_3ZrSi_2O_9$ host material indicating the position of the atoms in the lattice. (b–d) crystal structure data of the $Ca_3ZrSi_2O_9$ system from various directions along the (001), (100) and (010) planes.

the host lattice. Hence, in the present crystal structure, Ca^{2+} ($r = 1.12$ Å) ions are quite suitable for substitution by Tb^{3+} ($r = 1.18$ Å) ions rather than Zr^{4+} ($r = 0.72$ Å) ions owing to their similar ionic radii.[25]

The XRD spectral profiles of the $Ca_{3-x}ZrSi_2O_9{:}xTb^{3+}$ phosphor samples are shown in Fig. 3 and confirm their crystalline phase. These patterns are well matched with the standard JCPDS cards (83–0365), although minute amounts of impurities such as $CaZrO_3$ and β-Ca_2SiO_4 were present in the samples, and their corresponding peaks are marked in Fig. 3. The host structure does not appear to have been greatly affected by doping with Tb^{3+}, as the XRD peaks of all the samples are similar to the profile of $Ca_3ZrSi_2O_9$ phosphor. The major phase formation of all the prepared phosphors exhibits a monoclinic system corresponding to space group $P2_1/c$ (no.14).[26] The crystallite diameter of the prepared samples can be evaluated by employing the Debye–Scherrer (DS) formula:

$$D = \frac{0.9\lambda}{\beta\cos\theta} \tag{1}$$

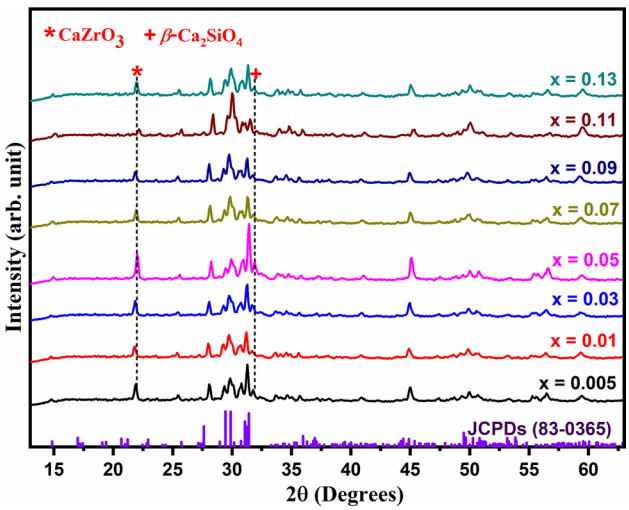


Fig. 3 XRD patterns of $Ca_3ZrSi_2O_9{:}xTb^{3+}$ ($x = 0.005 \le x \le 0.13$) phosphors.

where D = particle diameter size, λ = wavelength of x-ray source, β = full width at half maximum of the diffraction peak in radians and θ = Bragg angle in radians. From the

above equation, the average crystallite diameter was calculated as 0.78 nm from the XRD spectra using the DS formula.

SEM

The surface morphology of the $Ca_3ZrSi_2O_9$:xTb^{3+} phosphor samples is shown in Fig. 4 at different magnifications. The luminescence properties of phosphors are affected by surface morphology, such as size, shape, size distribution, and defects. The standard phosphor sample ranges in size from 1 µm to 2 µm and has strong crystallinity but an irregular surface morphology.[1,27] The present phosphor samples showed chunky solids interlocked in a tile-like pattern. It is widely assumed that the candle power can be enhanced by improving the morphological structure of the sample from chunky solids until it becomes closely packed.[28] The prepared samples should meet these criteria in the experiment; as a result, the particle size is measured in microns and the samples are composed of accumulated particles with a diameter of 1 to 2 microns.

FT-IR Analysis

The chemistry functional group of the $Ca_3ZrSi_2O_9$:xTb^{3+} phosphors was studied using FT-IR spectroscopy, and the findings are shown in Fig. 5. The low-lying bands at

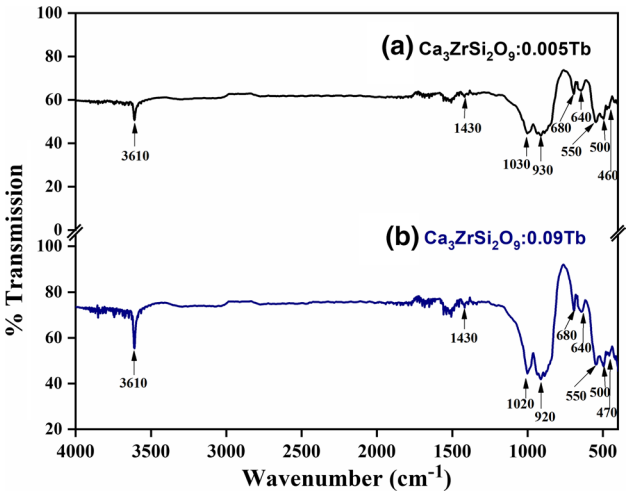


Fig. 5 FT-IR spectra of (a) $Ca_3ZrSi_2O_9$:$0.005Tb^{3+}$ and (b) $Ca_3ZrSi_2O_9$:$0.09Tb^{3+}$ phosphors.

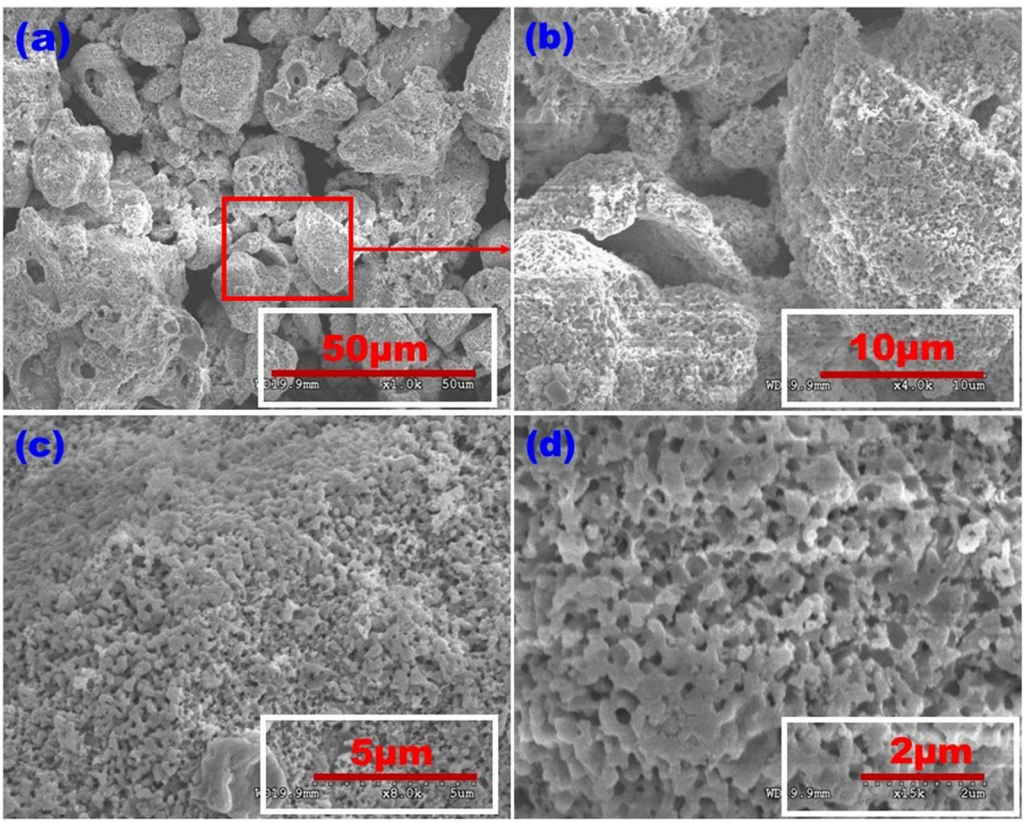


Fig. 4 Scanning electron microscopy (SEM) images of $Ca_3ZrSi_2O_9$:$0.09Tb^{3+}$ phosphor at low magnification (a, b) [where (b) is the magnified view of the red zone marked in image (a)] and at higher magnification (c, d) (Color figure online).

460 cm^{-1} and 500 cm^{-1} are due to bending modes of SiO_4 tetrahedra and the 550 cm^{-1} attributed to O–Si–O bending vibrations.[29] Zr–O–Zr and Zr–OH stretching trigger the bands at 470 cm^{-1} and 640 cm^{-1}, respectively. The presence of the M–O bond, where M is a metallic constituent, was established by the development of a peak absorption band at 680 cm^{-1}.[30] The vibration of Si–NBO–Ca in the band at about 1030 cm^{-1} and 930 cm^{-1} is due to bending.[31] As a result, certain oxygen atoms in the phosphor network are not coupled to a Si atom directly. Based on this criterion, SiO groups can be classed as "bridging" or "non-bridging."[32] Furthermore, the band at 1430 cm^{-1} is assigned to C=O vibrations, and that at 1540 cm^{-1} to C–C vibrations,[33] but the high-frequency band at about 3610 cm^{-1} is attributed to O–H stretching in calcium hydroxide, indicating the creation of a CaO phase.[34] The presence of many peaks generated by substantial crystallization of the phosphor matrix enriched the FT-IR result, which agreed with the XRD data.

PL Excitation and PL Emission Spectral Analysis

The PL performance of the prepared phosphors can be explored through their excitation and emission spectra. Figure 6 shows the PL excitation spectra of the $Ca_3ZrSi_2O_9$:xTb^{3+} phosphor materials recorded from 220 nm to 350 nm at an emission wavelength of 545 nm ($^5D_4 \rightarrow {}^7F_6$) corresponding to green light. These excitation spectra exhibit a broad hump ranging from 220 nm to 350 nm, with a peak at 243/249 nm and a shoulder at 283 nm, both resulting from the inter-configurational spin-allowed ($\Delta S = 0$) $4f^8 \rightarrow 4f^7 5d^1$ transition of trivalent terbium ions.[35] The shoulder in the present excitation spectra represents the absorption through charge transfer states. In addition, the excitation peak is shifted from 243.3 nm to 249.2 nm as the concentration of Tb^{3+} ions is increased from 0.005 mol to 0.13 mol owing to the redshift of 6 nm exhibited by the charge transfer band. Moreover, the intensity of the transitions alone varies without any change in slope, demonstrating that the variation of Tb^{3+} ion concentration may not affect the crystal field strength. However, 0.09 mol Tb^{3+} ion-doped $Ca_3ZrSi_2O_9$ phosphor exhibits the maximum absorption.

Figure 7 shows the PL emission spectra emitted by the $Ca_3ZrSi_2O_9$:xTb^{3+} phosphor samples in the wavelength range of 350–650 nm under an UV excitation of 243 nm ($4f^8 \rightarrow 4f^7 5d^1$ transition). The spectra contain multiple emission peaks related to *f–f* transitions from two excited energy levels $^5D_3$ and $^5D_4$ to lower energy levels ^7F_J ($J = 6, 5, 4, 3$). The luminescence bands were found at wavelengths of 383 nm (violet), 418 nm (violet), 438 nm (indigo), 490 nm (blue), 545 nm and 550 nm (green), 584 nm (yellow), and 622 nm (red), which were the characteristic emissions of Tb^{3+} ions over the total visible region. These peaks were allocated to the typical transitions $^5D_3 \rightarrow {}^7F_6$, $^7F_5$, $^7F_4$, $^5D_4 \rightarrow {}^7F_6$, $^7F_5$, $^7F_4$ and $^7F_3$ of Tb^{3+} ions according to Carnall.[36] The first three transitions are from the $^5D_3$ level, and the next four bands are from the $^5D_4$ level.[37] Among these transitions, the magnetic dipole transition $^5D_4 \rightarrow {}^7F_5$ at 545 nm strongly signifies the intensified luminescence in

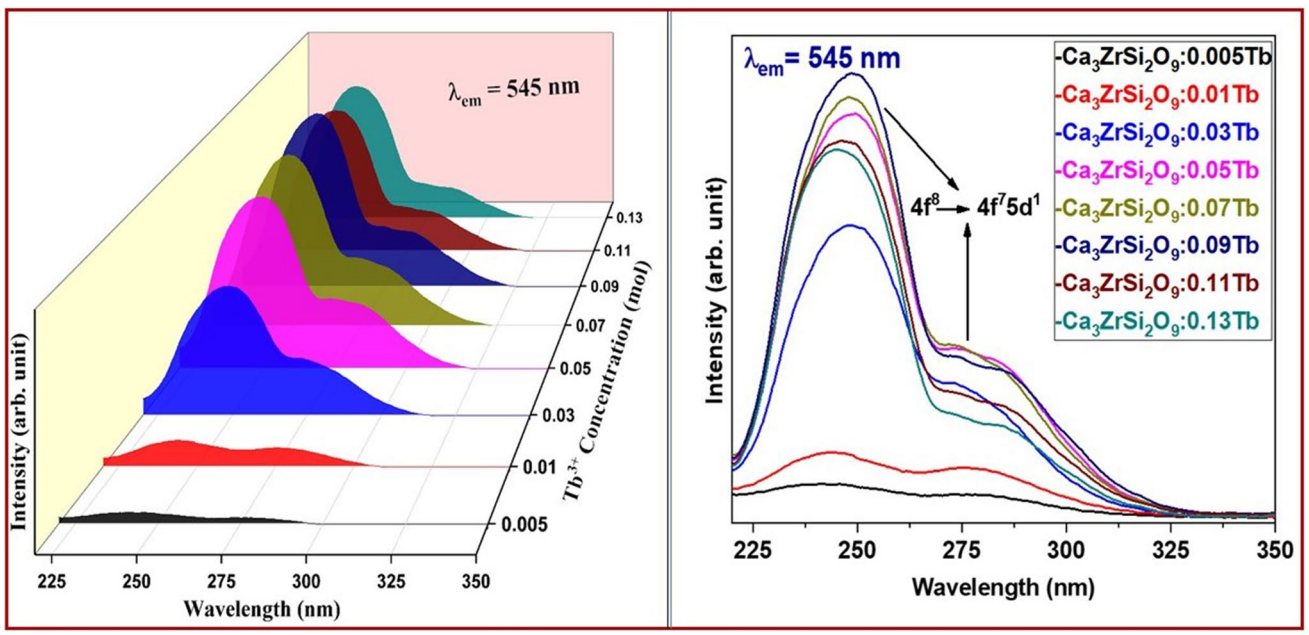


Fig. 6 Photoluminescence excitation 3D (left side) and 2D (right side) spectra of $Ca_3ZrSi_2O_9$:xTb^{3+} ($x = 0.005 \le x \le 0.13$) phosphors ($\lambda_{em} = 545$ nm).

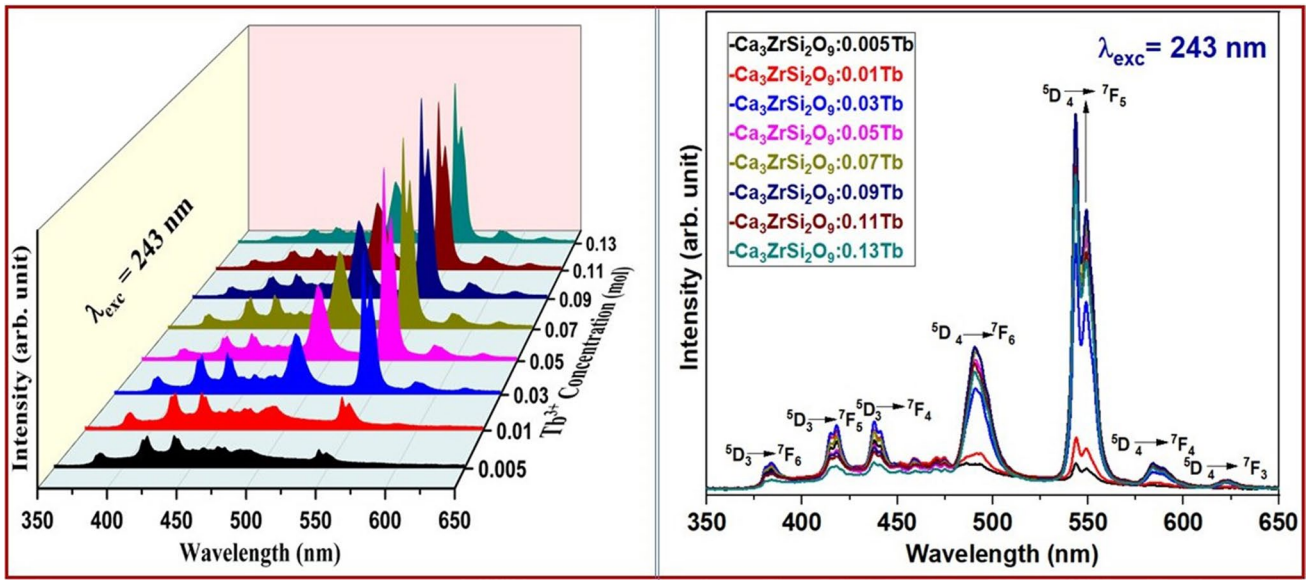


Fig. 7 Photoluminescence emission 3D (left side) and 2D (right side) spectra of $Ca_3ZrSi_2O_9:xTb^{3+}$ ($x=0.005\leq x\leq 0.13$) phosphors ($\lambda_{exc}=243$ nm) (Color figure online).

the green region by Tb^{3+} ions in all prepared phosphors.[38] The excitation and emission spectral features reveal that the $Ca_3ZrSi_2O_9$:0.09Tb^{3+} phosphor is useful for the production of LEDs and in solid state display applications.

Concentration Quenching Mechanism

The emission intensities of all the prepared phosphors relevant to the strong transition ($^5D_4\rightarrow{}^7F_5$) around 545 nm are plotted against their concentrations and are shown in the bar graph of Fig. 8. The emission intensity increases from a minimum at low concentrations (0.005 mol and 0.01 mol) to a maximum at 0.09 mol concentration, after which it decreases slightly because of concentration quenching. This concentration quenching can be attributed to cross-relaxation between closely spaced Tb^{3+} ions at their higher concentrations.[39]

To elucidate the concentration quenching mechanism in the as-synthesized phosphor, the critical distance (R_c) is measured using the Blasse equation[40]:

$$R_c = 2\left(\frac{3V}{4\pi x_c N}\right)^{1/3} \quad (2)$$

where V = volume of the unit cell, x_c = optimum doping concentration, and N = number of Tb^{3+} cations in the host. For $Ca_3ZrSi_2O_9:Tb^{3+}$, $V=784.75$ Å, $x_c=0.09$ and $N=4$. Using these values in the above equation, the critical distance between the activators (R_c) was calculated to be 16.09 Å, which confirmed the energy transfer mechanism. If $R_c<5$ Å, the non-radiative (NR) energy transfer among the activators occurs due to exchange interaction, whereas if $R_c>5$ Å, it

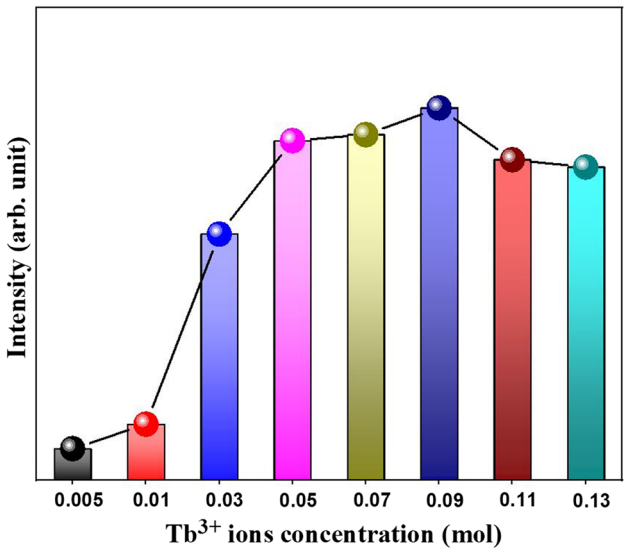


Fig. 8 Tb^{3+} concentration variation on emission intensity by bar diagram and increase in intensity with concentration (at 545 nm strong emission).

occurs because of multipole–multipole interaction.[41] As R_c is more than 5 Å for the phosphors of this study, the energy transfer may occur due to multipole–multipole interaction across the activators. Moreover, Dexter's theory of energy transfer re-affirms that multipole–multipole interaction alone causes the NR energy transfer among Tb^{3+} ions in the synthesized phosphors.[37]

In addition, the strength of the electric multipolar interaction can be determined by the Van Uitert theory, using the following equation[42]:

$$\frac{I}{x} = K\left[1 + \beta(x)^{Q/3}\right]^{-1} \quad (3)$$

where K and β are constants for a given host under the same excitation condition, x is the concentration of activator ion, I/x is the emission intensity (I) per activator (x). $Q = 6$, 8 and 10 is for dipole–dipole (d–d), dipole–quadrupole (d–q) and quadrupole–quadrupole (q–q) interactions, respectively. To investigate the type of interaction between Tb^{3+} ions, the Q value has to be evaluated. The above equation is modified to plot $\log(I/x)$ against $\log(x)$ and it was observed to be linear with slope $(-Q/3)$. To evaluate the α value, $\log(x)$ is plotted against $\log(I/x)$ for x ranging from 0.05 to 0.13. From Fig. 9, 'Q' was found to be 5.46 which is nearer to 6. This analysis demonstrated that the dipole–dipole interaction is predominant in the concentration quenching of Tb^{3+} ions emission in the $Ca_3ZrSi_2O_9$ host.

The partial energy level diagram representing the excitation and emission transitions of the trivalent terbium ions is depicted in Fig. 10. The samples exhibit only one excitation transition at 243 nm because of the transition from ground state $^7F_6$ to an excited energy band with $4f^7 5d^1$ configuration, followed by the non-radiative transition to $^5D_3$ and $^5D_4$ levels. This leads to three emission transitions occurring at 383 nm, 418 nm, and 438 nm from $^5D_3$ energy level to $^7F_6$, $^7F_5$ and $^7F_4$, respectively, whereas four emission transitions occur from $^5D_4$ energy level to $^7F_6$, $^7F_5$, $^7F_4$ and $^7F_3$ at 490 nm, 545 nm, 584 nm, and 622 nm, respectively, in the visible region. All these excitation and emission transitions are explained in the schematic energy level diagram of Tb^{3+} ions.

Photometric Analysis

To determine the color released by a phosphor under various illuminations, the photographs of the prepared phosphor powders under room light are compared with their photographs under 254-nm UV light. In the present work, Fig. 11 shows the characteristic images of (1) $Ca_3ZrSi_2O_9$:0.005Tb^{3+} and (2) $Ca_3ZrSi_2O_9$:0.09Tb^{3+} phosphors under room light (Fig. 11a and a′) and 254-nm UV light (Fig. 11b and b′), respectively. These photographs clearly demonstrate that the abovementioned phosphors are white in color at room light and appear green in color when they are illuminated with UV light.

The color coordinates (x, y) of all the prepared phosphors were evaluated by correlating the PL emission data ($\lambda_{exc} = 243$ nm) with Commission International de I'Eclairage (CIE) color matching functions[43,44] and the data are presented in Table I. The color purity percentages of the current phosphors were determined using the standard literature procedure[11] and are shown in Table I, along with their symbols. Most of the calculated chromaticity coordinates are spotted on the CIE chromaticity diagram with their color purity signs, as shown in Fig. 12. As Tb^{3+} ion concentration increases in the figure, the color emitted by the phosphors shifts from bluish green (0.2081, 0.3571) to bright green (0.2487, 0.5347) (as shown by the arrow in the diagram). Among all the prepared phosphors, the $Ca_3ZrSi_2O_9$:0.09Tb^{3+} phosphor's color coordinates (0.2487, 0.5347) in particular are comparable to the ideal green coordinates [0.29, 0.60; as per EBU (European Broadcasting Union)] and it has

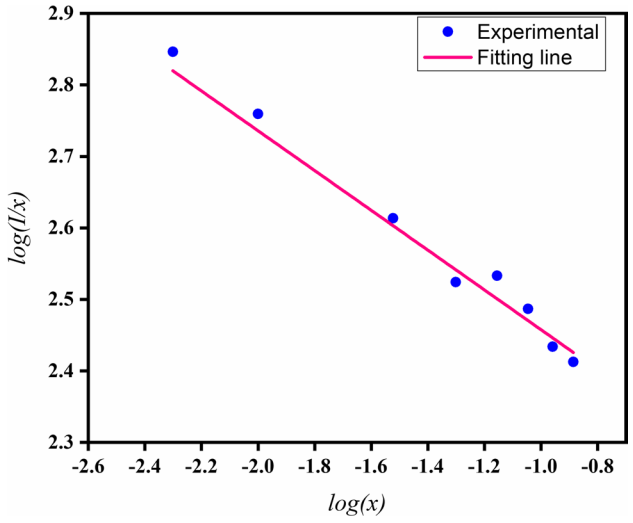


Fig. 9 Dependence of $\log(I/x)$ on $\log(x)$ in $Ca_3ZrSi_2O_9$:$x$$Tb^{3+}$ phosphors.

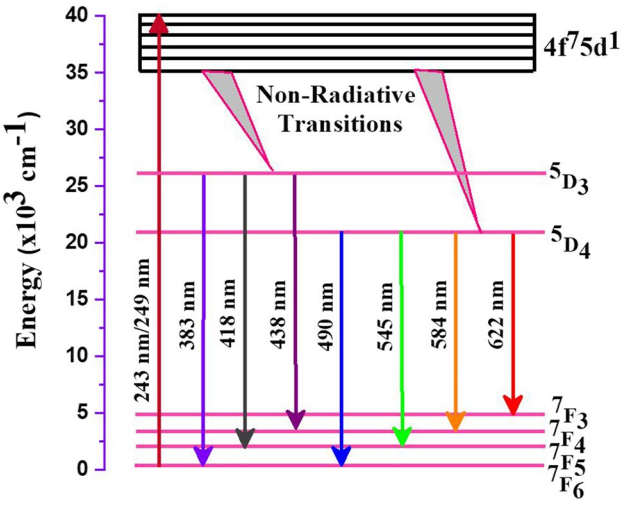


Fig. 10 Partial energy level diagram showing the luminescence mechanism of trivalent terbium ion.

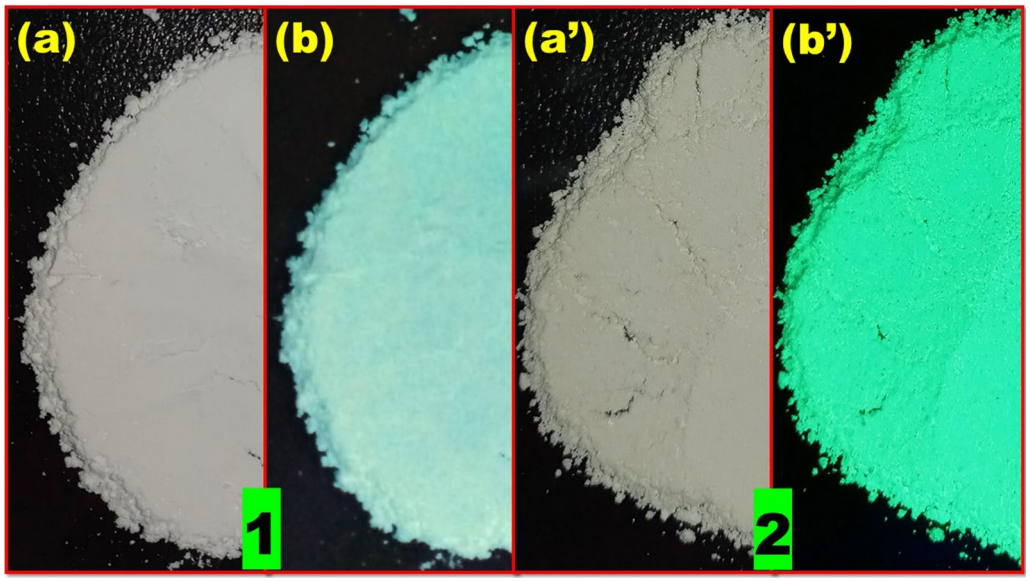


Fig. 11 Typical photographs of (1) $Ca_3ZrSi_2O_9$:0.005Tb^{3+} and (2) $Ca_3ZrSi_2O_9$:0.09Tb^{3+}. (a) and (a′) white powder image in direct room light and (b) and (b′) green visible image under a 254-nm ultraviolet lamp (Color figure online).

Table I CIE color coordinates for $Ca_3ZrSi_2O_9$:xTb^{3+} phosphors utilizing photoluminescence emission results (λ_{exc} = 243 nm)

Phosphors $Ca_3ZrSi_2O_9$:xTb^{3+}	λ_{exc} = 243 nm		Color purity %: (sign)
	x	y	
x = 0.005	0.2081	0.3571	27.08 (●)
x = 0.01	0.2150	0.3866	29.15 (●)
x = 0.03	0.2392	0.4957	47.55 (●)
x = 0.05	0.2472	0.5292	54.74 (●)
x = 0.07	0.2452	0.5233	53.49 (●)
x = 0.09	0.2487	0.5347	55.93 (●)
x = 0.11	0.2448	0.5213	53.05 (●)
x = 0.13	0.2474	0.5323	55.44 (●)

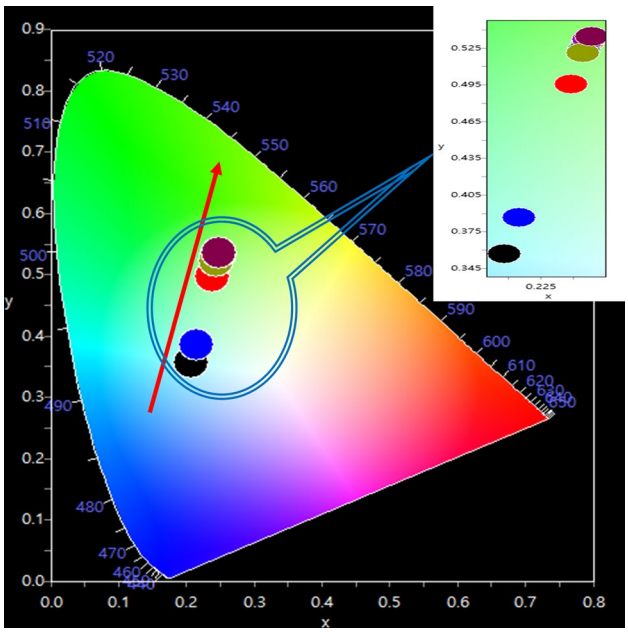


Fig. 12 CIE chromaticity diagram for $Ca_3ZrSi_2O_9$:xTb^{3+} phosphors (Color figure online).

good color purity (55.93%). Based on these study results, the $Ca_3ZrSi_2O_9$:0.09Tb^{3+} phosphor can serve as a potential source of green color emission and should prove useful as a green component to fabricate solid-state display devices.

Conclusions

To summarize, a new group of $Ca_3ZrSi_2O_9$:xTb^{3+} phosphors was synthesized by the sol–gel method and their structural and luminescence properties were investigated through XRD, SEM, FT-IR and PL characterization. The XRD analysis revealed that all the prepared phosphors are crystallized in a monoclinic structure and related to the $P2_1/c$ (no.14) space group. The SEM images showed the morphology of $Ca_3ZrSi_2O_9$:Tb^{3+} phosphors as chunky solids interlocked in a tile-like pattern. The FT-IR studies revealed the presence of phonon vibrational modes of SO_4 and confirmed the presence of a metallic element.

The PL excitation spectra of the $Ca_3ZrSi_2O_9$:xTb^{3+} samples under 545 nm emission exhibited only one broad band at 249 nm representing the characteristic transition $4f^8 \rightarrow 4f^7 5d^1$ of Tb^{3+} ions. PL emission spectra recorded under 243 nm excitation wavelength presented three emissions due to $^5D_3 \rightarrow {}^7F_J$ ($J = 6, 5, 4$) transitions and four emissions due to $^5D_4 \rightarrow {}^7F_J$ ($J = 6, 5, 4, 3$) transitions, of which $^5D_4 \rightarrow {}^7F_5$ was relatively more intense and lies in the green region (545 nm). When $Ca_3ZrSi_2O_9$:xTb^{3+} phosphors with concentrations 0.005 and 0.09 were excited with UV light (254 nm), the $Ca_3ZrSi_2O_9$:$0.09Tb^{3+}$ phosphor powder emitted the best green color. CIE chromatic analysis also validated the green color emission of the $Ca_3ZrSi_2O_9$:$0.09Tb^{3+}$ phosphor. The present investigation, though preliminary in nature, suggests that these photoluminescence excitation and emission spectral results, along with the CIE coordinates, support the potential application of the $Ca_3ZrSi_2O_9$:$0.09Tb^{3+}$ phosphor as a green emitter in the fabrication of solid-state display devices.

Acknowledgments This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1092509). This paper was supported by the KU Research Professor Program of Konkuk University.

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

References

1. Z.Q. Peng, R. Chen, and W.L. Feng, Photoluminescence properties of $Ca_3Si_2O_7$: Pr^{3+} orange-red phosphors prepared by high-temperature solid-state method. *Z. Naturforsch.* 73, 555 (2018).
2. P. Du, X. Huang, and J.S. Yu, Facile synthesis of bifunctional Eu^{3+}-activated $NaBiF_4$ red-emitting nanoparticles for simultaneous white light-emitting diodes and field emission displays. *Chem. Eng. J.* 337, 91 (2018).
3. M. Rajendran and S. Vaidyanathan, A novel highly efficient host-sensitized red emitting ($Ba_2YV_3O_{11}$:Eu^{3+}) phosphor for hybrid white LEDs. *Chem. Sel.* 5, 5128 (2020).
4. V. Singh, C.B.A. Devi, A.S. Rao, and J.L. Rao, Narrowband ultraviolet B emitting Gd^{3+} activated $NaSrPO_4$ phosphor prepared by co-precipitation method. *Optik* 208, 163632 (2020).
5. A.S. Rao, J.L. Rao, and S.V.J. Lakshman, Electron paramagnetic resonance and optical absorption spectra of Cu^{2+} ions in alkali cadmium borosulphate glasses. *J. Phys. Chem. Solids* 53, 1221 (1992).
6. V. Singh, S. Kaur, C.B.A. Devi, A.S. Rao, and J.B. Joo, Electron paramagnetic resonance and optical properties of green-emitting Mn (II)-doped hardystonite phosphor prepared by sol-gel method. *Optik* 266, 169553 (2022).
7. R.R. Pokam, Ch.B.A. Devi, S. Koneru, M. Shaik, R. Talewar, and S.R. Allam, NIR photoluminescence studies of Nd^{3+} doped B_2O_3–BaF_2–PbF_2–Al_2O_3 glasses for 1.063 μm laser applications. *J. Lumin.* 229, 117701 (2021).
8. V. Singh, A. Yadav, Ch.B.A. Devi, A.S. Rao, and N. Singh, Luminescence properties of Sm^{3+} doped LaP_3O_9 phosphors. *Optik* 242, 167264 (2021).
9. V. Singh, C.B.A. Devi, B.R.V. Rao, A.S. Rao, N. Singh, and B.M. Mistry, Narrow-band ultraviolet B (UVB) emitting $CaZr_4(PO_4)_6$ doped with Gd^{3+} phosphor. *Optik* 226, 165932 (2021).
10. R. Schmidt, M. Ströbele, C.P. Romao, D. Enseling, T. Jüstel, and H.J. Meyer, Synthesis, structure and properties of a calcium oxonitridosilicate phosphor showing green or red luminescence upon doping with Eu^{2+} or Ce^{3+}. *Dalton Trans.* 48, 14069 (2019).
11. V. Singh, Ch.B.A. Devi, S. Kaur, A.S. Rao, and N. Singh, Optical properties of Sr_2La_8 $(SiO_4)_6O_2$ doped with Ho^{3+} phosphor. *Optik* 242, 167268 (2021).
12. D. Halmurat, T. Yusufu, Q. Wang, J. He, and A. Sidike, Rare earth ion Tb^{3+} doped natural sodium feldspar ($NaAlSi_3O_8$) luminescent properties and energy transfer. *Sci. Rep.* 9, 14637 (2019).
13. V. Singh, C.B.A. Devi, S. Kaur, A.S. Rao, and N. Singh, Green emission of Er^{3+}-doped $Sr_2La_8(SiO_4)_6O_2$ obtained by a sol-gel method. *Optik* 243, 167322 (2021).
14. J.H. Schulman, R.J. Ginther, and E.W. Claffy, Note on the properties of calcium silicate phosphors. *J. Opt. Soc. Am.* 43, 318 (1953).
15. R.K. Tamrakar, D.P. Bisen, and K. Upadhyay, Photoluminescence behavior of ZrO_2: Eu^{3+} with variable concentration of Eu^{3+} doped phosphor. *J. Radiat. Res. Appl. Sci.* 8, 11 (2015).
16. S. Mahamuda, F. Syed, C.B.A. Devi, K. Swapna, M.V.V.K.S. Prasad, M. Venkateswarlu, and A.S. Rao, Spectral characterization of Dy^{3+} ions doped phosphate glasses for yellow laser applications. *J. Non-Cryst. Solids* 555, 120538 (2021).
17. V.M. Krishna, S. Mahamuda, P. Sailaja, K. Swapna, M. Venkateswarlu, and C.B.A. Devi, Near-IR luminescence in Nd^{3+} ions doped Na_2O-BaF_2-CaF_2-B_2O_3-TeO_2 glasses for 1064 nm laser and fiber amplifier applications. *J. Non-Cryst. Solids* 590, 121671 (2022).
18. G.H. Lee, C.K. Kim, D.H. Lee, and Y.J. Song, A study on the leaching of rare earth elements from waste phosphor powder. *Korean J. Met. Mater.* 59, 459 (2021).
19. R.A. Talewar, S. Mahamuda, K. Swapna, M. Venkateswarlu, and A.S. Rao, Spectroscopic studies of Sm^{3+} ions doped alkaline-earth chloro borate glasses for visible photonic applications. *Mater. Res. Bull.* 105, 45 (2018).
20. N. Deopa, S. Kaur, A. Prasad, B. Joshi, and A.S. Rao, Spectral studies of Eu^{3+} doped lithium lead alumino borate glasses for visible photonic applications. *Opt. Laser Technol.* 108, 434 (2018).
21. H. George, N. Deopa, S. Kaur, A. Prasad, M. Sreenivasulu, M. Jayasimhadri, and A.S. Rao, Judd-Ofelt parametrization and radiative analysis of Dy^{3+} ions doped sodium bismuth strontium phosphate glasses. *J. Lumin.* 215, 116693 (2019).
22. P.M. Tan, T. Ngoc, V.D. Nguyen, N.T. Hien, V.X. Hoa, N.X. Truong, V.T.K. Oanh, N.D. Tam, N.X. Ca, S. Bharti, and Y. Peng, Study of optical properties and energy transfer mechanism of Tb^{3+}, Sm^{3+} singly doped and co-doped ZnS quantum dots. *Opt. Mater.* 114, 110901 (2021).
23. J.R. Plaister, J. Jansen, R.A.G. de Graaff, and D.J.W. Ijdo, Structure determination of $Ca_3HfSi_2O_9$ and $Ca_3ZrSi_2O_9$ from powder diffraction. *J. Solid State Chem.* 115, 464 (1995).
24. Y. Sato, R. Miyake, A. Tanigaki, S. Akiyama, K. Tomita, and M. Kakihana, A novel Eu^{2+}-activated calcium zirconium silicate phosphor: $Ca_3ZrSi_2O_9$:Eu^{2+}. *J. Lumin.* 231, 117752 (2021).
25. G. Blasse, G. Bernardi, D.J.W. Ijdo, and J.R. Plaisier, Yellow zirconate luminescence in $Ca_3ZrSi_2O_9$. *J. Alloys Compd.* 217, 29 (1995).
26. Z. Zhu, Z. Sun, Z. Guo, and X. Zhang, Luminescence of $Ca_3ZrSi_2O_9$:Ce^{3+} blue phosphor with good thermal stability. *J. Lumin.* 207, 430 (2019).

27. P. Du, L.K. Bharat, X.-Y. Guan, and J.S. Yu, Synthesis and luminescence properties of color-tunable Dy^{3+}-activated $CaWO_4$ phosphors. *J. Appl. Phys.* 117, 083112 (2015).
28. X. Yan, R. Yu, Y. Xu, L. Li, P. Xu, D. Wang, J. Deng, J. Chen, and X. Xing, Morphology-tailored synthesis of flower-like Y_2O_3:Eu^{3+} microspheres. *Mater. Res. Bull.* 47, 2135 (2012).
29. C. Zuo, A. Xiao, Z. Zhou, Y. Chen, X. Zhang, X. Ding, X. Wang, and Q. Ge, Spectroscopic properties of Ce^{3+}–doped BaO-Gd_2O_3-Al_2O_3-B_2O_3-SiO_2 glasses. *J. Non-Cryst. Solids.* 452, 35 (2016).
30. Q. Chen, J. Li, and W. Wang, Synthesis and luminescence properties of Tb^{3+}/Eu^{3+} co-doped $GdAlO_3$ phosphors with enhanced red emission. *J. Rare Earth* 36, 924 (2018).
31. S. Kashyap, K. Griep, and J.A. Nychka, Crystallization kinetics, mineralization and crack propagation in partially crystallized bioactive glass 45S5. *Mater. Sci. Eng. C* 31, 762 (2011).
32. A. Balamurugan, G. Sockalingum, J. Michel, J. Faure, V. Banchet, L. Wortham, S. Bouthors, D.L. Maquin, and G. Balossier, Synthesis and characterisation of sol gel derived bioactive glass for biomedical applications. *Mater. Lett.* 60, 3752 (2006).
33. F. Bollino, E. Armenia, and E. Tranquillo, Zirconia/hydroxyapatite composites synthesized via sol-gel: influence of hydroxyapatite content and heating on their biological properties. *Materials* 10, 757 (2017).
34. A. Beganskiene, O. Dudko, R. Sirutkaitis, and R. Giraitis, Water based sol-gel synthesis of hydroxyapatite. *Mater. Sci.* 9, 383 (2003).
35. G.S.R. Raju, E. Pavitra, G. Nagaraju, X.-Y. Guan, and J.S. Yu, UV-A and UV-B excitation region broadened novel green color-emitting $CaGd_2ZnO_5$:Tb^{3+} nanophosphors. *RSC Adv.* 5, 22217 (2015).
36. W.T. Carnall, P.R. Fields, and K. Rajnak, Electronic energy levels of the trivalent lanthanide aquo ions. III. Tb^{3+}. *J. Chem. Phys.* 49, 4447 (1968).
37. C. Zeng, H. Huang, Y. Hu, S. Miao, and J. Zhou, A novel blue-greenish emitting phosphor $Ba_3LaK(PO_4)_3F$: Tb^{3+} with high thermal stability. *Mater. Res. Bull.* 76, 62 (2016).
38. A. Raja, G. Annadurai, D.J. Daniel, and P. Ramasamy, Synthesis, optical and thermal properties of vovel Tb^{3+} doped $RbCaF_3$ fluoroperovskite phosphors. *J. Alloys Compd.* 683, 654 (2016).
39. E. Cavalli, P. Boutinaud, R. Mahiou, M. Bettinelli, and P. Dorenbos, Luminescence dynamics in Tb^{3+}-doped $CaWO_4$ and $CaMoO_4$ crystals. *Inorg. Chem.* 49, 4916 (2010).
40. C. Zeng, H. Liu, Y. Hu, L. Liao, and L. Mei, Color-tunable properties and energy transfer in $Ba_3GdNa(PO_4)_3F$:Eu^{2+}, Tb^{3+} phosphor pumped for *n*-UV *w*-LEDs. *Opt. Laser Technol.* 74, 6 (2015).
41. D.L. Dexter, A theory of sensitized luminescence in solids. *J. Chem. Phys.* 21, 836 (1953).
42. L.G. Van Uitert, Characterization of energy transfer interactions between rare earth ions. *J. Electrochem. Soc.* 114, 1048 (1967).
43. W.M. Yen and S. Shionoya, *Phosphor handbook* (New York: CRC Press, 1999), p. 805.
44. F.W. Billmeyer and M. Saltzman Jr., *Principles of color technology* (New York: Wiley, 1981).

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

Advanced Production and Industrial Engineering
R.M. Singari and P.K. Kankar (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE220727

Image Caption Generator in Hindi Using Attention

Abhishek SETHI [a,1], Aditya JAIN [b], Chhavi DHIMAN [c]
[a, b, c] *Electronics and Communication Engineering (ECE)*
Delhi Technological University (formerly DCE)
Delhi, India
abhisheksethi_2k18ec015@dtu.ac.in, adityajain_2k18ec018@dtu.ac.in, chhavi.dhiman@dtu.ac.in

Abstract. Image Captioning has gained tremendous spotlight in recent years. However, the captioning models generate captions in English language. In this paper, we present an image caption generator for our regional language that is Hindi using Resnet50 and LSTM with attention module. An experimental study is shown highlighting the effect of attention-based learning on generated Hindi captions. Flickr8k dataset in Hindi language is used to validate the performance of proposed work in terms of BLEU score.

Keywords. Image captions, Pre-trained models, LSTM, BLEU scores

1. Introduction

Captioning images is a difficult undertaking since it requires not only identifying the many items in the image but also building a meaningful relationship between them. Solving this task effectively will have a lot of benefits across different domains. Be it healthcare, retail, IT or many more sectors that we are not yet aware of.

People generate image descriptions incorporating subjective properties like sentiment or stylistic effects(2). Also, when multiple objects are in an image, the traditional structure fails to capture and express the relationship among multiple objects. Recently, some image captioning model started using attention mechanisms(3), that highlight salient features dynamically as needed rather than compressing an entire image into a static representation, and also, improved the accuracy of description.

2. LITERATURE REVIEW

The challenge of creating natural language descriptions from visual data has been researched in computer vision for a long time, although mostly for video. This has resulted in complicated systems that combine visual basic recognizers with a structured formal language and then transform them to plain language using rule-based systems.(1) The systems constructed are substantially hand-crafted, brittle, and have only been demonstrated in a few areas, such as traffic or sports.(4)

The difficulty of describing static images with natural prose has recently piqued people's curiosity. We can drive natural language generation systems by using recent breakthroughs in object identification, characteristics, and locations. Farhadi et al.(5) employs detections to create a triplet of scene elements viz. {object, action and scene} and used these three elements to map it into a sentence which is further utilised to estimate the likelihood of the target sentence being related to the image, and to what degree. S. Li et al.(6) does the same thing, starting with detections and piecing together a final description using 2 steps viz. phrase selection and phrase fusion generated from identified items and relationships.

The initial work on image captioning was primarily focused on two study methods: retrieval-based and template-based. These solutions were reliant on hard-coded language structure, which had the drawback of being inflexible(7). As a result, these techniques have become obsolete, and Deep Neural Networks (DNNs) have taken their place. DNN has achieved major results in the fields of Computer Vision (CV) and Natural Language Processing (NLP)(8). DNN has exhibited a state-of-the-art outcome, particularly in the problem pertaining to captioning of images. Many alternative ways (7) to describe a picture are being explored as a result of the recent spike in research interest in image captioning utilising DNNs. Image captioning is defined in Vinyals et al.(9) encoder-decoder framework as estimating the likelihood of a phrase based on an input image feature.

भूरा और सफेद कुत्ता बर्फ से गुजर रहा है।
कुत्ता बर्फ में दौड़ रहा है
कुत्ता जो बर्फ से गुजर रहा है।
सफेद और भूरे रंग का कुत्ता बर्फ से ढंके मैदान से गुजर रहा है।
सफेद और भूरे रंग का कुत्ता बर्फ की सतह पर चल रहा है।

Figure 1: A sample image and its reference captions

3. METHODOLOGY

The proposed methodology in this research has a parallel architecture which consists of two pipelines. The first pipeline takes in the image data and generates features out of it. We have utilized multiple pre – trained well known CNN models viz VGG16, Resnet50. The second pipeline takes in processed reference captions and feeds them into an LSTM. The outputs from both the pipelines are concatenated together to generate captions for the image using a combination of dense layers consisting of 256 neurons. The architecture of the model is explained in fig 2.

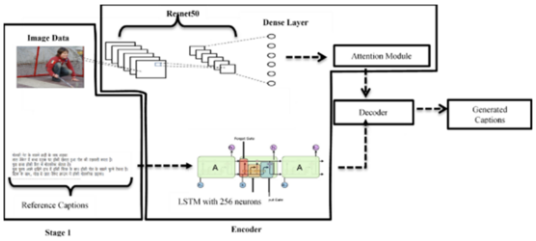


Fig. 2: Architecture of Proposed model

Before sending the reference captions through the pipeline, text data is pre-processed as follows:

3.1. Text Wrapping

For the pre-processing purposes, the training dataset has been wrapped in between two tags: <startseq> and <endseq>. This is done because the model predicts the captions based on image features and previous words sequences. So the tags are just an indicator for the model to stop because as soon as it predicts the <endseq>. It will stop predicting new words.

3.2. Text Tokenization

We have done text tokenization and padding, as deep learning algorithms cannot understand text, so we replace text by floating point values and they get processed by the deep learning algorithm (RNN), in our case LSTM. The pre-processed captions were fed into the LSTM model which were then concatenated with features generated from CNN models.

To further enhance the results of the model, attention module was introduced into the proposed architecture of the image captioning model. Various different versions of attention model have been tried out in the past(10). created a very efficiency weight-based model but the model started missing out on small parts of the image(11) introduced an attention module which focused on specific actions in the image.(12) came up with an adaptive attention module which switches between two different forms of attention - visual attention and language model. It automatically switches between the two methods and relies on combined result of both of them.

For feature extraction, we have used two pre-trained CNN models namely VGG16 and Resnet50 which are explained below.

3.3. VGG 16

Simonyan et al(13) built the VGG16 convolutional neural network (CNN) model in 2014. The authors after performing experimentation realised that the CNN was performing better as they increased the layers and they achieved appreciable results when the layer count reached 16.

Their conclusions were based on their entry to the ImageNet competition. The model had an appreciable test accuracy of 92.7 percent. Stride 2 MaxPool layers are used to supplement the convolutional layers. These layers are then connected to FC (fully-connected) layers that process and classify the data from the convolved output, using the SoftMax function for activation.

3.4. Resnet 50

The term "residual network architecture" refers to the architecture of residual networks. In 2012, the residual learning technique was highlighted using this model at the ILSVRC2012 classification challenge.

Instead of connecting layers and learning features, residual learning involves the learning of residuals, which is accomplished through the use of shortcut connections. That is, the nth layer's input is connected to the (n+x) th layer's input.

In parallel to feature extraction, the reference captions are processed using LSTM, explained below

3.5. LSTM (Long Short-Term Memory)

Recurrent neural networks (RNNs) are neural networks that repeat themselves. RNNs have a difficulty with long-term dependencies, which LSTM networks were designed to solve.

LSTMs differ from more standard feedforward neural networks in that they have feedback connections. This property allows LSTMs to handle whole data sequences (such as time series) without having to deal with each point individually.

3.6. Attention Mechanism

The attention mechanism was proposed to address the performance bottleneck of conventional encoder-decoder architectures, achieving significant improvements over the conventional approach.

The goal of this study is to create Hindi captions for images. There has been very less research on Hindi language and no research has been done on attention model along with encoder decoded framework. As a result, this work investigates implementing an encoder-decoder model in which picture features are encoded using pretrained CNN and text features are encoded using LSTM-RNN.

4. EXPERIMENTAL WORK

For the training and testing purpose, Flickr 8k Dataset has been used. It consists of 8000 different images along with 5 different captions for each images describing the events in the image in detail. The dataset contains all the captions in the original Flickr8k dataset but in our regional language i.e., Hindi. The Dataset has a split of 6000:1000:1000 for training, validation and testing purposes.

For checking the efficiency of the models, we have used Bleu scores.

We have compared multiple CNNs with the presence and absence of attention module to evaluate and compare the performance of different models constructed. The results obtained were at par and in some models even better than the existing work in English language of Image captioning.

BLEU or the Bilingual Evaluation Understudy(15), is a metric for checking the similarities between the predicted text using machine learning models and the originally classified text captions used as training labels. It is a score between 0 and 1. A perfect score of 1.0 indicates a human like translation and a score of 0 indicates a very poor translation. The more matches, the better the candidate translation is.

The weights are specified in a data structure where each value refers to the contribution of that index in overall text. You can give a weight of (1,0,0,0) for 1-gram matches and divide the weights correspondingly for 2,3,4-gram matches.

Table 1. Performance of the Proposed work on Flickr8K dataset

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
VGG16	0.425	0.183	0.207	0.246
Resnet - 50	0.461	0.192	0.212	0.298
VGG16 with attention	**0.564**	0.203	0.290	0.321
Resnet50 with attention	0.556	**0.219**	**0.304**	**0.335**

The weights for BLEU-1 are (1,0,0,0), for BLEU-2 are (0.5,0.5,0,0), for BLEU-3 are (0.33,0.33,0.33,0.33) and for BLEU-4 are (0.25,0.25,0.25,0.25) respectively.

The performance of the proposed work on Flickr8K dataset is provided in Table 1. It is observed that attention based deep features provided better BLEU scores than VGG-16 and ResNet-50 deep features. From the experiments it is also observed that ResNet50 based captioning model with attention layer performs best with highest BLEU score. A comparison with the existing works has also been shown in Table 2.

Table 2. Comparison with existing work

Model	Dataset	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Vinyals et al. (9)	Flickr8k	0.630	0.424	0.270	-
A. Rathi (3)	Flickr8k in Hindi	0.584	0.4	0.27	0.12
Our Model	Flickr8k in Hindi	0.556	0.219	**0.304**	**0.335**

Predicted captions along with actual captions are shown below.

5. CONCLUSION

The paper utilizes the flickr8k dataset which is a very reputed dataset in the field of image captioning and uses a Hindi version of the same which contains a Hindi translated version of English captions. The conversion cannot be 100% accurate as English and Hindi languages are fundamentally different in their grammar and vocabulary. The model trains on the given captions and generates results accordingly. It can be seen that the resnet50 with attention module generates the best results. The resnet50 is a clear winner when it comes to transfer learning. The results from this paper will contribute to the existing literature of image captioning in Hindi language and as we have seen that accurate image captioning has tremendous applications both in real life and in future research works. As English to Hindi conversion was done by a machine in Flickr8k. Doing so with the help of someone who speaks Hindi natively would definitely increase the results. Using a bigger Dataset so that model has more data points to train on.

(a) Real caption - तीन लोग छतरी के साथ बाहरी पिकनिक बेंच पर बैठे हैं।
Predicted - आदमी लोग अपने आदमी के साथ बाहर बैठा।

Real caption - काले बालों वाली लड़कियां भीड़ में हैं।
Predicted - महिला और महिला के साथ समूह में।

(b) Real caption - बाड़ के सामने कुत्ते बर्फ में हैं।
Predicted - कुत्ता कुत्ते के साथ बर्फ में।

Real caption - आदमी और महिला सजावटी वेशभूषा पहने और दर्शकों की भीड़ में नाच रहे थे।
Predicted - लड़का पानी पर अपने आदमी के साथ पानी में।

(c) Real caption - भूरे रंग का कुत्ता बाधा पर कूदता है।
Predicted - भूरे रंग का कुत्ता लाल और सफेद छड़ी पर।

Real caption - छोटी का हाथ पकड़े हुए महिला उसे दिखाने की कोशिश कर रही है।
Predicted - महिला छोटी लड़की का हाथ पकड़े है।

(d) Real caption - कुत्ते घास में नीले फ्रिसबी को पकड़ रहे हैं।
Predicted - फ्रिसबी के लिए कूदता है।

Real caption - बच्चों का झुंड स्कूल के खेल मैदान पर चलता है।
Predicted - लोग कई लोग मैदान में खेलते हैं।

Fig. 3: Sample results observed in Hindi for Flickr8K dataset using (a) VGG-16 + LSTM b) ResNet-50 + LSTM c) VGG-16+Attention+LSTM d) ResNet-50+Attention+LSTM

References

[1] Nezami OM, Dras M, Wan S, Paris C. Senti-Attend: Image Captioning using Sentiment and Attention. 2018; Available from: http://arxiv.org/abs/1811.09789

[2] Tova-izquierdo M. Deep learning approach for Image captioning in Hindi language Ankit Rathi National College of Ireland Supervisor :

[3] Rathi A. Deep learning apporach for image captioning in Hindi language. In: 2020 International Conference on Computer, Electrical Communication Engineering (ICCECE). 2020. p. 1–8.

[4] Gerber R, Nagel HH. Knowledge representation for the generation of quantified natural language descriptions of vehicle traffic in image sequences. IEEE International Conference on Image Processing. 1996;2:805–8.

[5] Farhadi A, Hejrati M, Sadeghi MA, Young P, Rashtchian C, Hockenmaier J, et al. Every picture tells a story: Generating sentences from images. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Infile:///D:/Abhishek/College/8th Sem/Major Project/New Paper/181109789.pdftelligence and Lecture Notes in Bioinformatics). 2010;6314 LNCS(PART 4):15–29.

[6] Li S, Kulkarni G, Berg TL, Berg AC, Choi Y. Composing simple image descriptions using web-scale N-grams. CoNLL 2011 - Fifteenth Conference on Computational Natural Language Learning, Proceedings of the Conference. 2011;220–8.

[7] Bai S, An S. A survey on automatic image caption generation. Neurocomputing. 2018;311:291–304.

[8] Zakir Hossain MD, Sohel F, Shiratuddin MF, Laga H. A comprehensive survey of deep learning for image captioning. ACM Computing Surveys. 2019;51(6).

[9] Vinyals O, Toshev A, Bengio S, Erhan D. Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(4):652–63.

[10] Xu K, Ba JL, Kiros R, Cho K, Courville A, Salakhutdinov R, et al. Show, attend and tell: Neural image caption generation with visual attention. 32nd International Conference on Machine Learning, ICML 2015. 2015;3:2048–57.

[11] You Q, Jin H, Wang Z, Fang C, Luo J. Image captioning with semantic attention. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016;2016-Decem:4651–9.

[12] Lu J, Xiong C, Parikh D, Socher R. Knowing when to look: Adaptive attention via a visual sentinel for image captioning. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017;2017-Janua:3242–50.

[13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. 2015;1–14.

[14] Jain A, Sethi A, Vishwakarma DK, Jain A. Ensembled Neural Network for Static Hand Gesture Recognition. In: 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021. 2021.

[15] Papineni K, Roukos S, Ward T, Zhu WJ. {B}leu: a Method for Automatic Evaluation of Machine Translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics [Internet]. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics; 2002. p. 311–8. Available from: https://aclanthology.org/P02-1040

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT).
India Habitat Centre, Lodhi Road, New Delhi, India. Sep 23-25, 2022

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT) | 978-1-6654-9365-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBCONPT57482.2022.9938303

Implementation of Secure Communication System Using Chaotic Masking

S Suchit
Dept. of Electronics and Communications Engineering
Delhi Technological University
Delhi, India
suchit981@gmail.com

Kriti Suneja
Dept. of Electronics and Communications Engineering
Delhi Technological University
Delhi, India
kritisuneja@dtu.ac.in

***Abstract*—In this paper, we have described our effort to develop a secure means of communication with the help of the Lorenz system. The design of the transmitter and the receiver, as well as the synchronization of the chaotic systems at both the ends have been explained in detail. The Lorenz system has been realized with the help of Runge Kutta's 4th Order method in the form of a Finite State Machine (FSM). These discussions are accompanied with results obtained from simulations carried out on Xilinx Vivado using Verilog, which validates our approach to the problem.**

Keywords—Lorenz system, synchronization, Runge Kutta's 4th Order method, Finite State Machine, Verilog

I. Introduction

Chaos theory was first introduced in 1963, as a study of random but deterministic behaviour of systems [1]. Since then, various chaotic systems like Lorenz, Rössler and Sprott chaotic systems have been studied [2-4]. Hardware implementations of such systems, both digital and analog [5-7], have been explored for their real-world applications, especially in communication [8]. However, due to the dependency of the analog circuits on age and temperature, digital realization has been preferred for its robustness [7].

The interests in the applications of chaotic systems have been on the rise due to some of their inherent characteristics, such as non-periodic and irregular nature, and susceptance to initial conditions. These characteristics are especially helpful in developing a secure means of communication [9]. The studies on this possible use of the chaotic systems were mainly encouraged by the idea that they can be synchronized [10]. Consequently, various synchronization and modulation schemes have been introduced and investigated since then [11].

This paper proposes a secure means of communication utilizing the Lorenz chaotic system in the configuration of Complete Replacement Synchronization [12], combined with the modulation technique called Chaotic Masking [13]. Here the Lorenz system generator has been designed using a Finite State Machine (FSM) to solve the differential equations. The digital realization of the system has been executed on Xilinx Vivado using Verilog Hardware Description Language (HDL).

The paper has been divided in sections in the following way: section II describes the Lorenz system, section III elucidates how such chaotic systems can be synchronized, section IV explains Chaotic masking for a system, section V describes the proposed design that combines the aforementioned ideas, the results for which we discuss and analyse in section VI.

II. Lorenz System

Lorenz system is one of the first chaotic systems introduced, which behaves as a chaotic attractor for certain values of the parameters. It's self-synchronization property lends itself to its potential use in private communications [14].

The chaotic system is represented by the following differential equations:

$$\frac{dx}{dt} = a(y - x) \tag{1a}$$

$$\frac{dy}{dt} = x(c - z) - y \tag{1b}$$

$$\frac{dz}{dt} = xy - cz \tag{1c}$$

Where x, y, z are state variables and a, b, c are parameters. The Lorenz system behaves as a chaotic attractor for the values a = 10, b = 28, c = 8/3.

III. Synchronization of Chaotic Systems

The successful encryption and decryption of a message at the transmitter and the receiver can only be achieved when the keys for encryption and decryption, i.e., the chaotic samples generated by the chaotic systems are identical.

Because of their sensitivity to initial conditions, at first it seems implausible for two chaotic systems with different initial conditions to converge, or synchronize over time. This issue is taken care of with the chaotic system in a master-slave configuration, where the master is a 3-variable system and the slave, a 2-variable system. Here, since the chaotic systems are supposed to be 3-variable systems, the variable in the master corresponding to the missing variable in the slave will help fulfil this condition by replacing it in any occurrences in the slave system.

This method of synchronization is called the complete replacement synchronization, where by only receiving partial information about the master (driver system), the slave (response system) now almost completely replicates the behaviour of the same [12].

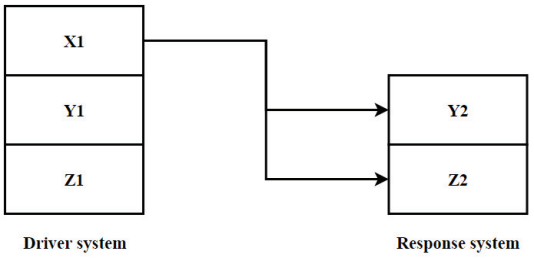


Fig. 1. Representation of two synchronized chaotic systems

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:12:29 UTC from IEEE Xplore. Restrictions apply.

Fig. 1 demonstrates the synchronization between two chaotic systems with the state variables X1, Y1, Z1 and Y2, Z2 for the drive and the response system respectively. Here, X1 is the drive signal. The differential equations for the response system can simply be written by only replacing X2 with X1.

We say that both the chaotic systems are synchronized when:

$$\lim_{n\to\infty} |Y1n - Y2n| = 0 \quad (2a)$$

$$\lim_{n\to\infty} |Z1n - Z2n| = 0 \quad (2b)$$

Where X1n, Y1n, Z1n and Y2n, Z2n are the nth samples generated by driver and response systems respectively.

IV. Chaotic Masking

There are several methods that can be employed for encryption of a message for establishing a reliable and safe means of communication. Chaotic masking, an analog modulation method is a simple way of doing so.

The message is encoded through arithmetic addition with the sample generated by the chaotic system at the transmitter. Similarly, we decode the message by arithmetic subtraction of the chaotic sample from the received signal [13]. Here, the synchronization between the chaotic systems at the transmitter and the receiver plays an essential role for the decoded information to be faithful to the original message. Without the synchronization of these systems, chaotic samples generated at both of these will not be identical, resulting in the decoded information to be different than the original message and hence, the communication to be unsuccessful.

Fig. 2 demonstrates the chaotic masking of a message signal *i* with a chaotic signal *x* generated by a chaotic system [12]. The transmitted signal when decoded at the slave side through arithmetic subtraction of chaotic signal, gives the signal *i'*.

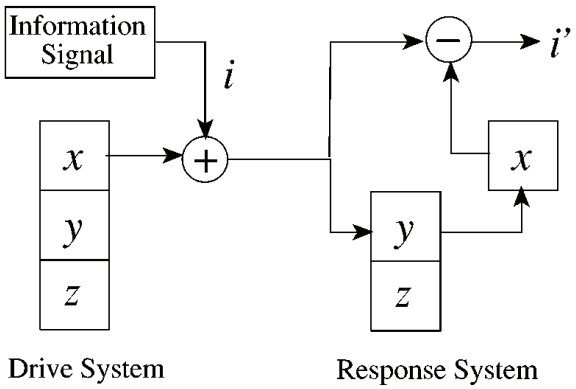


Fig. 2. Chaotic Masking of a message signal *i* using chaotic signal *x*

V. Proposed Design

The digital realization of our system involves first designing the Lorenz generator, a digital circuit whose output behaves as a Lorenz system when observed over time. We then develop the architectures of the transmitter and the receiver.

The following discussions also involve the design rationales required for the implementation of Complete Replacement Synchronization and Chaotic masking in our system.

A. Lorenz generator

To obtain the chaotic samples of the Lorenz system at different instants, we are required to solve the differential equations in (1). For this purpose, we may implement Euler's, Heun's or Runge - Kutta's 4th order method for solving a set of differential equations. We will be using the Runge - Kutta's 4th order method as it provides more accurate results, even though it requires relatively more mathematical calculations [5].

For the implementation of the Runge Kutta's 4th Order method, we have the following equations:

$$K_{X1} = F_x(X_n, Y_n, Z_n) \quad (3a)$$

$$K_{X2} = F_x(X_n + 0.5hK_{X1}, Y_n + 0.5hK_{Y1}, Z_n + 0.5hK_{Z1}) \quad (3b)$$

$$K_{X3} = F_x(X_n + 0.5hK_{X2}, Y_n + 0.5hK_{Y2}, Z_n + 0.5hK_{Z2}) \quad (3c)$$

$$K_{X4} = F_x(X_n + hK_{X3}, Y_n + hK_{Y3}, Z_n + hK_{Z3}) \quad (3d)$$

$$X_{n+1} = X_n + h(K_{X1} + 2K_{X2} + 2K_{X3} + K_{X4})/6 \quad (3e)$$

Where X, Y, Z are state variables, $F_x(X, Y, Z)$ is the derivative of the state variable X, h the step size, i.e., the interval between consecutive samples of the chaotic system and K_{X1}, K_{X2}, K_{X3}, K_{X4} the intermediate slopes.

Here, we generate the (n+1)th chaotic sample, i.e., X_{n+1} using nth chaotic samples X_n, Y_n and Z_n. Similarly, the above equations can be replicated for the state variables Y and Z as well.

To implement this method in Verilog HDL, we define the Lorenz generator both at the transmitter as well as the receiver as a Finite State Machine (FSM). This FSM consists of the states S0, S1, S2, S3, S4 and S5. At each of these states, we perform the following operations:

S0: Set initial conditions for the state variables
S1: Calculate the intermediate slopes K_{X1}, K_{Y1}, K_{Z1}
S2: Calculate the intermediate slopes K_{X2}, K_{Y2}, K_{Z2}
S3: Calculate the intermediate slopes K_{X3}, K_{Y3}, K_{Z3}
S4: Calculate the intermediate slopes K_{X4}, K_{Y4}, K_{Z4}
S5: Solve for solutions X, Y, Z

We first need to set the initial conditions for the state variables. So, initially, the FSM is at state S0. The FSM then constantly cycles through the states S1 to S5. We calculate the intermediate slopes when the FSM is at the states S1 to S4. Whenever state S5 is reached, the next chaotic sample is generated.

Fig. 3 shows the state diagram for the Lorenz generator described in this section, where, the computations of (n+1)th samples, i.e., the chaotic samples Y_{n+1} and Z_{n+1} have also been implemented using the derivatives F_y and F_z, the intermediate slopes K_{Y1}, K_{Y2}, K_{Y3} K_{Y4} and K_{Z1}, K_{Z2}, K_{Z3}, K_{Z4}, and Y_n and Z_n being the nth samples, respectively.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:12:29 UTC from IEEE Xplore. Restrictions apply.

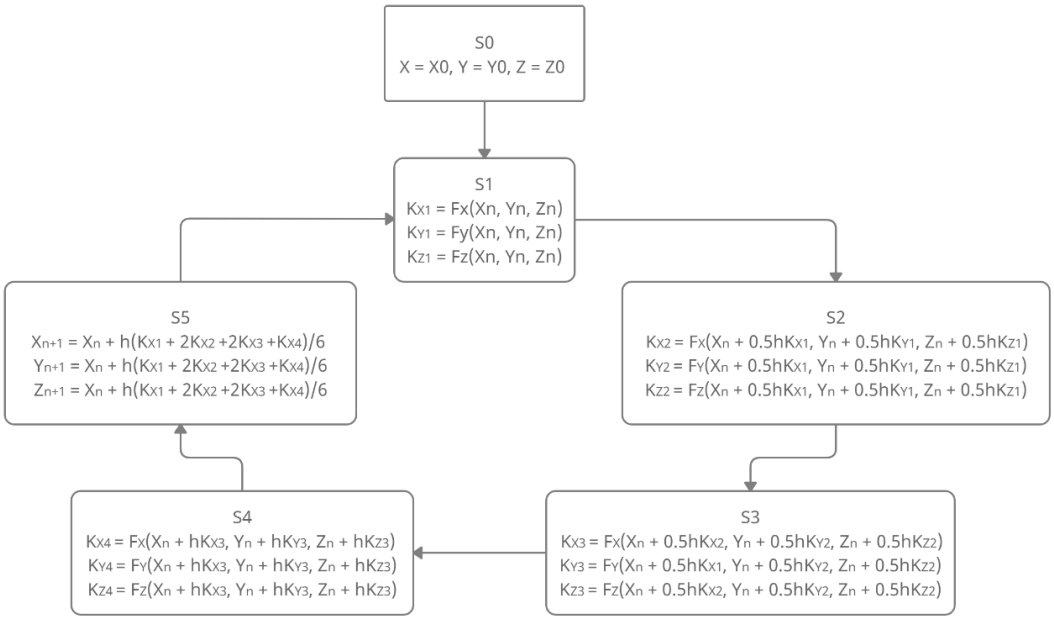


Fig. 3. State Diagram for the Lorenz Generator, where X = X0, Y = Y0 and Z = Z0 are the intial conditions

B. *Transmitter*

The Transmitter transmits the message signal encrypted using a chaotic signal generated from the Lorenz generator. A sine wave generator was designed using a Look-Up Table (LUT) to generate the message signal, initiating the communication.

Since the Lorenz generator here also behaves as a driver system, partial information of the system in the form of the drive signal is transmitted to the receiver as well.

The sine wave signal, generated by looking up a sample of the sine wave in the LUT at each clock pulse, is encrypted using the chaotic masking technique and hence is arithmetically added to the samples generated by the Lorenz system. The masked signal can then be transmitted via a communication channel.

C. *Receiver*

The Receiver decrypts the incoming signal from the communication channel using the chaotic samples generated by the Lorenz generator. The decrypted message is obtained from the arithmetic difference between the samples of the incoming signal and the samples generated by the Lorenz generator, corresponding to the one at the transmitter.

The receiver behaves as the response system, so it is a two-variable system. The missing variable is compensated for by the partial information received from the driver system, and helps in establishing synchronization between the driver and the response system.

VI. Results and Analysis

The system has been completely designed on Xilinx Vivado using Verilog HDL, where, the samples of the sine wave, chaotic or the transmitted signal have been encoded in 32 bits using fixed point data representation (16Q16). The waveforms for the various aspects of the design discussed in the previous sections have been obtained through simulations on the software. A 20MHz clock speed has been provided to the system in these simulations.

To first examine the working of the Lorenz generator, it was tested for the condition when a Lorenz system behaves

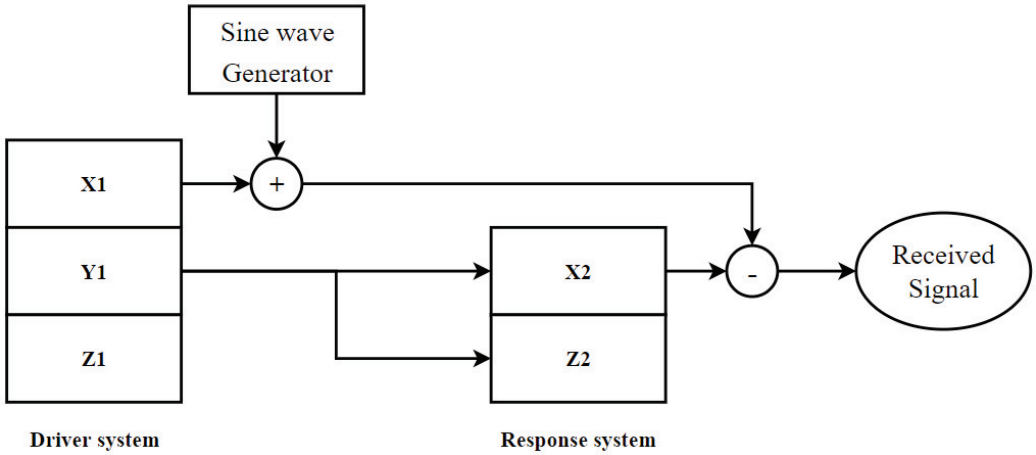


Fig. 4. Representation of the Signal Masking technique, used for communication between two synchronized chaotic systems, Y1 being the drive signal

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:12:29 UTC from IEEE Xplore. Restrictions apply.

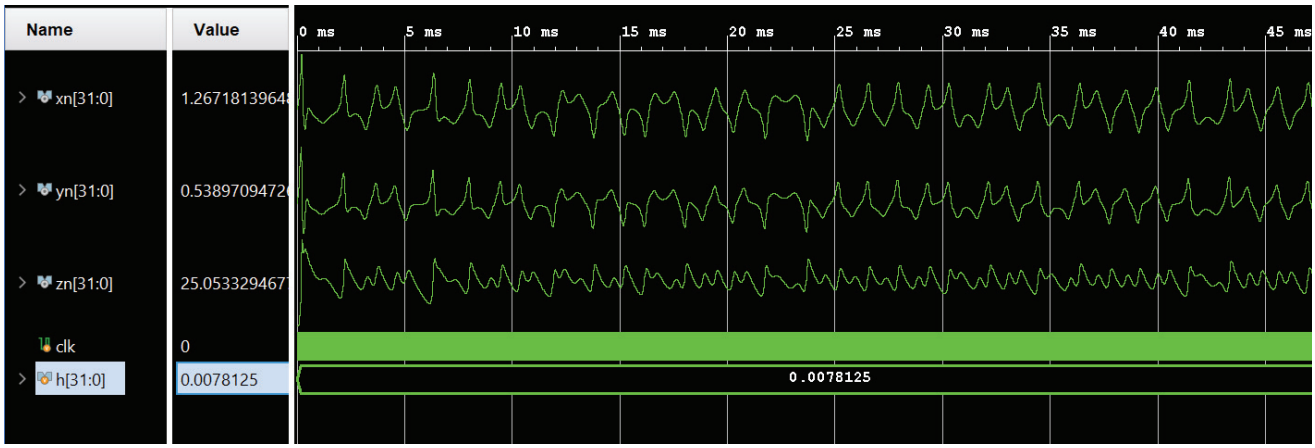


Fig. 5. Output waveforms obtained from the Lorenz Generator, for the step size h = 2^{-7} , and parameters a = 10, b = 28, c = 8/3

as a chaotic attractor, the initial conditions of the state variables being X = 0, Y = 2, Z = 0 and the parameters a = 10, b = 28, c = 8/3.

We have obtained the waveforms of the chaotic signals generated by the Lorenz generator, digitally realized using the FSM with the state diagram described in Fig. 3, applying the Runge Kutta's 4th order method. The results of the simulation in Xilinx Vivado have been shown in Fig. 5. A step size of 2^{-7} or 0.0078125 has been taken for the purpose of simulation.

Now, the synchronization of two such Lorenz generators has been achieved by using the chaotic signal generated through Y1 of the driver system as the drive signal. This means that all the occurrences of the Y2 in response system will be replaced by Y1.

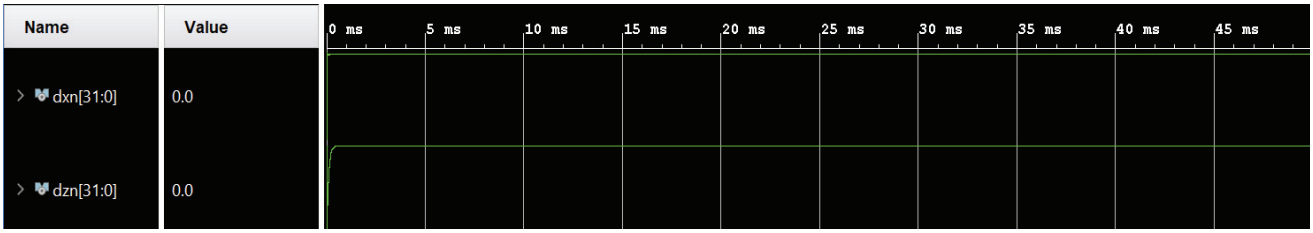


Fig. 6. Waveform obtained for difference between the state variables decreasing to zero, indicating synchronization

The resulting synchronization between the two systems is evident from the waveforms obtained in Fig. 6. Here, the differences *dxn* and *dzn* represent the differences between the chaotic signals generated by solving equations (1a) and (1c) for the driver and the response systems respectively. From (2), we know that when synchronization is achieved, the differences between the corresponding samples of the chaotic signals become zero over time which is also evident from these waveforms, validating our results.

While simulating the communication between the transmitter and the receiver, we generate a sine wave signal and transmit it to the receiver after encryption as mentioned in section V, which then gets decrypted to obtain the masked information.

The chaotic signal generated through the state variables X1 and X2 of the two systems are used for the encryption and decryption of the information.

The original, transmitted and the decoded signal are compared in Fig. 7, indicating the security provided by this form of communication through the chaotic masking technique as well as the fidelity of the decrypted information to the original signal.

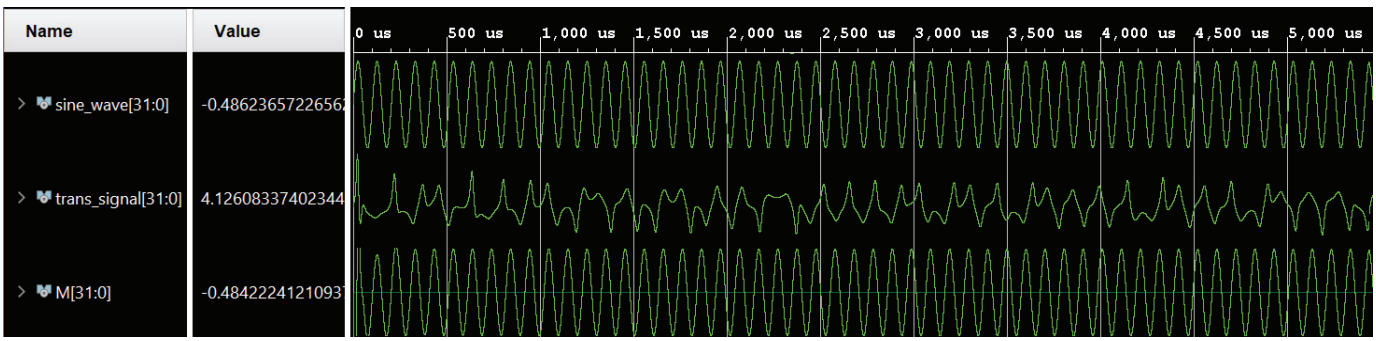


Fig. 7. Comparison of original, transmitted and decoded signal

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:12:29 UTC from IEEE Xplore. Restrictions apply.

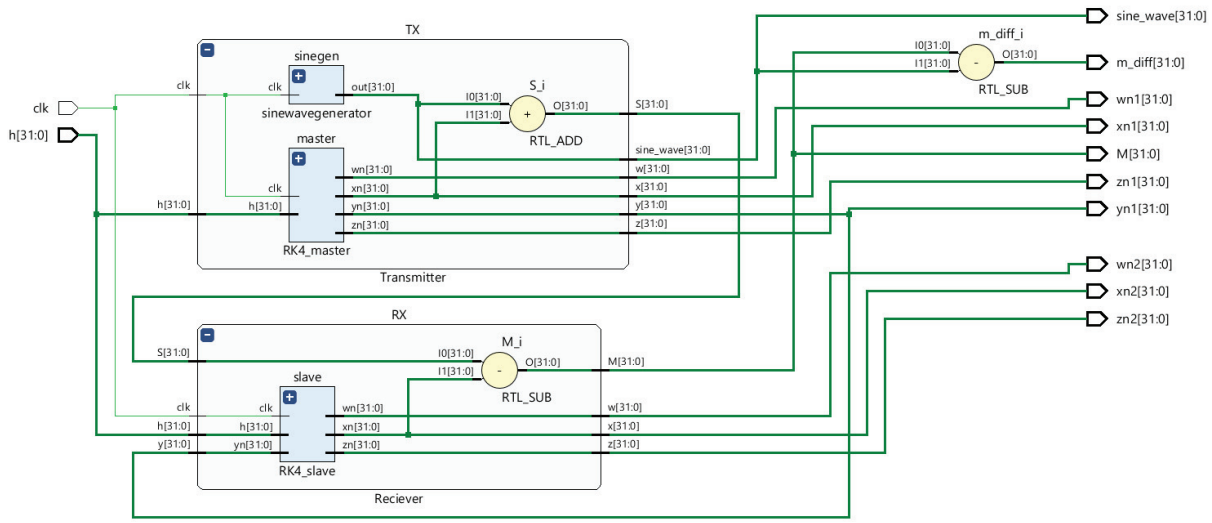


Fig. 8. RTL Design of the system

Initially, the received signal is unidentical to the test signal as the Lorenz generators are not synchronized yet, whereas, the latter section of the signal is completely identical to the original signal, indicating the successful synchronization, and hence the communication between the transmitter and the receiver.

The RTL design in Fig. 8, obtained based on the Verilog Code written in Xilinx Vivado gives us a visual of the overall configuration of the system, and clearly shows a transmitter and a receiver consisting of the required digital circuits as in the proposed design discussed in section V.

VII. Conclusion

The chaotic masking technique along with the implementation of complete replacement synchronization method provide us a simple, yet robust means of secure communication. In the process, we have also successfully performed the digital realization of the Lorenz system through a method which can be easily modified to form the generators for various other chaotic systems that may be of use according to the requirements of the communication. These claims have been validated by our experiments with the system, performed on Xilinx Vivado using Verilog HDL.

Although, the security aspects of this form of communication maybe questionable, the issues can be reduced to a minimum as has been suggested in [11]. For better security, we may implement the various other modulation techniques as one may find to be suitable.

References

[1] Lorenz, N. Edward “Deterministic Nonperiodic Flow". Journal of Atmospheric Sciences 20.2 (1963): 130-141.

[2] O. E. Rossler, An equation for continuous chaos, Phys. Lett. A 57 (1976) 397-398.

[3] J. C. Sprott, Some simple chaotic flows, Phys. Rev. E 50 (1994) 647-650.

[4] İ. Pehlivan and Y. Uyaroglu, A new chaotic attractor from general Lorenz system family and its electronic experimental implementation, Turk. J. Elec. Eng. Comp. Sci. 18 (2010) 171-184

[5] A. Garg, B. Yadav, K. Sahu, and K. Suneja , "An FPGA based Realtime Implementation of Nosé hoover Chaotic System using different numerical Techniques," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 108-113, doi10.1109/ICACCS51430.2021.9441923.

[6] S. Çiçek, A. Ferikoğlu and İ. Pehlivan, "Electronic circuit design of Sprott H chaotic system with CCII+," 2014 22nd Signal Processing and Communications Applications Conference (SIU), 2014, pp. 2015-2018, doi: 10.1109/SIU.2014.6830654.

[7] M. A. Aseeri, M. I. Sobhy and P. Lee, "Lorenz chaotic model using Filed Programmable Gate Array (FPGA)," The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002., 2002, pp. I-527, doi: 10.1109/MWSCAS.2002.1187274.

[8] W. Ji, "Chaos and Control of Nonlinear Dynamic Model in the Ecological System," 2010 International Workshop on Chaos-Fractal Theories and Applications, 2010, pp. 326-330, doi: 10.1109/IWCFTA.2010.41.

[9] A. Abel and W. Schwarz, "Chaos communications-principles, schemes, and system analysis," in Proceedings of the IEEE, vol. 90, no. 5, pp. 691-710, May 2002, doi: 10.1109/JPROC.2002.1015002.

[10] L. M. Pecora and T. L. Carroll, ‘‘Synchronization in chaotic systems,’’ Phys. Rev. Lett., vol. 64, pp. 821–823, Feb. 1990

[11] G. Kaddoum, "Wireless chaos-based communication systems: A comprehensive survey," in IEEE Access, vol. 4, pp. 2621-2648, 2016, doi: 10.1109/ACCESS.2016.2572730.

[12] L. M. Pecora, T. L. Carroll, G. Johnson, and D. Mar, “Fundamentals of synchronization in chaotic systems, concepts,and applications.” Chaos., vol. 7(4), pp. 520–543, 1997.

[13] K. M. Cuomo and A. V. Oppenheim, ``Circuit implementation of synchronized chaos with applications to communications," *Phys. Rev. Lett.*, vol. 71, pp. 65_68, Jul. 1993. Signal masking

[14] K. M. Cuomo, A. V. Oppenheim and S. H. Strogatz, "Synchronization of Lorenz-based chaotic circuits with applications to communications," in IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 40, no. 10, pp. 626-633, Oct. 1993, doi: 10.1109/82.246163.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:12:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

Improving the Scalability of GPU Synchronization Primitives

Preyesh Dalmia*, Rohan Mahapatra†, Jeremy Intan‡, Dan Negrut* and Matthew D. Sinclair*§

*University of Wisconsin-Madison, Email: [pdalmia, negrut, msinclair]@wisc.edu
†University of California, San Diego, Email: rohan@ucsd.edu
‡University of Illinois at Urbana-Champaign, Email: jintan2@illinois.edu
§AMD Research

Abstract

General-purpose GPU applications increasingly use synchronization to enforce ordering between many threads accessing shared data. Accordingly, recently there has been a push to establish a common set of GPU synchronization primitives. However, the expressiveness of existing GPU synchronization primitives is limited. In particular the expensive GPU atomics often used to implement fine-grained synchronization make it challenging to implement efficient algorithms. Consequently, as GPU algorithms scale to millions or billions of threads, existing GPU synchronization primitives either scale poorly or suffer from livelock or deadlock issues because of heavy contention between threads accessing shared synchronization objects. We seek to overcome these inefficiencies by designing more efficient, scalable GPU barriers and semaphores. In particular, we show how multi-level sense reversing barriers and priority mechanisms for semaphores can be designed with the GPUs unique processing model in mind to improve performance and scalability of GPU synchronization primitives. Our results show that the proposed designs significantly improve performance compared to state-of-the-art solutions like CUDA Cooperative Groups and optimized CPU-style synchronization algorithms at medium and high contention levels, scale to an order of magnitude more threads, and avoid livelock in these situations unlike prior open source algorithms. Overall, across three modern GPUs the proposed barrier algorithm improves performance by an average of 33% over a GPU tree barrier algorithm and improves performance by an average of 34% over CUDA Cooperative Groups for five full-sized benchmarks at high contention levels; the new semaphore algorithm improves performance by an average of 83% compared to prior GPU semaphores.

1 Introduction

Traditionally GPUs focused on providing high performance for streaming, data parallel workloads with limited data reuse or data sharing and coarse-grained synchronization that usually only happened at kernel boundaries. Thus, these applications traditionally exploited reuse through the GPUs specialized memories including constant cache, scratchpads, and texture caches. However, recently applications with different characteristics have also started taking advantage of the parallelism exposed by the GPUs. For example, recent work in scientific and high performance computing (including machine learning) use GPUs for persistent kernels [14], [36], [109], buddy allocation [23], and particle partitioning [10] while other work fuses GPU kernels [1], [26], [55] and uses concurrent streams [5], [63]. Unlike traditional, streaming, data parallel general-purpose GPU (GPGPU) applications, these next generation GPU applications repeatedly access shared data in global memory across threads from different thread blocks (TBs). Thus, they use device-wide synchronization primitives such as mutex locks, semaphores, and global barriers [28], [89], [92], [94] to avoid data races and ensure correctness.

Listing 1 shows a high-level example of how parallel reductions [34] use global barriers for device-wide synchronization, which we use as an exemplar of how GPU applications use device-wide synchronization. This algorithm (discussed further in Section 5) iteratively reduces data by assigning each TB on each streaming multiprocessor (SM) a subset of the data, which the TB performs a partial reduction on using shared memory (line 4). As the TBs complete their partial reductions, they update the corresponding global output array with the reduced value (line 9). Although each TB reduces an independent portion of the input data, if additional reduction steps remain (line 1), in the next loop iteration TBs will often be operating on data produced from a different TB in the previous iteration (line 17). Thus, to ensure the updated data is visible to all TBs before the next iteration, a global barrier (discussed further in Section 2.2.1) must be performed to ensure all TBs reach this point before subsequent iterations operate on the data (line 14). Since many threads repeatedly use these synchronization primitives, and many of the accesses contend with one another, we refer to this as fine-grained synchronization. Although some prior CPU work classifies global barriers as coarse-grained synchronization [80], we use fine-grained synchronization to distinguish from the standard GPU definition of coarse-grained synchronization, which often refers to synchronizing via an implicit CPU-side global barrier at the end of a kernel [87], [92], [94], [106].

Traditionally, developers relied on complex GPU solutions [27] or utilized CPU-side solutions [98] for providing device-wide synchronization. Prior work has also shown

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508


```
while (additionalReductionSteps) {
    // perform local reduction: read from
    // global memory, write to shared memory
    perTB_sharedMem_reduction(global_inputData,
        sharArr);
    __syncthreads();

    // write TB reduced output to global memory
    if (tid == 0) {
        global_outputData[blockId] = sharArr[0];
    }
    __syncthreads();
    // Ensure all TBs have finished their portion
    // of the reduction
    globalBarrier();

    // Swap pointers for next reduction step
    swap(global_inputData, global_outputData);
}
```

Listing 1: High level example for performing a parallel reduction using explicit device-wide synchronization.

that even simple forms of synchronization can enable programmers to design GPU solutions with better utilization than more complex, synchronization-free algorithms [56], [57]. Unfortunately, fine-grained synchronization is inefficient on modern GPUs. Unlike multi-core CPUs, which have significant OS support and complex coherence protocols like MOESI that make synchronization relatively cheap, GPUs have limited OS support and use simple, software-driven coherence protocols. Although this approach is efficient when synchronization is infrequent, it requires heavyweight actions when synchronization occurs to ensure correctness [3], [25], [30], [39], [87], [88], [90]. To partially reduce this overhead, GPU memory consistency models utilize scoped synchronization, which allows programmers to specify the level at which threads need to synchronize [22], [32], [33], [43], [47]. In a scope-based memory consistency model, if the synchronizing threads belong to the same TB, then the synchronization occurs locally with reduced overhead. However, if the synchronizing threads are in different TBs, which is often the case for these workloads, expensive global synchronization is still required. We discuss GPU coherence and consistency further in Section 2.

To reduce atomic overhead, hardware vendors recently added better atomic support to improve their performance by an order of magnitude [4], [61], [65], [66]. However, this support only accelerates the atomic operations themselves, not the heavyweight operations that accompany them. Consequently, fine-grained synchronization is still prohibitively expensive, and is often a bottleneck for emerging workloads that utilize it (discussed further in Section 5). Additionally, synchronization overheads are exacerbated by the level of parallelism on GPUs: GPUs typically run kernels with millions or billions of threads. Thus, synchronization variables are heavily contended (even after common optimizations such as using one master thread per TB to perform the synchronization). Accordingly, it is important to improve support for fine-grained GPU synchronization, especially at high contention levels.

Recent work also improved software synchronization support. For example, NVIDIA proposed CUDA Cooperative Groups (CCG) [28], a hierarchy of synchronization methods with support ranging from small groups of threads to multi-GPU devices. CCG is tightly integrated with CUDA and performs well, especially at low contention levels. However, as we show in Section 6, CCG suffers from high contention for shared synchronization variables as the number of threads the synchronization is performed across increases [108]. NVIDIA also introduced libcu++, which provides fully compatible C++ synchronization support [64]. Researchers have also developed device-wide software barriers, which either use global counters with atomic operations, lock-free synchronization [106], or portable atomics to implement GPU barriers [92]. However, these approaches limit the number of TBs that can be launched on an SM to avoid deadlock. Another popular open source device-wide barrier is HeteroSync's tree barrier [96]. However, HeteroSync uses two consecutive barriers in order to properly support context switching, which significantly increases global atomics to update the shared counter variables [89]. We discuss related work further in Section 3.

However, these existing techniques are inefficient as contention increases. Thus we propose optimizations for both global barriers and semaphores. First, we design a multi-level sense-reversing tree barrier (SRB). Although tree barriers and SRBs are widely used in CPUs [31], [96], in Section 6 we show that naively applying CPU SRB concepts to GPUs performs poorly. Instead, like prior work on Xeon Phi's [82], GPU SRB algorithms must be designed to fit the GPU's processing model in order to successfully meet scalability demands. Accordingly, our multi-level SRB naturally conforms to the GPU memory hierarchy and scoped consistency model. As a result, most threads use cheap, locally scoped atomics, improving scalability. Second, we optimize GPU reader-writer semaphores to remove scalability bottlenecks. For example, when a writer attempts to exit the semaphore and multiple readers attempt to enter it simultaneously, the contention causes current semaphores to scale poorly and frequently livelock. Thus, we propose a priority mechanism that favors TBs exiting the semaphore. Overall, we make the following contributions:

1) We propose a two-level GPU SRB that reduces atomic transactions required by 50% compared to HeteroSync's two-level tree barrier.
2) We propose a priority mechanism for GPU semaphores, which reduces the time spent on synchronization and eliminates livelock compared to the HeteroSync's reader-writer semaphores, thereby improving scalability.
3) We first evaluate our optimized synchronization primitives across several microbenchmarks. Compared to CPU-style SRBs and HeteroSync, our work improves average performance by 34% on Volta, 38% on Turing, and 27% on Ampere GPUs for barriers, and by 89% for Volta, 68% for Turing, and 90% for Ampere GPUs for semaphores. Moreover, for five larger benchmarks that utilize global barriers on average G-SRB improves performance by 36% on Volta, 34% on Turing, and 32% on Ampere at medium and high contention levels ($\geq$ 8 threads joining the barrier per SM, where each thread joining the barrier is a master thread per TB) compared to CUDA's CCG. G-SRB also outperforms an op-

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

timized CPU-style SRB by 12% on Volta, 7% on Turing, and 8% on Ampere for medium and high contention levels, showing the need for intelligently designing SRBs for GPUs – especially as contention increases.

4) We also propose a hybrid GPU barrier that combines the best of CCG and G-SRB to deliver high performance at all levels of contention.[1]

2 Background

2.1 GPU Coherence & Consistency

Traditional GPGPU applications are highly parallel and use synchronization infrequently (often only at the beginning and end of a kernel), which enables GPUs to use a simple, software-driven coherence protocol [3], [30], [39], [78], [87], [88], [90]. The coherence protocol invalidates the cache on acquires (synchronization reads) while releases (synchronization writes) must wait for all previous writes to complete. Moreover, both acquires and releases must be performed at a memory that is shared by all SMs (usually the L2). Although these accesses are heavyweight, since traditional GPGPU applications synchronized infrequently, the overheads were acceptable.

However, emerging applications require fine-grained synchronization [1], [10], [14], [23], [26], [36], [41], [55], [76], [77], which are often implemented using atomics. Since fine-grained synchronization is extremely expensive, programmers often use scope-based synchronizations [22], [32], [33], [43], [47]. Both AMD and NVIDIA's consistency models extend the sequentially consistent for data-race-free (SC-for-DRF) consistency model that is widely used in CPUs [8], [48] with scoped synchronization. Thus, it is called the sequentially consistent for heterogeneous-race-free (SC-for-HRF) memory consistency model. Although SC-for-HRF has multiple scopes, we focus on the two most widely used scopes: local and global. A locally scoped atomic or fence (e.g., CUDA's `__threadfence_block()`) is only guaranteed to be visible to other threads in the same TB and requires all prior loads and stores in the TB complete before the atomic or fence completes. Similarly, a globally scoped atomic or a globally scoped fence (e.g., CUDA's `__threadfence()`) is visible to all threads across the GPU by invalidating all valid data in local caches (e.g., L1) at the beginning of a synchronization pair (e.g., a load acquire) and waiting for all writes to complete at the end of a synchronization pair (e.g., a store release). Therefore, locally scoped synchronization is significantly cheaper than globally scoped synchronization. Thus, programmers synchronize locally whenever possible. However, global scope is required for inter-TB synchronization.

2.2 GPU Synchronization Primitives

Several recent projects sought to develop a common set of GPU synchronization primitives. For example, Synchronization Primitives (SyncPrims) released a set of mutexes, barriers, and semaphores focused on GPU atomic performance [94]. HeteroSync extended SyncPrims to model memory accesses that required fine-grained synchronization,[2] improve the algorithms, and added both local and globally scoped versions of most algorithms [89]. NVIDIA also introduced CCG [28] and libcu++ [64], which allows kernels to dynamically organize groups of threads across all levels of granularity from a small group of threads in a GPU to a multi-GPU device. These primitives enable new patterns of cooperative parallelism within CUDA, including producer-consumer parallelism. Thus, we next describe HeteroSync's barrier and semaphore and CCG's barrier implementations.

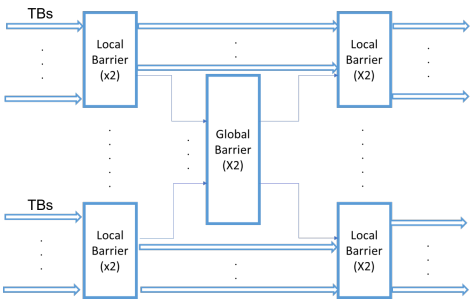


Fig. 1: Baseline Two-Level Tree Barrier [89] where a per-SM leader TB joins the global barrier.

2.2.1 Barriers

Figure 1 illustrates HeteroSync's two-level centralized tree barrier, which uses a hybrid local-global scope. In a centralized barrier, TBs increment a shared counter as they reach the barrier and spin until the counter indicates that all TBs are present. HeteroSync also includes a decentralized, two-level lock-free barrier, which extends a prior decentralized, single-level lock-free barrier [106]. Decentralized barriers trade off increased memory consumption for reduced contention, improved efficiency, and improved scalability. However, unlike our approach (and CCG), they consume significantly more memory. In HeteroSync's implementation all TBs on a Streaming Multiprocessor (SM) access unique data before joining a local barrier (lines 1-2, Figure 2c). In each TB, assuming no races between threads in the same TB, a single leader thread joins the local barrier (line 4, Figure 2a). Next each TB spins (using a per-TB scratchpad variable, `done_local`) until all TBs on the SM join the local barrier (lines 8-11, Figure 2a). Once all TBs have reached the local barrier, a statically designated leader TB from each SM proceeds to join the global barrier (Figure 2b), while all other TBs join a second group of local barriers (line 7, Figure 2c).

However, in order to ensure correctness across potential context switches [2], [70], [99], [102], [103], [104], [107], HeteroSync's tree barrier uses a second barrier (lines 2, 5, 8, Figure 2c), similar to a sense-reversing barrier. To do this, HeteroSync passes unique variables for the counts of the two barriers from Figure 2c to Figure 2a (`localBarr1` and `localBarr2` per SM are passed to `perSMBarr`) and Figure 2b (`globalBarr1` and `globalBarr2` are passed to `globBarr`). Without these unique variables, if a TB i was context switched out after it joined a barrier and not scheduled again until after the barrier completed, it may deadlock since TB i would not be able to distinguish the next barrier from the one it joined previously. Once the expected number of leader TBs reach the global barrier (`numJoin`), these leader TBs join their respective second group of local

1. We will release our code at: https://github.com/hal-uw/.

2. SyncPrims' critical section memory accesses used local scratchpads, and thus did not require synchronization to ensure ordering.

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508


```
1  __shared__ bool done_local = false;
2  numTBsThisSM = gridDim.x / NUM_SM;
3  __threadfence_block();
4  if (threadIdx == (0,0,0)) { // leader thread
5    // perSMBarr = localBarr1 or localBarr2
6    // inc by 1 with wraparound
7    atomicInc(&perSMBarr, 0x7FFFFFFF);
8    while (!done_local) {
9      if (atomicCAS(&perSMBarr,numTBsThisSM,0)
           == numTBsThisSM) {
10       __threadfence_block();
11       done_local = true;
12 } } }
13 __syncthreads(); // other threads wait here
```

(a) Baseline per SM local barrier.

```
1  __shared__ bool done = false;
2  numJoin = (gridDim.x < NUM_SM) ? gridDim.x :
                                    NUM_SM;
3  __threadfence();
4  if (threadIdx == (0,0,0)) { // leader thread
5    // inc by 1 with wraparound
6    atomicInc(&globBarr, 0x7FFFFFFF);
7    while (!done) {
8      // globBarr = globalBarr1 or globalBarr2
9      if (atomicCAS(&globBarr,numJoin,0)==numJoin) {
10       __threadfence();
11       done = true;
12 } } }
13 __syncthreads(); // other threads wait here
14 if (!done) { DoBackoff(); }
```

(b) Baseline GPU global barrier.

```
1  localBarrier(localBarr1[smID]); // Part A
2  localBarrier(localBarr2[smID]); // Part A
3  if (perSMLeaderTB) {
4    globalBarrier(globalBarr1);   // Part B
5    globalBarrier(globalBarr1);   // Part B
6  }
7  localBarrier(localBarr1[smID]); // Part A
8  localBarrier(localBarr1[smID]); // Part A
```

(c) Overall baseline tree barrier.

Fig. 2: Pseudo-code for components of baseline GPU two-level tree barrier [89].

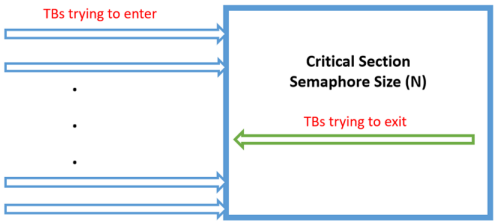


Fig. 3: Baseline Semaphore Implementation [89].

barriers (line 7, Figure 2c, where other TBs from the same SM are spinning). Unfortunately, this approach significantly increases the number of atomics, which adversely impacts performance. Using a tree barrier partially mitigates this overhead by making many of the atomics locally scoped, but tree barriers do not completely mitigate the overhead.

Although CCG is closed source, and thus not all of its implementation details are known, we used NVIDIA's NVBit to disassemble and study the SASS for a Volta `grid_group.sync()` (barrier) [62], [100]. Based on the SASS, CCG appears to use multiple different barrier implementations. Although we could not determine when each barrier is dynamically selected, all of them use a single-level global memory barrier. Specifically, the barrier CCG selected for our configurations (Section 5) is an aggressive, single-level global memory barrier, similar to open source barrier implementations [92], [94]. Like HeteroSync, each TB elects a leader thread and the remaining threads spin waiting for the barrier to complete. CCG also uses a sense direction counter that exploits integer overflow to avoid requiring a second barrier to ensure correctness in the presence of context switches. Interestingly, CCG does not appear to perform any backoff,[3] likely because it may increase latency at low contention levels. However, this approach struggles when contention increases (e.g., as the number of threads or TBs joining a barrier increase) [108].

2.2.2 Semaphores

Figure 3 provides a high-level overview of HeteroSync's reader-writer semaphores. Each SM has one writer TB that tries to write all the data, and $N-1$ reader TBs that try to read a subset of the data. When a TB tries to enter the critical section (CS, Figure 4a `post` sub-routine), it first acquires a mutex lock and checks to see if there is enough capacity in the semaphore for the TB (line 8, Figure 4a). If there is capacity, the TB updates the semaphore and releases the lock. Similarly, when a TB leaves the CS (the `wait` sub-routine), it acquires the mutex lock and updates the semaphore (lines 8-9, Figure 4b) to remove itself from the semaphore before releasing the lock. When the size of the semaphore is greater than one, multiple reader TBs can enter the CS simultaneously. Unlike on CPUs, which utilize better OS support, both the post and wait sub-routines must utilize a lock to ensure ordering for accesses to the semaphore count ($semSize$, lines 4-9 in Figure 4a and lines 4-10 in Figure 4b. Thus, as the number of TBs increases, the mutex lock is a bottleneck and causes contention for reader and writer TBs trying to enter and exit the semaphore simultaneously. For example, a TB leaving the semaphore may not be able to exit if it cannot access the mutex variable because other TBs are repeatedly trying to enter the semaphore. To reduce contention, we can add backoff to make the TBs wait for a short period of time between each unsuccessful acquire.

SyncPrims [94] also contained a lock-free GPU semaphore ("SleepSem") that eschews a mutex. In reader-only semaphores like SyncPrims' studied, this approach had races, but did not affect correctness since all data was read-only. However, the lack of load linked store conditional (LLSC)-type mechanisms in GPUs made this version very livelock-prone for reader-writer semaphores like the ones we are studying. In a reader-writer semaphore, each thead attempting to join the semaphore must a) check if there is space in the semaphore and, once there is space b) update the semaphore to indicate that it has joined. Without GPU LLSC-type memory accesses or a single, combined atomic operation [16], these two operations in a lock-free semaphore must be performed separately. Consequently, the accesses may be *repeatedly* interleaved such that another thread (thread B) may update (and fill) the semaphore in between thread A's a) and b) operations – thus preventing forward progress. Accordingly, we do not include a lock-free semaphore in our study.

3 Related Work

CPU Synchronization Primitives: CPUs have long utilized efficient synchronization primitives for fine-grained synchronization, including centralized and decentralized mutexes [6], [52], [53], [73], [74], [84], ticket locks [29], [42], barriers [31], [96], and semaphores [15]. Modern OSs often support these synchronization primitives [20]. For example, a simple realization of a CPU mutex is a spin lock

3. Backoff represents a period of time where, after an access fails, a thread waits before attempting to perform the access again. This can significantly improve scalability by reducing contention [6].

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508


```
1  __shared__ bool acqLock = false;
2  decAmt = isWriter ? maxSemCnt : 1;
3  if (threadIdx == (0,0,0)) { // leader thread
4    while (!acqLock) {
5      if (atomicCAS(&lock,0,1) == 0) {
6        __threadfence();
7        acqLock = true;
8        if (semSize >= decAmt) {semSize -= decAmt;}
9    } }
10    __threadfence();
11    atomicExch(&lock,0);
12  }
13  __syncthreads(); // other threads wait here
```

(a) Post routine of the baseline GPU semaphore

```
1  __shared__ bool acqLock = false;
2  incAmt = isWriter ? maxSemCnt : 1;
3  if (threadIdx == (0,0,0)) { // leader thread
4    while (!acqLock) {
5      if (atomicCAS(&lock,0,1) == 0) {
6        __threadfence();
7        acqLock = true;
8        // no need to check semSize when exiting
9        semSize += incAmt;
10   } }
11   __threadfence();
12   atomicExch(&lock,0);
13 }
14 __syncthreads(); // other threads wait here
```

(b) Wait routine of the baseline GPU semaphore

Fig. 4: Pseudo-code for baseline GPU semaphore

Feature	HeteroSync Tree Barrier [88]	CPU-style SRBs	CCG [28]	G-SRB
Scalable, low contention synch variables	✓	✓	X	✓
Optimized for GPU processing model	✓	X	✓	✓
Efficient context switching support	X	✓	✓	✓

TABLE 1: Comparing G-SRB to prior work.

implemented with atomics [79], [81], [95]. However, these synchronization primitives often do not scale well on GPUs because GPUs do not have robust OS support and most CPU spin lock primitives are implemented using linked data structures which causes warp divergence [105]. Further, as threads increase these mutexes heavily use globally scoped atomics that degrade performance (Section 2.1). Moreover, GPU applications traditionally use fine-grained synchronization sparingly, and the level of GPU parallelism necessitates simpler coherence protocols and less OS involvement.

GPU Synchronization Primitives: SyncPrims and HeteroSync developed GPU synchronization primitives microbenchmarks, including locally and globally scoped atomic variants [88], [94]. However, as shown in Section 6, HeteroSync's barrier and semaphore implementations scale poorly and suffer from livelock. The proposed approaches address these shortcomings, scale better, and perform significantly fewer atomics. Prior work has also shown that ticket locks can be used instead of mutex locks in some GPU algorithms to improve scalability [56], [57], [89]. However, we focus on mutex locks because they are more commonly used in GPU applications and because ticket locks imply an ordering on when a TB accesses the CS that was not implied in either CCG or the benchmarks. Nevertheless, this is an interesting alternative for future work.

Other prior work dynamically estimates a GPU kernel's occupancy for barriers [92]. This approach restricts the number of TBs based on a discovery protocol estimate but guarantees fair scheduling and ensures deadlock-freedom. Finally, GPU manufactures such as AMD and NVIDIA have started to add hardware support for inter-block synchronization [4], [40], [44], [65]. However, this hardware support is limited in scope and does not significantly impact our results on NVIDIA GPUs.

Overall, Table 1 compares G-SRB to prior work on synchronization primitives across three key metrics: scalability of accessing synchronization variables as contention increases, compatibility with the GPU's processing model, and efficient support for context switching. As discussed above, HeteroSync was designed both for the GPU's processing model and with scalability in mind, but requires two barriers to support context switches. In contrast, CPU-style SRBs efficiently support context switching and scale efficiently, but perform poorly when applied to GPUs (as we show in Section 6). Finally, CCG is designed explicitly for GPUs and efficiently supports context switches (Section 2.2.1), but suffers scalability issues as the number of TBs joining the barrier increase (Section 6).

Accelerator Synchronization Primitives: Other prior work has also explored how to design optimized barriers for Xeon Phi co-processors [82]. Like our work, Rodchenko, et al. identify the importance of designing multi-level barrier algorithms with the accelerator's memory hierarchy in mind. However, designing barriers for Xeon Phi's and GPUs requires different considerations because the accelerators have different memory hierarchies, coherence protocols, consistency models, and access patterns. For example, Xeon Phi barrier algorithms must consider inter-core communication since atomics may be performed locally. In comparison, both locally- and globally-scoped GPU atomics must be performed at the last level cache. Thus, GPU barrier algorithms utilize different optimizations. Moreover, unlike our work, Rodchenko, et al. did not explore semaphore optimizations.

4 Design

4.1 Sense Reversing Barriers

As discussed in Section 2.2.1, in HeteroSync's atomic tree barrier every TB joins the local barrier twice (Figure 2c). Although HeteroSync's second barrier is necessary for correctness, it causes significant overhead (as we show in Section 6). To overcome this we design *G-SRB*, a hierarchical SRB [31]. Although SRBs already exist in CPUs, in Section 6 we show that directly applying CPU SRB concepts

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

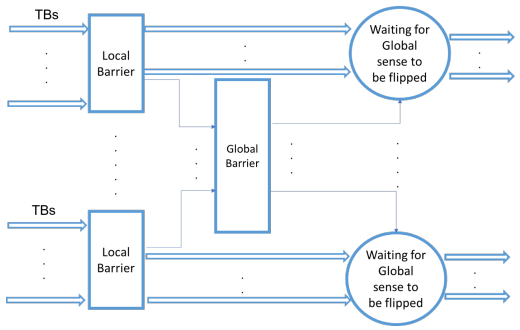


Fig. 5: Overview of GPU Sense Reversing Barrier design with statically elected per-SM leader that joins the global barrier and flips the sense.

to GPUs results in performance and scalability issues. This also differs from CCG, which uses aggressive, single-level barriers without backoff – which exploit integer overflow to avoid needing a second global barrier (like HeteroSync), but which do not scale well because of high contention for the single, shared global variable. Thus, our key insight is how to design SRBs while taking the GPU's unique memory hierarchy and higher level of parallelism into account (also discussed in Section 7).

Figure 5 shows a high-level overview of G-SRB. Like HeteroSync, G-SRB utilizes a tree barrier to reduce contention (Figure 6). We chose two levels for our tree barrier because this naturally conforms to the memory hierarchy and scoped synchronization of modern GPUs – locally scoped atomics avoid frequent, expensive cache invalidations and flushes for the local, per SM barriers, while the more expensive globally scoped atomics are only necessary for the second level, global barrier [17], [19], [21], [25], [32], [35], [45], [68]. Like Figure 2c, G-SRB has two parts: a local barrier per SM (Figure 6a) and a global barrier across all SMs (Figure 6b). Moreover, like CCG and HeteroSync (Section 2.2.1), we elect a leader thread per TB to join the global barrier to reduce contention (Figure 6b).

Local Barrier: All TBs on a SM share a sense variable (`sense`) initialized to false the first time the barrier is called. Moreover, each TB also has a local sense variable (`s`) initialized to the inverse of `sense` and stored in the scratchpad to reduce latency. TBs spin until `sense`'s value matches `s` (line 9, Figure 6a). As in CPU SRBs, `sense` matches (i.e., is flipped) when all TBs on a SM join the local barrier. Afterwards, like HeteroSync (lines 4-7, Figure 2c), a statically designated leader TB (lines 2-4, Figure 6) proceeds to the global barrier. The remaining TBs on each SM wait for the global barrier to complete (lines 6-9, Figure 6c). `sense` and `done_local` (Figure 2a) are used to identify when the local barrier is complete. However, in combination with `s`, `sense` serves an additional purpose, similar to SRBs in CPUs. If TB i is context switched out after it joins the local barrier and is rescheduled after the other TBs on this SM join and exit another local barrier (and flip `sense`), TB i's local copy of `s` will match `sense`. As a result, TB i can immediately progress, without causing deadlock as in Section 2.2.1.

Global Barrier: All leaders increment a global counter for the global barrier across TBs. When this count reaches the total number of active SMs, all leader TBs have reached the global barrier and the global sense (`global_sense`) is flipped (line 10, Figure 6b – like `sense` `global_sense` is initialized to false the first time the barrier is called). Next, these TBs advance to their final local, per-SM barriers (line 5, Figure 6c), and the overall barrier completes. Like the local SRB, the global SRB's `global_sense` variable serves a similar purpose to `done` from Figure 2b – identifying when all the leader TBs per SM have joined the global barrier. However, by tracking `sense` and `global_sense` separately, any TB that is context switched out will see that `sense` and `global_sense` now match (line 7, Figure 6b) and proceed to the next barrier. Moreover, if the other TBs have advanced to the next barrier, as in CPU SRBs, they must wait for this TB to arrive at the corresponding barrier before they can proceed. Thus, unlike HeteroSync G-SRB does not require a second set of barriers to ensure correctness on context switches (Figure 6c). Furthermore, G-SRB retains the benefits of HeteroSync's tree barrier by keeping most thread's atomics locally scoped.

4.1.1 *Hybrid Barrier*

G-SRB's two-level barrier and use of locally scoped atomics helps reduce contention as the number of threads joining the barrier scale. However, at low contention levels G-SRB requires additional atomics relative to a single-level global barrier like the one CCG uses. Conversely, CCG's single-level global barrier does not scale very well as the the number of threads or TBs in the program increases [108], because a single level barrier without backoff can result in added contention, especially when many threads or TBs join the barrier. Although some optimizations, such as eliding G-SRB's local barriers when there is a single thread (from a single TB) per SM, are simple, a more general approach is needed. Accordingly, since each design is optimized for different design points, we also create a hybrid global barrier implementation that uses CCG at low contention levels and G-SRB at medium and high contention levels to examine the benefits of integrating our design into future libraries.

4.2 Semaphores

Semaphores allow multiple TBs to enter the CS simultaneously based on the semaphore size. Typically, to avoid data races semaphores either require each thread to access unique data or only allow readers in the semaphore simultaneously. As discussed in Section 2.2.2, due to the lack of OS and hardware support in modern GPUs, current semaphore implementations often use mutex locks to prevent multiple threads from updating the semaphore count simultaneously. However, this centralized mutex is highly contented in GPU programs and creates a bottleneck for threads attempting to exit the semaphore, since they contend with threads acquiring the lock while trying to enter the semaphore. We demonstrate that this approach livelocks when the number of threads accessing the semaphore scaled beyond 32 or 64 (Section 6.2).[4] Even software backoff does not solve this

4. Recent work uses similar GPU reader-writer semaphores [54]. Although we have not found significant differences in their post and wait routines compared to Figures 4b and 8a, it scales to more GPU threads. However, our approach still improves performance over this style of reader-writer semaphore (Section 6.2).

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508


```
1  __shared__ bool s = !(*sense);
2  numTBsThisSM = gridDim.x / NUM_SM;
3  __threadfence_block();
4  if (threadIdx == (0,0,0)) { // leader thread
5    // perSMBarr = localBarr[smID]
6    // inc by 1 with wraparound
7    atomicInc(&perSMBarr,0x7FFFFFFF);
8    while (*sense != s) {
9      if (atomicCAS(&perSMBarr,numTBsThisSM,0) ==
numTBsThisSM){
10       __threadfence_block();
11       *sense = s; // flip sense
12 } } }
13 __syncthreads(); // other threads wait here
```

(a) Local (per SM) phase of G-SRB.

```
1  numJoin = (gridDim.x < NUM_SM) ? gridDim.x :
NUM_SM;
2  __threadfence();
3  if (threadIdx == (0,0,0)) { // leader thread
4    // globBarr = globalBarr
5    // inc by 1 with wraparound
6    atomicInc(&globBarr,0x7FFFFFFF);
7    while (*global_sense != *sense) {
8      if (atomicCAS(&globBarr,numJoin,0) == numJoin)
{
9        __threadfence();
10       *global_sense = *sense; // flip sense
11     } else { DoBackoff(); }
12 } }
13 __syncthreads(); // other threads wait here
```

(b) Global phase of G-SRB.

```
1  localSRBarrier(localBarr[smID]); // Fig 6a
2  if (perSMLeaderTB) {
3    globalSRBarrier(globalBarr); // Fig 6b
4  } else {
     // wait for global barrier to complete
5    if (threadIdx == (0,0,0)) { // leader thread
6      while(ld_glb_cg(global_sense) !=
ld_gbl_cg(sense)) { ; }
7  }
8    __syncthreads(); // other threads wait here
9  }
```

(c) Overall G-SRB design.

Fig. 6: Pseudo-code for components of proposed G-SRB

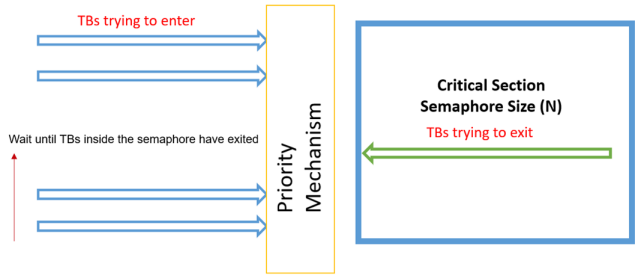


Fig. 7: Overview of proposed priority semaphore.

livelock, because it does not guarantee that threads trying to exit the semaphore can obtain the lock in a timely fashion.

Thus, we propose to add a priority mechanism to prioritize threads exiting the semaphore (*PriorSem*). This helps ensure forward progress and reduces the serialization penalty resulting from multiple attempts at acquiring the lock variable by a group of TBs in which some are trying to exit while others enter the semaphore. Although this approach temporarily favors threads exiting the CS, it does not cause starvation or unfairness in the long run because it is solely focused on releasing already held resources, not acquiring them – by prioritizing threads exiting the semaphore, other threads are able to enter the semaphore sooner. Figure 7 shows a high-level overview of the proposed design, and Figures 8a and 8b show the post and wait components of the proposed priority semaphore, respectively. Prioritizing the threads exiting the CS does not impact the TB or warp schedulers. Instead, our priority mechanism has other threads perform backoff instead of attempting to enter the CS (lines 5-8, Figure 8a. Once at least one thread exits the CS (lines 16-19, Figure 8b, the semaphore is no longer full so the priority flag is unset, and the other threads resume attempting to enter the CS again (lines 4-13, Figure 8a). These changes are the key differences from Section 2.2.2.

4.3 Assumptions

In general, we designed our barriers and semaphores to avoid significant reliance on the underlying system. For example, the code does not depend on the order TBs are launched. However, we do make two assumptions. First, like prior work, we assume that GPUs use a round robin (RR) arbitration scheme to schedule TBs across SMs [58], [75], [91]. Although to the best of our knowledge all modern GPUs use RR schemes [75], if future GPUs do not retain RR schemes [99] we can use inlined assembly with NVIDIA's

GPU Feature	Titan V	RTX 2080Ti	RTX 3090
Architecture	Volta	Turing	Ampere
# SMs	80	68	82
# CUDA Cores/SM	64	128	128
Max TBs/SM	32	16	32
Process	12 nm	12 nm FFN	8 nm

TABLE 2: GPUs used with their relevant system parameters.

`smid` PTX instruction to obtain per SM TB counts. Given this information, our algorithms could identify which and how many TBs are running on each SM, in order to correctly join the appropriate local and global barriers. Second, although `threadfence_block` only guarantees visibility to other threads in the same TB, in practice we find that the GPU coherence protocol makes these updates visible to all other threads on the same SM since they all use the same ordering point (the L1 cache). Thus we avoid expensive global `threadfence`'s for the local, per SM barriers in Figure 6a. Moreover, G-SRB is designed for applications with a fixed pattern where one thread from each TB joins the local barrier, then statically elected leaders per SM join the global barrier. We focused on this pattern because the GPU applications we study (Section 5) use this common pattern. Nevertheless, G-SRB can also be adapted for arbitrarily shaped groups (like CCG and libcu++) by modifying variables like `numJoin` (Figure 6b) and `numTBsThisSM` (Figure 6a). However, this may require adopting intra-warp deadlock-avoidance techniques for the semaphores [18], [46], which we avoid by electing a leader thread per TB.

5 Methodology

5.1 System Setup

We study our proposed algorithms (Section 4) on three NVIDIA GPUs: Titan V [66], RTX 2080Ti [60], and RTX 3090 [40] We focus on desktop-class GPUs because they have significant parallelism and are widely used. Table 2 lists the system configurations. We use CUDA 11 for all experiments [59]. Since GPUs do not have a dedicated, low latency OS, all of our algorithms are written at the user-level and do not involve the OS. We ran all benchmarks repeatedly with different input sizes to empirically ensure deadlock did not occur. Moreover, to ensure correctness we also check the outputs of each benchmark.

5.2 Benchmarks

We use barrier and semaphore microbenchmarks and benchmarks to evaluate our proposed algorithms. The mi-

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508


```
__shared__ bool acqLock = false;
decAmt = isWriter ? maxSemCnt : 1;
if (threadIdx == (0,0,0)) { // leader thread
  while (!acqLock) {
    // wait if threads exiting semaphore
    while (atomicCAS(&priority,0,0) != 0) {
      DoBackoff();
    }
    if (atomicCAS(&lock,0,1) == 0) {
       __threadfence();
       acqLock = true;
       if (semSize >= decAmt) { semSize -= decAmt; }
  } }
  __threadfence();
  atomicExch(&lock,0);
}
__syncthreads(); // other threads wait here
```

(a) Post routine of the proposed GPU priority semaphore

```
__shared__ bool acqLock = false;
__shared__ bool setPriority = false;
incAmt = isWriter ? maxSemCnt : 1;
if (threadIdx == (0,0,0)) { // leader thread
  while (!acqLock) {
    if (atomicCAS(&lock,0,1) == 0) {
      __threadfence();
      acqLock = true;
     // no need to check semSize when exiting
       semSize += incAmt;
    } else {
      if (!setPriority) {
        atomicOr(&priority,1);
        setPriority = true;
  } } }
  __threadfence();
  atomicExch(&lock,0);
  atomicAnd(&priority,0);
  setPriority = false;
}
__syncthreads(); // other threads wait here
```

(b) Wait routine of the proposed GPU priority semaphore

Fig. 8: Pseudo-code for baseline GPU semaphore

Applications	Description
Microbenchmarks	
Barriers	
atomTreeBarr [89] (Baseline)	Two-level atomic tree barrier.
G-SRB	Proposed two-level tree SRB.
G-CPUSRB	Two-level CPU-style tree SRB.
CCG [28]	NVIDIA's CCG barrier.
Hybrid	CCG (low); G-SRB (medium/high contention).
Semaphores	
SpinSem1, 10, 120 [89]	Semaphore with size [1, 10, 120].
SpinSemEBO1, 10, 120 [89]	SpinSem* with exponential software backoff.
PriorSem1, 10, 120	SpinSem* extended with proposed priority flag.
PriorSemEBO1, 10, 120	PriorSem* with exponential software backoff.
Benchmarks	
BFS [9]	Graph traversal algorithm.
PageRank [9]	Ranks search engine website results.
Parallel Scan [86]	Element i is sum of all elements up to i.
Reduce [34]	Reduces array elements into one result.
SSSP [9]	Computes shortest path of each node from a source node.

TABLE 3: Microbenchmarks and benchmarks studied.

crobenchmarks allow us to easily compare different algorithms and CS sizes (represented by number of instructions). For the baseline microbenchmarks, we use HeteroSync's tree barrier and semaphores with and without exponential software backoff, as described in Section 2. Since HeteroSync extends SyncPrims, we do not include a separate comparison against SyncPrims [94]. Table 3 lists the synchronization primitive microbenchmarks and benchmarks used in the experiments.

Barriers: We compare against CCG's barrier and a CPU-style SRB: *G-CPUSRB*, which we optimize to reduce contention by electing a leader thread per TB to join the global SRB (having all threads join the global SRB, like CPU algorithms do, resulted in extremely poor performance). Thus, the main difference between G-SRB and G-CPUSRB is that all TBs per SM join the global barrier in G-CPUSRB, where in G-SRB we select a leader TB per SM to join the global barrier. [5] Finally, we also compare against a hybrid global barrier (*Hybrid*), as discussed in Section 4.1.1.

We also compare the performance of each barrier for five modern GPGPU benchmarks [9], [11], which are representative of larger GPU programs that use barriers. BFS, SSSP, and PageRank are widely use in graph analytics, which often utilize GPUs [3], [9], [12], [69], [88], [101]. Since prior work has shown that different graph analytics algorithms provide the best GPU performance in different situations [38], [85], [93], we focus on implementations for these algorithms that require explicit synchronization. Similarly, performing Scan and Parallel Reduce on GPUs are foundational building blocks in Computing Aided Engineering (CAE), including computational fluid dynamics (CFD) [50], finite element analysis [37], multibody dynamics [49], [72], [97], and granular dynamics (GD) [51], [71]. Many CAE simulations repeatedly compute the norm-i of a vector, be it norm-1, norm-2 (Euclidean), or infinite norm, which are all Parallel Reduce operations. For instance, computing the kinetic energy of a granular system in its flow or the largest value of the residual in the momentum balance equation in the Finite Element Analysis method require a Parallel Reduce operation via the "+" or "max" operator, respectively. Likewise, any collision detection task in GD simulation requires a Scan operation at each time step (in one second of physics, there are approximately one million scans performed, which are sometimes performed on arrays as large as 50 million entries). Any Lagrangian method used for CFD also requires a Scan operation to determine the number of neighbors each fluid particle interacts with. Finally, both Scan and Parallel Reduce operations are used in gaming, performed in many cases at each time step of the simulation. Thus, Scan and Parallel Reduce are two of the most common operations done in CAE and gaming, and these operations often utilize GPUs to take advantage of their parallelism. For Scan and

5. We also compared against libcu++'s barrier implementations but found that CCG always provided equivalent or better performance, so we omit libcu++'s results for brevity. After discussing this finding with one of the developers, we believe CCG outperforms libcu++ because libcu++ focuses on full compatibility with C++, and thus cannot always optimize GPU code as well as CCG.

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

Parallel Reduce we use the CUDA Samples implementations as references, but replace the implicit CPU-side barrier with explicit grid synchronization [11]. As Zhang, et al. also observed, we saw comparable performance between the two methods [108].

Each thread in the barrier microbenchmarks accesses unique memory locations. To model various contention levels (TB/SM discussed in Section 5.3) we use a variety of input graphs [7], [9], [13], [83] for BFS, SSSP, and PageRank: bgg.gr (abbreviated 1-bgg, 1 TB/SM), USA-road-d.NY (2-NY, 2 TBs/SM), soc-academia (4-SA, 4 TBs/SM), USA-road-d.FLA (4-FLA, 4 TBs/SM), USA-road-d.W (8-W, 8 TBs/SM), web-google (8-WG, 8 TBs/SM), CoAuthorsDBLP (8-CO, 8 TBs/SM), and USA-road-dUSA-road-d.USA (16, 16 TBs/SM). Similarly, for Scan and Parallel Reduce, for maximum contention we use 163840 elements and halve the elements for each lower contention level.

Semaphores: Although some GPU workloads have started using semaphores [23], to the best of our knowledge they are not publicly available. Thus, we compare the baseline HeteroSync semaphores against our proposed semaphores (*PriorSem*). We also attempted to compare against libcu++'s semaphores, but its binary and counting semaphores do not support reader-writer semaphores. For all semaphores, the readers each access a fraction $(1/N)$ of the global memory data, while the writers write all N memory locations.

5.3 Configurations

For all experiments, we report steady state behavior by running the benchmarks in Table 3 ten times and averaging the results. The number of threads per TB varies across benchmarks. However, since all variants we study (including the CCG barriers we used) use the optimization where a single thread per TB joins the barrier or semaphore, we express contention in terms of the number of TBs per SM (TBs/SM). As the number of TBs/SM increase, contention for the synchronization primitives also increases, regardless of number of threads per TB. We set the maximum exponential backoff to 1024 for all results, because we empirically found this provided a good balance of reduced contention without sleeping too long.[6] Moreover, to ensure each SM has the same number of TBs, we utilize NVIDIA's RR scheduler arbitration scheme (Section 4.3) to schedule TBs across SMs.

To analyze the algorithm's sensitivity under various loads, we vary the CS size and contention level in isolation. First, we hold the contention level constant (by using 16 TBs/SM to model a high contention situation) and vary the CS size by varying the number of global loads and stores each thread performs from 100 through 10000 in multiples of 10. For the barrier implementations, these accesses use the barriers to ensure there are no data races, while for the semaphores these accesses are all performed in the CS. Although our primary focus is high contention cases, we also use weak scaling [24] to examine the effect of contention. Here, we hold the CS size constant (at 100 global loads and stores) and vary the number of TBs from 1 TB/SM to 32 TBs/SM for Volta, and from 1 TB/SM to 16 TBs/SM for Turing since Turing allows a maximum of 16 TBs/SM. We show the runtime for 10 iterations of the micro-benchmark across all data points. We also show the kernel GPU activity profiles for both the baseline barrier/semaphore and the proposed implementations [67].

For the semaphore, we compare the baseline's (with and without exponential backoff) and proposed implementation's (with and without exponential backoff) execution times across different contention levels. We create these different contention levels by varying either the semaphore size, number of TBs/SM, or number of load and store instructions while keeping the other two constant. For example, we use semaphore sizes of 1 (single reader or writer in CS), 10 (up to 10 readers or 1 writer in CS), and 120 (up to 120 readers or 1 writer in CS), as in prior work [89], [94], to model different reader-writer ratios.

Finally, we examine the microbenchmarks' synchronization and non-synchronization breakdown, to analyze how well our optimizations reduce the number of atomics and global synchronization for the barriers, and the degree of CS contention for the semaphores. To estimate this percentage we measured the number of clock cycles spent in the synchronization call relative to the number of clock cycles spent in the entire kernel, averaged across all TBs.

6 Evaluation

6.1 Barriers

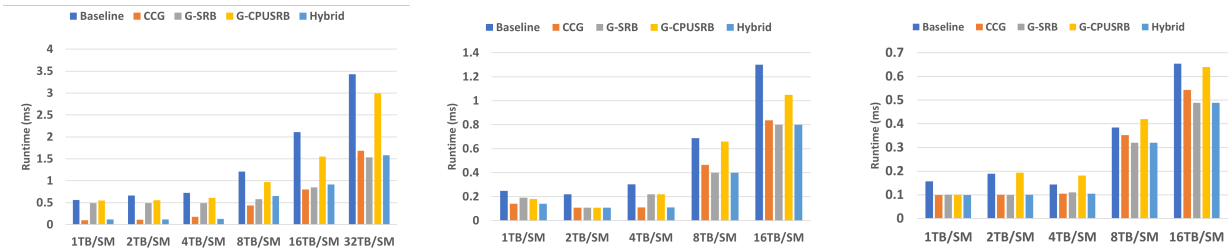
Microbenchmarks: Figure 9 compares the barrier's execution times as the number of TBs vary. Overall, CCG and our sense reversing barrier (*G-SRB*) significantly outperform the baseline. The CPU-style SRB (*G-CPUSRB*) also outperforms the baseline but is less efficient than CCG and G-SRB, especially at high contention levels. On average, G-SRB improves performance by 34% on Volta, 38% on Turing, and 27% on Ampere while G-CPUSRB improves performance by 15% on Volta, 26% on Turing, and 3% on Ampere respectively, compared to the baseline. By removing the second barrier call, G-SRB significantly reduces the number of atomic accesses versus the baseline by over 50%. G-SRB's gains over the baseline also increase as contention increases; although the baseline also uses a tree barrier, its additional atomics (relative to G-SRB) hurt performance. Thus, G-SRB improves scalability. To further examine why G-SRB improves on the baseline, Figure 10 breaks down their execution time into synchronization and non-synchronization time for the Titan V GPU. On average, G-SRB reduces synchronization time by 78% over the baseline. Moreover, non-synchronization time only increases a little for G-SRB as contention increases. Thus, the synchronization time reduction is not replaced by power-hungry active waiting.

CCG also provides significant benefits over the baseline. Notably, CCG outperforms G-SRB at lower contention levels because at low contention levels G-SRB's software backoff and hierarchical design offer less benefit (Section 4.1.1). Although Ampere provides better hardware support for barriers, this support focuses on barriers that can use shared memory [65], which our barriers cannot use since they synchronize across multiple TBs which are often running on different SMs. G-CPUSRB also provides similar benefits at lower contention levels. However, as contention increases, G-SRB's more scalable design approaches and then outperforms CCG and G-CPUSRB. Thus, *Hybrid*, which

6. We also examined inlined assembly with CUDA's `nanosleep` PTX instruction. The results showed $< 2\%$ improvement over software backoff, so we exclude them due to space constraints.

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508



(a) Volta Architecture (b) Turing Architecture (c) Ampere Architecture

Fig. 9: Execution time for barrier microbenchmarks as contention increases.

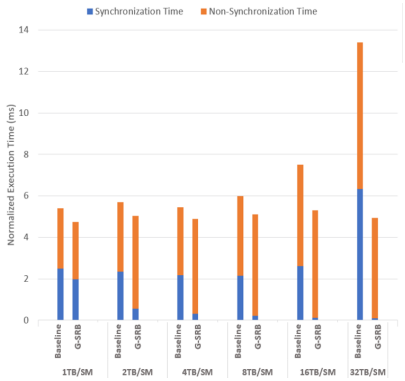


Fig. 10: Sync versus non-sync time for Volta barriers.

utilizes CCG at low contention levels and G-SRB at medium and high contention levels, provides high performance regardless of contention level. We set *Hybrid*'s threshold in between selecting CCG and G-SRB at 8 TBs/SM, as it provides a good transition point to switch between the two approaches. Moreover, since CCG, G-CPUSRB, G-SRB, and Hybrid significantly outperform the baseline, we focus on these barriers for the full-sized benchmarks.

Benchmarks: Figure 11 examines the barrier's normalized performance as contention varies for the five full-sized benchmarks. Although Figures 9 and 10 used absolute execution time in – to demonstrate how the performance scaled as contention increased, Figure 11 uses normalized performance because each benchmark's runtime differs significantly. Similar to the microbenchmarks, at lower contention levels CCG often outperforms G-SRB. However, unlike the microbenchmarks G-SRB consistently outperforms CCG and G-CPUSRB at much lower levels of contention (e.g., 8 TBs/SM). This happens because G-SRB's locally scoped atomics reduce the number of global flushes and invalidations, which improves performance relative to CCG, which must globally flush and invalidate more frequently due to its single-level design. As discussed in Section 2.2.1, every atomic in CCG performs a `threadfence` that flushes and invalidates (Section 2). In comparison, in G-SRB only one TB per SM that takes part in the second level barrier (Figure 6b) must perform a `threadfence`, and the remaining TBs can perform the cheaper `threadfence_block` at the local, per-SM barriers (Figure 6a). This difference is further magnified since CCG does not perform backoff. Thus, as contention increases, G-SRB reduces unnecessary accesses to the shared global synchronization variables – which is magnified by the degree of synchronization in some workloads. For example, usually the gains were slightly lower for social media and similar for web networks and road networks at the same contention level because the social media networks require less synchronization.

For all five benchmarks, G-SRB's performance improvement over CCG is closely tied to the percentage of total execution time spent performing global synchronization (Section 5.3). As expected, global synchronization is a major component in *Reduce* and *Scan* where on average 90% and 82%, respectively, of kernel runtime is spent synchronizing at the maximum contention level. As a result, G-SRB outperforms CCG by an average of 62% and 49%, respectively, for Reduce and Scan at contention levels $\geq$ 8 TBs/SM. In contrast, on average PageRank spends only 3% of kernel time synchronizing. Consequently, G-SRB only improves performance by 0.1% on average over CCG. BFS and SSSP represent a middle point between Reduce, Scan, and SSSP: they spend 36% and 47%, respectively, of their time synchronizing at maximum contention. As a result, CCG mostly performs better than G-SRB for lower contention levels (e.g., 3% and 8% better for BFS and SSSP, respectively, on average at contention levels $<$ 8 TBs/SM). For these lower contention levels G-CPUSRB also does worse than CCG but is slightly better than G-SRB. Since G-CPUSRB has a more centralized design, at lower contention levels, scalability is less important, so G-SRB's additional local barriers add overhead relative to G-CPUSRB. However, as contention increases, G-SRB's improved scalability again enable it to outperform CCG and G-CPUSRB: at contention levels $\geq$ 8 TBs/SM G-SRB is 19% and 19% better than CCG on average across the three GPUs, respectively, for BFS and SSSP. Interestingly, for high contention levels G-CPUSRB also often outperforms CCG. Finally, Hybrid again gives the best performance in almost every case, effectively blending CCG and G-SRB. Note that the full-sized benchmarks utilize a wide variety of input sizes, including ones that cause higher contention. Moreover, we expect that input sizes will continue to increase in the future. Thus, identifying more scalable synchronization solutions is important.

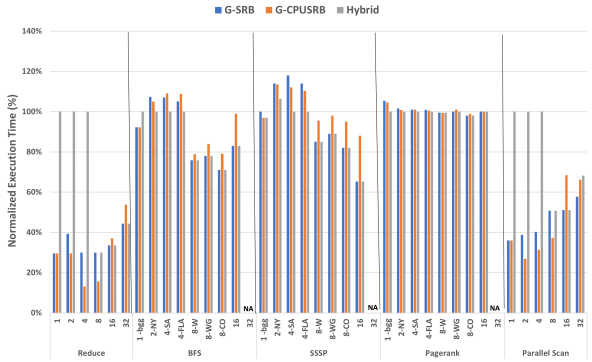
Varying CS Size: Figures 12 evaluates how the barriers scale as the CS size varies from 100-10000 instructions; we evaluate these CS sizes at 1 TB/SM, 16 TBs/SM, and 32 TBs/SM for Volta, and 1 TB/SM, 8 TBs/SM, and 16 TBs/SM for Turing. These results further demonstrate that as contention increases, G-SRB again outperforms CCG. Interestingly, as CS size increases, CCG and G-SRB provide smaller benefits over the baseline because an increasing amount of time is spent performing data accesses.

6.2 Semaphores

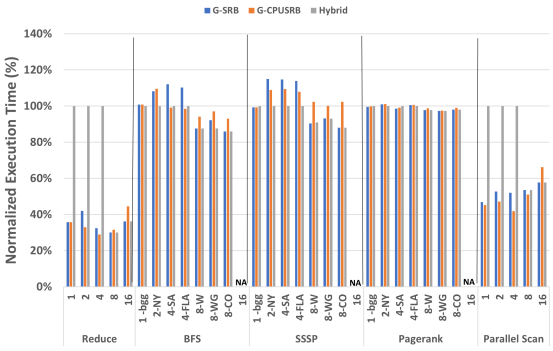
Microbenchmarks: Figures 13, 14, and 15 show the execution times of all semaphore algorithms across different contention levels ranging from 1 TB/SM to the maximum

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

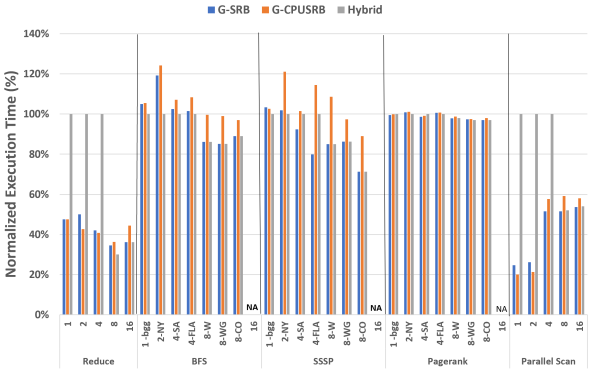
This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPDS.2022.3218508



(a) Volta Architecture

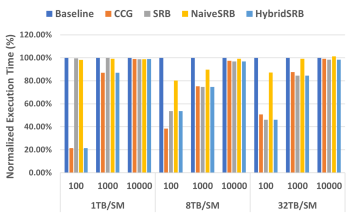


(b) Turing Architecture

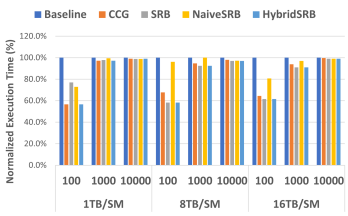


(c) Ampere Architecture

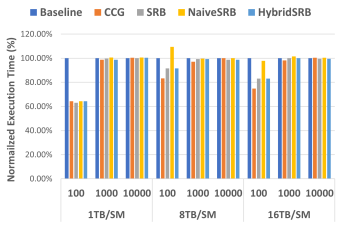
Fig. 11: Benchmark's with barriers performance as contention varies, normalized to CCG. NA's denote configurations the benchmark cannot run due to input size. Since Baseline is much worse for the microbenchmarks we do not include it here.



(a) Volta Architecture

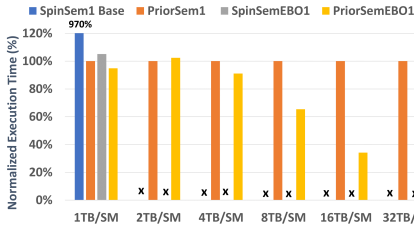


(b) Turing Architecture

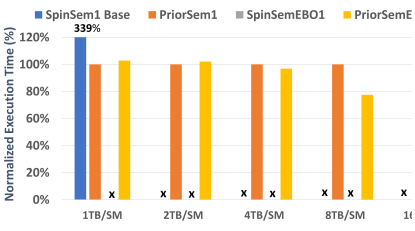


(c) Ampere Architecture

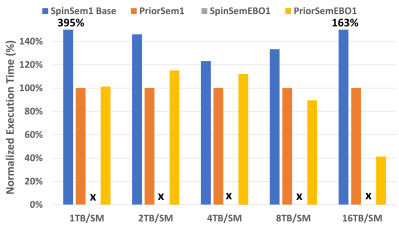
Fig. 12: Barrier microbenchmark's execution time as CS size varies from 100 to 10000 instructions, normalized to the baseline.



(a) Volta Architecture



(b) Turing Architecture



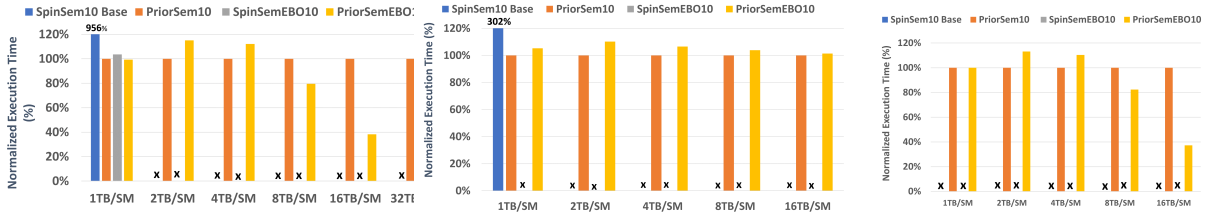
(c) Ampere Architecture

Fig. 13: Size 1 semaphore's execution time as contention increases, normalized to *PriorSem1*. *X*'s denote deadlocks.

TB/SM across different GPU architectures (Table 2). We create these different contention points by either varying the semaphore size, number of TBs/SM, or number of load and store instructions while keeping the other two constants. Overall, on average for semaphore of size 1 our proposed implementation improves performance by 89% over the baseline implementation on the Volta, 70% on the Turing GPU, and 90% on Ampere GPU. For the semaphores with less contention (10 and 120), the gains are similar; for example for size 10 the average performance gain is 89% for Volta and 66% for Turing, while for Ampere the Baseline livelocks even for 1 TB/SM. The baseline implementation with exponential backoff shows much better performance and is only 5% slower on the Volta GPU for a semaphore of

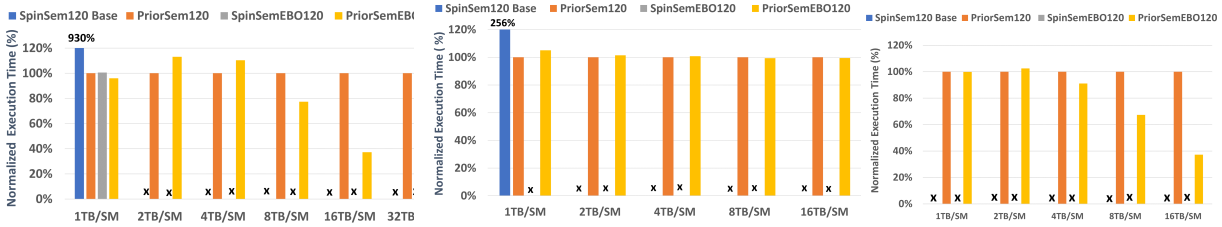
© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508



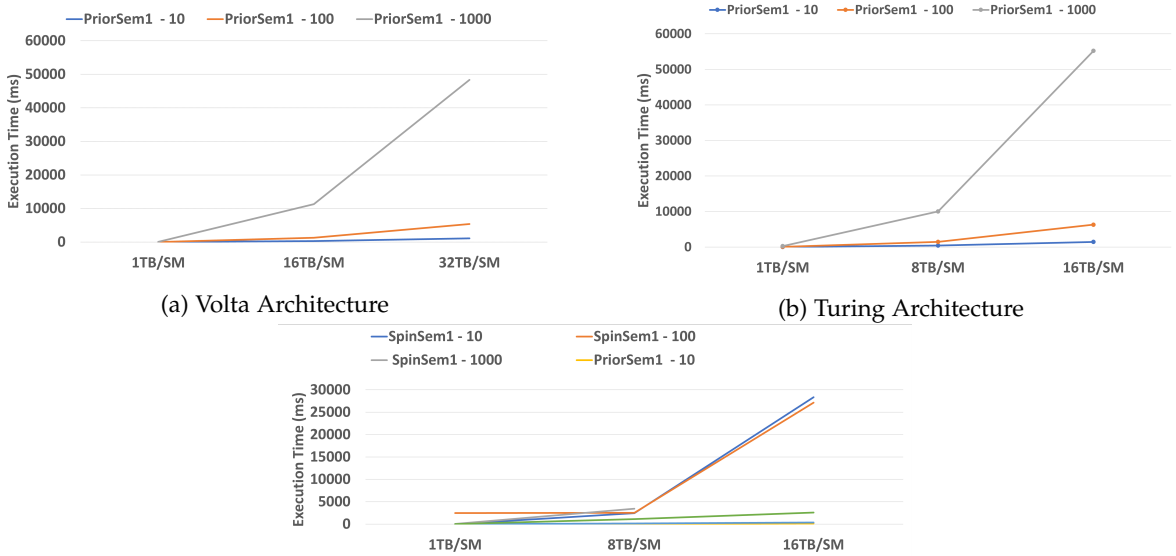
(a) Volta Architecture (b) Turing Architecture (c) Ampere Architecture

Fig. 14: Size 10 semaphore's execution time as contention increases, normalized to *PriorSem10*. *X*'s denote deadlocks.



(a) Volta Architecture (b) Turing Architecture (c) Ampere Architecture

Fig. 15: Size 120 semaphore's execution time as contention increases, normalized to *PriorSem120*. *X*'s denote deadlocks.



(a) Volta Architecture (b) Turing Architecture

(c) Ampere Architecture

Fig. 16: Semaphore's execution time as CS size varies. The labels show the microbenchmark name suffixed with the CS size. We do not include the baseline semaphores because they deadlock beyond 1 TB/SM.

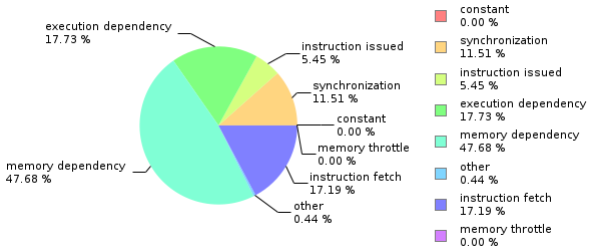


Fig. 17: GPU kernel profile for the baseline semaphore.

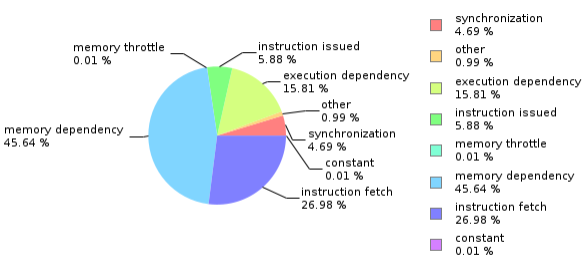


Fig. 18: GPU kernel profile for the priority semaphore.

size 1. However, *SpinSem* livelocks as contention increases for its centralized semaphore on the Volta and Turing GPUs, and is so significant that it always deadlocks on the Ampere GPU. As a result, we were only able to obtain *SpinSem* data for 1 TB/SM on the Volta and Turing GPUs. Although *SpinSemEBO* reduces contention compared to *SpinSem*, it also livelocks for $\geq$ 1 TB/SM. In comparison, *PriorSem* and *PriorSemEBO* avoid livelock for all the evaluated data points. Thus, the priority mechanism successfully prevents TBs entering the semaphore from stopping TBs trying to exit the semaphore and ensures forward progress. Moreover, the GPU kernel activity profiles (Figures 17 and 18) show that the priority mechanism also reduces redundant attempts to

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

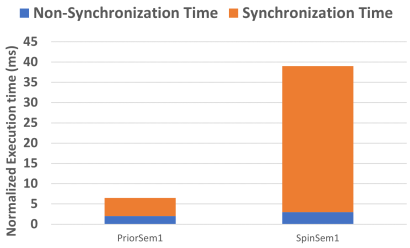


Fig. 19: Semaphore sync versus non-sync time for 1 TB/SM.

acquire the lock, reducing the overall synchronization time from 12% in baseline to 5% in *PriorSem*. Further, Figure 19 shows the priority mechanism also reduces CS synchronization time by 88% over the baseline.

Varying CS Size: Finally, like G-SRB, we evaluate the semaphores across different CS sizes and for different numbers of TBs/SM. The semaphores (Figure 16) and G-SRB have similar trends: as CS size increases the overall gains decrease because the percentage of time spent in synchronization decreases as we increase the size of the CS. Overall, these results show that our proposed modifications improve performance and avoid livelock.

7 Discussion & Conclusion

Modern GPU applications increasingly utilize fine-grained synchronization. However, existing solutions scale poorly or suffer from livelock at high contention levels. We propose optimizations to state-of-the-art GPU barriers and semaphores. Although our techniques extend well known approaches like SRBs, we demonstrate how utilizing these concepts on GPUs requires careful consideration of the GPU memory hierarchy, coherence protocol, consistency models, and threading model. Directly applying CPU-style SRBs is sub-optimal, especially as contention increases. We show how synchronization primitives should be designed with the GPU's parallelism, memory hierarchy, and consistency models in mind. We design G-SRB accordingly, utilizing a two-level barrier that scales efficiently by mirroring the GPU's memory hierarchy and exploiting the GPU's scoped consistency to frequently perform synchronization locally. We also utilize the GPU's scratchpads to spin on local, per-TB variables to improve performance – without affecting SM utilization because each thread only store a couple of booleans (e.g., Figure 6a) in the scratchpad.

Overall, our algorithms significantly improve the performance and scalability of GPU barriers and semaphores for Volta, Turing, and Ampere GPUs, and avoid livelocks. On average our proposed techniques improve performance by 36% relative to the baseline tree barrier and 79% relative to the baseline semaphore. Since our barrier scales better than CCG and CPU-style SRBs, at levels of contention $\geq$ 8 TBs/SM, for five full-sized benchmarks, it outperforms CCG by 36% on Volta, 34% on Turing, and 32% on Ampere; likewise it outperforms CPU-style SRBs by 12% on Volta, 7% on Turing, and 8% on Ampere GPUs. Collectively, this makes a case for G-SRB to replace CCG for higher levels of contention. Accordingly, our hybrid barrier combines the best of CCG and G-SRB to provide high performance for all contention levels. Moreover, although we focus on single GPU synchronization, the ideas are also easily extensible to multi-GPU and multi-chiplet setups by adding an additional hierarchy level to keep most synchronization local per GPU.

References

[1] A. Abdolrashidi, D. Tripathy, M. E. Belviranli, L. N. Bhuyan, and D. Wong, "Wireframe: Supporting Data-Dependent Parallelism through Dependency Graph Execution in GPUs," in *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, 2017, p. 600–611.

[2] J. T. Adriaens, K. Compton, N. S. Kim, and M. J. Schulte, "The Case for GPGPU Spatial Multitasking," in *IEEE International Symposium on High-Performance Computer Architecture*, ser. HPCA, 2012, pp. 1–12.

[3] J. Alsop, M. S. Orr, B. M. Beckmann, and D. A. Wood, "Lazy Release Consistency for GPUs," in *49th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, 2016, pp. 26:1–26:13.

[4] AMD, "AMD Graphics Core Next (GCN) Architecture, Generation 3," http://developer.amd.com/wordpress/media/2013/12/AMD_GCN3_Instruction_Set_Architecture_rev1.1.pdf, 2016.

[5] AMD, "OpenCL Programming Guide," http://developer.amd.com/wordpress/media/2013/07/AMD_Accelerated_Parallel_Processing_OpenCL_Programming_Guide-rev-2.7.pdf, 2016.

[6] T. E. Anderson, "The Performance of Spin Lock Alternatives for Shared-Memory Multiprocessors," *IEEE Trans. Parallel Distrib. Syst.*, vol. 1, no. 1, p. 6–16, Jan. 1990.

[7] D. A. Bader, A. Kappes, H. Meyerhenke, P. Sanders, C. Schulz, and D. Wagner, *Benchmarking for Graph Clustering and Partitioning*. Springer New York, 2018, pp. 161–171.

[8] H.-J. Boehm and S. V. Adve, "Foundations of the C++ Concurrency Memory Model," in *Proceedings of the 2008 ACM SIGPLAN Conference on Programming Language Design and Implementation*, ser. PLDI, 2008, p. 68.

[9] M. Burtscher, R. Nasre, and K. Pingali, "A Quantitative Study of Irregular Programs on GPUs," in *IEEE International Symposium on Workload Characterization*, ser. IISWC, Nov 2012, pp. 141–151.

[10] D. Cederman and P. Tsigas, "On Dynamic Load Balancing on Graphics Processors," in *Proceedings of the 23rd ACM SIGGRAPH/EUROGRAPHICS Symposium on Graphics Hardware*, 2008, pp. 57–64.

[11] N. D. Center, "NVIDIA Cuda samples," https://docs.nvidia.com/cuda/cuda-samples/index.html, 2022.

[12] S. Che, B. M. Beckmann, S. K. Reinhardt, and K. Skadron, "Pannotia: Understanding Irregular GPGPU Graph Applications," in *IEEE International Symposium on Workload Characterization*, ser. IISWC, Sept 2013, pp. 185–195.

[13] C. Demetrescu, "9th DIMACS Implementation Challenge," http://users.diag.uniroma1.it/challenge9/download.shtml, 2006.

[14] G. Diamos, S. Sengupta, B. Catanzaro, M. Chrzanowski, A. Coates, E. Elsen, J. Engel, A. Y. Hannun, and S. Satheesh, "Persistent RNNs: Stashing Recurrent Weights On-Chip," in *Proceedings of the 33nd International Conference on Machine Learning*, ser. ICML, 2016, pp. 2024–2033.

[15] E. W. Dijkstra, "A tutorial on the split binary semaphore," in *Theoretical Foundations of Programming Methodology: Lecture Notes of an International Summer School, directed by F. L. Bauer, E. W. Dijkstra and C. A. R. Hoare*, 1982, pp. 555–564.

[16] A. Dutu, M. D. Sinclair, B. M. Beckmann, D. A. Wood, and M. Chow, "Independent Forward Progress of Work-groups," in *Proceedings of the 47th International Symposium on Computer Architecture*, ser. ISCA, May 2020.

[17] I. J. Egielski, J. Huang, and E. Z. Zhang, "Massive Atomics for Massive Parallelism on GPUs," in *Proceedings of the 2014 International Symposium on Memory Management*, ser. ISMM, 2014, p. 93–103.

[18] A. ElTantawy and T. M. Aamodt, "MIMD Synchronization on SIMT Architectures," in *49th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, 2016, pp. 1–14.

[19] M. Elteir, H. Lin, and W.-C. Feng, "Performance Characterization and Optimization of Atomic Operations on AMD GPUs," in *IEEE International Conference on Cluster Computing*, 2011, pp. 234–243.

[20] J. I. Fiestas and R. E. Bustamante, "A Survey on the Performance of Different Mutex and Barrier Algorithms," in *IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing*, ser. INTERCON, 2019, pp. 1–4.

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

[21] S. Franey and M. Lipasti, "Accelerating Atomic Operations on GPGPUs," in *Seventh IEEE/ACM International Symposium on Networks-on-Chip*, ser. NoCS, 2013, pp. 1–8.

[22] B. R. Gaster, D. Hower, and L. Howes, "HRF-Relaxed: Adapting HRF to the Complexities of Industrial Heterogeneous Memory Models," *ACM Trans. Archit. Code Optim.*, vol. 12, no. 1, pp. 7:1–7:26, Apr. 2015.

[23] I. Gelado and M. Garland, "Throughput-Oriented GPU Memory Allocation," in *Proceedings of the 24th Symposium on Principles and Practice of Parallel Programming*, ser. PPoPP, 2019, p. 27–37.

[24] J. L. Gustafson, "Gustafson's law," in *Encyclopedia of Parallel Computing*, 2011, pp. 819–825.

[25] A. Gutierrez, B. M. Beckmann, A. Dutu, J. Gross, M. LeBeane, J. Kalamatianos, O. Kayiran, M. Poremba, B. Potter, S. Puthoor, M. D. Sinclair, M. Wyse, J. Yin, X. Zhang, A. Jain, and T. Rogers, "Lost in Abstraction: Pitfalls of Analyzing GPUs at the Intermediate Language Level," in *IEEE International Symposium on High Performance Computer Architecture*, ser. HPCA, 2018, pp. 608–619.

[26] I. E. Hajj, J. Gomez-Luna, C. Li, L. Chang, D. Milojicic, and W. Hwu, "KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism," in *49th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, 2016, pp. 1–12.

[27] M. Harris, "Optimizing Parallel Reduction in CUDA," *Nvidia Developer Technology*, vol. 2, no. 4, p. 70, 2007.

[28] M. Harris and K. Perelygin, "Cooperative Groups: Flexible CUDA Thread Programming," https://devblogs.nvidia.com/cooperative-groups/, 2017.

[29] B. He, W. N. Scherer, and M. L. Scott, "Preemption Adaptivity in Time-Published Queue-Based Spin Locks," in *High Performance Computing*, ser. HIPC, 2005, pp. 7–18.

[30] B. Hechtman, S. Che, D. Hower, Y. Tian, B. Beckmann, M. Hill, S. Reinhardt, and D. Wood, "QuickRelease: A Throughput-Oriented Approach to Release Consistency on GPUs," in *20th International Symposium on High Performance Computer Architecture*, ser. HPCA, Feb 2014, pp. 189–200.

[31] D. Hensgen, R. Finkel, and U. Manber, "Two Algorithms for Barrier Synchronization," *Int. J. Parallel Program.*, vol. 17, no. 1, p. 1–17, Feb. 1988.

[32] D. R. Hower, B. A. Hechtman, B. M. Beckmann, B. R. Gaster, M. D. Hill, S. K. Reinhardt, and D. A. Wood, "Heterogeneous-race-free Memory Models," in *Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS, 2014, pp. 427–440.

[33] HSA Foundation, "HSA Platform System Architecture Specification," http://www.hsafoundation.com/?ddownload=4944, 2015.

[34] W. A. R. Jradi, H. A. Dantas do Nascimento, and W. Santos Martins, "A Fast and Generic GPU-Based Parallel Reduction Implementation," in *Symposium on High Performance Computing Systems*, ser. WSCAD, 2018, pp. 16–22.

[35] M. Khairy, Z. Shen, T. M. Aamodt, and T. G. Rogers, "Accel-Sim: An Extensible Simulation Framework for Validated GPU Modeling," in *ACM/IEEE 47th Annual International Symposium on Computer Architecture*, ser. ISCA, 2020, pp. 473–486.

[36] F. Khorasani, H. A. Esfeden, N. Abu-Ghazaleh, and V. Sarkar, "In-Register Parameter Caching for Dynamic Neural Nets with Virtual Persistent Processor Specialization," in *Proceedings of 51st IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, 2018, pp. 377–389.

[37] N. Khude, I. Stanciulescu, D. Melanz, and D. Negrut, "Efficient Parallel Simulation of Large Flexible Body Systems With Multiple Contacts," *ASME Journal of Computational and Nonlinear Dynamics*, vol. 8, no. 4, pp. 041 003–041 003, 2013.

[38] J. Kim and C. Batten, "Accelerating Irregular Algorithms on GPGPUs Using Fine-Grain Hardware Worklists," in *Proceedings of the 47th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, 2014, p. 75–87.

[39] K. Koukos, A. Ros, E. Hagersten, and S. Kaxiras, "Building Heterogeneous Unified Virtual Memories (UVMs) Without the Overhead," *ACM Trans. Archit. Code Optim.*, vol. 13, no. 1, pp. 1:1–1:22, Mar. 2016.

[40] R. Krashinsky, O. Giroux, S. Jones, N. Stam, and S. Ramaswamy, "NVIDIA Ampere Architecture In-Depth," https://developer.nvidia.com/blog/nvidia-ampere-architecture-in-depth/, 2020.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NeurIPS, 2012, pp. 1097–1105.

[42] L. Lamport, "A New Solution of Dijkstra's Concurrent Programming Problem," *Commun. ACM*, vol. 17, no. 8, p. 453–455, Aug. 1974.

[43] Lee Howes and Aaftab Munshi, "The OpenCL Specification, Version 2.0," Khronos Group, 2015.

[44] B. A. Lelbach, O. Giroux, J. Bastien, D. Vollmann, and D. Olsen, "P1135R5: The C++20 Synchronization Library," http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2019/p1135r5.html, 2019.

[45] J. Lew, D. Shah, S. Pati, S. Cattell, M. Zhang, A. Sandhupatla, C. Ng, N. Goli, M. D. Sinclair, T. G. Rogers, and T. M. Aamodt, "Analyzing Machine Learning Workloads Using a Detailed GPU Simulator," in *IEEE International Symposium on Performance Analysis of Systems and Software*, ser. ISPASS, 2019.

[46] A. Li, G.-J. van den Braak, H. Corporaal, and A. Kumar, "Fine-Grained Synchronizations and Dataflow Programming on GPUs," in *Proceedings of the 29th ACM on International Conference on Supercomputing*, ser. ICS, 2015, p. 109–118.

[47] D. Lustig, S. Sahasrabuddhe, and O. Giroux, "A Formal Analysis of the NVIDIA PTX Memory Consistency Model," in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '19, 2019, p. 257–270.

[48] J. Manson, W. Pugh, and S. V. Adve, "The Java Memory Model," in *Proceedings of the 32nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, ser. POPL '05, 2005, p. 378–391.

[49] H. Mazhar, T. Heyn, A. Tasora, and D. Negrut, "Using Nesterov's method to accelerate multibody dynamics with friction and contact," *ACM Trans. Graph.*, vol. 34, no. 3, pp. 32:1–32:14, 2015.

[50] H. Mazhar, A. Pazouki, M. Rakhsha, P. Jayakumar, and D. Negrut, "A differential variational approach for handling fluid-solid interaction problems via smoothed particle hydrodynamics," *Journal of Computational Physics*, vol. 371, pp. 92–119, 2018.

[51] H. Mazhar, T. Heyn, and D. Negrut, "A scalable parallel method for large collision detection problems," *Multibody System Dynamics*, vol. 26, pp. 37–55, 2011, 10.1007/s11044-011-9246-y.

[52] J. M. Mellor-Crummey and M. L. Scott, "Algorithms for Scalable Synchronization on Shared-Memory Multiprocessors," *ACM Trans. Comput. Syst.*, vol. 9, no. 1, p. 21–65, Feb. 1991.

[53] R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed packet switching for local computer networks," *Commun. ACM*, vol. 19, no. 7, p. 395–404, Jul. 1976.

[54] d. Miller, J. Nelson, A. Hassan, and R. Palmieri, "KVCG: A Heterogeneous Key-Value Store for Skewed Workloads," in *Proceedings of the 14th ACM International Conference on Systems and Storage*, ser. SYSTOR '21, 2021.

[55] M. Munford, V. Lima, T. Vieira, G. Heinzelmann, T. Creczynski-Pasa, and A. Pasa, "Afm in-situ characterization of supported phospholipid layers formed by vesicle fusion," *Microscopy and Microanalysis*, vol. 11, no. S03, pp. 90–93, 2005.

[56] J. Nelson, d. Miller, and R. Palmieri, "Don't forget about synchronization! Guidelines for using locks on graphics processing units," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 2, p. e5757, 2022.

[57] J. Nelson and R. Palmieri, "Don't Forget About Synchronization!: A Case Study of K-Means on GPU," in *Proceedings of the 10th International Workshop on Programming Models and Applications for Multicores and Manycores*, ser. PMAM, 02 2019, pp. 11–20.

[58] A. Nere, A. Hashmi, and M. Lipasti, "Profiling Heterogeneous Multi-GPU Systems to Accelerate Cortically Inspired Learning Algorithms," in *Proceedings of the 2011 IEEE International Parallel & Distributed Processing Symposium*, ser. IPDPS, 2011, p. 906–920.

[59] NVIDIA, "CUDA programming guide," https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html, 2016.

[60] ——, "Pascal P100," https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf, 2016.

[61] ——, "Pascal P102," https://international.download.nvidia.com/geforce-com/international/pdfs/GeForce_GTX_1080_Whitepaper_FINAL.pdf, 2016.

[62] ——, "Cuda-gdb," http://docs.nvidia.com/cuda/cuda-gdb/index.html/, NVIDIA, 2017, accessed Nov 15, 2017.

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

[63] NVIDIA, "CUDA Stream Management," http://developer.download.nvidia.com/compute/cuda/2_3/toolkit/docs/online/group__CUDART__STREAM.html, 2018.

[64] ——, "libcu++: The C++ Standard Library for Your Entire System," https://nvidia.github.io/libcudacxx/, 2020.

[65] ——, "Split Arrive/Wait Barrier," https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html\#aw-barrier, 2020.

[66] NVIDIA Corp., "Inside Volta: The World's Most Advanced Data Center GPU," https://devblogs.nvidia.com/parallelforall/inside-volta/, May 2017.

[67] NVIDIA Corp, "Profiler User's Guide," https://docs.nvidia.com/cuda/profiler-users-guide/index.html, 2018.

[68] L. Nyland and S. Jones, "Understanding and Using Atomic Memory Operations," in *Proceedings of GPU Technology Conference*, ser. GTC, 2013.

[69] M. S. Orr, S. Che, A. Yilmazer, B. M. Beckmann, M. D. Hill, and D. A. Wood, "Synchronization Using Remote-Scope Promotion," in *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS, 2015, p. 73–86.

[70] J. J. K. Park, Y. Park, and S. Mahlke, "Chimera: Collaborative Preemption for Multitasking on a Shared GPU," in *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '15, 2015, p. 593–606.

[71] A. Pazouki, M. Kwarta, K. Williams, W. Likos, R. Serban, P. Jayakumar, and D. Negrut, "Compliant contact versus rigid contact: A comparison in the context of granular dynamics," *Physical Review E*, vol. 96, no. 4, p. 042905, 2017.

[72] A. Pazouki, H. Mazhar, and D. Negrut, "Parallel collision detection of ellipsoids with applications in large scale multibody dynamics," *Mathematics and Computers in Simulation*, vol. 82, no. 5, pp. 879–894, 2012.

[73] G. L. Peterson and M. J. Fischer, "Economical Solutions for the Critical Section Problem in a Distributed System (Extended Abstract)," in *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing*, ser. STOC, 1977, p. 91–97.

[74] G. Peterson, "Myths about the mutual exclusion problem," *Information Processing Letters*, vol. 12, no. 3, pp. 115 – 116, 1981.

[75] S. Puthoor, X. Tang, J. Gross, and B. M. Beckmann, "Oversubscribed Command Queues in GPUs," in *Proceedings of the 11th Workshop on General Purpose GPUs*, ser. GPGPU-11, 2018, p. 50–6.

[76] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR, 2017, pp. 7263–7271.

[77] ——, "YOLOv3: An Incremental Improvement," *CoRR*, vol. abs/1804.02767, 2018. [Online]. Available: http://arxiv.org/abs/1804.02767

[78] X. Ren, D. Lustig, E. Bolotin, A. Jaleel, O. Villa, and D. Nellans, "HMG: Extending Cache Coherence Protocols Across Modern Hierarchical Multi-GPU Systems," in *26th IEEE International Symposium on High Performance Computer Architecture*, ser. HPCA, 2020, pp. 582–595.

[79] I. Rhee, "Optimizing a FIFO, scalable spin lock using consistent memory," in *17th IEEE Real-Time Systems Symposium*, ser. RTSS, 1996, pp. 106–114.

[80] M. C. Rinard, "Effective Fine-Grain Synchronization for Automatically Parallelized Programs Using Optimistic Synchronization Primitives," *ACM Transactions on Computer Systems*, vol. 17, no. 4, p. 337–371, Nov. 1999.

[81] M. A. Rivas and M. G. Harbour, "Evaluation of new POSIX real-time operating systems services for small embedded platforms," in *Proceedings of the 15th Euromicro Conference on Real-Time Systems*, 2003, pp. 161–168.

[82] A. Rodchenko, A. Nisbet, A. Pop, and M. Luján, "Effective Barrier Synchronization on Intel Xeon Phi Coprocessor," in *Euro-Par 2015: Parallel Processing*, J. L. Träff, S. Hunold, and F. Versaci, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 588–600.

[83] R. A. Rossi and N. K. Ahmed, "The Network Data Repository with Interactive Graph Analytics and Visualization," in *AAAI*, 2015. [Online]. Available: https://networkrepository.com

[84] L. Rudolph and Z. Segall, "Dynamic Decentralized Cache Schemes for Mimd Parallel Processors," in *Proceedings of the 11th Annual International Symposium on Computer Architecture*, ser. ISCA '84, 1984, p. 340–347.

[85] G. Salvador, W. H. Darvin, M. Huzaifa, J. Alsop, M. D. Sinclair, and S. V. Adve, "Specializing Coherence, Consistency, and Push/Pull for GPU Graph Analytics," in *IEEE International Symposium on Performance Analysis of Systems and Software*, ser. ISPASS, 2020.

[86] S. Sengupta, M. Harris, Y. Zhang, and J. D. Owens, "Scan primitives for GPU computing," 2007.

[87] M. D. Sinclair, J. Alsop, and S. V. Adve, "Efficient GPU Synchronization without Scopes: Saying No to Complex Consistency Models," in *Proceedings of the 48th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, December 2015, pp. 647–659.

[88] ——, "Chasing Away RAts: Semantics and Evaluation for Relaxed Atomics on Heterogeneous Systems," in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, ser. ISCA, 2017, pp. 161–174.

[89] ——, "HeteroSync: A Benchmark Suite for Fine-Grained Synchronization on Tightly Coupled GPUs," in *IEEE International Symposium on Workload Characterization*, ser. IISWC, October 2017.

[90] I. Singh, A. Shriraman, W. W. L. Fung, M. O'Connor, and T. M. Aamodt, "Cache Coherence for GPU Architectures," in *19th International Symposium on High Performance Computer Architecture*, ser. HPCA, 2013, pp. 578–590.

[91] D. O. Son, C. T. Do, H. J. Choi, J. Nam, and C. H. Kim, "A Dynamic CTA Scheduling Scheme for Massive Parallel Computing," *Cluster Computing*, vol. 20, no. 1, p. 781–787, Mar. 2017.

[92] T. Sorensen, A. F. Donaldson, M. Batty, G. Gopalakrishnan, and Z. Rakamarić, "Portable Inter-Workgroup Barrier Synchronisation for GPUs," in *Proceedings of the 2016 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications*, ser. OOPSLA, 2016, pp. 39–58.

[93] T. Sorensen, S. Pai, and A. F. Donaldson, "One Size Doesn't Fit All: Quantifying Performance Portability of Graph Applications on GPUs," in *IEEE International Symposium on Workload Characterization*, ser. IISWC, November 2019.

[94] J. A. Stuart and J. D. Owens, "Efficient Synchronization Primitives for GPUs," *CoRR*, vol. abs/1110.4623, 2011.

[95] H. Takada and K. Sakamura, "Predictable spin lock algorithms with preemption," in *Proceedings of 11th IEEE Workshop on Real-Time Operating Systems and Software*, 1994, pp. 2–6.

[96] P. Tang and P.-C. Yew, "Software combining algorithms for distributing hot-spot addressing," *Journal of Parallel and Distributed Computing*, vol. 10, no. 2, pp. 130 – 139, 1990.

[97] A. Tasora, R. Serban, H. Mazhar, A. Pazouki, D. Melanz, J. Fleischmann, M. Taylor, H. Sugiyama, and D. Negrut, "CHRONO: An open source multi-physics dynamics engine," in *High Performance Computing in Science and Engineering – Lecture Notes in Computer Science*, T. Kozubek, Ed. Cham: Springer, 2016, pp. 19–49.

[98] S. Tokui, K. Oono, S. Hido, and J. Clayton, "Chainer: A Next-Generation Open Source Framework for Deep Learning," in *Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NeurIPS)*, vol. 5, 2015, pp. 1–6.

[99] Y. Ukidave, X. Li, and D. Kaeli, "Mystic: Predictive Scheduling for GPU Based Cloud Servers Using Machine Learning," in *IEEE International Parallel and Distributed Processing Symposium*, ser. IPDPS, 2016, pp. 353–362.

[100] O. Villa, M. Stephenson, D. Nellans, and S. W. Keckler, "NVBit: A Dynamic Binary Instrumentation Framework for NVIDIA GPUs," in *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO, 2019, p. 372–383.

[101] Y. Wang, A. Davidson, Y. Pan, Y. Wu, A. Riffel, and J. D. Owens, "Gunrock: A High-Performance Graph Processing Library on the GPU," in *Proceedings of the 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, ser. PPOPP, 2016.

[102] Z. Wang, J. Yang, R. Melhem, B. Childers, Y. Zhang, and M. Guo, "Simultaneous Multikernel GPU: Multi-tasking Throughput Processors via Fine-Grained Sharing," in *IEEE International Symposium on High Performance Computer Architecture*, ser. HPCA, 2016, pp. 358–369.

[103] Z. Wang, J. Yang, R. Melhem, B. Childers, Y. Zhang, and M. Guo, "Quality of Service Support for Fine-Grained Sharing on GPUs," in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, ser. ISCA '17, 2017, pp. 269–281.

[104] B. Wu, X. Liu, X. Zhou, and C. Jiang, "Flep: Enabling flexible and efficient preemption on gpus," in *Proceedings of the Twenty-Second International Conference on Architectural Support for Programming*

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Parallel and Distributed Systems. 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/TPDS.2022.3218508

Languages and Operating Systems, ser. ASPLOS '17, 2017, pp. 483–496.

[105] P. Xiang, Y. Yang, and H. Zhou, "Warp-level divergence in GPUs: Characterization, impact, and mitigation," in *20th IEEE International Symposium on High Performance Computer Architecture*, ser. HPCA, 2014, pp. 284–295.

[106] S. Xiao and W. Feng, "Inter-Block GPU Communication via Fast Barrier Synchronization," in *IEEE International Parallel and Distributed Processing Symposium*, ser. IPDPS, 05 2010, pp. 1 – 12.

[107] Q. Xu, H. Jeon, K. Kim, W. W. Ro, and M. Annavaram, "Warped-Slicer: Efficient Intra-SM Slicing through Dynamic Resource Partitioning for GPU Multiprogramming," in *ACM/IEEE 43rd Annual International Symposium on Computer Architecture*, ser. ISCA, 2016, pp. 230–242.

[108] L. Zhang, M. Wahib, H. Zhang, and S. Matsuoka, "A Study of Single and Multi-device Synchronization Methods in Nvidia GPUs," in *IEEE International Parallel and Distributed Processing Symposium*, ser. IPDPS, 2020, pp. 483–493.

[109] F. Zhu, J. Pool, M. Andersch, J. Appleyard, and F. Xie, "Sparse Persistent RNNs: Squeezing Large Recurrent Networks On-Chip," in *Proceedings of 6th International Conference on Learning Representations*, ser. ICLR, 2018.

Preyesh Dalmia is a PhD student in the Electrical Engineering department at the University of Wisconsin-Madison. His research interests include synchronization bottlenecks in GPGPUs from HW/SW co-design perspective. Preyesh has a Bachelors in Technology in Electronics and Communication Engineering from Delhi Technological University.

Rohan Mahapatra is a PhD student in the Computer Science and Engineering department at the University of California San Diego. His research interests include GPGPUs, hardware accelerators and Systems for Data Centers. Rohan has a MS in Electrical Engineering from the University of Wisconsin-Madison.

Jeremy Intan is a second year graduate student in the Computer Sciences Department at the University of Illinois at Urbana-Champaign. His research interests include computer architecture and robotics. Intan has a BS in Computer Engineering from the University of Wisconsin-Madison.

Dan Negrut is the Bernard A. and Frances M. Weideman Professor of Mechanical Engineering, with joint appointments in Computer Science and the Department of Electrical and Computer Engineering. His research is in high performance computing as applied in computational dynamics; robotics and autonomous vehicle modeling and simulation; and fluid-solid interaction problems. Negrut has a PhD in Mechanical Engineering from the University of Iowa. Since 2010, he is an NVIDIA CUDA Fellow.

Matthew D. Sinclair is an Assistant Professor in the Computer Sciences Department at the University of Wisconsin-Madison, with a joint appointment in the Department of Electrical and Computer Engineering. His research interests include synchronization, simulation, and efficient general-purpose accelerator design. Sinclair has a PhD in Computer Science from the University of Illinois at Urbana-Champaign, was a post-doctoral researcher in the Compute-Oriented Graphics Research Group at AMD Research, and is an IEEE Member.

]

© 2022 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 November 03,2022 at 05:55:55 UTC from IEEE Xplore. Restrictions apply.

arXiv:2211.09683v1 [cs.SI] 17 Nov 2022

INFLUENCE MAXIMIZATION IN SOCIAL NETWORKS USING DISCRETIZED HARRIS' HAWKS OPTIMIZATION ALGORITHM AND NEIGHBOR SCOUT STRATEGY

Inder Khatri*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
inderkhatri999@gmail.com

Arjun Choudhry*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
choudhry.arjun@gmail.com

Aryaman Rao*
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
aryaman26601@gmail.com

Aryan Tyagi
Department of Mechanical Engineering
Delhi Technological University
New Delhi, India
tyagiaryan82@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
dinesh@dtu.ac.in

Mukesh Prasad
School of Computer Science
University of Technology Sydney
Ultimo, Australia
mukesh.prasad@uts.edu.au

November 18, 2022

ABSTRACT

Influence Maximization (IM) is the task of determining k optimal influential nodes in a social network to maximize the influence spread using a propagation model. IM is a prominent problem for viral marketing, and helps significantly in social media advertising. However, developing effective algorithms with minimal time complexity for real-world social networks still remains a challenge. While traditional heuristic approaches have been applied for IM, they often result in minimal performance gains over the computationally expensive Greedy-based and Reverse Influence Sampling-based approaches. In this paper, we propose the discretization of the nature-inspired Harris' Hawks Optimisation meta-heuristic algorithm using community structures for optimal selection of seed nodes for influence spread. In addition to Harris' Hawks' intelligence, we employ a neighbour scout strategy algorithm to avoid blindness and enhance the searching ability of the hawks. Further, we use a candidate nodes-based random population initialization approach, and these candidate nodes aid in accelerating the convergence process for the entire populace. We evaluate the efficacy of our proposed DHHO approach on six social networks using the Independent Cascade model for information diffusion. We observe that DHHO is comparable or better than competing meta-heuristic approaches for Influence Maximization across five metrics, and performs noticeably better than competing heuristic approaches.

Keywords Influence Maximization · Harris' Hawks Optimization Algorithm · Social Networks · Meta-heuristics · Complex Networks · Community Structure

1 Introduction

Over the years, Social Networks (SNs) have gained immense popularity in our modern-day society. Large Social Networks like Twitter, Facebook, and Instagram, which have billions of users and connections, serve as viable communication tools as well as important marketing platforms for advertisers due to their significant user base. Recent studies (Leskovec, Adamic and Huberman, 2006; Peng, Zhou, Cao, Yu, Niu and Jia, 2018; Brown and Reingen, 1987)

*Equal Contribution

show that people tend to believe and trust the recommendations by individuals in their social circles, rather than the ads from more conventional means like television advertisements, billboards, etc. This 'word-of-mouth' approach tends to make information have a cascading effect to reach a wider range of people, gradually increasing the oral publicity effect. Utilizing this concept, advertisers choose a set of initial influencers and make them promote their product through 'word-of-mouth' to produce the largest cascade spread possible and influence a wide demographic of people on social networks. This approach is commonly known as viral marketing. Therefore, choosing the optimal initial users so that the information is propagated as widely in the network as possible is the core objective of influence maximization as a task.

Influence Maximization (IM) can be formally defined as the task of identifying the optimal set of nodes in a network that, when activated, can maximize the spread of information over the entire social network. Domingos and Richardson (2001) were the first group of researchers who studied IM. They introduced a novel solution by modeling the network as a Markov random field. Kempe, Kleinberg and Tardos (2003) officially defined the problem of IM, proving that, when simulated using Independent Cascade (IC) and Linear Threshold (LT) models, IM behaves like an NP-hard problem. They further introduced a hill-climbing strategy-based greedy approach to identify optimal initial spreader nodes for maximum information spread. Their approach was evaluated for hundreds of Monte-Carlo simulations for every node, making it computationally expensive. This seriously restricts its usability in larger networks.

To minimize the computational cost and space complexity, the optimality of IM had to be compromised. Later, certain approaches were proposed to meet the required purpose of less complexity and great accuracy. Leskovec, Krause, Guestrin, Faloutsos, VanBriesen and Glance (2007c) presented *Cost-Efficient Lazy Forward (CELF)*, an enhanced greedy approach that reduced the computational cost of the greedy algorithm by approximately 700 times. Chen, Wang and Yang (2009) further introduced two novel algorithms called NewGreedy and MixGreedy, which brought substantial improvements over Kempe et al. (2003)'s greedy approach. While NewGreedy and MixGreedy brought substantial computational cost reductions as compared to the approach introduced by Kempe et al. (2003), they still were unviable for large networks.

Jiang, Song, Cong, Wang, Si and Xie (2011) proposed the Expected Diffusion Value (EDV) approach, which was used to approximate the spread of influence in the Independent Cascade model, calculated using Monte Carlo simulations. They further used a simulated annealing-based approach to optimize EDV, which was found to give substantially better results with millions of iterations. Their experiments showed that the EDV algorithm is more accurate and 100 to 1000 times faster than the greedy algorithm. Li, Wang, Zhang, Zhou, Chu and Wu (2017) further proposed an alternative method for Simulated Annealing and two heuristic algorithms. They showed that their simulated annealing algorithm led to significantly faster convergence, making their approach more applicable to larger networks.

Some recent approaches for IM have also incorporated Community Structures (Venkatakrishna Rao and Chowdary, 2022; Li, Cheng, Su and Sun, 2018; He, Wang, Mao, Lv, Cai, Huang and Xu, 2019; Khomami, Rezvanian, Meybodi and Bagheri, 2021), and depend on the proposition that a social network can be divided into groups based upon how many communities can be made in that network. A Community structure is a group of interconnected in close vicinity. A closed and clustered community is generally easier to completely influence than one which is open and sparsely connected. It is much easier to spread influence rapidly in a community with only a few seed individuals. These community structures are found to be more efficient than other centrality measure-based approaches and can be applied more efficiently to larger social networks. In this work, we propose the discretization of Harris' Hawks Optimization (DHHO) algorithm with community structures for Influence Maximization. We employ the Louvain algorithm for detecting significant communities in the networks, which are further used by the DHHO algorithm to detect optimal nodes for maximum influence spread. We also introduce a novel neighbor scout strategy and an adapted randomized population initialization approach for improved performance. We evaluate DHHO against seven novel approaches on five metrics across six datasets. Experimental results verify DHHO performs comparably or better than competing meta-heuristic approaches for IM. DHHO further outperforms competing heuristic and centrality measures-based approaches. Our contributions are summarized below.

- We introduce DHHO, a discretized Harris' hawks optimization algorithm to optimize the Local Influence Estimator (LIE) function for Influence Maximization, using the intelligence of the hawk population. We propose the approach to maximize the expected influence spread by a set of extracted seed nodes, across any given online social network. We have employed a random population initialization strategy, along with a novel neighbor scout strategy to accelerate the process of convergence and achieve further improved results.
- We employ community structures for narrowing down the search space for our DHHO algorithm to find optimal candidate nodes, thus increasing the efficiency of our approach as well as reducing the execution time. Community clusters aid in information propagation within the community and reduce the time complexity significantly. We detect community structures in social networks using the Louvain Algorithm and assign a seed node budget to each significant community on the basis of its size and degree centrality. Further,

a candidate pool of optimal nodes is created for the purpose of influence maximization using the DHHO algorithm.

- We evaluate DHHO on six real-world social networks and compare it with seven baseline approaches across five different performance metrics. Our experimental results verify that DHHO outperforms competing meta-heuristic approaches, as well as the other novel heuristics algorithms and centrality measures. Statistical tests further verify our results, proving the superiority of DHHO over competing approaches.

The remaining paper is structured as follows: Section 2 delineates the existing research works distributed by the type of algorithm employed by the said approaches. Section 3 contains the preliminaries needed for our approach. Section 4 contains our proposed approach and algorithms. Section 5 contains information about various experimental details like the datasets used, the performance metrics, and the implementation environment. Section 6 consists of our experimental results, their analysis, and broad outcomes. Section 7 elucidates our statistical tests to evaluate the performance of our approach. Section 8 contains our concluding statement.

2 Related Works

Since Domingos and Richardson (2001) treated Influence Maximization as an algorithmic task using Markov Random Field Theory, various approaches have been introduced for Influence Maximization in social networks. These can be broadly classified into four classes, which have been discussed below in significant detail.

2.1 Greedy Algorithms-based Approaches

Following Domingos and Richardson (2001), Kempe et al. (2003) showed that Influence Maximization can be simulated using information diffusion models like Linear Threshold (LT), Independent Cascade (IC), and Weighted Independent Cascade (WIC), implying that the task of selecting the most influential nodes in a network is an NP-hard problem. Thus, Kempe et al. (2003) proposed a hill-climbing search strategy-based greedy algorithm on the sub-modular function to find an optimal solution within a factor of $1 - 1/(e - \epsilon)$. Here, e represents the natural logarithm, while ϵ is the error obtained due to the Monte Carlo simulation. One major drawback of greedy algorithms is that they need to choose the best k seed nodes in the network iteratively. They select a single seed node at a time, and then the remaining nodes in the network need to be traversed. Thus, tens of thousands of Monte-Carlo simulations are required to accurately assess the spread of influence due to a given seed node set. This makes the greedy algorithm computationally expensive.

The high time complexity of greedy algorithms severely limits their ability to deal with large social networks, and thus they are primarily used for small networks with lesser nodes. To overcome these barriers, Leskovec et al. (2007c), by exploiting the sub-modularity property, stored nodes with maximum marginal gain in a priority queue, and introduced the Cost-Effective Lazy-Forward (CELF) algorithm for finding the optimal seed nodes more efficiently. CELF was found to achieve solution accuracy similar to traditional greedy algorithms, while reducing the number of Monte-Carlo simulations by nearly 700 times when evaluating the marginal gain of nodes. Goyal, Lu and Lakshmanan (2011) later proposed the CELF++ strategy to reduce the time complexity of CELF, and was found to be up to 55% faster than CELF.

Chen et al. (2009) introduced two novel algorithms called NewGreedy and MixGreedy. The NewGreedy algorithm reduces the time complexity of the algorithm by scraping those edges from the graph which don't contribute to the dissemination of information from the original network to construct a smaller network. The MixGreedy algorithm, on the other hand, integrates the ideas from NewGreedy and CELF to further boost the efficiency of the algorithm. Cheng, Shen, Huang, Zhang and Cheng (2013) later presented StaticGreedy, a new greedy algorithm that stored the static information of Monte-Carlo simulations within the network, decreasing the Monte-Carlo simulations by about two orders of the total magnitude.

Heidari, Asadpour and Faili (2015) proposed a machine greedy algorithm, that evaluates the magnitude of traversing nodes during the estimation propagation process. They further constructed a Monte Carlo graph in the simulation process. Lu, Zhou and Wu (2016) presented a novel approach to evaluate the spread of influence in the network recursively through the utilization of the reachability probability between the nodes. By applying the greedy strategy, nodes with the highest estimate were recursively selected as seed nodes in the network. Contrary to more conventional greedy algorithms, the computational cost of these algorithms is greatly diminished. However, these approaches are still computationally expensive, especially for larger networks.

2.2 Reverse Influence Sampling-Based Algorithms

Borgs, Brautbar, Chayes and Lucier (2012) introduced Reverse Influence Sampling (RIS) for the selection of influential nodes in a network, an approach derived from the random sampling theory. It generates $1 - 1/(e - \epsilon)$ viable solutions with a computational cost of $(m + n)\epsilon - 3logn$. RIS works on the fundamental idea that if a node frequently occurs in the randomly generated sub-graphs, then it has a higher probability of being a seed node. Tang, Xiao and Shi (2014) later proposed TIM and TIM+ algorithms to improve upon the RIS algorithm with greater algorithmic efficiency and significantly lesser time complexity by introducing improved sampling mechanisms. Tang, Shi and Xiao (2015) later introduced IMM, a novel approach derived from the Martingale estimation. IMM generates similar optimal solutions as TIM+ and TIM while having significantly higher efficiency, thus mitigating their drawbacks.

Nguyen, Thai and Dinh (2016) recently introduced SSA and D-SSA, two RIS-based algorithmic approaches, with 1200 times the computational efficiency of IMM. They further adhere to the same $1 - 1/(e - \epsilon)$ approximation guarantee. While the IMM algorithm has nearly linear time complexity, its practical sampling time noticeably exceeds the threshold for the minimum sampling time in actual operation. Since SSA and D-SSA are derived from the stop-and-stare strategy, they alleviate the shortcomings of IMM. Both of these approaches exhibited the same approximate optimal result guarantee as that of IMM but with around 1200 times lesser execution time.

Cohen, Delling, Pajor and Werneck (2014) developed SKIM, a new greedy Sketch-based algorithm based on building combined reachability sketches. While it maintains the same approximation guarantee as that of TIM+, it can easily be applied to extremely large networks with billions of nodes and edges. Some other sketch-based RIS frameworks that were proposed are BKRIS (Wang, Zhang, Zhang, Lin and Chen, 2017) and RSRS (Kim, Hyeon, Oh, Han and Yu, 2017). They were introduced to ameliorate the drawbacks of SKIM and TIM+. While RIS-based approaches typically exhibit lesser computational costs as compared to greedy approaches, they sometimes return sub-optimal results on increasing the number of seeds.

2.3 Heuristics Methods

Recent works on IM have introduced a variety of heuristic approaches involving node centrality (Berahmand, Bouyer and Samadi, 2019; Samadi and Bouyer, 2019). These involve a simple protocol to explicitly choose the top k nodes out of the set of total nodes when ranked according to the given centrality methods. Some prominent Heuristic approaches are Closeness Centrality (Okamoto, Chen and Li, 2008), Betweenness centrality (Estrada and Rodríguez-Velázquez, 2005), Degree centrality (Freeman, 1978), and PageRank (Brin and Page, 1998).

Over the years, more advanced heuristic algorithms have been introduced by researchers. Kitsak, Gallos, Havlin, Liljeros, Muchnik, Stanley and Makse (2010) proposed K-shell decomposition, a global centrality measure, which divides the nodes into different levels on the basis of their location with respect to the core of the network. Kundu, Murthy and Pal (2011) proposed the Diffusion Degree algorithm using the two-hop neighbourhood region to gauge a node's influence. Bae and Kim (2014) proposed an improved algorithm called Neighbourhood Coreness measure, which ranks nodes on the basis of the sum of the K-shell of the neighbourhood nodes. This concept was further extended to get Extended Neighbourhood Coreness by taking the sum of the Neighbourhood Coreness values of a node's neighbours. Zeng and Zhang (2013) introduced Mixed Degree Decomposition (MDD), an iteration of K-shell that further distinguishes nodes in the same shell by using the removed and residual neighbours during the K-shell decomposition process. Inspired by the concept of gravity, Ma, Ma, Zhang and Wang (2016) introduced GC-Index (gravity centrality index) for IM. Centrality-measures-based heuristic algorithms aid in the selection of influential nodes efficiently. However, they sometimes suffer from low accuracy and stability.

Community structures often hide subtly in online social networks. Nodes belonging to the same community are often closely connected with each other. Consequently, nodes from different communities have relatively sparse connections (Newman, 2006). Thus, information propagation is much easier within a community structure (Lin, Hu, Wang and Yu, 2015). Some of the most relevant community-based IM approaches introduced in the last few years are CAOM (He et al., 2019), CFIN (Khomami et al., 2021), and GLR (Salavati, Abdollahpouri and Manbari, 2018). The community-based works focus on finding out the core nodes in the significant communities, and further leverage the higher connection density of communities to increase the spread using these nodes as initial spreaders.

Chen, Zhu, Peng, Lee and Lee (2014) tackled IM using community structures in the heat propagation model. They introduced the CIM approach, which performed very well for IM. However, its efficacy in other common information diffusion models remains to be tested. While IM approaches based on community structures are reasonably efficient, they frequently use infection propagation of nodes amongst their own communities to approximate the node influence in the whole social network. Thus, their accuracy is critically dependent on the validity of the community partitions.

2.4 Metaheuristic Methods

Recently, researchers have applied evolutionary optimization algorithms for IM. For these, a fitness function needs to be defined. Jiang et al. (2011) introduced the Expected Diffusion Value (EDV), which is shown in Eq. 1.

$$EDV(S) = k + \sum_{v \epsilon N_S^{(1)} \setminus S} (1 - (1-p)^{r(v)}) \tag{1}$$

Here, $N_S^{(1)}$ signifies the one-hop area of the candidate set S, $N_S^{(1)}$ signifies the nodes in the one-hop area except for the candidate nodes S, $r(v)$ is the number of edges connected to node v in $N_S^{(1)}$, and p represents the activation probability.

Jiang et al. (2011) further adopted the simulated annealing approach to identify a seed set S that maximizes the EDV fitness function. Their experiments showed that the algorithm is more accurate and 100 to 1000 times faster as compared to the greedy algorithm.

Gong, Yan, Shen, Ma and Cai (2016) introduced a Local Influence Estimator (LIE) function, which evaluates a set S's influence on its neighbours within a two-hop area. They introduced a Discrete Particle Swarm Optimization (DPSO) framework to maximize the LIE fitness function. We delineate the LIE function in Section 3.4. Cui, Hu, Yu, Yan, Ming, Wen and Lu (2018) proposed the Degree-Descending Search Evolution (DDSE) algorithm for IM. It utilises the EDV function to overcome the inefficiencies shown by the greedy algorithm. They underscored the reduced computational cost of DDSE as compared to greedy approaches in their experimental results. Zareie, Sheikhahmadi and Jalili (2020) used the Gray Wolf optimization algorithm for IM, along with a new fitness function inspired by the concept of entropy.

The Discrete Bat Optimization algorithm for IM was introduced by Tang, Zhang, Yao, Zhao, Wang, Li and Yuan (2018). Ma et al. (2016) pointed out that the LIE fitness function ignores interactions between neighbours in the same-hop surroundings from the seed node. They introduced a three-layer influence evaluation approach and adopted an improved bee colony optimization framework for the maximization of the three-layer evaluation model. The biggest advantage of using meta-heuristic algorithms is that running the Monte-Carlo simulations can be avoided. This increases efficiency and reduces the time complexity of the problem. However, it is crucial that the evolution mechanism chosen is effective in producing a high-quality and high-efficiency solution.

3 Preliminary Information

3.1 Influence Maximization

IM, as defined by Kempe et al. (2003), is the process of choosing k initial seed nodes in a social network graph G, such that when these nodes are considered as influential spreader nodes, the information diffusion over the entire network is maximum for any association of N users (Kempe et al., 2003). The seed set S of most influential nodes contains k elements $(k \leq n)$. The influence spread from a set of seed nodes S is maximum at the end of the diffusion process and is denoted as $\sigma(S)$. The aim is to increase information diffusion and expand the set of nodes influenced. Generally, the value of k is minuscule as compared to N, considering the assets and time constraints. Mathematically, IM can be defined as in Eq. 2.

$$S = \arg\max \sigma(S) \; where \; S \subset V, |S| = k \tag{2}$$

3.2 Information Diffusion Models

Information Diffusion (ID) can be defined as the process of propagating information in a social network (Guille, Hacid, Favre and Zighed, 2013; Chen, Lakshmanan and Castillo, 2013). It aids in evaluating various spreader-ranking methods. To model information propagation, usually, epidemic models are used. In a network, nodes act as the population, while the information propagation is represented as an infectious disease traversing through the population on interaction. The Linear Threshold (LT) (Granovetter, 1978), Weighted Independent Cascade (WIC) (Wang, Wang and Li, 2015), and Independent Cascade (IC) models (Goldenberg, Libai and Muller, 2001) are some of the most common diffusion models used to simulate information propagation through a network. We employ the Independent Cascade model for simulating information diffusion in our evaluations.

In the Independent Cascade model, every node is either in an infected (or active) state or in a susceptible (or inactive) state. An active node would be a user who has been influenced by or accepted new ideas. Users who have not been influenced by the idea yet are inactive nodes. An active node cannot switch to an inactive node, however, the vice-versa is possible. The edges in the network are assigned probabilities that correspond to the possibility of an active node activating its neighboring inactive nodes.

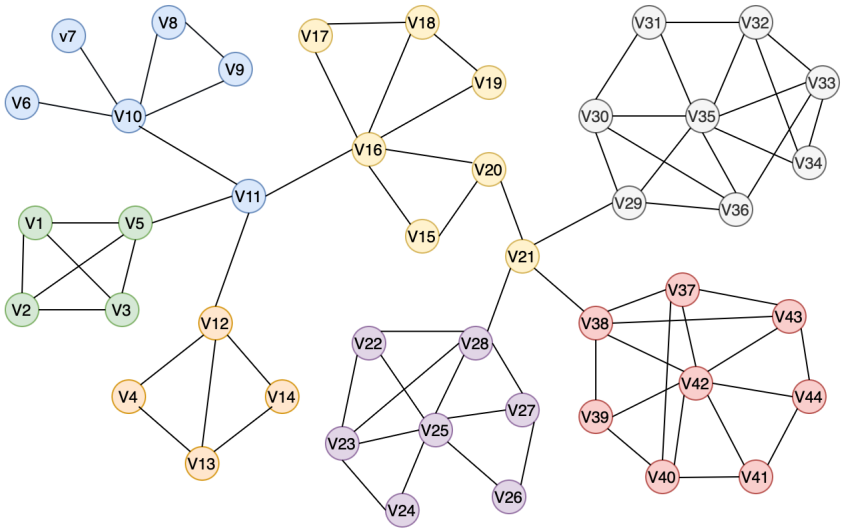


Figure 1: Illustration showing the detection of communities using the Louvain algorithm.

3.3 Community Structure

Community structures often convey prominent features in most complex social networks. They indicate a part of a network with similar characteristics and behavior. They are characterized by the fact that the connections inside a community are noticeably denser than those between different communities (Girvan and Newman, 2002; Venkatakrishna Rao and Chowdary, 2022). Communities in social networks can exist due to various reasons, including but not limited to location, common interests, or occupation.

Detecting communities in complex networks makes it easier to study and analyze the underlying features in the network. It further provides insight into the functioning of systems that are represented by social networks. For example, in metabolic networks, communities represent cycles and pathways. Another instance is the existence of communities in a citation network, representing the group of research papers on the same topic or citing the same article. Community structures are also found in the network of web pages, where each community corresponds to a group of links on a similar topic. Another prominent use of community structures in graphs is that communities often have properties that greatly differ from the standard characteristics of the entire network. In such cases, analyzing only these general characteristics can result in the negligence of important and relevant features. For example, in a social network, both sociable and reserved individuals might exist. Figure 1 shows communities that are identified in social networks using the Louvain algorithm.

We employ community structures in our approach in the following way:

- First, the graph is divided into its constituent communities using the Louvain algorithm (Blondel, Guillaume, Lambiotte and Lefebvre, 2008) for community detection. It is an efficient approach for community detection previously used in a variety of IM works (Huang, Shen, Meng, Chang and He, 2019; Shang, Wu, Zhou, Zhong, Feng and Qiang, 2018; Gong et al., 2016).
- Then, the communities detected by the Louvain algorithm are detached from each other by eliminating the connections between different communities.
- We define a threshold, and select only those communities larger than the defined threshold. These communities are known as significant communities.
- We then assign a seed node budget to each significant community, depending on their size.
- The communities are input into the DHHO algorithm to obtain k most influential seed nodes.

3.4 Influence Estimator

The Local Influence Estimator (LIE) function is used to approximate the spread of influence in a two-hop area around a set of nodes S. The set of seed nodes that achieves the maximum value for the LIE function is the optimal seed node set.

For the maximization task, the Harris' Hawks optimization algorithm is employed, which has not been used before for the problem of influence maximization. The Three Degree Theory states that a person's influence over his/her's friends gradually dissipates over 3 degrees of separation. Thus, a node has significant influence only over a three-hop area around it. According to Pei, Muchnik, , Zheng and Makse (2015), a node's influence in its three-hop area is a valid indicator of its influence over the entire network. In our study, we use the Local Influence Estimator (LIE) function, which is a two-hop area influence estimator developed by (Gong et al., 2016). It helps quantify the influence spread for a node set S in its two-hop surroundings. It is calculated as shown in Eq. 3.

$$\begin{aligned} LIE(s) &= \sigma_0(S) + \sigma_1^*(S) + \tilde{\sigma_2}(S) \\ &= k + \sigma_1^*(S) + \frac{\sigma_1^*(S)}{|N_S^{(1)} \backslash S|} \sum_{u \in N_S^{(2)} \backslash S} p_u^* d_u^* \\ &= k + (1 + \frac{1}{|N_S^{(1)} \backslash S|} \sum_{u \in N_S^{(2)} \backslash S} p_u^* d_u^*) \sigma_1^*(S) \\ &= k + (1 + \frac{1}{|N_S^{(1)} \backslash S|} \sum_{u \in N_S^{(2)} \backslash S} p_u^* d_u^*) \sum_{i \in N_S^{(2)} \backslash S} (1 - \prod_{(i,j) \in E, j \in S} (1 - p_{ij})) \end{aligned} \tag{3}$$

where $\sigma_0(S)$ represents the initial set of seed nodes, $\sigma_1^*(S)$ and $\sigma_2\tilde{(S)}$ represent the expected one-hop and two-hop information propagation of the node set S. p_u^* represents the diffusion model's propagation probability. d_u^* represents the number of edges of node u with nodes in $N_S^{(1)}$ and $N_S^{(2)}$. We achieve the maximization of the above objective function the Harris-Hawks optimization algorithm. Thus, IM is transformed into an objective function optimization problem.

3.5 Harris' Hawks Optimization

Harris' Hawks Optimisation (HHO) (Heidari, Mirjalili, Faris, Aljarah, Mafarja and Chen, 2019) algorithm is a novel gradient-free, population-based meta-heuristic approach applicable to a variety of optimization tasks subjected to a proper formulation of the objective function. The Harris' Hawk Optimisation (HHO) is inspired by the explorative nature of the prey, as well as different attacking strategies like the surprise pounce of Harris' hawks. This optimization technique can be divided into two phases: the explorative phase and the exploitative phase.

In the exploration mechanism, the hawks use their powerful eyes to detect and track the prey. However, occasionally the prey is not visible to the hawks. Hence, the hawks patiently wait and observe the deserted sites to detect their prey. Once their prey is detected, they switch to their exploitation phase and plan their pounce accordingly. The prey (usually a rabbit) represents the best candidate solution, while the hawks are the other candidate solutions. The Hawks strategize their exploitation technique depending upon the escaping energy and jump strength left with the prey. This energy and jump strength can be modeled as shown in Eq. 4 and Eq. 5, respectively. The value of E decreases upon increasing the number of iterations increases. E indicates the escaping energy of the prey, J indicates the jump strength, T represents the maximum number of iterations, and E_0 is the initial energy state. The exploitation phases can be broadly categorized into two varieties of pounces: soft besiege and hard besiege. The former focuses on following the prey until it has run out of its escaping energy and then finally encircling and attacking the prey, while the latter aims at ambushing the prey with a surprise pounce once it is exhausted.

$$E = 2E_0(1 - \frac{t}{T}) \tag{4}$$

$$J = 2 \times (1 - rand()) \tag{5}$$

4 Proposed Methodology

We introduce Discrete Harris' Hawks Optimisation (DHHO), our proposed approach for Influence Maximization. DHHO skillfully combines discretized Harris' hawks, community structures, and a neighbor scout strategy to extract optimal and most effective seed nodes for maximum information propagation in a given social network. We employ the Louvain algorithm to initialize several communities in social networks, which aids in information propagation and reduces the execution time complexity in our proposed approach. DHHO further utilizes an adapted random

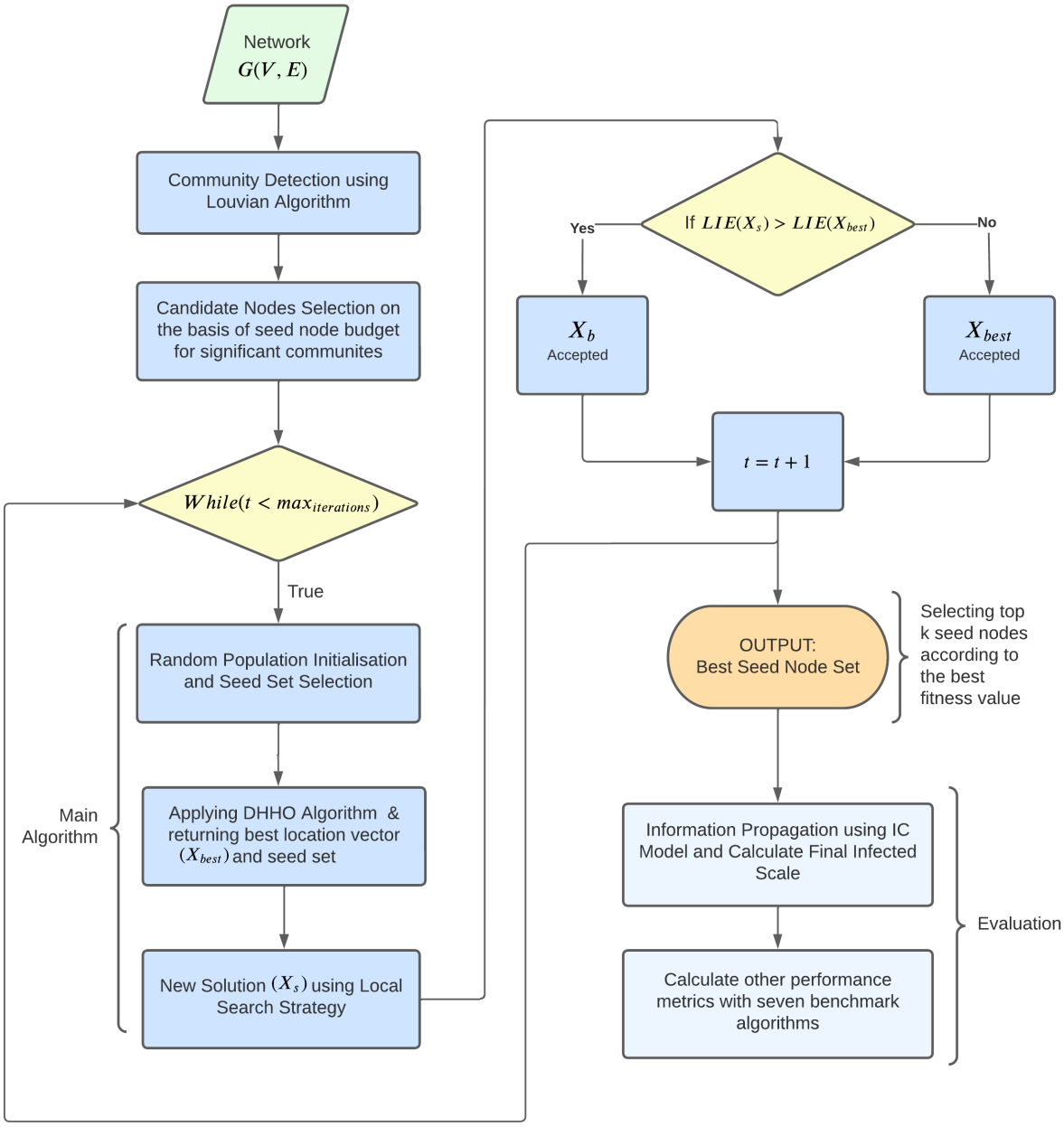


Figure 2: Flowchart representation of the steps involved in our proposed DHHO approach.

probability-based initialization strategy to increase the convergence of the hawk population at the early stages of Harris' hawk manipulation. Figure 2 shows the flow of process in our proposed DHHO framework for IM.

4.1 Proposed Algorithm

IM involves finding a group of initial spreader nodes in a network, which can maximize the influence spread across the network. This is an optimization problem, which we aim to solve using the Local Influence Estimator (LIE) function. The LIE function enables us to gauge the expected information diffusion for any set of nodes in a given network. Swarm and population intelligence-based algorithms utilize efficient evolution mechanisms for optimizing the objective

function. Here, we propose a discretized form of Harris' hawk optimization (DHHO) algorithm for IM. Since the solution space of the IM task is discrete, the unmodified Harris' hawk framework cannot be used for IM, as it is suitable only for continuous problems. Thus, we propose the redefinition of the solution vector in a discrete form. The rules for initializing the location vector of Harris' hawks are given below in Eq. 6.

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2X(t)|, & \text{if } q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)), & \text{if } q < 0.5 \end{cases} \tag{6}$$

$$X_m(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t) \tag{7}$$

Here, q is the probability of selecting different perching strategies for the hawks, $X(t+1)$ denotes the position vector of hawks in the upcoming iteration t, $X_{rabbit}(t)$ is the rabbit's position (best solution), and $X(t)$ is the current position vector of hawks. The variables r_1, r_2, r_3, r_4, and q are randomly chosen and have values between 0 and 1, which are updated in each iteration. LB and UB represent the lower and upper bounds of the variables. $X_{rand}(t)$ is a randomly selected hawk from the current populace, and the $X_m(t)$ is the mean position of all of the hawks from the current populace.

There are four different possible exploitation techniques. Let there be a variable r indicating whether a prey escapes the hawks. ($r < 0.5$) depicts the possibility of the prey escaping the hawks, while ($r \geq 0.5$) depicts the possibility of the hawks catching the prey.

- If the energy level $|E| \geq 0.5$ and $r \geq 0.5$, then the prey has sufficient energy to escape the hawks by some random, misleading movements. In this strategy, the Harris' hawks gather around the prey till it gets exhausted and then perform a pounce. This is called soft besiege.
- If the energy level $|E| < 0.5$ and $r \geq 0.5$, then the prey doesn't have sufficient energy to escape and is exhausted. The Harris' hawks simply perform a pounce to catch the prey. This strategy is called hard besiege. These two approaches are demonstrated mathematically below in Eq. 8 and Eq. 9.

$$X(t+1) = \Delta X(t) - E|JX_{rabbit}(t) - X(t)|, \Delta X(t) = X_{rabbit}(t) - X(t) \tag{8}$$

$$X(t+1) = X_{rabbit}(t) - E|\Delta X(t)| \tag{9}$$

- If the energy level $|E| \geq 0.5$ and $r < 0.5$, then the prey has sufficient energy to successfully escape the hawks by some leapfrog movement, which can be depicted mathematically using the Levy Flight function. In this condition, the Harris' hawks try to aim at the prey perfectly but do not manage to catch it. It makes multiple rapid dives considering the deceptive motions of the prey. This strategy is called soft besiege with rapid progressive dives.
- If the energy level $|E| < 0.5$ and $r < 0.5$, the prey is exhausted and has no energy left to escape. In this condition, a hard besiege is constructed before the surprise pounce to catch and kill the prey, at the same time decreasing the distance of the average location of the prey. This strategy is called hard besiege with rapid progressive dives. Eq. 10 and Eq. 11 correspond to both besieges, where S is a random vector of dimension $(1, D)$, while $LF(D)$ represents the Levy flight function, computed using Eq. 14. Eq. 13 signifies the updated location vector for both besieges with rapid progressive dives using Y obtained from Eq. 10 and 11 (for soft and hard besieges respectively), and Z from the Eq. 12.

$$Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X(t)| \tag{10}$$

$$Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X_m(t)| \tag{11}$$

$$Z = Y + S \times LF(D) \tag{12}$$

$$X(t+1) = \begin{cases} Y, & \text{if } F(Y) < F(X(t)) \\ Z, & \text{if } F(Z) < F(X(t)) \end{cases} \tag{13}$$

$$LF(x) = 0.01 \times \frac{u\sigma}{|v|^{(1/\beta)}}, \sigma = \left(\frac{\Gamma(1+\beta) \times sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{(\frac{\beta-1}{2})}} \right) \tag{14}$$

Our proposed method follows a unique discretization procedure for influence maximization, involving the judgment of candidate nodes on the basis of their selection probability. According to Zareie, Sheikhahmadi and Khamforoosh (2018), the nodes with degree 1 have minimal possibility of being selected as seed nodes. Thus, to reduce the computational cost of the DHHO algorithm and to make sure it does not get trapped with lesser-degree nodes in a loop, only nodes with degrees greater than one are considered possible seed nodes.

In DHHO, each hawk (solution) has two properties: a location vector and a corresponding seed set. The position of hawk i is represented using a vector with elements. The j^{th} index conveys the possibility of the node being selected as a seed node. The corresponding seed set of hawk i is shown as S_i. It contains k nodes with the highest probability of becoming the final seed nodes. The k nodes are then selected according to their decreasing probability. The optimal solution is chosen as the k seed set with the highest probability. This solution is then used to calculate the value of other hawk solutions X_i. Once X_i is updated according to the rules of Harris' hawk optimization, then their corresponding seed set values are ranked with respect to their fitness value. We use a neighbor scout strategy to choose an alternate seed set, which is then compared with the main solution. The best seed set is then selected as the global best solution. This process is repeated after each iteration, till the maximum number of iterations is reached.

Algorithm 1 Pseudocode of Discrete Harris' Hawks Optimisation

1: **Input:** Undirected Graph G = (V,E), seed set size k, population size N, number of maximum iterations T
2: **Output:** S ▷ Set of k nodes as initial seedset
3: Initialize E, J, and $|V'| = \{v_i \in V | d_i > 1\}$
4: Initialize X_1 to X_n randomly and determine seed set S_i corresponding to each X_i
5: Calculate fitness value for each $S_i (i = 1, 2, ...N)$
6: Assign the best solution as X_{best} (also known as X_{rabbit}) according to the fitness value
7: Set t = 0
8: **while** $(t < T)$ **do**
9: **for** each hawk i (except X_{best}) **do**
10: Update E and J using Eq. 4, 5
11: **if** $(|E| \geq 1)$ **then** ▷ Exploration phase
12: Update the location vector using Eq. 6
13: **end if**
14: **if** $|E| < 1$ **then** ▷ Exploitation phase
15: **if** $(r \geq 0.5$ and $|E| \geq 0.5)$ **then** ▷ Soft besiege
16: Update the location vector using Eq. 8
17: **else if** $(r \geq 0.5$ and $|E| \leq 0.5)$ **then** ▷ Hard besiege
18: Update the location vector using Eq. 9
19: **else if** $(r < 0.5$ and $|E| \geq 0.5)$ **then** ▷ Soft besiege
20: Update the location vector using Eq. 10,12,13
21: **else if** $(r < 0.5$ and $|E| < 0.5)$ **then** ▷ Hard besiege
22: Update the location vector using Eq. 11,12,13
23: **end if**
24: **end if**
25: Update X_i and S_i according to the values of E and q
26: **end for**
27: Calculate the best location vector
28: Calculate fitness value for each S_i
29: Find new solution using neighbor scout strategy as F_{new}
30: **if** $(F_{new} > Fitness_i)$ **then**
31: g_{best} = new solution
32: **end if**
33: $t = t + 1$
34: **end while**
35: **Return** X_{best}

The framework for our DHHO algorithm for IM is shown in Algorithm 1. In line 4, the primary solution vectors are generated according to the initialization strategy. A random position function generates each random solution. In line 5, the LIE function is used to compute the fitness value of each solution. The optimal solution X_{best} is determined using the fitness value of the solutions. In lines 8-34, the operations are iterated for the maximum number of iterations to maximize the fitness value. In each iteration, by using the update strategy and seed set selection function, the positions of the hawks are updated iteratively. Seed set S_i is determined for each hawk in line 25. Line 28 computes the fitness

value for each S_i, and based on the fitness value, the best hawk particle is chosen. In the end, a neighbor scout strategy is used to find nearby solutions which can perform better than the updated solutions. If the solution generated by the neighbor scout method has a greater fitness value than the original solution, then the former solution is picked accordingly. It aids in preventing being trapped in the local optimum solution.

4.2 Initialization

In our proposed DHHO algorithm, we employ an adapted random position initialization function (Zareie et al., 2020) to generate the initial solution for the hawk i and its associated seed set. The pseudo-code of this function is shown in Algorithm 2. Each node V_j ($V_j \subset V$) in the set V is accessed, and the position X_{ij} is assigned with the help of a random variable r and the degree of node V_j, as shown in lines 3-5. Then, the corresponding seed set is initialized for each node V_j, in lines 8-10. The final location vector and seed set values are returned.

Algorithm 2 Random position function

1: **Input:** Undirected Graph G = (V,E), V', the seed set size k
2: **Output:** X_i, the position of hawk i, and S_i, corresponding seed set
3: **for** j = 1 to $|V'|$ **do**
4: $r = rand(1...d_j)$
5: $X_{ij} = r/max(d)$
6: **end for**
7: Set $S_i = \{\}$
8: **for** i = 1 to k **do**
9: Find the next maximum X_{ij}
10: Add v_j to S_j
11: **end for**
12: Return X_i and S_i

4.3 Update Strategy

Let us take an example to explain the working of the update strategy of the location vector for DHHO clearly. Assuming that all the necessary constants are already defined, let us consider a network in which nodes 1, 5, 6, 9, 13, 15, 20, 26, 29, and 33 have degrees greater than 1. We have set the seed set number k as 3 ($k = 3$) while the nodes of the graphs chosen are $V' = \{1, 5, 6, 9, 13, 15, 20, 26, 29, 33\}$ and $|V'| = 10$. Consider the position vector assigned for each hawk i has been initialized according to the random initialization strategy as:

Node ID	1	5	6	9	13	15	20	26	29	33
X_{ij}	0.436	0.213	0.121	0.456	0.746	0.589	0.322	0.234	0.886	0.055

For each X_{ij}, the i^{th} index represents the population size, and j^{th} node represents the chance of the node in V to be selected as the seed node. For example, the chance of the 1^{st} node is 0.436, while it is 0.213 for the 5^{th} node and 0.161 for the 6^{th} node, and so on. The k nodes with the highest probability of being selected as the corresponding seed set, that is, $S_i = 13, 15, 29$.

Let us consider the max_{iter} = 10, and we have to update the location of hawk i for the 6^{th} iteration, i.e., t = 6. The value E (refer to Eq. 4) is calculated as $E = E_0(1 - \frac{6}{10})$ and $J = (1 - rand())$, where $E_0 = (2 \times rand() - 1)$. A random value is drawn from the domain $[0, 1]$ as r.

Let us suppose $r = 0.75$, $E = 0.2$ and $J = 0.25$. For this condition, $|E| < 1$, $|E| < 0.5$ and $rand() > 0.5$, hard besiege is activated. Calculating the value of X, we get $X(t) = X_{best} - E * (J * X_{best} - X(t))$.

$X(t)$ turns out to be 0.345. Similarly, this process is carried out till the max_{iter} is reached for all the hawks. Then, the seed set and fitness value are calculated for each hawk in the population accordingly. If the fitness value of the current solution exceeds the X_{rabbit} solution, they get replaced with the best global solution.

4.4 Neighbor Scout Strategy

We present a novel Neighbor Scout Strategy (NSS) that searches for a better solution vector by replacing a few nodes with their neighboring nodes. This strategy aims to find a new solution vector near the vicinity of the current location vector in the search space. It ensures that the solution does not get trapped in the algorithm's local optimum solution

of the search space. The nodes are replaced with their neighbor nodes only if the size of their one-hop neighbor is large enough. The original Harris' hawks algorithm tends to have a slower convergence speed for computationally demanding optimization tasks. Thus, this strategy further assists in accelerating the convergence speed of the proposed algorithm. The neighbor scout strategy (NSS) used in our proposed DHHO is illustrated in Algorithm 3.

Algorithm 3 Neighbor Scout Strategy (x_i,G, L)

1: **Input:** Graph G = (V,E), position x_i.
2: $d_{x_i} \leftarrow Degree(x_i)$
3: $X_i' \leftarrow Order(x_i, d_{x_i})$
4: $N \leftarrow Neighbours(X_i)$
5: **for** each element $X_{ij}' \in X_j'$ **do**
6: **for** each node N_i in $N_{X_{ij}'}^{(1)}$ **do**
7: **if** $r > 0.5$ and $len(N_{X_{ij}'}^{(1)}) > L$ **then**
8: $X_i' \leftarrow Replace(X_{ij}', N_i)$
9: **if** $LIE(X_i') > LIE(x_i)$ **then**
10: $x_i \leftarrow X_i'$
11: **end if**
12: **end if**
13: **end for**
14: **end for**
15: **Output:** The new position x_i

First, the current solution of nodes is sorted in ascending order based on their degree of centrality. This order ensures that the nodes of lower influence get replaced by more influential nodes in their neighborhood (as shown in lines 1-2). Each node is accessed iteratively and replaced by a node in its one-hop neighbor depending upon a probability r. The nodes are only replaced if the size of their neighborhood is bigger than a given threshold L, then the LIE fitness value for the whole solution vector is calculated. If the LIE value of the replaced solution vector is greater than the LIE value of the original vector, then the original node gets replaced (demonstrated in lines 5-8).

5 Datasets, Baseline Models and Performance Metrics

5.1 Datasets

We evaluate our DHHO approach on six social network datasets of varying domains and sizes. Table 1 provides various details about the datasets like the number of nodes, edges and communities detected by Louvain's algorithm. They are described below:

- Wiki-Vote (Leskovec, Huttenlocher and Kleinberg, 2010a): This dataset contains voting information from Wikipedia till January 2008. Each node signifies a user. Directed edges between node u and node v indicate that person u voted for person v.
- Power-law Cluster Graph (PCG) (Holme and Kim, 2002): PCG is a technique used for the generation of random graphs using approximate average clustering and power-law degree distribution. Parameters n, m, and p are required to generate a graph, where n represents the number of nodes in the graph, m denotes the number of edges per node in the graph, and p represents the probability of a triangle is created upon adding a new edge to the graph.
- Email-univ (Leskovec, Kleinberg and Faloutsos, 2007a): This dataset contains a network obtained from a research institution in Europe, using email data from October 2003 to May 2005. The nodes signify the users, while the directed edges in the graph denote the flow of emails.
- Jazz (Kunegis, 2013): The Jazz dataset contains a network denoting collaborations in the jazz music community collected in 2003. The nodes signify Jazz musicians, while edges represent collaborations between two Jazz musicians.
- Hamsterster (Hamsterster): The hamsterster dataset represents a network obtained from hamsterster.com. The nodes represent users on the website, while edges signify a friendship between the nodes.

Table 1: The different datasets used for experimentation along with the various characteristics such as the Number of Nodes, Edges, and Communities detected by the Louvain Algorithm

Dataset	Nodes (n)	Edges (m)	Communities
Wiki-Vote (Leskovec et al., 2010a) (Leskovec, Huttenlocher and Kleinberg, 2010b)	889	2914	40
PCG (Holme and Kim, 2002)	2000	9963	21
Jazz (Kunegis, 2013)	198	2,742	40
Email-univ (Leskovec et al., 2007a)	1100	5500	11
Hamsterster (Hamsterster)	2,400	16,600	169
p2p-Gnutella (Leskovec et al., 2007b)	8,846	31,839	23

- P2P-Gnutella (Leskovec, Kleinberg and Faloutsos, 2007b): This dataset contains snapshots of the Gnutella file-sharing network. The nodes indicate a host in the network, while edges signify connections between the hosts.

5.2 Baseline Approaches

We compare DHHO with seven existing baseline approaches for IM over five evaluation metrics for a comprehensive comparative analysis. A brief overview of these approaches is shown below.

- Extended Neighborhood Coreness (ENC) (Bae and Kim, 2014): ENC is a centrality measure used for IM. It is based on the K-shell decomposition (Kitsak et al., 2010). It extends over K-shell decomposition to a two-hop measure by taking the sum of both one and two-hop measures.
- Gateway Local Rank (GLR) (Salavati et al., 2018): GLR is an extension of the closeness centrality measure and simplifies it by reducing the search set to the local and gateway nodes.
- PageRank (PR) (Brin and Page, 1998): PageRank measures the significance of a given node in the network based on its outgoing degree and damping factor.
- H-index (HI) (Liu, Zhu, Jia, Deng, Zhou, Zhu and Zou, 2018): H-index is a node centrality measure that considers the neighborhood measures, the quality, as well as the quantity of the neighboring nodes simultaneously.
- Discrete Particle Swarm Optimisation (DPSO) (Gong et al., 2016): DPSO involves the use of the discretized Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995) for IM with a local search strategy and degree-based heuristic initialization.
- Discrete Bat Algorithm (DBA) (Tang et al., 2018): DBA meta-heuristic approach is used for Influence Maximization using a probabilistic local search strategy and random walk strategy.
- Efficient Discrete Differential Evolution (E-DDE) (Li, Zhang and Liu, 2022): E-DDE was used for Influence Maximization with community structures and population-based initialization.

5.3 Performance Metrics

- Final Infected Scale (FIS): FIS is the ratio of the infected nodes after the simulation ends and the total nodes in the graph. The activation of nodes is caused due to the activation probability working in the favor of the nodes. The initial spreader fraction plays a critical role in the final number of infected nodes in the graph. Generally, a higher spreader fraction leads to a higher FIS.
- LIE and log(LIE) function vs Spreader Fractions: The LIE function has a prominent contribution in evaluating the FIS in a social network. The LIE function value represents the fitness value of the objective function for IM. Usually, the LIE fitness value has a large magnitude hence, we use a natural logarithm (log)function for convenience. The greater the fitness value of the objective function, the higher is the influence of the seed node. LIE and log(LIE) function values are plotted against each spreader function, logically a higher spreader fraction leads to a higher LIE fitness value. The LIE and log(LIE) values indicate the accuracy of the approach.
- Activation Probability (P) vs Final Infected Scale (FIS): Activation Probability (P) plays a crucial role in the total final spread. Typically, as P increases, the FIS value also increases. FIS indicates the number of infected nodes at the end of the spread time. Increasing the activation probability means increasing the chances of activating more nodes for information propagation, hence increasing the final influence diffusion.

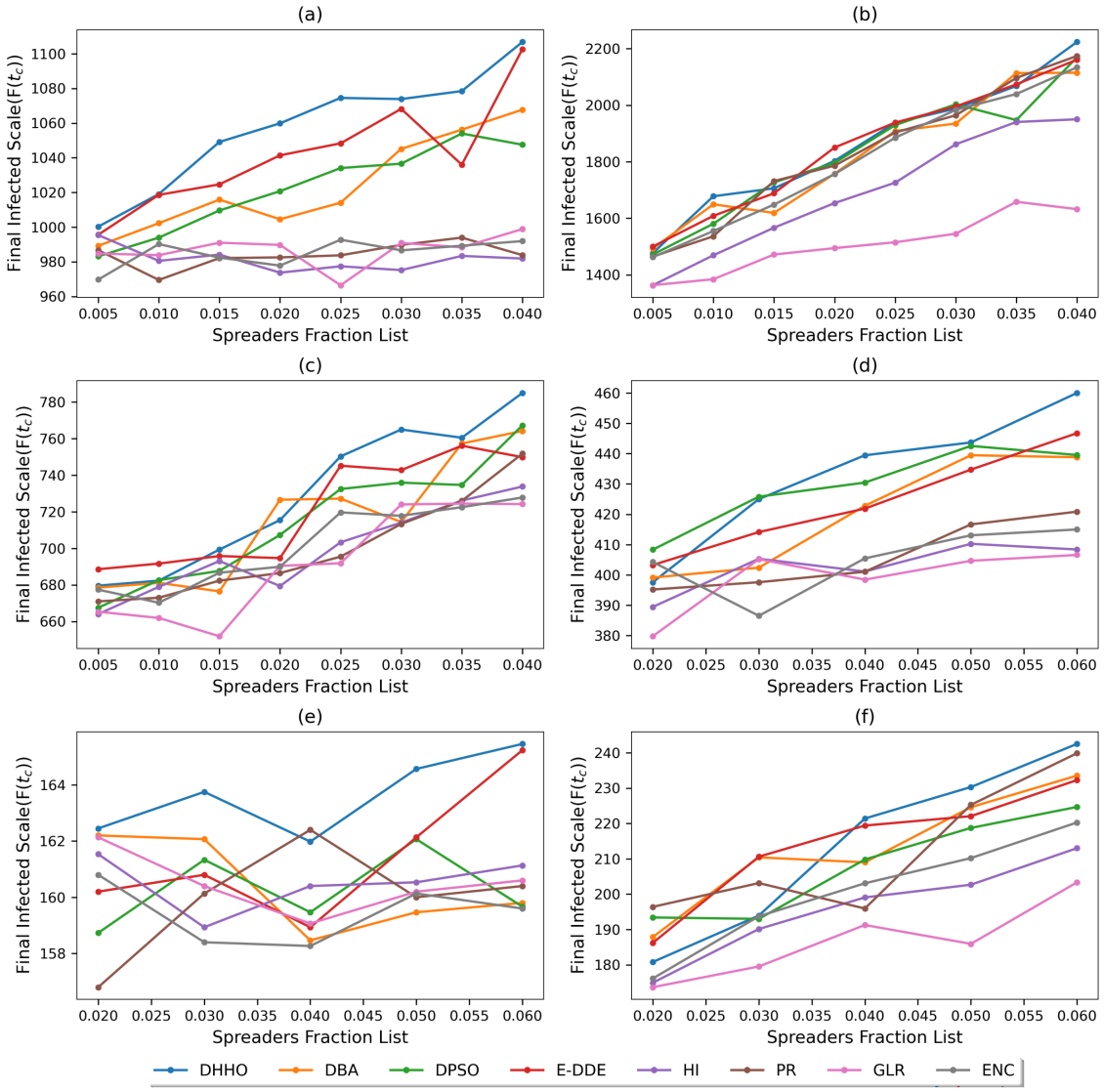


Figure 3: Final infected scale plots for various values for initial spreader fraction on the (a) hamsterster dataset (b) P2P-gnutella dataset (c) PCG dataset (d) Email-univ dataset (e) Jazz dataset (f) Wiki-vote dataset.

- Running Time: In order to compare our proposed approach with other baseline approaches, we also take into account the amount of time it takes to run the whole algorithm and rank nodes based on their ability to influence the network. The time consumption was evaluated while running all approaches in a Jupyter Environment, on a system with a Tesla K80 GPU and 12GB RAM available to the environment.

6 Experimental Results and Analysis

6.1 Final Infected Scale (FIS)

The FIS plots the final infection at the end of the simulation corresponding to the various values of the spreader fraction. We ran the Independent Cascade (IC) simulation 50 times for our proposed DHHO approach as well as for other competing approaches on six publicly available datasets and plotted the average values for various values of the spreader fraction. The spreader fraction value lies in the set 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04 for larger datasets (nodes > 2000), and in the set 0.02, 0.03, 0.04, 0.05, 0.06 for smaller datasets. For smaller datasets, we have taken relatively larger spreader fractions because smaller values lead to a very small absolute number of initial spreaders in

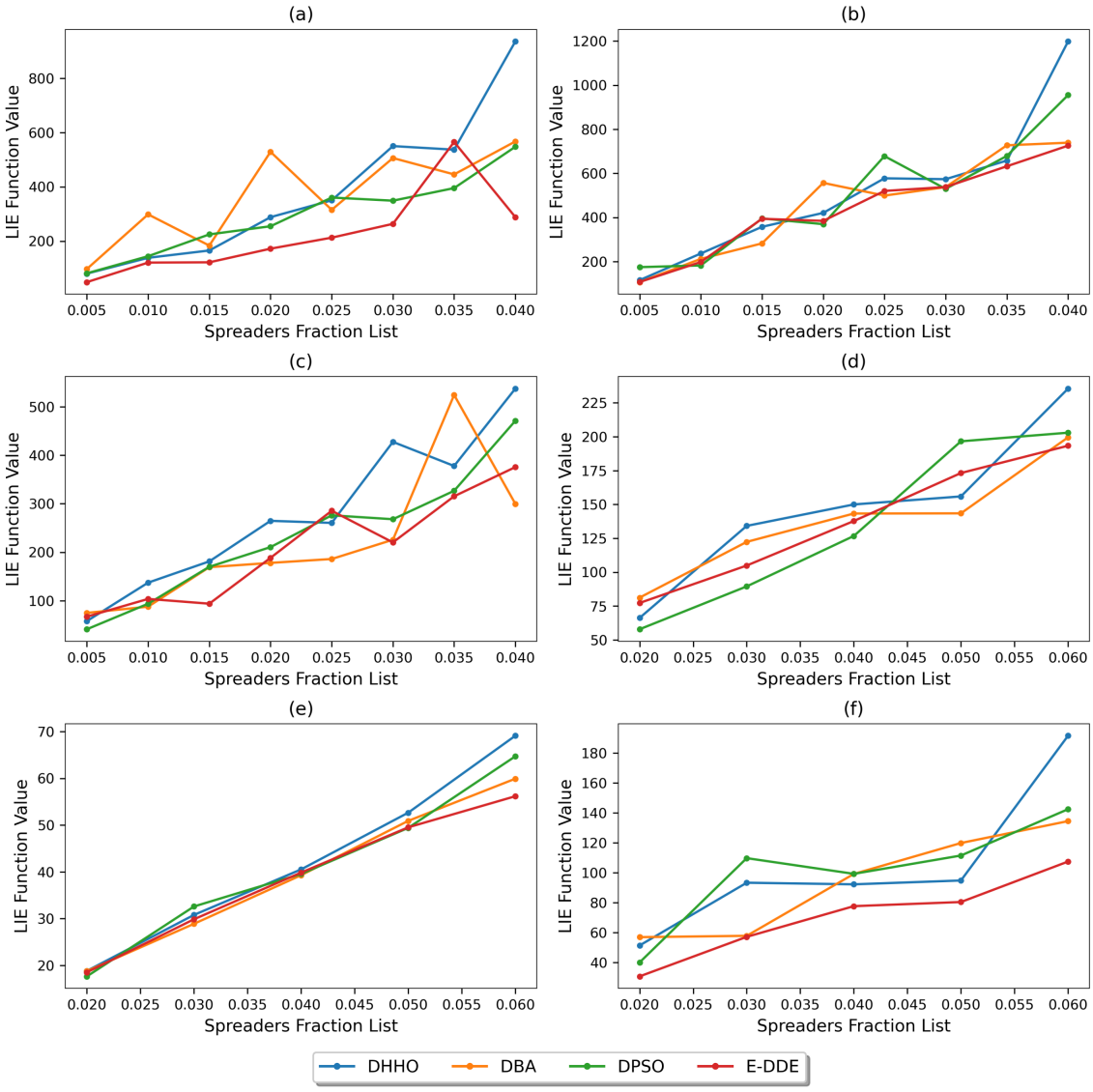


Figure 4: LIE Function Value plots for various values of initial spreader fraction for the (a) hamsterster dataset (b) P2P-gnutella dataset (c) PCG dataset (d) Email-univ dataset (e) Jazz dataset (f) Wiki-vote dataset.

smaller datasets. The activation probability was taken to be 0.1 for all methods and datasets. Figure 3 shows the plots for the FIS vs Spreader fraction for DHHO and the baseline approaches. The initial spreader fractions are plotted along the x-axis, while the corresponding FIS values are plotted along the y-axis.

We observed that for most datasets, as the spreader fraction increases, the final infected scale value also increases correspondingly. However, for a few approaches like PR, GLR, and HI, the final infected scale did not increase continuously upon increasing the spreader fraction for some datasets. Our proposed approach, DHHO, appeared to perform significantly better than the other baseline approaches. Other meta-heuristic approaches like DBA and E-DDE performed well on all datasets too and sometimes outperformed DHHO for a given value of the spreader fraction. However, in general, DHHO performs noticeably better than other meta-heuristic and heuristic algorithms across all datasets and all values of spreader fraction.

6.2 LIE vs Spreader Fractions and log(LIE) vs Spreader Fraction

We evaluated the performance of the fitness value of the Local Influence Estimator (LIE) objective function against the different values of the spreader fractions for DHHO and other baseline meta-heuristic approaches. We plotted LIE and

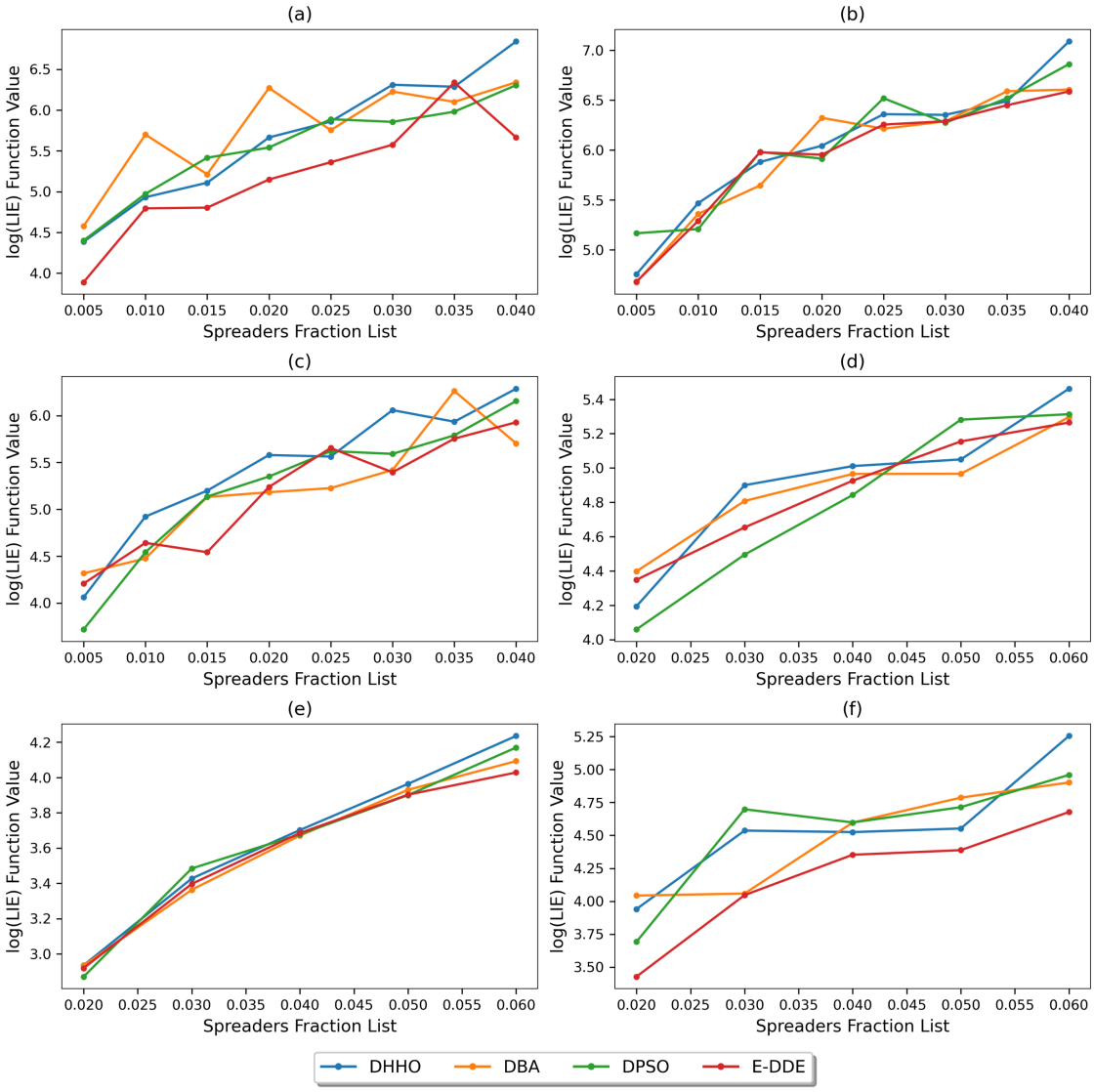


Figure 5: log(LIE) Function Value plots for various values of initial spreader fraction on the (a) hamsterster dataset (b) P2P-gnutella dataset (c) PCG dataset (d) Email-univ dataset (e) Jazz dataset (f) Wiki-vote dataset.

log(LIE) Function Values for different values of spreader fraction as shown in Figures 4 and 5. The spreader fraction values are plotted along the x-axis and the corresponding LIE function values are plotted along the y-axis. The initial spreader fraction values for datasets having greater than 2000 nodes were chosen from the set 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, whereas for smaller datasets with less than 2000 nodes, the initial spreaders fraction were taken from the set 0.02, 0.03, 0.04, 0.05, 0.06. Once again, for smaller datasets, we have taken relatively larger spreader fractions because small spreader fractions will lead to an insignificant number of initial spreaders in small datasets.

The log(LIE) metric is similar to the LIE function metric but the slope or the curvature of the graph changes significantly. As the spreader fraction increases, the fitness value increases as well, and so does the final influence diffusion. Our proposed approach DHHO performs as well or better than the competing approaches on all datasets. While DBA and DPSO sometimes performed slightly better than DHHO at some value of the spreader fraction for some datasets, DHHO performed more consistently and was noticeably better for higher values of spreader fraction. Thus, we can conclude that DHHO performs as well, if not better than competing novel meta-heuristic approaches for IM in terms of LIE-vs-Spreader-Fraction and log(LIE)-vs-Spreader-Fraction.

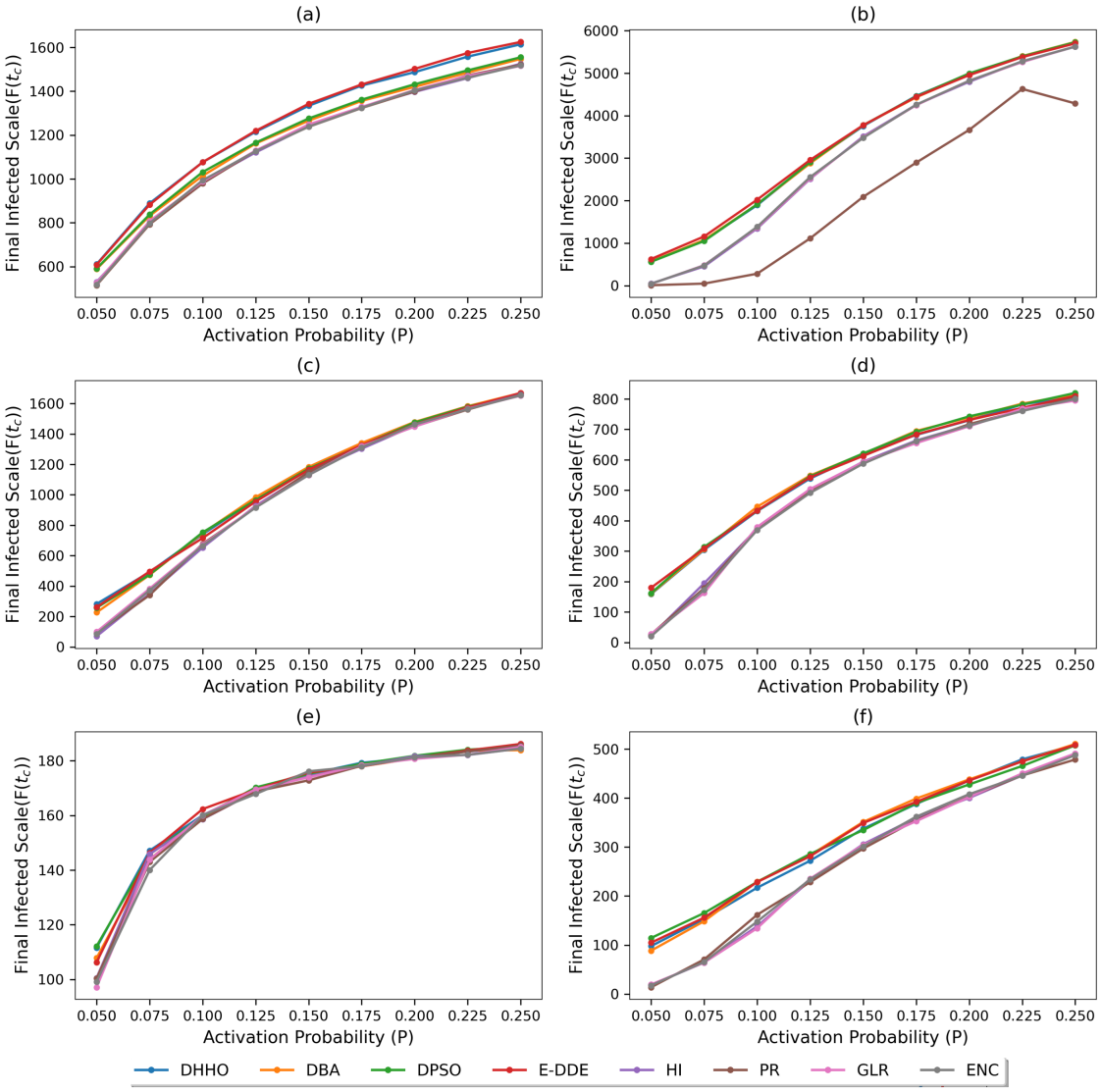


Figure 6: Final Infected Scale plots for different activation probability (p) values on the (a) hamsterster dataset (b) P2P-gnutella dataset (c) PCG dataset (d) Email-univ dataset (e) Jazz dataset (f) Wiki-vote dataset. The initial spreader fraction was set as 0.10.

6.3 Activation Probability (P) vs Final Infected Scale (FIS)

We compared the performance of DHHO with other baseline approaches on the basis of FIS for different values of activation probability (P). We took the different activation probability values in the range 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25. The initial spreader fraction for each case was taken to be 0.10, and for each dataset, the simulation was performed 50 times. Figure 6 shows the FIS vs P plots for the different methods on all six datasets. The values for the activation probabilities are taken along the x-axis, whereas the Final Infected State is taken along the y-axis. From the given plots, we can observe that as per our expectation, the FIS increased upon increasing the value of P. Our proposed method DHHO and E-DDE perform significantly better than the competing approaches on all datasets. Other methods like DBA and DPSO performed sometimes perform comparably to DHHO, but do not perform better than DHHO at any spreader fraction. HI, PR, GLR, and ENC are noticeably behind the meta-heuristic approaches on all six datasets.

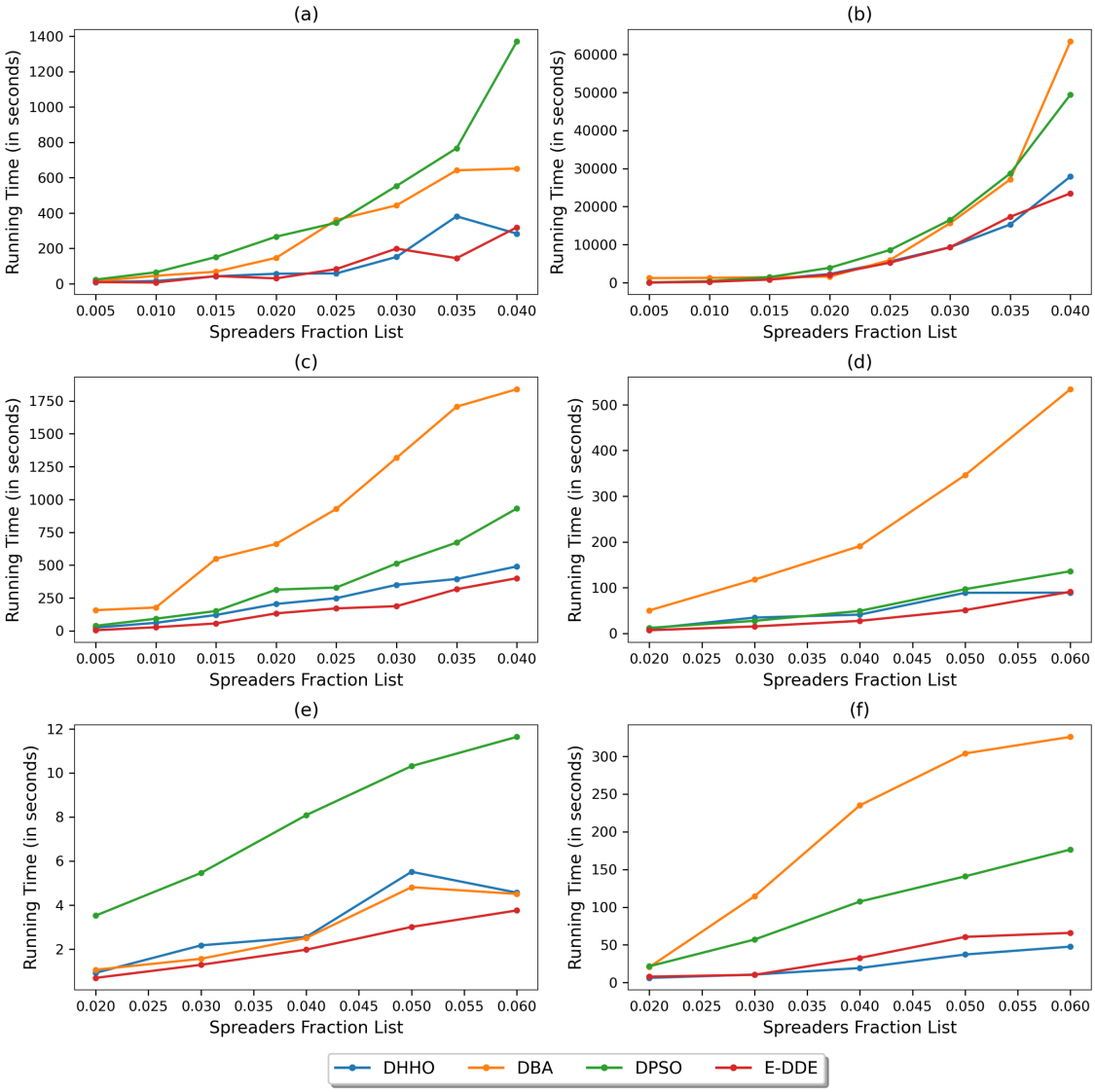


Figure 7: The execution time plots for different methods for ranking the spreaders for the (a) hamsterster dataset (b) P2P-gnutella dataset (c) PCG dataset (d) Email-univ dataset (e) Jazz dataset (f) Wiki-vote dataset.

6.4 Running Time

We also evaluated our proposed approach DHHO against the other competing meta-heuristic approaches on the basis of time execution for compiling the final influence diffusion. Figure 7 shows our plots for execution time vs spreader fraction. The various values of the spreader fraction are plotted along the x-axis, while the execution time (in seconds) is plotted along the y-axis. The initial spreader fraction values again for datasets having greater than 2000 nodes were chosen from the set 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, whereas for smaller datasets with less than 2000 nodes, the initial spreaders fraction were taken from the set 0.02, 0.03, 0.04, 0.05, 0.06.

We observed that DPSO and DBA usually have a longer execution time than DHHO and E-DDE. Since DHHO and E-DDE employ community structures and do not employ lengthy functions such as random walk strategy, the execution time for DHHO and E-DDE is significantly lesser than the other meta-heuristic approaches. DHHO and E-DDE are comparable across three datasets, DHHO performs better than E-DDE on one dataset, while E-DDE outperforms all approaches on two datasets in terms of execution time. However, E-DDE does not have the best final infected scale value. Thus, DHHO, with a similar execution time and a higher final influence spread across almost all datasets and spreader fractions, is a better approach when taking into account the holistic performance across different metrics.

7 Statistical Testing

To further evaluate the performance of DHHO using statistical tests, we have conducted the Friedman test (Friedman, 1937), which is a non-parametric statistical test used for multiple comparisons (more than two given methods). We have performed this test to determine whether the performance of our proposed method is significantly different than the other approaches for two or more data sets. The Friedman test attempts to detect significant differences based on the ranking of the methods, rather than their errors. It consists of two hypotheses: the null (H_0) and the alternate hypothesis (H_1). The former states that there are no prominent differences between these algorithms (equality of medians) and the latter insists that the algorithms have significant differences in their median of the population, thus negating the null hypothesis. We have performed the following test on the Final Infected Scale (FIS) vs Spreader Fractions metric. In order to perform the Friedman Statistical Test, we execute the following procedure:

- Gather the generated results for each problem pair.
- Rank the values in ascending order from 1 (best value) to n (worst value) for each problem i for a particular algorithm j.
- For algorithm j, the average rank for each problem i is calculated using the Eq. 15:

$$R_j = \frac{1}{n}\sum_{i=1}^{n}(r_i^j) \tag{15}$$

  where r^j is the rank ($1 < j < k$) and R_j is the average rank.
- Now that all algorithms are ranked according to their priority, compute the Friedman Statistic F_f using the following equation. F_f is based on a Chi-square distribution, with $k-1$ degree of freedom. F_f is computed as shown in Eq. 16. In the following equation, n represents number of rows and k represents number of columns ($n = 8$ and $k = 6$).

$$F_f = \frac{12n}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{16}$$

- The previous statistic produced relatively conservative results, which were not desired. Thus, Iman and Davenport (1980) introduced a new statistic F_{id}, as shown below, which follows the $F - distribution$ with the degree of freedom as $k-1$ and $(n-1)(k-1)$, given in Eq. 17:

$$F_{id} = \frac{(n-1)\chi_F^2}{n(k-1) - \chi_F^2} \tag{17}$$

Table 2 shows the average ranking calculated from the Friedman test for all the approaches used in this paper. DHHO clearly has a better average ranking than the other competing approaches.

Table 2: Average ranking of algorithm calculated using the Friedman test

Sr No.	Algorithm	Average Ranking
1	DHHO	**1.867**
2	DBA	3.375
3	DPSO	3.250
4	E-DDE	3.083
5	HI	6.416
6	PR	5.208
7	GLR	6.375
8	ENC	5.667

The unadjusted P-value or Holm P-value, obtained from the Iman-Davenport statistic, corresponding to the performance of DHHO, advocates the rejection of the null hypothesis H_0. All of the computed P-values are less than the standard significance level $\alpha = 0.05$. This indicates that there is a significant difference between the performance of the baseline approaches and our proposed approach. Hence, the obtained P-values conclude the negation of null hypothesis H_0 but are not appropriate for comparison with different methods. In order to compare these approaches with each other, we must calculate their respective adjusted P-values keeping DHHO as the control algorithm.

Hence, we use adjusted P-values (APVs) for evaluation in order to perform a comparative study between the control method and benchmark methods on a statistical ground. Adjusted P-value provides the correct correlation between these

algorithms, taking into account the accumulated family error with respect to the DHHO control algorithm. Moreover, the adjusted P-values can be directed compared with the significance level α, which is equivalent to 0.05. In order to calculate the adjusted P-values, a few post-hoc procedures need to be defined. Various post-hoc procedures such as Dunn (1961) and Holland and Copenhaver (1987) differ as they adjust the value of α to compensate for multiple comparisons for multiple methods. In this paper, we have used the common Holm's procedure (Holm, 1979) to evaluate the respective APVs. The values of Holm P-values and APVs are always sorted in ascending order. The equation for the same is given below, where indices i and j refers to the main hypothesis whose APVs are being computed and different hypothesis in the set respectively. P_j is the P-value for the j^{th} hypothesis. Holm APV is computed as shown in Eq. 18.

$$HolmAPV_i = \min\{v, 1\}, \text{where } v = \max\{(k-j)p_j : 1 \leq j \leq i\} \tag{18}$$

Table 3: Adjusted P-values (APVs) using Holm procedure

Sr No.	Algorithm (DHHO is control algorithm)	Z score	Holm p-value	Adjusted p-value
1	HI	-7.03	9.89e-13	6.92e-12
2	GLR	-6.97	1.56e-12	9.41e-12
3	ENC	-5.87	2.12e-09	1.06e-08
4	PR	-5.16	1.20e-07	4.71e-07
5	DBA	-2.32	1.00e-02	2.79e-02
6	DPSO	-2.13	1.65e-02	3.31e-02
7	E-DDE	-1.87	3.06e-02	4.99e-02

The results of the APVs using Holm's procedure are shown in Table 3. The intended APVs are less than the significance level, thus rejecting the null hypothesis. It is clearly observed from the results obtained that DHHO is significantly better than the competing approaches on these statistical tests as well.

8 Conclusion

The tremendous popularity of social networks in today's world has made influence maximization a crucial research problem in today's world. The selection of optimal seed nodes for maximum influence spread is critical for viral marketing. In this paper, we presented a Discretized Harris' Hawks Optimization approach for IM using community structures for optimal selection of seed nodes for influence spread. We employed a neighbor scout strategy algorithm to avoid blindness and enhance the searching ability of the hawks in the DHHO algorithm. We used a random population initialization strategy based on candidate nodes to accelerate the process of convergence of the population. We verified the performance of our approach by evaluating it on six social network datasets across five metrics. Our proposed DHHO approach outperformed the three competing meta-heuristic approaches and the four competing heuristic approaches. Community structures helped enable our approach to perform extremely efficiently and converge in minimal running time compared to other approaches. Statistical tests done on our obtained results further verified the superiority of our proposed DHHO approach compared to other approaches, indicating that the seed nodes selected by DHHO enable a wider propagation of information in social networks as compared to competing approaches.

References

Bae, J., Kim, S., 2014. Identifying and ranking influential spreaders in complex networks by neighborhood coreness. Physica A Statistical Mechanics and its Applications 395, 549–559. doi:10.1016/j.physa.2013.10.047.

Berahmand, K., Bouyer, A., Samadi, N., 2019. A new local and multidimensional ranking measure to detect spreaders in social networks. Computing 101. doi:10.1007/s00607-018-0684-8.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008, P10008. URL: https://doi.org/10.1088%2F1742-5468%2F2008%2F10%2Fp10008, doi:10.1088/1742-5468/2008/10/p10008.

Borgs, C., Brautbar, M., Chayes, J., Lucier, B., 2012. Maximizing social influence in nearly optimal time. URL: https://arxiv.org/abs/1212.0884, doi:10.48550/ARXIV.1212.0884.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30, 107–117. doi:10.1016/S0169-7552(98)00110-X.

Brown, J.J., Reingen, P.H., 1987. Social Ties and Word-of-Mouth Referral Behavior*. Journal of Consumer Research 14, 350–362. URL: https://doi.org/10.1086/209118, doi:10.1086/209118, arXiv:https://academic.oup.com/jcr/article-pdf/14/3/350/5151897/14-3-350.pdf.

Chen, W., Lakshmanan, L., Castillo, C., 2013. Information and influence propagation in social networks. Synthesis Lectures on Data Management 5, 1–177. doi:10.2200/S00527ED1V01Y201308DTM037.

Chen, W., Wang, Y., Yang, S., 2009. Efficient influence maximization in social networks, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 199–208. URL: https://doi.org/10.1145/1557019.1557047, doi:10.1145/1557019.1557047.

Chen, Y.C., Zhu, W.Y., Peng, W.C., Lee, W.C., Lee, S.Y., 2014. Cim. ACM Transactions on Intelligent Systems and Technology 5, 1–31. doi:10.1145/2532549.

Cheng, S., Shen, H., Huang, J., Zhang, G., Cheng, X., 2013. Staticgreedy: Solving the scalability-accuracy dilemma in influence maximization, in: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, Association for Computing Machinery, New York, NY, USA. p. 509–518. URL: https://doi.org/10.1145/2505515.2505541, doi:10.1145/2505515.2505541.

Cohen, E., Delling, D., Pajor, T., Werneck, R.F., 2014. Sketch-based influence maximization and computation: Scaling up with guarantees, in: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Association for Computing Machinery, New York, NY, USA. p. 629–638. URL: https://doi.org/10.1145/2661829.2662077, doi:10.1145/2661829.2662077.

Cui, L., Hu, H., Yu, S., Yan, Q., Ming, Z., Wen, Z., Lu, N., 2018. Ddse: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks. Journal of Network and Computer Applications 103, 119–130. doi:10.1016/j.jnca.2017.12.003.

Domingos, P., Richardson, M., 2001. Mining the network value of customers, in: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 57–66. URL: https://doi.org/10.1145/502512.502525, doi:10.1145/502512.502525.

Dunn, O.J., 1961. Multiple comparisons among means. Journal of the American statistical association 56, 52–64.

Estrada, E., Rodríguez-Velázquez, J.A., 2005. Subgraph centrality in complex networks. Physical Review E 71, 056103. doi:10.1103/PhysRevE.71.056103.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Social Networks 1, 215–239. doi:10.1016/0378-8733(78)90021-7.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 32, 675–701. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1937.10503522, doi:10.1080/01621459.1937.10503522, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1937.10503522.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99, 7821–7826. URL: https://www.pnas.org/doi/abs/10.1073/pnas.122653799, doi:10.1073/pnas.122653799, arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.122653799.

Goldenberg, J., Libai, B., Muller, E., 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters 12, 211–223. doi:10.1023/A:1011122126881.

Gong, M., Yan, J., Shen, B., Ma, L., Cai, Q., 2016. Influence maximization in social networks based on discrete particle swarm optimization. Information Sciences 367-368, 600–614. doi:10.1016/j.ins.2016.07.012.

Goyal, A., Lu, W., Lakshmanan, L.V., 2011. Celf++: Optimizing the greedy algorithm for influence maximization in social networks, in: Proceedings of the 20th International Conference Companion on World Wide Web, Association for Computing Machinery, New York, NY, USA. p. 47–48. URL: https://doi.org/10.1145/1963192.1963217, doi:10.1145/1963192.1963217.

Granovetter, M., 1978. Threshold models of collective behavior. American Journal of Sociology 83, 1420–1443. doi:10.1086/226707.

Guille, A., Hacid, H., Favre, C., Zighed, D.A., 2013. Information diffusion in online social networks: A survey. ACM Sigmod Record 42, 17–28.

Hamsterster, . Hamsterster social network. Http://www.hamsterster.com.

He, Q., Wang, X., Mao, F., Lv, J., Cai, Y., Huang, M., Xu, Q., 2019. Caom: A community-based approach to tackle opinion maximization for social networks. Information Sciences 513. doi:10.1016/j.ins.2019.10.064.

Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., Chen, H., 2019. Harris hawks optimization: Algorithm and applications. Future Generation Computer Systems 97, 849–872. URL: https://www.sciencedirect.com/science/article/pii/S0167739X18313530, doi:https://doi.org/10.1016/j.future.2019.02.028.

Heidari, M., Asadpour, M., Faili, H., 2015. Smg: Fast scalable greedy algorithm for influence maximization in social networks. Physica A: Statistical Mechanics and its Applications 420, 124–133. doi:10.1016/j.physa.2014.10.088.

Holland, B.S., Copenhaver, M.D., 1987. An improved sequentially rejective bonferroni test procedure. Biometrics , 417–423.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65–70. URL: http://www.jstor.org/stable/4615733.

Holme, P., Kim, B.J., 2002. Growing scale-free networks with tunable clustering. Physical Review E 65. URL: https://doi.org/10.1103%2Fphysreve.65.026107, doi:10.1103/physreve.65.026107.

Huang, H., Shen, H., Meng, Z., Chang, H., He, H., 2019. Community-based influence maximization for viral marketing. Applied Intelligence 49, 2137–2150. doi:10.1007/s10489-018-1387-8.

Iman, R.L., Davenport, J.M., 1980. Approximations of the critical region of the fbietkan statistic. Communications in Statistics - Theory and Methods 9, 571–595. URL: https://doi.org/10.1080/03610928008827904, doi:10.1080/03610928008827904, arXiv:https://doi.org/10.1080/03610928008827904.

Jiang, Q., Song, G., Cong, G., Wang, Y., Si, W., Xie, K., 2011. Simulated annealing based influence maximization in social networks, in: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI Press. p. 127–132.

Kempe, D., Kleinberg, J., Tardos, E., 2003. Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 137–146. URL: https://doi.org/10.1145/956750.956769, doi:10.1145/956750.956769.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks, pp. 1942–1948 vol.4. doi:10.1109/ICNN.1995.488968.

Khomami, M., Rezvanian, A., Meybodi, M., Bagheri, A., 2021. Cfin: A community-based algorithm for finding influential nodes in complex social networks. The Journal of Supercomputing 77. doi:10.1007/s11227-020-03355-2.

Kim, D., Hyeon, D., Oh, J., Han, W.S., Yu, H., 2017. Influence maximization based on reachability sketches in dynamic graphs. Information Sciences 394-395, 217–231. doi:10.1016/j.ins.2017.02.023.

Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A., 2010. Identification of influential spreaders in complex networks. Nature Physics 6, 888–893. URL: https://doi.org/10.1038%2Fnphys1746, doi:10.1038/nphys1746.

Kundu, S., Murthy, C.A., Pal, S.K., 2011. A new centrality measure for influence maximization in social networks, in: Kuznetsov, S.O., Mandal, D.P., Kundu, M.K., Pal, S.K. (Eds.), Pattern Recognition and Machine Intelligence, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 242–247.

Kunegis, J., 2013. KONECT – The Koblenz Network Collection, in: Proc. Int. Conf. on World Wide Web Companion, pp. 1343–1350. URL: http://dl.acm.org/citation.cfm?id=2488173.

Leskovec, J., Adamic, L.A., Huberman, B.A., 2006. The dynamics of viral marketing, in: Proceedings of the 7th ACM Conference on Electronic Commerce, Association for Computing Machinery, New York, NY, USA. p. 228–237. URL: https://doi.org/10.1145/1134707.1134732, doi:10.1145/1134707.1134732.

Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010a. Predicting positive and negative links in online social networks, in: Proceedings of the 19th International Conference on World Wide Web, Association for Computing Machinery, New York, NY, USA. p. 641–650. URL: https://doi.org/10.1145/1772690.1772756, doi:10.1145/1772690.1772756.

Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010b. Signed networks in social media, in: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 1361–1370.

Leskovec, J., Kleinberg, J., Faloutsos, C., 2007a. Graph evolution: Densification and shrinking diameters. ACM Trans. Knowl. Discov. Data 1, 2–es. URL: https://doi.org/10.1145/1217299.1217301, doi:10.1145/1217299.1217301.

Leskovec, J., Kleinberg, J., Faloutsos, C., 2007b. Graph evolution: Densification and shrinking diameters. ACM Trans. Knowl. Discov. Data 1, 2–es. URL: https://doi.org/10.1145/1217299.1217301, doi:10.1145/1217299.1217301.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N., 2007c. Cost-effective outbreak detection in networks, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery

and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 420–429. URL: https://doi.org/10.1145/1281192.1281239, doi:10.1145/1281192.1281239.

Li, D., Wang, C., Zhang, S., Zhou, G., Chu, D., Wu, C., 2017. Positive influence maximization in signed social networks based on simulated annealing. Neurocomputing 260, 69–78. URL: https://www.sciencedirect.com/science/article/pii/S0925231217304848, doi:https://doi.org/10.1016/j.neucom.2017.03.003.

Li, H., Zhang, R., Liu, X., 2022. An efficient discrete differential evolution algorithm based on community structure for influence maximization. Applied Intelligence 52, 1–19. doi:10.1007/s10489-021-03021-x.

Li, X., Cheng, X., Su, S., Sun, C., 2018. Community-based seeds selection algorithm for location aware influence maximization. Neurocomputing 275, 1601–1613. URL: https://www.sciencedirect.com/science/article/pii/S0925231217316326, doi:https://doi.org/10.1016/j.neucom.2017.10.007.

Lin, S., Hu, Q., Wang, G., Yu, P.S., 2015. Understanding community effects on information diffusion, in: Cao, T., Lim, E.P., Zhou, Z.H., Ho, T.B., Cheung, D., Motoda, H. (Eds.), Advances in Knowledge Discovery and Data Mining, Springer International Publishing, Cham. pp. 82–95.

Liu, Q., Zhu, Y.X., Jia, Y., Deng, L., Zhou, B., Zhu, J.X., Zou, P., 2018. Leveraging local h-index to identify and rank influential spreaders in networks. Physica A: Statistical Mechanics and its Applications 512, 379–391. URL: https://www.sciencedirect.com/science/article/pii/S0378437118309932, doi:https://doi.org/10.1016/j.physa.2018.08.053.

Lu, W.X., Zhou, C., Wu, J., 2016. Big social network influence maximization via recursively estimating influence spread. Knowledge-Based Systems 113, 143–154. URL: https://www.sciencedirect.com/science/article/pii/S0950705116303513, doi:https://doi.org/10.1016/j.knosys.2016.09.020.

Ma, L.L., Ma, C., Zhang, H.F., Wang, B.H., 2016. Identifying influential spreaders in complex networks based on gravity formula. Physica A: Statistical Mechanics and its Applications 451, 205–212. doi:10.1016/j.physa.2015.12.162.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences 103, 8577–8582. doi:10.1073/pnas.0601602103.

Nguyen, H.T., Thai, M.T., Dinh, T.N., 2016. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks, in: Proceedings of the 2016 International Conference on Management of Data, Association for Computing Machinery, New York, NY, USA. p. 695–710. URL: https://doi.org/10.1145/2882903.2915207, doi:10.1145/2882903.2915207.

Okamoto, K., Chen, W., Li, X.Y., 2008. Ranking of closeness centrality for large-scale social networks, in: Frontiers in Algorithmics, pp. 186–195. doi:10.1007/978-3-540-69311-6_21.

Pei, S., Muchnik, L., , J.S.A., Zheng, Z., Makse, H.A., 2015. Searching for superspreaders of information in real-world social media. Scientific Reports 4, 5547. doi:10.1038/srep05547.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., Jia, W., 2018. Influence analysis in social networks: A survey. Journal of Network and Computer Applications 106, 17–32. URL: https://www.sciencedirect.com/science/article/pii/S1084804518300195, doi:https://doi.org/10.1016/j.jnca.2018.01.005.

Salavati, C., Abdollahpouri, A., Manbari, Z., 2018. Ranking nodes in complex networks based on local structure and improving closeness centrality. Neurocomputing 336. doi:10.1016/j.neucom.2018.04.086.

Samadi, N., Bouyer, A., 2019. Identifying influential spreaders based on edge ratio and neighborhood diversity measures in complex networks. Computing 101. doi:10.1007/s00607-018-0659-9.

Shang, J., Wu, H., Zhou, S., Zhong, J., Feng, Y., Qiang, B., 2018. Impc: Influence maximization based on multi-neighbor potential in community networks. Physica A: Statistical Mechanics and its Applications 512, 1085–1103. doi:10.1016/j.physa.2018.08.045.

Tang, J., Zhang, R., Yao, Y., Zhao, Z., Wang, P., Li, H., Yuan, J., 2018. Maximizing the spread of influence via the collective intelligence of discrete bat algorithm. Knowledge-Based Systems 160. doi:10.1016/j.knosys.2018.06.013.

Tang, Y., Shi, Y., Xiao, X., 2015. Influence maximization in near-linear time: A martingale approach, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, New York, NY, USA. p. 1539–1554. URL: https://doi.org/10.1145/2723372.2723734, doi:10.1145/2723372.2723734.

Tang, Y., Xiao, X., Shi, Y., 2014. Influence maximization: Near-optimal time complexity meets practical efficiency, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, New York, NY, USA. p. 75–86. URL: https://doi.org/10.1145/2588555.2593670, doi:10.1145/2588555.2593670.

Venkatakrishna Rao, K., Chowdary, C.R., 2022. Cbim: Community-based influence maximization in multilayer networks. Information Sciences 609, 578–594. URL: https://www.sciencedirect.com/science/article/pii/S0020025522007927, doi:https://doi.org/10.1016/j.ins.2022.07.103.

Wang, X., Zhang, Y., Zhang, W., Lin, X., Chen, C., 2017. Bring order into the samples: A novel scalable method for influence maximization. IEEE Transactions on Knowledge and Data Engineering 29, 243–256. doi:10.1109/TKDE.2016.2624734.

Wang, Y., Wang, H., Li, J., 2015. Efficient influence maximization in weighted independent cascade model. URL: https://arxiv.org/abs/1510.06201, doi:10.48550/ARXIV.1510.06201.

Zareie, A., Sheikhahmadi, A., Jalili, M., 2020. Identification of influential users in social network using gray wolf optimization algorithm. Expert Systems with Applications 142, 112971. doi:10.1016/j.eswa.2019.112971.

Zareie, A., Sheikhahmadi, A., Khamforoosh, K., 2018. Influence maximization in social networks based on topsis. Expert Systems with Applications 108. doi:10.1016/j.eswa.2018.05.001.

Zeng, A., Zhang, C.J., 2013. Ranking spreaders by decomposing complex networks. Physics Letters A 377, 1031–1035.

J Mater Sci: Mater Electron

Influence of calcination and sintering temperature on the microstructure, dielectric, ferroelectric and piezoelectric properties of the lead-free KNN ceramics

Komal Verma[1], Sahil Goel[2], and Richa Sharma[1,*]

[1] Department of Applied Physics, Delhi Technological University, Delhi 110042, India
[2] Department of Physics and Electronics, Hansraj College, University of Delhi, Delhi 110007, India

Received: 12 May 2022
Accepted: 7 October 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

ABSTRACT

In this study, the aim is to determine optimized calcination and sintering temperatures for the synthesis of lead-free $K_{0.5}Na_{0.5}NbO_3$ (KNN) ceramics to achieve excellent di-/piezo-/ferro-electric characteristics. The solid-state reaction method was used to synthesize KNN ceramics. Initially, KNN ceramic powders were prepared by calcination at different temperatures in the range 850–1050 °C. The crystalline structure and phase analysis of calcined KNN powdered samples were done by X-Ray Diffraction (XRD) combined with Rietveld Refinement. Based on XRD and Rietveld refinement data, KNN powder calcined at 1050 °C was chosen as the optimized calcination temperature. Keeping the calcination temperature to be fixed at 1050 °C, KNN pellets were sintered at temperatures in the range 1075–1150 °C and were characterized for their structural, morphological, dielectric, piezoelectric and ferroelectric properties. Scanning Electron Microscopy (SEM) images showed increase in average grain size with increase in sintering temperature. Furthermore, significant influence of densification and average grain size were observed on all the measured electrical properties. The KNN ceramics which were calcined at 1050 °C and sintered at 1125 °C displayed enhanced di-/piezo-/ferro-electrical characteristics in comparison to other sintered KNN samples. The results obtained in the present work suggest that the optimized calcination and sintering temperatures for the preparation for KNN ceramics were 1050 °C and 1125 °C, respectively. For KNN sample sintered at 1125 °C, the values of various measured parameters are: $\varepsilon_r \sim 460$ (@ RT, @10 kHz), $d_{33} \sim 290$ pC/N, $d_{33}^*{}_{max} \sim 718$ pm/V, $d_{33}^*{}_{avg.} \sim 321$ pm/V, $E_c \sim 5.73$ kV/cm and $P_r \sim 98.21$ μC/ making them a viable lead-free piezo-/ferro-electric contender.

Address correspondence to E-mail: richasharma@dtu.ac.in

https://doi.org/10.1007/s10854-022-09295-2
Published online: 30 October 2022

1 Relevance Summary

- Present work gives a detailed study of influence of processing parameters of solid-state reaction on various di-/piezo-/ferro-electric properties
- In present work both sintering and calcination temperatures are optimized, which is rarely reported in the literature
- Obtained results indicate influence of processing conditions on various properties
- For achieving high di-/piezo-/ferro-electric properties synthesis conditions optimization is very crucial in addition to composition modification of KNN ceramics

2 Introduction

For many years, piezoelectric and ferroelectric materials have been technologically important and have received a lot of attention due to their wide range of practical applications in ultrasonic transducers, actuators, sensors, energy harvesting devices, and other electronic equipment [1–5]. Many Piezoelectric materials have been designed, manufactured, described, and exploited for various purposes, which includes Lead Zirconate Titanate (PZT), Potassium Sodium Niobate (KNN), Barium Titanate (BT), Bismuth Sodium Titanate (BNT), Bismuth Potassium Titanate (BKT) ceramics and rare-earth doped Zinc Oxide (ZnO) nano-/micro-architectures [6–16]. Besides, many complex solid solutions with excellent di-/piezo-/ferro-electric characteristics, such as $Pb(Mg_{1/3}Nb_{2/3})$ O_3-$PbTiO_3$ [PMN-PT] [17–19], $Pb(Mg_{1/3}Nb_{2/3})$ O_3 − $Pb(In_{1/2}Nb_{1/2})$ O_3 − $PbZrO_3$- − $PbTiO_3$ [PMN-PIN-PZ-PT] [20], Pb $(Yb_{1/2}Nb_{1/2})$ O_3-$Pb(Mg_{1/3}Nb_{2/3})O_3$-$PbTiO_3$ [PYN-PMN-PT] [21], $BiFeO_3$-$Pb(Mg_{1/3}Nb_{2/3})$ O_3-$PbTiO_3$ [BF-PMN-PT] [22, 23], $Pb(Mg_{1/3}Nb_{2/3})O_3$–0.10Pb $(Fe_{1/2}Nb_{1/2})$ O_3–$0.05PbZrO_3$–$x$$PbTiO_3$ [PMN-PFN-PZ-PT] [24], (1–x) Bi $(Zn_{2/3}Nb_{1/3})$ O_3-$x$$BaTiO_3$ [BZN-BT] [25], (1–x) $BiScO_3$-$x$$PbTiO_3$ [BS-PT] [26], and (0.95–x) Pb $(Mg_{1/3}Nb_{2/3})$ O_3-0.05Bi $(Zn_{2/3}Nb_{1/3})$ O_3-x $PbTiO_3$ [PMN-BZN-PT], (1– x) $Bi(Zn_{2/3}Nb_{1/3})$ O_3-$x$$PbTiO_3$ [(1-x)BZN-xPT] [27] and $(Bi_{0.5}Na_{0.5})$ TiO_3-$BaTiO_3$-Pb $(Mg_{1/3}Nb_{2/3})$ O_3-$PbTiO_3$ [BNT-BT-PMN-PT] [28] have also been successfully synthesized by various researchers across the globe. Further, Lead- based ceramics have emerged as the most useful piezoelectric materials in a variety of applications, however, because of the toxicity of lead, lead-based products have a considerable environmental disadvantage. As a result, there is an urgent need to develop lead-free piezoelectric materials which are environment friendly and also possesses enhanced piezoelectric properties to replace lead-based ceramics efficiently in various applications [29].

In recent years, significant research has been conducted in this arena to synthesize lead-free piezoelectric ceramics with better characteristics than lead based materials. KNN, BNT, and BKT-based lead-free piezoelectric ceramics have pleased the curiosity of researchers. Among these, Potassium Sodium Niobate (K_xNa_{1-x}) NbO_3 ceramics are the most promising lead-free compounds to substitute PZT families due to its high piezoelectric performance (d_{33} = 50–500 pC/N), relatively high curie temperature (T_c = 150–500 °C) [29–34] and environment friendly nature, which is used in many practical applications such as actuators and transducer [35]. KNN ceramics with a perovskite structure, were initially manufactured in the 1950s by *Nader* et al. [36]. KNN is a full binary solid solution system consisting of $KNbO_3$ (KN, with a low melting temperature of 1090 °C) and $NaNbO_3$ (NN, with a higher melting temperature of 1420 °C). At 1200 °C, KNN material with a 50/50 composition is in equilibrium between the solid phase of $K_{0.27}Na_{0.73}NbO_3$ and the liquid phase of $K_{0.73}Na_{0.27}NbO_3$ [37, 38].

One of the major problems concerning the KNN ceramics is getting dense ceramics. It was claimed that high-density KNN ceramics with repeatable qualities were difficult to generate using the usual solid-state reaction approach because of potassium and sodium volatility, a small sintering temperature range, and sodium and potassium ion segregation [39]. Therefore, to solve the aforementioned problems, numerous synthesis procedures including hot-pressing [36], spark plasma sintering [40, 41], and the use of different sintering aids [42, 43] have been utilized. Although, these techniques have shown to raise ceramic relative density (> 97%) and improved properties in comparison to KNN prepared by solid state reaction method, the relationship between the processing conditions, microstructure and properties remain unsolved [44]. There is still a need of careful investigation and optimization of calcination and sintering parameters in order to

obtain reproducible and high-quality ceramics. Moreover, these techniques are very expensive in comparison to the solid-state reaction method. Additionally, for a long time, the interaction between the four elements composition—structure—properties—processing has been at the heart of material science. Unfortunately, researchers appeared to be much more concerned with KNN compositions and less attention was paid in improving processing conditions for KNN ceramics, which is yet to be documented [45–47]. Before analyzing the influence of compositions, dopants and solid solution on KNN ceramics di-/piezo-/ferro-electric properties, it is very crucial that the optimized (calcination and sintering) temperature conditions for the preparation of KNN ceramics must be focused on, since this may make a significant difference, as will be documented in the present work.

In view of this, in the present work efforts have been made to find the optimized calcination and sintering temperatures of KNN ceramics prepared by solid state reaction method that will demonstrate high ferro- to para-electric phase transition temperature (T_c) value and excellent electrical properties. Furthermore, the influence of calcination and sintering temperature on the densification, microstructure, dielectric, ferroelectric and piezoelectric properties are analyzed in detail.

3 Experimental details

3.1 Materials

Potassium Carbonate (K_2CO_3, 99%), Sodium Carbonate (Na_2CO_3, 98%), Niobium pentaoxide (Nb_2O_5, 99.5%) and propanol (C_3H_8O) were purchased from Alfa-Aesar, India. All these chemicals employed for the synthesis of pure KNN and were used as received without any further purification.

3.2 Synthesis of KNN ceramic powder

Lead-free $(K_{0.5}Na_{0.5})NbO_3$ (KNN) ceramics were synthesized using the conventional solid-state reaction method. Potassium Carbonate, Sodium Carbonate and Niobium Pentoxide were used as the raw materials and first mixed in an agate mortar for 10 h using propanol as the medium in stoichiometric quantities (given in Table 1) according to the following reaction for the synthesis of 20 gm of KNN ceramic powder:

$$(K_2CO_3) + 0.25\ (Na_2CO_3) + 0.5\ (Nb_2O_5) \rightarrow (K_{0.5}Na_{0.5}NbO_3) + CO_2\uparrow$$

The thoroughly mixed powder mixture was then dried at 60 °C for 24 h in an oven. Finally, the dried powder was calcined for 4 h in air at 850 °C, 900 °C, 950 °C, 1000 °C and 1050 °C, following the temperature profile illustrated in Fig. 2a for temperature 1050 °C. In Fig. 1 various steps involved in the calcination of KNN powder are depicted.

Figure 1 illustrates the procedure that was followed for preparing sintered KNN pellets from the above-synthesized KNN powder calcined at 1050 °C. For the preparation of green pellets, KNN powder calcined at 1050 °C were combined with 5 wt.% polyvinyl alcohol (PVA). Thereafter, the material was dried at 120 °C for 2 h in an oven. Now, the dried powder was grounded thoroughly in agate mortar to get fine powder. With the help of a hydraulic press machine, the powder was uniaxially pressed at 150 MP into discs of diameter 10 mm. Finally, the pressed pellets were sintered in air at 1075 °C, 1100 °C, 1125 °C, 1150 °C and 1175 °C for 3 h at 5 °C/min heating rates and 2 °C/min cooling rates (Fig. 2b showing temperature profile of KNN sintered at 1125 °C for 3 h). To our surprise, we observed that KNN pellets sintered at 1175 °C were melted which can be clearly seen in the bottom left of Fig. 1, thereby, giving an idea that 1175 °C sintering temperature is very high for the preparation of KNN pellets. In view of this, all the characterizations have been done for the KNN samples sintered at 1075 °C, 1100 °C, 1125 °C and 1150 °C.

3.3 Characterization techniques

X-ray powder diffraction (Cu-K_α radiation, λ = 1.5406 Å, Make: Bruker, Model: D8 Discover) was used to analyze the crystal structure and phase purity of the KNN ceramic powder calcined and sintered at different temperatures. In addition to it, Rietveld refinement was also done by Fullprof Software to get refined cell parameters of the calcined and sintered KNN samples. Scanning electron microscope (SEM) was used to examine the morphology of the sintered KNN samples using TESCAN instrument (Model: MIRA3LMH). The Archimedes technique was used to determine the density of sintered specimens. Furthermore, the polarization (P) *vs*. Electric field

Table 1 List of various raw materials used in the preparation of 20 gm of KNN ceramic powder

S. No	Various Precursor raw molecules used	Molecular Weight (in gm)	Coefficient of precursor molecule[#]	Weight (in gm)	Calculation for the weight[#] (In gm)
1	K_2CO_3	138.21	0.25	4.018	$\left(\frac{138.21\times0.25}{171.955}\right)\times 20 = 4.018$
2	Na_2CO_3	105.99	0.25	3.081	$\left(\frac{105.99\times0.25}{171.955}\right)\times 20 = 3.081$
3	Nb_2O_5	265.82	0.50	15.458	$\left(\frac{265.82\times0.50}{171.955}\right)\times 20 = 15.458$

[#]Molecular weight of a $(K_{0.5}Na_{0.5})\ NbO_3$ molecule is 171.955 gm and total amount of KNN ceramic which needs to be prepared is 20 gm

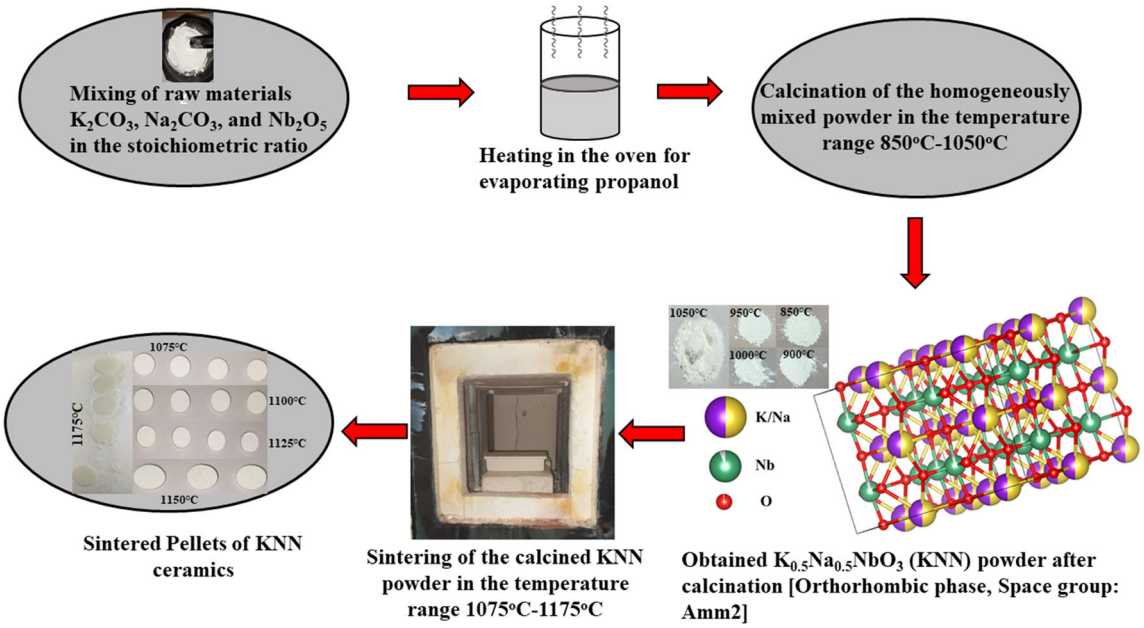


Fig. 1 Schematic diagram of the preparation of $K_{0.5}Na_{0.5}NbO_3$ powder at different calcined and sintered temperatures

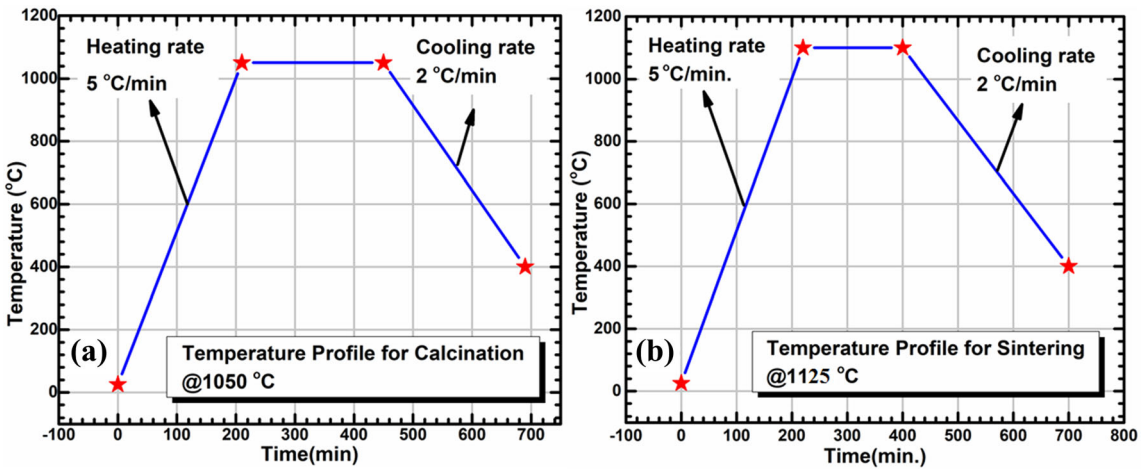


Fig. 2 Temperature profiles used for preparing **a** calcined KNN ceramic powder at 1050 °C **b** sintered KNN ceramic pellets at 1125 °C

(*E*) hysteresis loops of the KNN ceramics sintered at different temperatures were traced (@200, 250, 300, 350, 400, and 450 Hz) on a Marine India Instrument (Model: 20PE 1 kHz 1 N). The dielectric properties of sintered KNN pellets were measured as a function of frequency and temperature using a Broadband

Dielectric/Impedance Analyzer (Model: C-50 Alpha A; Make: Novocontrol Germany) in a frequency range of 20 Hz to 2 MHz and with temperature ranging from 30 to 490 °C. For piezoelectric measurement, samples were poled. The procedure for poling the sintered samples is following. For poling, samples were immersed in the silicone oil and then poled in an electric field in the range 3–4 kV/mm using a DC poling unit. The electric field was applied to the sample at a temperature of 120 °C for 30 min and then the samples were cooled down to room temperature in the presence of an electric field. The direct piezoelectric charge coefficient (d_{33}) of KNN sintered samples (polled at 3 kV/cm for 25 min) was measured on a piezometer (Model: PM 300 Piezotest) at 0.25 N tapping force with a frequency of 110 Hz. In addition to it, Converse piezo-charge coefficients (d_{33}^*) of KNN sintered pellets were also measured using displacement *vs*. applied voltage (*D-V*; butterfly) loops that were recorded with the help of linear variable differential transform (LVDT) sensor (Model: 0.1 SM 100HZ 18; Make: Marine India).

4 Results and discussions

4.1 XRD analysis

In order to get the better understanding of the phase formation and to determine the optimized calcination temperature for the KNN ceramics calcination was done in the temperature range 850–1050 °C. Figure 3 represents the room temperature powder XRD patterns recorded for KNN powder calcined at 850 °C, 900 °C, 950 °C, 1000 °C and 1050 °C. The diffraction patterns show the Bragg peaks resulting from the (011), (100), (111), (102), (022), (200), (122), (211), (131), (220), (040), (222), (300), (024) and (311) crystallographic planes for all the calcined samples.

The (022) and (200) peaks at roughly 2θ = 45°–46.5° can be used to determine the phase structures of KNN-based ceramics. The splitting of the peak is clearly evident from the XRD plots, with higher intensity of (022) diffraction peak in comparison to (200) diffraction peak indicating that the calcined ceramics are in an orthorhombic phase [48]. The obtained XRD data matches well with the standard JCPDS card no. 98-024-7572 of KNN ceramic with orthorhombic structure having space group *Amm2*. Moreover, no extra peaks were seen, which indicates

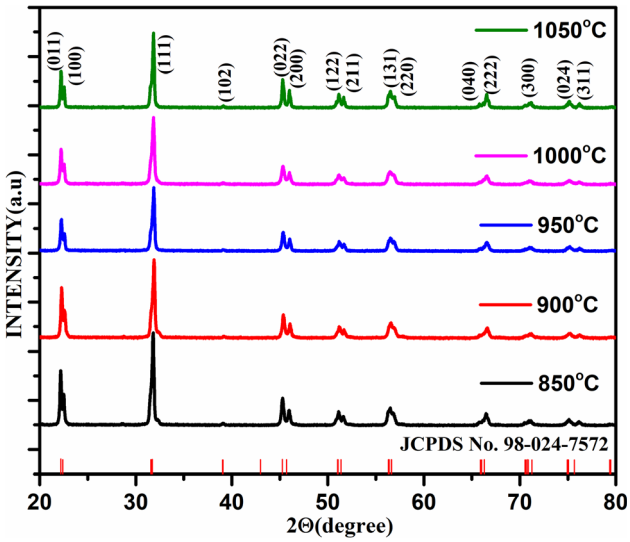


Fig. 3 XRD patterns of KNN ceramics calcined at different temperatures (850 °C, 900 °C, 950 °C, 1000 °C and 1050 °C)

that KNN ceramic powder was successfully synthesized using the solid-state reaction route. At all the calcination temperatures, peaks corresponding to KNN phase were always present, only the peaks of X-ray were becoming narrower and sharper with increasing calcination temperatures from 850 to 1050 °C confirming more crystallinity as expected.

Further, the acquired XRD data of KNN ceramic powder calcined at different temperature were Rietveld refined using Fullprof software. The fitted XRD patterns using the software are presented in Fig. 4 and the corresponding values of lattice parameters along with profile parameters deduced from the refinement were summarized in Table 2. The crystal structure of all the calcined powders matched well with the orthorhombic phase. It can be noticed from the Table 2, that the goodness of fit (χ^2) is lowest for the sample calcined at temperature 1050 °C indicating that the obtained results from refinement were rational and reliable for this sample. The sample calcined at 1050 °C has the lattice constants (a = 3.990 Å, b = 5.608 Å, and c = 5.908 Å). Based on the XRD combined with Rietveld refinement, temperature 1050 °C was chosen as the optimized calcination temperature for the preparation of KNN ceramic powders.

Figure 5 represents the indexed XRD patterns recorded for KNN ceramic powder calcined at 1050 °C and sintered at 1075 °C, 1100 °C, 1125 °C, and 1150 °C and further analyzed by Rietveld Refinement. Peak splitting at ~ 45° in all the

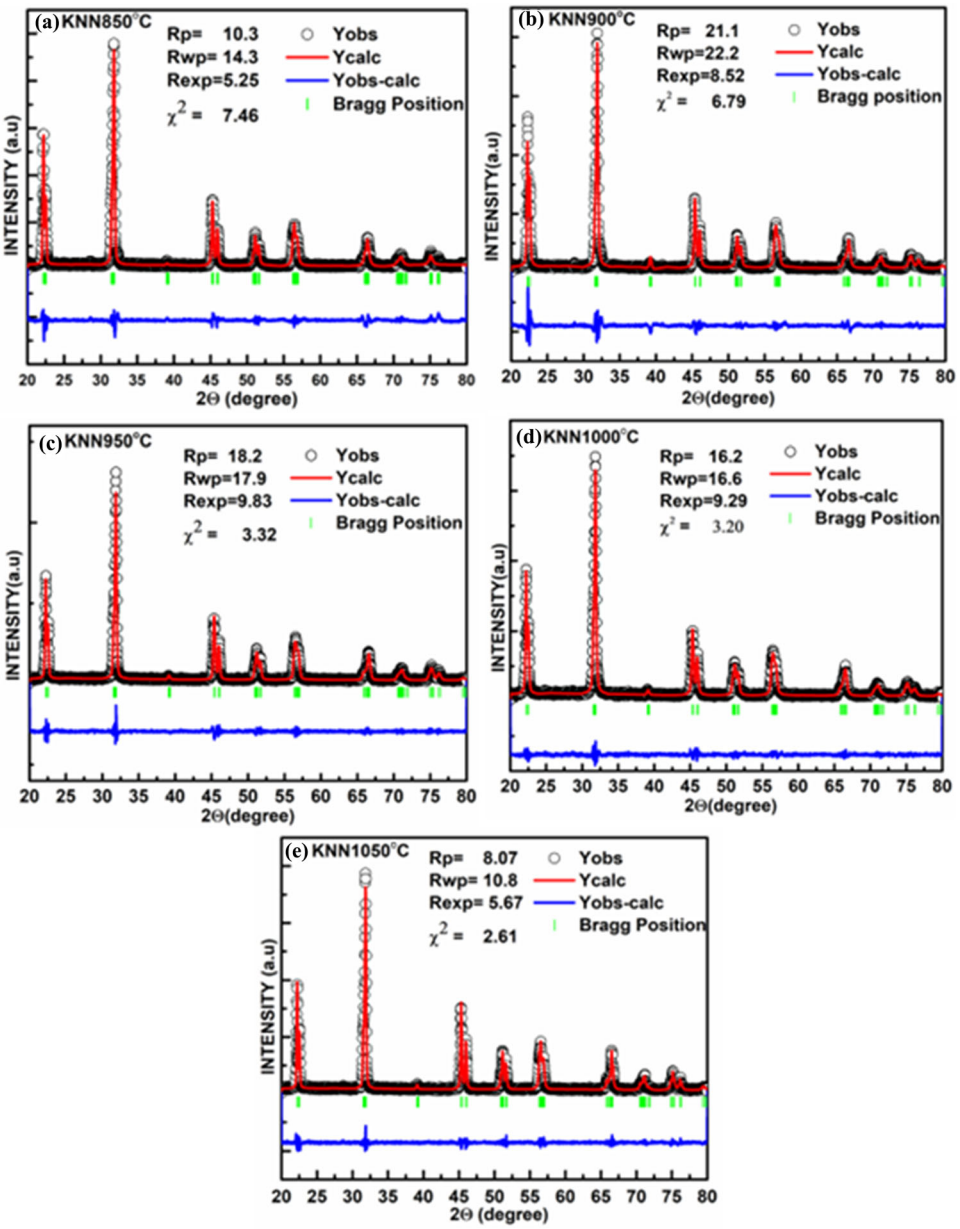


Fig. 4 Rietveld refinement of XRD patterns of KNN ceramics calcined at **a** 850 °C, **b** 900 °C, **c** 950 °C, **d** 1000 °C, and **e** 1050 °C

sintered ceramics can be seen from the XRD plots, with increasing intensity of the (022) peak in comparison to (200) as sintering temperature increases from 1075 to 1150 °C indicating that all the sintered ceramics are in an orthorhombic phase which gradually enhances.

Further, the acquired XRD data of KNN sintered pellets were Reitveld refined using Fullprof software. The fitted XRD patterns are presented in Fig. 6 and the corresponding values of refined lattice parameters along with profile parameters are given in Table 3. From the Table 3, it can be seen that goodness of fit (χ^2) has the lowest value for the sample sintered at 1125 °C.

4.2 Microstructural analysis

SEM analysis of the top surface of the KNN pellets sintered at different temperatures was done to examine the influence of sintering temperature on the microstructure and obtained results indicate considerable effect of sintering temperature on the microstructure and corresponding images were shown in Fig. 7a, d, g, and j, and their respective zoomed images were shown in Fig. 7b, e, h, and k,

Table 2 Refined XRD data of KNN powder using Fullprof Software for samples calcined at 850 °C, 900 °C, 950 °C, 1000 °C and 1050 °C

Parameters	Calcination Temperatures				
	850 °C	900 °C	950 °C	1000 °C	1050 °C
Crystal System	Orthorhombic	Orthorhombic	Orthorhombic	Orthorhombic	Orthorhombic
Cell length, a (Å)	3.949	3.940	3.944	3.946	3.990
Cell length, b (Å)	5.677	5.635	5.639	5.644	5.608
Cell length, c (Å)	5.651	5.664	5.669	5.672	5.908
Cell angle $\alpha = \beta = \gamma$ (°)	90	90	90	90	90
Volume, V (Å^3)	126.7	125.7	126.1	126.3	132.2
Goodness of Fit (χ^2)	7.45	6.79	3.32	3.20	2.61

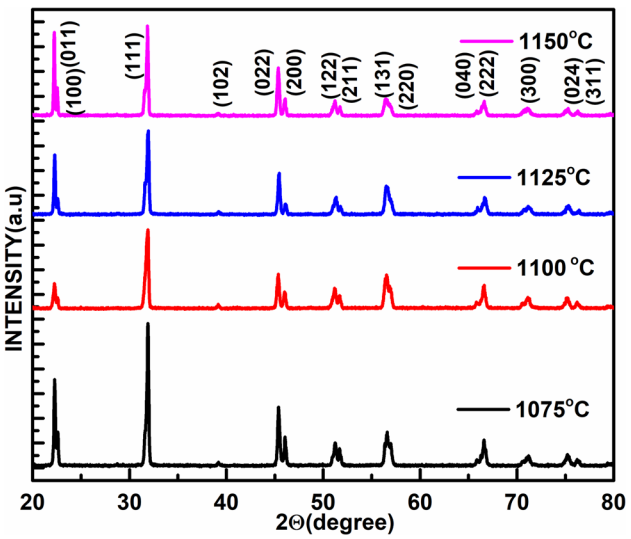


Fig. 5 X-ray diffraction patterns of KNN powders calcined at 1050 °C and sintered at 1075 °C, 1100 °C, 1125 °C and 1150 °C

respectively. Furthermore, the bar histograms shown in Fig. 7c, f, i, and l for sintering temperatures 1075 °C, 1100 °C, 1125 °C, and 1150 °C, respectively indicate the frequency versus particle size distribution which were used to calculate the average grain size using ImageJ Software.

From the Fig. 7a and d slight porosity can be seen for the sample sintered at 1075 °C and 1100 °C, which suggest that these sintering temperatures was not high enough to obtain dense ceramics. On the other hand, when sintering temperature was increased to 1125 °C, formation of very dense and uniform microstructure can be seen from the SEM micrograph (Fig. 7g and h) with average grain size of 3.33 µm. In addition, uniformly distributed grains can also be seen in the Fig. 7g and h. With further increasing the sintering temperature to 1150 °C, significant grain growth with average grain size of 7.16 µm is evident from the SEM images (Fig. 7j and k). It is reported in the literature that, in KNN based ceramics, during the sintering process alkali metals will volatize resulting in the formation of vacancies, which can promote ion migration and making the grains grow up easily [49]. Further, at sintering temperature 1175 °C, partially melting of the sample in the alumina tray is clearly evident in the Fig. 1. The observed melted behavior in the samples sintered at 1150 °C and 1175 °C might be attributed to the high sintering temperature which is close to the solidus temperature of stoichiometric KNN at which liquid phase formation began [50, 51]. Therefore, KNN samples sintered in the temperature range 1075 °C and 1150 °C, were characterized further using various techniques.

In order to analyze the effect of sintering temperature on the density of the KNN pellets sintered at different temperatures, Archimedes principle was employed to measure the density of sintered KNN samples. Table 4 shows the measured density of KNN ceramics sintered at different temperatures. It can be noticed from the Table 4 that the density has increased from 3.89 to 4.38 g/cm^3 with an increase in sintering temperature from 1075 to 1150 °C. The variation of relative density of KNN pellets sintered at different temperatures is shown in Fig. 8. For samples sintered at temperature > 1100 °C the, relative density achieved was > 95% to the reported theoretical density (4.51 g/cm^3) [52]. In addition, the obtained density results of the sintered samples are in agreement with the SEM results which show reduction in pores and better welding of the grains as a function of sintering temperature.

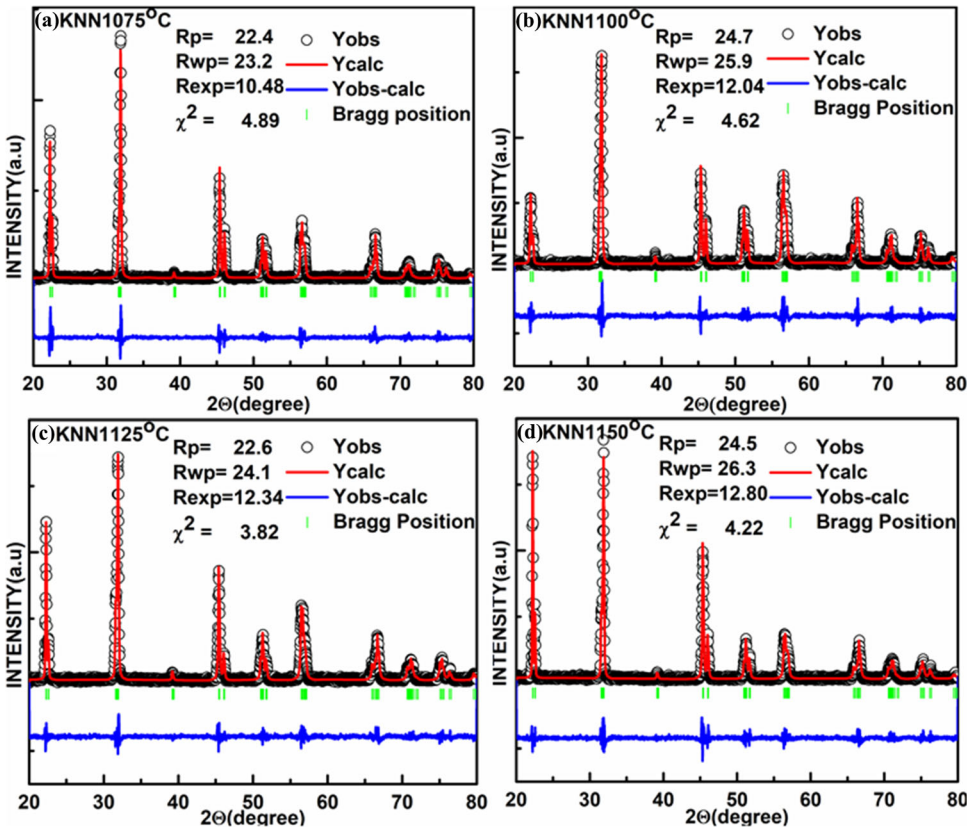


Fig. 6 Rietveld refinement of pure KNN ceramics calcined at 1050 °C and sintered at **a** 1075 °C, **b** 1100 °C, **c** 1125 °C and **d** 1150 °C

Table 3 Refined XRD Data of KNN powder calcined at 1050 °C and sintered at 1075 °C, 1100 °C, 1125 °C and1150 °C

Parameters	Sintering Temperatures			
	1075 °C	1100 °C	1125 °C	1150 °C
Crystal System	Orthorhombic	Orthorhombic	Orthorhombic	Orthorhombic
Cell length, a (Å)	3.940	3.943	3.936	3.942
Cell length, b (Å)	5.633	5.640	5.629	5.639
Cell length, c (Å)	5.663	5.669	5.660	5.669
Cell angle $\alpha = \beta = \gamma$ (°)	90	90	90	90
Volume, V (Å^3)	125.739	126.110	125.452	126.073
Goodness of Fit (χ^2)	4.89	4.62	3.82	4.22

Furthermore, as reported in the literature that low sintering temperature of KNN based ceramics results in the formation of pores and thus low density which is also observed in the present work for the samples sintered at temperature ≤ 1100 °C having relative density $< 95\%$ to the reported theoretical density (4.51 g/cm^3). On the other hand, very high sintering temperature for the KNN based ceramics can easily cause the loss of Na and K in the samples [39]. Based on detailed XRD analysis combined with Rietveld refinement results, SEM analysis and density measurements, it is expected that the KNN pellet sintered at 1125 °C will demonstrate improved piezoelectric properties as it showed uniform microstructure and high density ($> 95\%$ to the reported theoretical density (4.51 g/cm^3)) and lowest value of goodness of fit (χ^2).

4.3 Dielectric Study

Figure 9a and b depicts the variation of the real component of dielectric constant; ε_r and dielectric loss ($tan\ \delta$; a measure of dissipated energy) at room temperature as a function of the frequency of applied oscillating field (in the frequency range from 20 Hz to 200 kHz) of various KNN pellets sintered at different

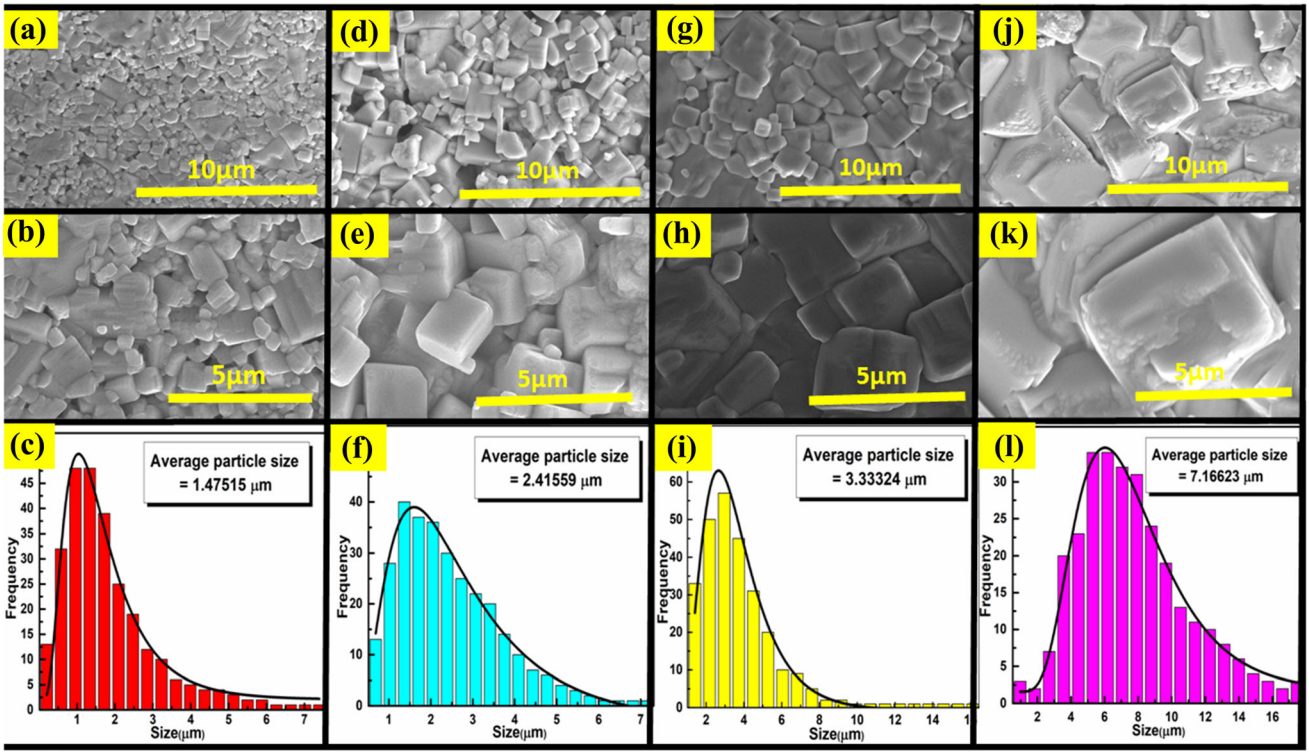


Fig. 7 SEM micrographs of $K_{0.5}Na_{0.5}NbO_3$ pellets sintered at different temperatures **a & b** 1075 °C, **d & e** 1100 °C, **g & h** 1125 °C and **j & k** 1150 °C.**c, f, i & l** represents the histogram distribution curve of KNN ceramic sintered at 1075 °C, 1100 °C, 1125 °C and 1150 °C

Table 4 Average grain size and Density of $K_{0.5}Na_{0.5}NbO_3$ ceramics sintered at different temperatures

Sintering Temperature (°C)	Average Grain size (μm)	Density (g/cm^3)
1075	1.475 ± 0.061	3.89 ± 0.07
1100	2.415 ± 0.306	4.23 ± 0.17
1125	3.333 ± 0.141	4.30 ± 0.15
1150	7.166 ± 0.139	4.38 ± 0.13

temperatures (1075 °C, 1100 °C, 1125 °C and 1150 °C).

From the Fig. 9a it can be seen that the value of the real component of the dielectric constant (ε_r) for all the sintered KNN pellets were found to decrease with increasing frequency followed by a plateau region (i.e., frequency-independent response) for all the sintered samples. The decrease in dielectric constant with increase in frequency may be attributed to decrease in polarization with increase in frequency. In dielectric materials, polarization is the sum of contributions coming from ionic, electronic, dipolar and interfacial polarization [53]. All the polarizations respond easily to the time varying field at low frequencies but as frequency increases contribution from different polarization filters out, resulting in the decrease of net polarization which causes decrease in the dielectric constant values.

It is evident from the Fig. 9a that with increase in sintering temperature from 1075 to 1125 °C dielectric constant increases and then decreases with further increase in the sintering temperature to 1150 °C. The observed increase in dielectric constant (ε_r) with increase in sintering temperature are related to grain size distribution [54]. The observed increase in the dielectric constant with increase in the sintering temperature is attributed to the increase in grain size of the samples Fig. 7. The increased grain size results in a decrease in thickness of the relatively more insulating grain boundary [55] and thus dielectric constant increases with increase in grain size. However, decrease in the value of dielectric constant for the sample sintered at 1150 °C suggests that the grain

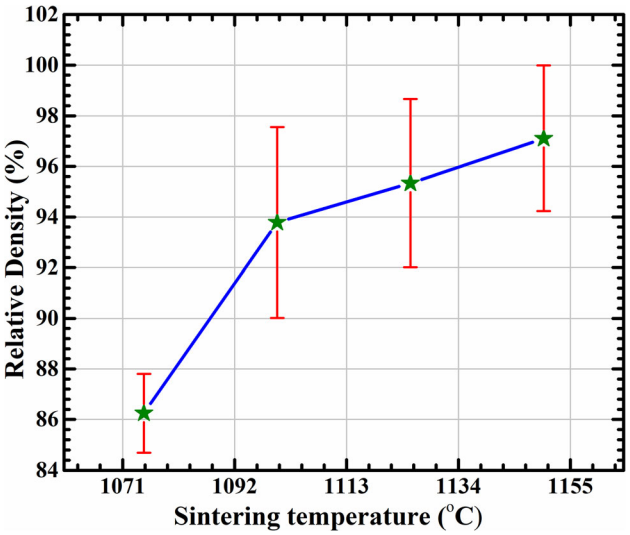


Fig. 8 The relative density (with error bars) of $K_{0.5}Na_{0.5}NbO_3$ pellets at different sintering temperature

size is not the only factor which affects the dielectric constant. In addition to grain size, there are various other factors which includes densification, domain size and lattice distortion which can also affect the dielectric constant of the ceramics [56]. Besides, loss of alkali ions (K^+ and Na^+) might also contribute to the decrease in dielectric constant for the sample sintered at 1150 °C.

The ε_r values of sintered KNN ceramics samples at 10 kHz frequency are ~ 14, ~ 185, ~ 460 and ~ 263 corresponding to the sintering temperatures 1075 °C, 1100 °C, 1125 °C and 1150 °C, respectively. For all the samples (see Fig. 9b), the dielectric loss goes on falling with rising frequency up to 100 kHz according to Debye formula. The room temperature *tan δ* values at 10 kHz frequency are 0.086, 0.108, 0.293 and 0.121 for KNN pellets sintered at 1075 °C, 1100 °C, 1125 °C and 1150 °C, respectively.

The temperature dependence of the real component of dielectric constant (ε_r) for KNN pellet sintered at 1125 °C at selected frequencies (10 KHz to 2 MHz) in the temperature range from 30 to 490 °C is displayed in Fig. 10. The dielectric constant (ε_r) increases with increasing temperature and achieves a maximum value at temperature (T_c ~ 376 °C) and thereafter rapidly decreases on further increasing the temperature. From the Fig. 10, two transitions can be clearly seen in the plot at temperatures ~ 183 °C and ~ 376 °C which corresponds to the orthorhombic phase to the tetragonal phase transition and a ferroelectric (tetragonal) phase to the paraelectric (cubic) phase transition of KNN ceramics, respectively. The observed two transition temperatures for the KNN sample sintered at 1125 °C are in agreement with the earlier reported literature [57].

4.4 Ferroelectric Hysteresis Analysis

The room temperature *P-E* hysteresis loops of the synthesized KNN ceramics sintered at different temperatures (1075 °C, 1100 °C, 1125 °C and 1150 °C) tested at different frequencies (@200, 250, 300, 350, 400, and 450 Hz) are shown in Fig. 11. All samples at different electric fields below their breakdown strength exhibits typical *P-E* loops suggesting that all the sintered samples are ferroelectrics. It was noted that the remnant as well as spontaneous polarization

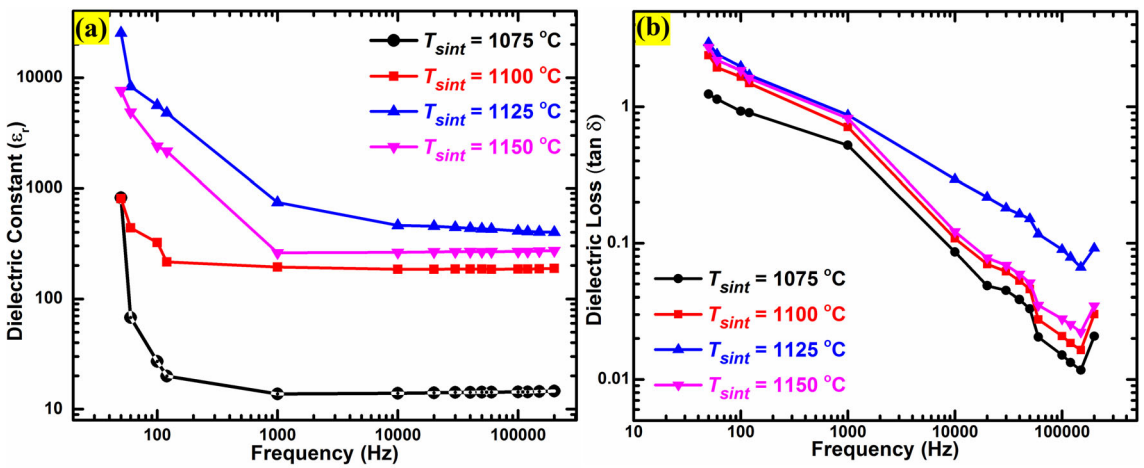


Fig. 9 Variation of **a** dielectric constant and **b** dielectric loss at room temperature as a function of frequency for KNN ceramics sintered at different temperatures

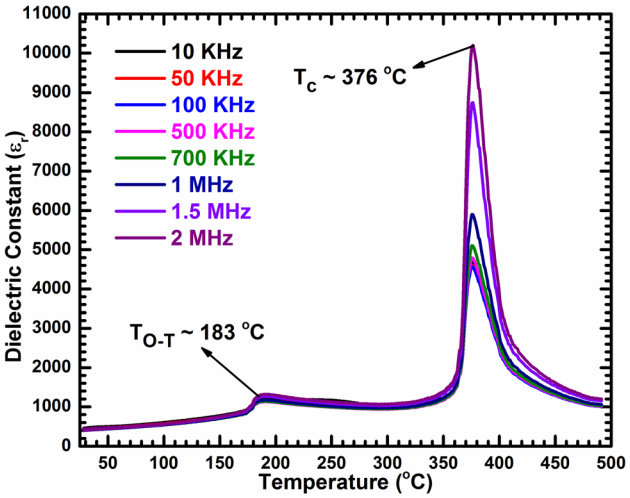


Fig. 10 Temperature dependence of dielectric constant (ε_r) of KNN ceramics calcined at 1050 °C & sintered at 1125 °C at selected frequencies

decreases on increasing the frequency. Conceptually, as frequency is increased, some of the domains in KNN ceramic are not able to reverse their orientation in the response to such a frequent oscillation of applied electric field, so a fall in polarization is observed. Further, there are a few difficulties faced by domain walls in switching such as viscosity or resistance during movement, which becomes more prominent at higher frequencies [58]. Furthermore, it is clearly evident from Fig. 11 that sintering temperature has a significant effect on the shapes of the *P-E* hysteresis loops at all the measured frequencies which may be ascribed to the microstructure and grain size of the samples. Figure 12 shows the variation of remnant polarization (P_r) and coercive field (E_c) as a function of sintering temperature at different frequencies.

It can be seen from the plots, that in general with increase in sintering temperature from 1075 to 1125 °C remnant polarization (P_r) first increases and then decreases when sintering temperature increased to 1150 °C at all frequencies. Almost similar trend was observed for coercive field (E_c).

The observed high value of P_r for the sample sintered at 1125 °C is due to its highly dense, uniform microstructure and average grain size of 3.33 μm.

It is reported in the literature, that the ferroelectric properties are significantly influenced by the domain structure, domain nucleation and domain mobility which in turn are affected by the grain size [54]. The ferroelectric property is substantially suppressed in ceramics with small grain size due to grain boundary effects which causes hindrance in polarisation switching as reversal of polarization of a domain is much easier in a large grain in comparison to small grain [59]. The observed variation of P_r is in accordance with the grain size variation noticed in present work. However, for the sample sintered at 1150 °C despite larger grain size and higher density, lower value of P_r was obtained in comparison to the sample sintered at 1125 °C which is similar to the behavior demonstrated by dielectric constant as a function of sintering temperature and can be interpreted on the similar lines.

Table 5 depicts the values of remnant polarization, coercive field, and maximum polarization at frequency 250 Hz for the KNN ceramics sintered at different temperatures. The values of ferroelectric parameters (E_c & P_r) obtained for sintered KNN ceramics were found to be in synchronization with the earlier reported literature [6, 17, 25, 26, 28, 40, 60–70]. Further, it can be seen from the table that the KNN sample sintered at 1125 °C shows the highest remnant polarization (P_r) of 98.21 μC/cm^2 and a coercive field (E_c) of 5.73 kV/cm at 250 Hz frequency.

The squareness factor (R_{sq}) of the recorded ferroelectric *P-E* loops of all the sintered KNN sample was evaluated with the help of empirical relation [71, 72],

$$R_{sq} = \frac{P_r}{P_s} + \frac{P_{1.1Ec}}{P_r}$$

where, $P_{1.1Ec}$ denotes the polarization measured at an electric field 1.1 times E_c. For a perfect saturated *P-E* hysteresis loop, the value of R_{sq} must be equal to 2. The values of R_{sq} are depicted in Table 5. It can be seen from the table that for the sample sintered at 1125 °C, the value of R_{sq} is closest to 2, thereby, indicating excellent 'switching' applications of KNN ceramic sintered at 1125 °C in switching devices.

4.5 Piezoelectric study

Table 6 lists the direct piezoelectric charge coefficient (d_{33}; measured by a direct piezoelectric effect) and converse piezoelectric charge coefficient (d_{33}*; measured by a converse piezoelectric effect) for KNN ceramics calcined at 1050 °C and sintered at different temperatures.

It can be observed from the Table 6 that d_{33} of the KNN pellets increases with increase in sintering temperature from 1075 to 1125 °C and then decreases

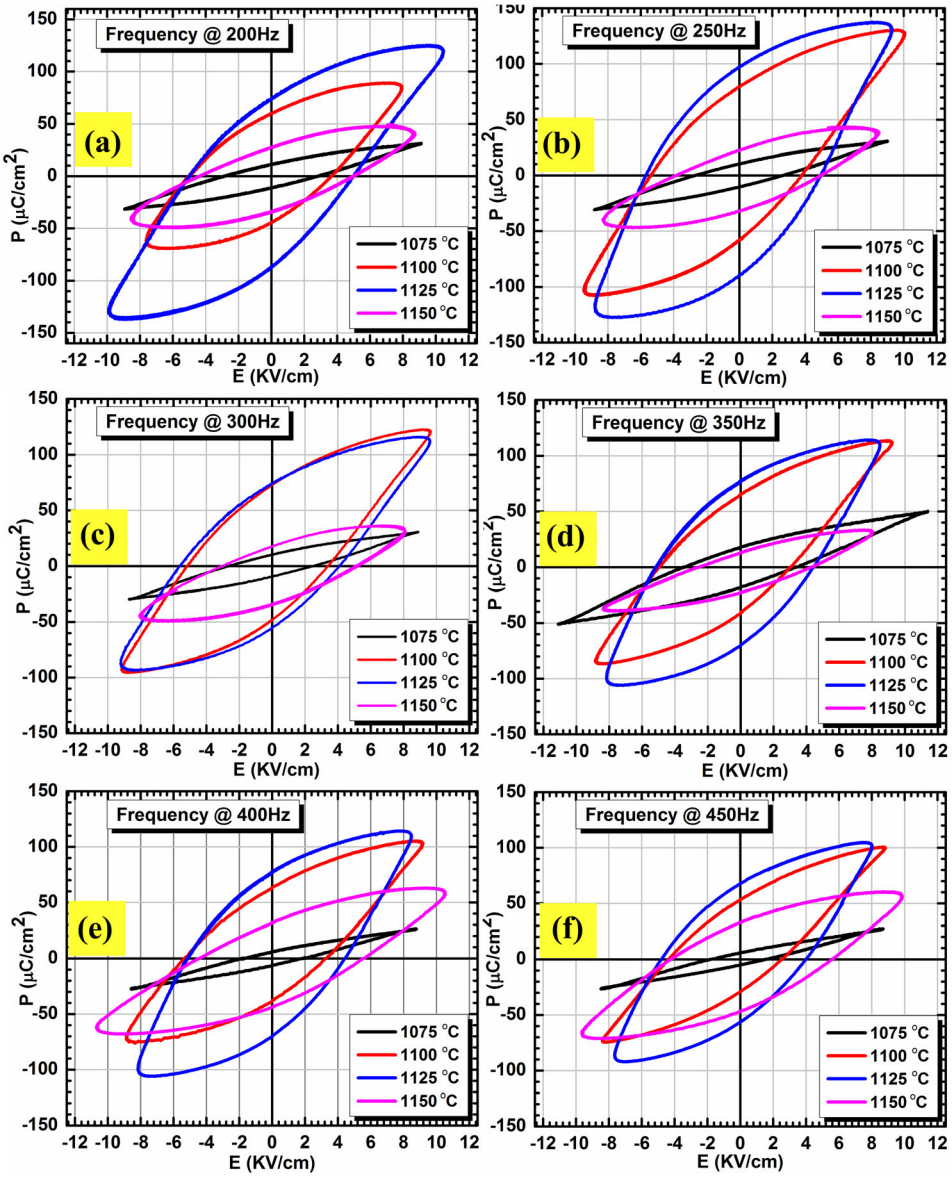


Fig. 11 Room temperature *P-E* hysteresis loops for KNN ceramic pellets prepared at different sintering temperatures (@1075 °C, 1100 °C, 1125 °C and 1150 °C) measured at different frequencies **a** 200 Hz, **b** 250 Hz, **c** 300 Hz, **d** 350 Hz, **e** 400 Hz and **f** 450 Hz

when sintering temperature further increased to 1150 °C. According to the empirical relationship $d_{33} \propto \varepsilon_r P_r$, enhanced ferroelectric and dielectric properties are beneficial for getting the improved piezoelectric properties [73]. The observed variation of d_{33} and $\varepsilon_r P_r$ as a function of sintering temperature is shown in Fig. 13. It is clearly evident from the plot that d_{33} and $\varepsilon_r P_r$ show similar trend, both of them first increases with increase in sintering temperature and then decreases, with d_{33} reaching the maximum value of 290 pC/N at 1125 °C. Thus, the observed variation of d_{33} as a function of sintering temperature for KNN ceramics can be mainly ascribed to the associated change in P_r and ε_r.

Further, Fig. 14 displays the butterfly loop, i.e., displacement versus voltage (*D-V*) loops of KNN ceramic pellets prepared at different sintering temperatures (@1075 °C, 1100 °C, 1125 °C and 1150 °C). An AC bias voltage was applied to the KNN pellets and the corresponding displacement was recorded with the help of linear variable differential transform (LVDT) sensor. The maximum values obtained for displacement at the corresponding applied voltages for the KNN samples sintered at different temperatures are listed in the Table 7.

In addition, the converse piezoelectric effect was used to evaluate the effective converse piezoelectric charge coefficient (d_{33}^*) for KNN pellets from the

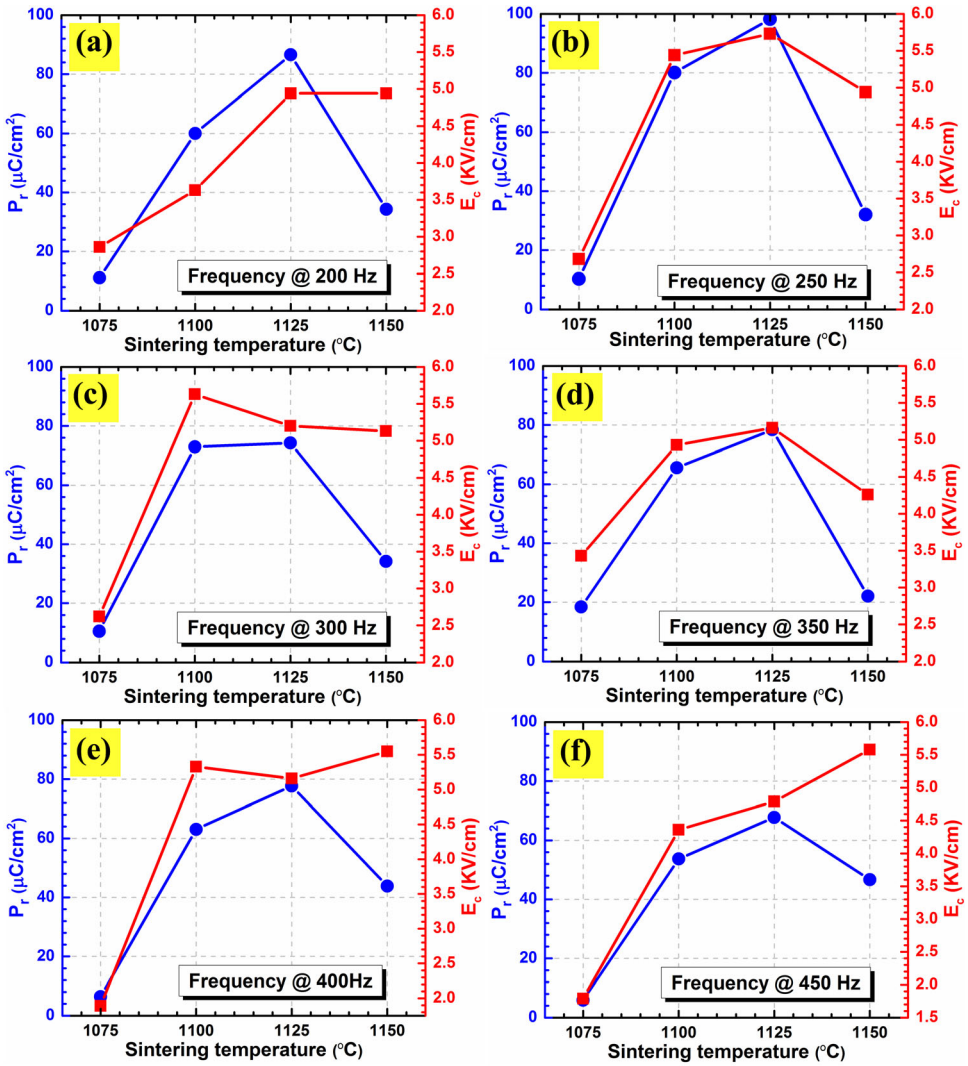


Fig. 12 Variation of ferroelectric characteristics (P_r & E_c) with sintering temperature at different frequencies **a** 200 Hz, **b** 250 Hz, **c** 300 Hz, **d** 350 Hz, **e** 400 Hz & **f** 450 Hz

Table 5 Ferroelectric parameters (P_r, E_c, P_{max} and R_{sq}) of KNN ceramics sintered at different temperatures at a frequency of 250 Hz

Sintering Temperature (°C)	P_r (μC/cm^2)	E_c (KV/cm)	$P_{max.}$ (μC/cm^2)	R_{sq}
1075	10.31	2.68	30.61	2.51
1100	80.10	5.44	130.42	2.34
1125	98.21	5.73	137.95	2.11
1150	32.08	4.94	46.50	2.33

Table 6 Piezoelectric charge coefficients (direct; d_{33} and converse d_{33}*) of KNN ceramic pellets prepared at a calcination temperature of 1050 °C and at different sintering temperatures (@1075 °C, 1100 °C, 1125 °C and 1150 °C)

Sintering Temperature (°C)	d_{33} (pC/N)	d_{33}* (pm/V) Max.	d_{33}* (pm/V) Avg.
1075	80 ± 2.4	196	90 ± 110.67
1100	240 ± 12.2	695	296 ± 313.3
1125	290 ± 14.3	718	321 ± 182.30
1150	210 ± 8.4	535	234 ± 144.18

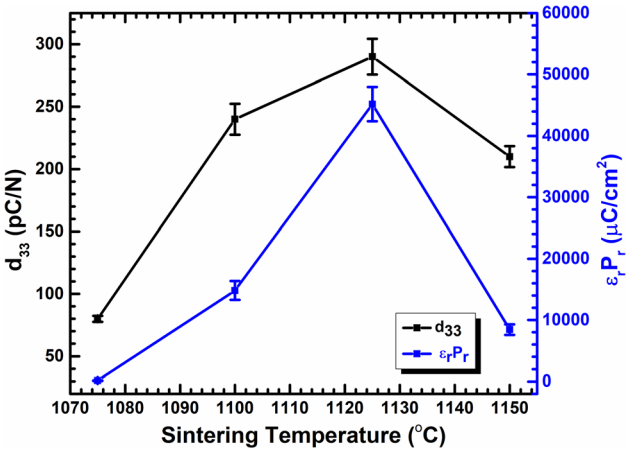


Fig. 13 Variation of d_{33} and $\varepsilon_r P_r$ (with error bars) of KNN ceramics as a function of sintering temperature

slope of *D-V* (butterfly) curves (see Fig. 14) [11–16], which is as follows:

$$d_{33}^{*} = \frac{(D - D_0)}{(V - V_0)}$$

where D is the displacement (measured in μm) of KNN pellet in response to the applied voltage V (in volts), V_o and D_o represent the applied voltage and displacement at the intersection point of the *D-V* loop that must be deducted from each (D, V) point to remove the effect of unwanted origin reallocation. The maximum and average values of effective d_{33}^{*} for KNN ceramic pellets sintered at 1075 °C, 1100 °C, 1125 °C and 1150 °C are tabulated in Table 6. It can be seen from the Table 6 that the giant piezoelectric

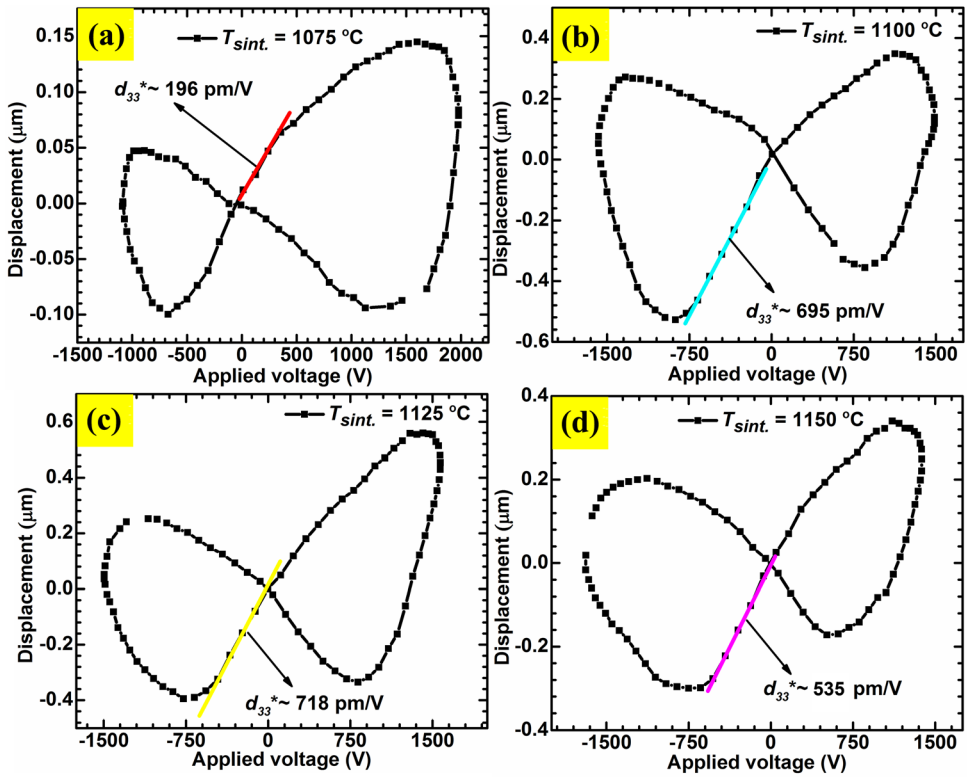


Fig. 14 Room temperature displacement (D) vs. voltage (V) (butterfly loops) for KNN pellets prepared at different sintering temperatures: **a** 1075 °C, **b** 1100 °C, **c** 1125 °C & **d** 1150 °C

Table 7 Maximum displacement of KNN ceramics sintered at different temperatures

Sintering Temperature (°C)	Maximum Displacement (μm)	Applied field (V)
1075	144.93	1595
1100	529.69	875
1125	556.59	1425
1150	341	1117

Table 8 A brief comparison of ferro- to para-electric phase transition temperature (T_c), piezoelectric charge coefficients (direct; d_{33} and converse d_{33}*) and ferroelectric characteristics (P_r, E_c) of best-optimized KNN ceramics (in present work) with that of earlier reported values for various other KNN samples by some other research groups across the globe

S. No	KNN ceramic	E_c (kV/cm)	P_r (μC/cm^2)	T_c (°C)	d_{33} (pC/N)	d_{33}* (pC/N)	Reference
1	$K_{0.5}Na_{0.5}NbO_3$ T_{cal} = 825 °C T_{sint} = 1114 °C	20	20	400	100	–	[40]
2	$K_{0.5}Na_{0.5}NbO_3$ T_{cal} = – T_{sint} = 1100 °C	24	15	411	–	80	[60]
3	$K_{0.5}Na_{0.5}NbO_3$ T_{cal} = 900 °C T_{sint} = 1200 °C	5	33	–	160	240	[61]
4	$K_{0.31}Na_{0.69}NbO_3$ T_{cal} = 450 °C T_{sint} = 1120 °C	8.31	5.8	420	84	–	[62]
5	$K_{0.5}Na_{0.5}NbO_3$ T_{cal} = 860 °C T_{sint} = 1075 °C	4.57	25.5	409	117	–	[63]
6	$K_{0.5}Na_{0.5}NbO_3$-SnO_2 T_{cal} = 950 °C T_{sint} = 1100 °C	11.5	17	405	108	–	[64]
7	$K_{0.5}Na_{0.5}NbO_3$-Sc_2O_3 T_{cal} = 950 °C T_{sint} = 1100 °C	12.5	28	394	100	–	[64]
8	$K_{0.5}Na_{0.5}NbO_3$-ZnO T_{cal} = 950 °C T_{sint} = 1100 °C	15	23	418	117	–	[64]
9	$K_{0.5}Na_{0.5}NbO_3$ (ZnO_2 + SnO_2) T_{cal} = 860 °C T_{sint} = 1100–1120 °C	17.33	24.5	420	124	–	[65]
10	$K_{0.5}Na_{0.5}NbO_3$-CuO (Textured) T_{cal} = 600 °C T_{sint} = 1175 °C	–	–	–	68	–	[66]
11	$K_{0.5}Na_{0.5}NbO_3$ (Spark Plasma Texturing) T_{cal} = 900 °C T_{sint} = 1000 °C	11.9	19.7	–	108	–	[67]
12	$K_{1-x}Na_xNbO_3$ (x = 20-80 mol%) (Spark Plasma Sintering) T_{cal} = 750 °C T_{sint} = 920 °C	–	–	–	148	–	[68]
13	$K_{0.5}Na_{0.5}NbO_3$ (Spark Plasma Sintering) T_{cal} = 750 °C T_{sint} = 900 °C	13	6.4	400	148	–	[69]
14	$K_{0.5}Na_{0.5}NbO_3$ (Oscillatory Pressure-assisted hot-pressing) T_{cal} = 850 °C T_{sint} = 860 °C	–	–	–	142	–	[70]

Table 8 continued

S. No	KNN ceramic	E_c (kV/cm)	P_r (μC/cm^2)	T_c (°C)	d_{33} (pC/N)	d_{33}* (pC/N)	Reference
15	$(Bi_{0.5}Na_{0.5})TiO_3$ T_{cal} = 900 °C T_{sint} = 1200 °C	37	40	–	230	–	[6]
16	$0.47(Bi_{0.5}Na_{0.5})TiO_3$-$0.04BaTiO_3$ T_{cal} = 870 °C T_{sint} = 1030 °C	16.56	12.19	–	–	177	[17]
17	$0.95BaTiO_3$-$0.05Bi(Zn_{2/3}Nb_{1/3})O_3$ T_{cal} = 1125 °C T_{sint} = 1300 °C	8	5	120	22.7	–	[25]
18	(1-x)$BiScO_3$-x$PbTiO_3$ (x = 60 mol%) T_{cal} = 1125 °C T_{sint} = 1300 °C	20	32	450	260	–	[26]
19	$0.47(Bi_{0.5}Na_{0.5})TiO_3$-$0.04BaTiO_3$-$0.31Pb(Mg_{1/3}Nb_{2/3})$-$0.18PbTiO_3$ T_{cal} = 870 °C T_{sint} = 1030 °C	16.56	13.09	190	–	177	[28]
20	$K_{0.5}Na_{0.5}NbO_3$ T_{cal} = 1050 °C T_{sint} = 1125 °C	5.73	98.2	376	290	321	[Present Work]

characteristics (d_{33} ~ 290 pC/N, $d_{33}^*{}_{max.}$ ~ 718 pm/V and $d_{33}^*{}_{avg.}$ ~ 321 pm/V) have been obtained for the KNN sample sintered at temperature 1125 °C. Moreover, at same sintering temperature (1125 °C), maximum values of dielectric constant and remnant polarization (P_r) with dense and uniform microstructure were achieved.

Overall, it may be concluded that a calcination temperature of 1050 °C along with a sintering temperature of 1125 °C represent the best-optimized conditions for the preparation of KNN ceramic with high dielectric, ferroelectric and piezoelectric properties which was also expected from the density and SEM analysis in the present work. In Table 8 the di-/piezo-/ferro-electric response of the $K_{0.5}Na_{0.5}NbO_3$ ceramic prepared under best-optimized conditions (T_{cal} = 1050 °C and T_{sint} = 1125 °C in present work) are compared with KNN ceramics prepared under different conditions by several other researchers. The obtained enhanced dielectric, ferroelectric and piezoelectric properties of the KNN ceramics sintered at 1125 °C in the present work, might be due to the large grain size, high density and uniform distribution of the grains and absence of any impurity phases in its XRD pattern. In addition, the high value of piezoelectric coefficient obtained is also associated with large remnant polarization and dielectric constant. Furthermore, large ferroelectric, piezoelectric and dielectric properties have also been reported in the literature which are obtained by employing quite expensive techniques such as spark plasma sintering, hot pressing and spark plasma texturing technologies [67–70]. But due to their high cost, and no massive production, makes it difficult to commercialize to industrial production. On the other hand, conventional sintering technique which has been used in the present work is an effective method which can be commercialized with ease, owing to its low cost and massive production.

5 Conclusions

In the present work, $K_{0.5}Na_{0.5}NbO_3$ (KNN) ceramics were prepared by conventional solid state reaction method followed by sintering at different temperatures in the range 1050–1150 °C. Influence of sintering temperature on the densification, microstructure,

dielectric, ferroelectric and piezoelectric properties were investigated and from the obtained experimental findings following conclusions can be extracted:

1 The XRD results imply that the KNN ceramic powder were successfully prepared at different calcination temperatures in the range 850–1050 °C. In all the samples, no impurity phase was detected. Further, (022) and (200) peaks at roughly $2\theta = 45°–46.5°$ with higher intensity of (022) diffraction peak in comparison to (200) diffraction peak clearly indicating that the calcined ceramics are in an orthorhombic phase. Based on the XRD analysis combined with Rietveld refinement, temperature 1050 °C was chosen as the optimized calcination temperature for the preparation of KNN ceramic powders.
2 SEM images showed the presence of slight porosity in samples sintered at temperatures < 1125 °C. At sintering temperature 1125 °C a very dense and uniform microstructure was obtained. With further increase in sintering temperature, abnormal grain growth and melting behavior of KNN ceramics were observed. Besides, increase in average grain size and density with increase in sintering temperature from 1050 to 1150 °C were also observed.
3 Significant effect of sintering temperature was observed on the dielectric, ferroelectric and piezoelectric properties of the prepared KNN ceramics. The dielectric constant, remnant polarization and piezoelectric constant all exhibited increasing trend with increase in sintering temperature from 1050 to 1125 °C and then decreases when sintering temperature increased to 1150 °C. The observed variation was explained on the basis of grain size distribution.
4 The sample calcined at 1050 °C and sintered at 1125 °C demonstrated the highest values of all the measured properties: dielectric constant ($\varepsilon_r \sim 460$ @10 kHz), giant piezoelectric charge coefficient (d_{33}) ~ 290 pC/N, converse piezoelectric charge coefficient ($d_{33}{}^*_{avg.}$) ~ 321 pC/N and excellent ferroelectric parameters, $P_r \sim 98.21$ μC/cm^2 and $E_c \sim 5.73$ kV/cm.
5 Based on the systematic study done in the present work, it can be concluded that the calcination temperature of 1050 °C along with sintering temperature of 1125 °C represent the best-optimized conditions for the preparation of KNN ceramic with highly dense and uniform microstructure with excellent di-/piezo-/ferroelectric characteristics.

Acknowledgements

The authors wish to thank the DTU Project Research Grant (F.NO.DTU/IRD/619/2019/2113) for providing the research grant to carry out this work. The authors would also like to thank the Delhi Technological University (DTU), India for providing the DTU fellowship (2K19/PhDAP/06).

Author contributions

KV contributed in conceptualization, methodology, investigations, data analysis and has written the original manuscript draft. RS contributed in conceptualization, writing, reviewing, editing, project administration, and supervision of the manuscript at different stages. SG did the data collection and analysis writing, reviewing, editing, and helped in writing of original manuscript draft. All authors agreed to the final version of the manuscript.

Funding

The funding was provided by DTU Project Research Grant (F.NO.DTU/IRD/619/2019/2113) to carry out this research work.

Data availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

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.

Supplementary Information: The online version contains supplementary material available at https://doi.org/10.1007/s10854-022-09295-2.

References

1. S. Goel, B. Kumar, J. Alloys Compd. **816**, 152491 (2020)
2. X. Zhang, J. Chen, Y. Wang, A.C.S. Appl, Nano Mater. **1**, 1461 (2018)
3. L. Qi, S. Ruan, Y.-J. Zheng, Adv. Mater. **33**, 2005098 (2021)
4. T. Akutagawa, Bull. Chem. Soc. Jpn. **94**, 1400–1420 (2021)
5. S. Zhou, L. You, H. Zhou, Y. Pu, Z. Gui, J. Wang, Front. Phys. **16**(1), 13301 (2021)
6. D.Q. Xiao, D.M. Lin, J.G. Zhu, P. Yu, J. Electroceram. **21**, 34 (2008)
7. S.T. Lau, C.H. Cheng, S.H. Choy, D.M. Lin, K.W. Kwok, H.L.W. Chan, Journal of Applied Physics **103**, 104105 (2008)
8. T.R. Shrout, S.J. Zhang, J. Electroceram. **19**, 111 (2007)
9. M.H. Lee, R. Guo, A.S. Bhalla, *Pyroelectric Sensors* (1998).
10. S. Goel, N. Sinha, H. Yadav, A.J. Joseph, B. Kumar, J. Mater. Sci. Mater. Electron. **29**, 13818 (2018)
11. H. Yadav, N. Sinha, S. Goel, B. Kumar, J. Alloys Compd. **689**, 333 (2016)
12. S. Goel, N. Sinha, A. Hussain, A.J. Joseph, B. Kumar, Ionics (Kiel). **86**, 142907 (2018)
13. S. Goel, N. Sinha, B. Kumar, Phys. E Low-Dimensional Syst. Nanostruct. **105**, 29 (2019)
14. S. Goel, N. Sinha, H. Yadav, A.J. Joseph, B. Kumar, Phys. E Low-Dimensional Syst. Nanostruct **91**, 72 (2017)
15. S. Goel, N. Sinha, H. Yadav, S. Godara, A.J. Joseph, B. Kumar, Mater. Chem. Phys. **202**, 56 (2017)
16. N. Sinha, S. Goel, A.J. Joseph, H. Yadav, K. Batra, M.K. Gupta, B. Kumar, Ceram. Int. **44**, 8582 (2018)
17. A.J. Joseph, N. Sinha, S. Goel, A. Hussain, B. Kumar, Arab. J. Chem. **42**, 115421 (2018)
18. A. Hussain, N. Sinha, A.J. Joseph, K. Dhankhar, S. Goel, B. Kumar, J. Mater. Sci. Mater. Electron. **28**, 14298 (2017)
19. A. Hussain, N. Sinha, A.J. Joseph, S. Goel, B. Kumar, J. Mater. Sci. Mater. Electron. **29**, 19567 (2018)
20. C. He, Z. Wang, X. Li, Y. Liu, D. Shen, T. Li, X. Long, J. Phys. D. Appl. Phys. **45**, 105305 (2012)
21. N. Luo, S. Zhang, Q. Li, C. Xu, Z. Yang, Q. Yan, Y. Zhang, T.R. Shrout, A.C.S. Appl, Mater. Interfaces **8**, 15506 (2016)
22. A.J. Joseph, S. Goel, A. Hussain, B. Kumar, Ceram. Int. **43**, 16676 (2017)
23. A.J. Joseph, N. Sinha, S. Goel, A. Hussain, B. Kumar, Ceram. Int. **44**, 18633 (2018)
24. N. Luo, S. Zhang, Q. Li, Q. Yan, W. He, Y. Zhang, T.R. Shrout, Appl. Phys. Lett. **104**, 182911 (2014)
25. A. Paterson, H.T. Wong, Z. Liu, W. Ren, Z.-G. Ye, Ceram. Int. **41**, S57 (2015)
26. R.E. Eitel, C.A. Randall, T.R. Shrout, S.-E. Park, Jpn. J. Appl. Phys. **41**, 2099 (2002)
27. Z. Liu, A.R. Paterson, H. Wu, P. Gao, W. Ren, Z.-G. Ye, J. Mater. Chem. C **5**, 3916 (2017)
28. A.J. Joseph, N. Sinha, S. Goel, A. Hussain, H. Yadav, B. Kumar, J. Mater. Sci. Mater. Electron. **30**, 12729 (2019)
29. G.H. Khorrami, A. Kompany, A. KhorsandZak, Mater. Lett. **110**, 172 (2013)
30. X. Pang, J. Qiu, K. Zhu, J. Du, Ceram. Int. **38**, 2521 (2012)
31. J.B. Lim, Y.H. Jeong, M.H. Kim, D. Suvorov, J.H. Jeon, Ceram. Int. **38**, 2605 (2012)
32. Y. Li, Y. Li, C. Hui, M. Wu, Y. Wang, Ceram. Int. **38**, 2605–2608 (2012)
33. T. Takenaka, H. Nagata, Y. Hiruma, Y. Yoshii, K. Matumoto, J. Electroceram. **19**, 259 (2007)
34. E. Aksel, J.L. Jones, Sensors **10**, 1935 (2010)
35. J. Wu, D. Xiao, J. Zhu, Chem. Rev. **115**, 2559 (2015)
36. N.M. Hagh, B. Jadidian, A. Safari, J. Electroceram. **18**, 339 (2007)
37. D.W. Baker, P.A. Thomas, N. Zhang, A.M. Glazer, Appl. Phys. Lett. **95**, 091903 (2009)
38. D.I. Woodward, I.M. Reaney, Acta Crystallogr. B **61**, 387 (2005)
39. B. Malic, J. Koruza, J. Hrescak, J. Bernard, K. Wang, J. Fisher, A. Bencan, Materials **8**(12), 8117–8146 (2015)
40. H. Birol, D. Damjanovic, N. Setter, J. Eur. Ceram. Soc. **26**, 861 (2006)
41. J.F. Li, K. Wang, B.P. Zhang, L.M. Zhang, J. Am. Ceram. Soc. **89**, 706 (2006)
42. S.H. Park, C.W. Ahn, S. Nahm, J.S. Song, JPN J. Appl. Phys. 2: Lett. **43**, (2004).
43. M. Matsubara, T. Yamaguchi, W. Sakamoto, K. Kikuta, T. Yogo, S.I. Hirano, J. Am. Ceram. Soc. **88**, 1190 (2005)
44. S. Zhang, H. Lee, C. Ma, X. Tan, J. Am. Ceram. Soc. **94**, 3659–3665 (2011)
45. H.C. Thong, A. Payne, J.W. Li, Y.S. Cheng, J.L. Jones, K. Wang, Acta Mater. **211**, 116833 (2021)
46. H. He, X. Lu, E. Hanc, C. Chen, H. Zhang, L. Lu, J. Mater. Chem. C **8**, 1494 (2020)
47. K. Wang, J.F. Li, J. Am. Ceram. Soc. **93**, 1101 (2010)
48. J. Sharma, D. Kumar, A. Sharma, Solid State Communications, 334–335 (2021).
49. L. Xie, J. Xing, Z. Tan, L. Jiang, Q. Chen, J. Wu, W. Zhang, D. Xiao, J. Zhu, J. Alloy. Compd. **758**, 14–24 (2018)
50. A. Erdem, Y. Huseyin, Ceram. Int. **41**, 3659–3667 (2015)
51. A. Lacomini, S. Garroni, G. Mulas, S. Enzo, L. Cappai, M. Mureddu, C. Cau, A. Garcia, L. Pardo, Open Ceram. **9**, 100247 (2022)

52. R.E. Jaeger, L. Egerton, J. Am. Ceram. Soc. **45**, 209 (1962)
53. A. Singh, T. Goel, R. Mendirata, O. Thakur, C. Prakash, J. Appl. Phys. **91**(100), 6626–6629 (2002)
54. C. Jiten, R. Gaur, R. Laishram, K. Singh, Ceram. Int. **42**(12), 14135–14140 (2016)
55. P. Kumar, P. Patel, Ceram. Int. **36**(5), 1725–1729 (2010)
56. L. Xin, C. Xiao, L. Mei, Y. Zi, L. Jia, Z. Xiaoziyue, L. Jiang, J. Hong, Ceram. Int. **48**, 11615–11622 (2022)
57. P. Kumar et al., Ceram. Int. **39**, 65 (2013)
58. A. Hussain, N. Sinha, A.J. Joseph, S. Goel, B. Singh, I. Bdikin, B. Kumar, Arab. J. Chem. **120**, 064103 (2018)
59. C. Leu, C. Chen, M. Chang, F. Hsu, C. Hu, Appl. Phys. Lett. **82**(20), 3493–3495 (2003)
60. H. Uršič, A. Benčan, M. Škarabot, M. Godec, M. Kosec, J Appl. Phys. **107**, 033705 (2010)
61. S. Zhang, H.J. Lee, C. Ma, X. Tan, J. Am. Ceram. Soc. **94**, 3659 (2011)
62. P. Palei, M. Pattanaik, P. Kumar, Ceram. Int. **38**, 851 (2012)
63. D. Zhang, Z. Zhang, Ferroelectrics **466**, 8 (2014)
64. R. Zuo, J. Rodel, R. Chen, L. Li, J. Am. Ceram. Soc. **89**, 2010–2015 (2006)
65. Z. Li, G. Xu, Y. Li, A. Sun, L. Jiang, P. Cui, Phys. B Condens. Matter **405**, 296–299 (2010)
66. Y. Saito, H. Takao, J. Eur. Ceram. Soc. **27**, 4085–4092 (2007)
67. R. Pinho, A.T. Kach, S. Zlotnik, M.E. Costa, J. Noudem, I.M. Vilarinho, Appl. Mater. Today **19**, 100566 (2020)
68. B.P. Zhang, J.F. Li, K. Wang, J. Am. Ceram. Soc. **89**(5), 1605–1609 (2006)
69. J.F. Li, K. Wang, J. Am. Ceram. Soc. **89**(2), 706–709 (2006)
70. Z. Li, H. Sun, X. Liu, H. Sui, H. Guo, Ceram. Int. **46**, 11617–11621 (2020)
71. G. Ray, N. Sinha, S. Bhandari, B. Kumar, Mater. Chem. Phys. **159**, 107–113 (2015)
72. A. Hussain, N. Sinha, A.J. Joseph, S. Goel, B. Singh, I. Bdikin, B. Kumar, Arab. J. Chem. **13**, 1874 (2020)
73. C. Shi, J. Ma, J. Wu, X. Wang, F. Miao, Y. Huang, K. Chen, W. Wu, B. Wu, J Alloys Compounds **846**, 156245 (2020)

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.

Advanced Production and Industrial Engineering
R.M. Singari and P.K. Kankar (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE220725

Investigating the Economic and Environmental Sustainability of Logistic Operations in India Using ARDL Procedure

Akanksha SINGH[1], Shweta KUMARI[2], Sanjay kumar PANDEY[3], Bharat SINGH[4], Nand KUMAR[5]

[1]*Research Scholar, Department of Humanities, Delhi Technological University,*
[2]*Research Scholar, Department of Humanities, Delhi Technological University,*
[3]*Associate Professor, Department of Economics, College of Commerce, Arts & Science, Pataliputra University,*
[4]*Professor, Satyawati College, University of Delhi,*
[5]*Professor & Head of Department of Humanities, Delhi Technological University,*

Abstract. The study aims to examine the effect of environment and economic variables on logistic performance in India. In order to study the long run and short run association between the variables the study employed auto regressive distribution lag (ARDL) approach on a time series data from 2007 to 2018. The result revealed that foreign direct investment (FDI) has a positive relation with LPI whereas fossil fuel consumption in both the short and long run has a negative relation. On the other hand, GDP per capita has a negative relation with LPI while total greenhouse gas emissions has a positive relation, which is a sign of concern for environment sustainability. In the recent report published by the World Bank India's rank has slipped down from 35th to 44thposition worldwide whereby all the six dimensions have shown a downward trend. India being one of the largest customer oriented market would negatively affect the world economy if its logistics operation are poorly driven. This study highlights few reason why India lacks behind in its logistics performance and provide suggestion how India can improve its logistic operation at global level.

Keywords: Logistic performance index, ARDL, total greenhouse gas emission, foreign direct investment, GDP per capital

1. Introduction

In the 21st century, logistics assumes an indispensable part in associating urban communities, nations, and landmasses all around the world [Khan et al., 2018]. As of late, logistic industry has pulled in much consideration in academic studies because of its focal points and inconveniences on nations' monetary development and ecological execution. Firms around the world have begun to implement green practices in their logistics tasks to accomplish social, economic and environmental sustainability, while in many countries green practices have a negative impact on the firm's profit. In an

[5]Corresponding author: Nand Kumar
E-mail: nandkumar@dce.ac.in

examination to discover the influence of green supply chain network on firms' financial and natural exhibition, it was establish that green experiments including an environmental plan of items, participation with clients, green buying of crude materials and segments, green transportation, and dispersion have a critical positive relationship with nature. It was also analyzed that green practices particularly green buying, adversely affects a company's productivity because of weighty charges and import obligations on material and segments [Khan et al 2017]. Undeniably, worldwide logistics tasks intensely rely upon energy, particularly petroleum derivative, bringing about natural crumbling and grave repercussion on general wellbeing. An investigation on European nations to examine the connection between logistics and energy request, the outcomes indicated that logistical tasks have a solid relationship with energy interest [Zaman et al., 2017]. Logistics industry plays a critical role in the monetary advancement of a nation yet nations do not invest the required time and energy on the other hand green practices in logistical tasks are the lone answer to bring about control in air contamination, environmental change and a worldwide temperature alteration issue[Li et al., 2018] and [Khan et al., 2017].The BRICS nations should develop a model for green practices and business through appropriation green practices in logistics activities, [Aldakhil et al., 2018].

1.1 Research motivation and objective

Logistic operation play's a critical role in the economic development of the country but its percentage contribution in the GDP is comparatively less, the government need to focus on measures which promotes logistic performance as countries like Germany, Sweden and Belgium are leading the list of logistic performance index whereas India far lack behind. The motive behind the study is to highlight the parameter, which are good for logistics operation so that logistic operation can move forward in a sustainable manner.

2. Review of Literature

[Khan et al., 2019; Zhang et al., 2019; Khan et al., 2017; Mensah et al., 2020] studied the relationship among GLP and social, environmental and economic factors and suggested that constructive policy that promotes the logistic industry without compromising the Environment sustainability should be encouraged. [Yadav et al., 2016] examined the intension of the young consumer towards buying a green product. The research carried out in India using Theory of Planned Behavior (TPB) to check environmental concern and environmental knowledge. [Chhabra et al., 2017;Pishvaee et al., 2012] used MDCM to analyze the alternatives for green practices in the assembly and packaging processes, to obtain best alternative solution for the logistic company particularly in India. They concluded the best packing alternative is carbon positive material whereas the best alternative for assembly operation is clinch joint.[Karaman et al., 2020] investigates the relation between GLP and sustainability drawing from the signaling theory and tested the link with the corporate governance. [Zhang et al., 2020; Oliverira et al., 2016] used LCA for a better understanding of remanufacturing and assessment of 3Es. [Kilkis et al., 2017] introduced the nearly zero exergy airport and used the law of thermodynamics to analyze the relationship between the 3 Es. [Pao et al., 2015] employed Lotkae Volterra model to examine the competitive interaction among the three Es and laid emphasis on EKC hypothesis. [Sarkodie et al., 2019;Wang

et al., 2019] revealed the existence of inverse U-shaped relationship between the increasing population's increasing demand for energy consumption, food production and economic development that has triggered climate change. [Menash et al., 2020, Kwak et al., 2020] employed partial least square structural equation modeling to check the relationship among green logistics management practices, social-environmental sustainability and business performance (BP). [Karaman et al., 2020] employed the signaling theory to analyze the relationship between GLP and sustainability reporting, it was discovered there is positive and relevant relationship between them. [Mensah et al., 2020] used Heterotrait- Monotrait ratio and Fornell-Larcker criteria to analyze the relationship among GLMP on social, environmental, market and financial performance and it was noted that environmental performance was positively impacted by GLMP whereas, social, market and financial performance has insignificant impact on EP. [Baah et al., 2019] employed partial least square structural equation modeling to analyze the effects of regulatory stakeholder pressure and organizational stakeholder pressure on financial performances and GLP. The results revealed GLP was influenced by organizational stakeholders and regulatory stakeholder's pressure. [Gupta et al., 2020] applied graph theory matrix approach to analyze the sustainability index of the logistics service provider and concluded that green procurement, renewable energy sources, recyclable packaging, carbon emission reduction, rainwater harvesting, use of CNG fleets, reduction in fuel consumption and resources optimization plays a major role in sustainable practices. [Zhang et al., 2020] conducted interviews based on the grounded theory method and concluded that the development level of logistics industry, the level of social supervision, the perfect of GLP system, the level of perception of logistics enterprises on GLP and the green governance capacity of government have a great influence on GLP.[Menash et al., 2020] employed partial least square structural equation model software to study the impact of green warehousing logistics optimization and social values and ethics on economic performances and supply chain sustainability. [Abdullahi et al., 2020] proposed a weighted sum model and an opsilion-constraint model that combines three dimensions of the 3Es. [Paul D.larson., 2021] employedHayes Process macro to analyze the indirect effects of LPI on economic and environment. [Richnak et al., 2021] used Pearson chi-square test to test the hypothesis. A total of 165 enterprises were surveyed and the result indicated that environmental policy were used by large enterprises in Slovakia. [Chawla et al., 2021] deployed fuzzy AHP to analyze green operations management practices and operation management. It laid emphasis that through GOM both nature and man-made world can coexist.[Stekelorum etal., 2020] used fuzzy set comparative analysis to survey 232 third party logistic provider to analyze the internal and external green supply chain management practices that influences TPLs operational and financial performances.[Nair et al., 2021] reviewed papers from 2000 to 2020 and collected data through survey and found that very few literature exist on social sustainability and also identified twenty six social issues in supply chain exist in two different economy USA and India.[Zhanget al., 2021;Guirong et al 2012; Alexandrova et al., 2021] used Grey relation analysis method GRA to predict urban logistic demand and concluded that subway can promote sustainability in urban logistic they highlighted the importance of government in order to achieve a circular economy. [Li et al., 2021] employed two stages least square and generalized method to analyze the environment and economic effects of green logistics and the results indicated that green logistics positively impact the economies of MENA, Asian and OBRI countries but promote environmental pollution. [Caggiani et al., 2021] proposed a two-echelon electric vehicle routing

problem (2E-EVRPTW-PR) model for last-mile urban deliveries where e-vans and e-cargo bikes can be used for customers' deliveries in the urban zone and second restricted traffic zone, respectively. [Adams et al., 2021] Concludes that firms cannot expertise in supply chain and sustainability at the same time and managers should have qualification and background of both. [Zheng et al., 2021] employed fuzzy AHP and analyzed that the growth of different logistic models has negative affect on agricultural product cold chain. [Managi et al., 2021] viewed that science, technology and innovation plays an important role in achieving the SDGs. [Naqvi et al., 2010] correlated renewable energy consumption, financial development and real income to ecological footprint by augmented panel algorithm method and concluded that financial development is positively related however per capita income for real income and high-income group have a negative outcome. [Majid, M. A., 2020] [Harjanne et al., 2019] objective was to present challenges, investment and employment opportunities in the field of renewable energy in India. [Zhang et al., 2019] examined the relationship between tourism, environment degradation and logistics in a time series data in Thailand using ARDL model.

3. Methodology

Logistic performance index was first calculated in 2007 by the world Bank based on the six key indicator i.e. speed of custom clearance process, quality of infrastructure related to trade and transport, ease of competitive price shipment, quality of logistic service, ability to track and trace consignments and expected delivery time of goods and services to destination. This study draws the relationship between India's logistic operations with economic (GDPPC, FDI) and environmental factor (TGHGE, Fossil). Undoubtedly logistic operation are very important for a country's economic growth but due to their high dependence on fossil fuel they enhances carbon emission and global warming, a sustainable policies in the field of logistics operation is needed to solve the various environmental and social problems[Khan et al.,2018]. In this study, we have used secondary data downloaded from the World Bank development indicators (WDI) for testing hypothesis and achieving research objective. Equation: 1 is formulated to check the relationship of overall logistic operation with environment and economic factors in the context of India during the period 2007-2018.

$$\ln LPI = f(\ln GDPPC, \ln FDI, \ln TGHGE, \ln FOSSIL) \quad \text{.....(1)}$$

In the above equation, ln is the natural log and the descriptions of the variables are in table 1.

$$\Delta \ln LPI_t = a_0 + \sum_{i=1}^{n_1} a_{1i} \Delta \ln GDPPC_{t-i} + \sum_{i=1}^{n2} a_{2i} \Delta FDI_{t-i} + \sum_{i=1}^{n3} a_{3i} \Delta \ln TGHGE_{t-i}$$
$$+ \sum_{i=1}^{n4} a_{4i} \Delta \ln FOSSIL_{t-i} + \beta_1 \ln GDPPC_{t-1} + \beta_2 \ln FDI_{t-1} + \beta_3 \ln TGHGE_{t-1} + \beta_4 \ln FOSSIL_{t-1} + \mu_t \quad \text{.....(2)}$$

In equation 2 the short run relationship is represented by α_1 to α_4 whereas the long run relationship is presented by β_1 to β_4 and the drift component isα_0.While n_i is the optimal lag and μ_t is the error term.

$$\Delta \ln LPI_t = \beta_0 + \sum_{i=1}^{n_1} \beta_{1i} \Delta \ln GDPPC_{t-i} + \sum_{i=1}^{n2} \beta_{2i} \Delta \ln FDI_{t-i} + \sum_{i=1}^{n3} \beta_{3i} \Delta \ln TGHGE_{t-i}$$
$$+ \sum_{i=1}^{n4} \beta_{4i} \Delta \ln FOSSIL_{t-i} + \theta ECM_{t-1} + \mu_t \quad \text{.....(3)}$$

Co- integration among the variables was tested and then the error correction model was developed whereas θ is the speed of adjustment of the long run equilibrium.

Table 1: List of variables

Variables	Description
LPI	Logistics performance index(1=low to 5=high)
GDPPC	GDP per capita, PPP (constant 2017 international $)
TGHGE	Total greenhouse gas emissions (kt of CO_2)
FDI	Foreign direct investment, net inflows (% of GDP)
FOSSIL	Fossil fuel energy consumption (% of total)

Source: [World Bank Database, 2020]

As some variables were stationary at level and some at first difference, we found ARDL (autoregressive distribution lag) technique to be most appropriate for our research. Table 1 indicates the variable description where GDPPC, FDI represents the economic factor affecting LPI on the other hand TGHGE and FOSSIL are the factors affecting environment sustainability.

4. Results and discussion

Table 2: Illustrates the result of unit root test. Augmented Dickey Fuller (ADF) test was employed to check the stationarity at level and 1st difference. The results proved the stationarity of the dependent variable LPI at level whereas the independent variables were stationary at 1st difference. ARDL model allows having a mixture of variables at level and first difference, therefore, ARDL model is most appropriate for our study. Before applying ARDL co-integration, ARDL bound test is used to check the long run association between the dependent and independent variables. The bound test shows two critical values the upper bound and the lower bound where upper bound assumes all variables are at first difference and lower bound assumes that all variables are at level. If the value of upper bound is less than F statistic the null hypothesis is rejected showing the presence of co-integration vice-versa.

Table 2: ADF Unit Root test

Variables	level t-statistics	level Prob.	1st diff t-statistics	1st diff Prob.
LPI	-3.937927	0.0169	-4.092602	0.0156
GDPPC	-2.736086	0.2449	-4.064610	0.0531
TGHT	-2.555023	0.1324	-4.644740	0.0061
FOSSIL	-1.768203	0.3745	-3.162729	0.0539
FDI	-1.891757	0.3234	-4.761373	0.0052

Table 3:shows the results of ARDL bound test, the results indicates that there existence a significant co-integration between the dependent and the independent variables as F statistics is much greater than the upper bound statistic.

Table 3: Auto Regressive Distributed Lag bond test for co-integration

Variable	F statistics	Co-integration	
F(LPI,GDPPC,FDI,GHGE,FOSSIL)	23.16019***		
Critical value	1%	5%	10%
Lower bound	3.29	2.56	2.2
Upper bound	4.37	3.49	3.09

*significant at 10% level, **significant at 5%, ***significant at 1%

Table 4: illustrates ARDL short run results where the findings indicates that total greenhouse gas emission (TGHGE) and FDI are significant and has positive correlation with LPI (0.07<0.1,0.25<0.05) whereas fossil fuel emission has a negative relation with the logistic performance and is significant at 5% (0.04<0.05). On the other hand GDP per capita is significant and have a negative relation with the LPI (0.02 <0.05) indicating that government policies are not aligned with the logistic performance index in India.

Table 4: Auto Regressive Distributed Lag short-run estimates

Variables	Probability	t-statistic	Coefficient
GDPPC	0.0255*	-6.144854	-1.180468
FDI	0.0421**	4.720288	0.259092
TGHGE	0.0713*	3.540824	3.180970
FOSSIL	0.0650*	-3.727804	-2.924298

Note: Adjusted $R^2 = 0.90$ *significant at 10%, **significant at 5%, ***significant at 1%

Table 5: demonstrates long run result of ARDL, findings suggest that the value of F statistic is much greater than upper bound hence there exist a long run relationship at 5% level of significance between the dependent and independent variable. Total greenhouse gas emission (TGHGE) is statistically significant and is positively correlated with the logistic performance index which is a sign of concern for human health. On the other hand FDI is significant and have a positive impact on LPI performance in the long run. On the other hand Gross Domestic Product per capita (GDPPC) has a negative relation with the LPI performance and is significant in the long -run. This is mainly because government policies are notallied with logistic operations (World Bank, 2018). Lastly, in long-run Fossil fuel consumption has a negative relation with the logistic performance index, which has a positive sign for environmental sustainability.

Table 5: Auto Regressive Distributed Lag long-run estimates

Variables	Probability	t-statistic	Coefficient
GDPPC	0.0678*	-3.640979	-0.694821
FDI	0.0602*	3.888108	0.117010
TGHGE	0.0462**	4.488684	1.241758
FOSSIL	0.1064*	-2.422927	-1.096131

Note: Adjusted $R^2 = 0.90$ *significant at 10%, **significant at 5%, ***significant at 1%

Table 6: Illustrates the results of diagnostics test, we have used Jarque Bera test and results shows that residuals are normally distributed. In addition heteroscedasticity

Breusch – Pagan Godfrey test has been deployed to check where the data are homoscedastic or heteroskedastic the results indicates that the dataset is homoscedastic. Further Ramsey RESET test was employed and the result shows that the model is free from specification error. Lastly, Breush Godfrey serial correlation LM test revealed that there exists no serial correlation among the error terms in our model

Table 6: Daignostic tests

Jarque-Bera Normality Test, JB stat	0.6024	0.7399
Ramsey RESET test,Log likelihood ratio	61.2801	0.6646
Breusch-Godfery serial Correlation ,LM test	4.9906	0.5295
Heteroskedasticity Breusch-Pagan-Godfery test, Obs R-squared	5.9676	0.6509

Table 7 shows the results of Granger causality test. The test is used in time series data to assess whether there exits any potential predictability power from one variable to the other. The result indicates that there exists a unidirectional causality between LPI and FDI, FOSSIL and GDPPC, FDI and TGHDE and FOSSIL and FDI.

Table 7: Granger causality test

Direction of causality	F-statistics	probability
LPI-FDI	5.8833	0.0415*
FDI-TGHGE	4.3049	0.071**
FOSSIL-GDPPC	5.1398	0.053**
FOSSIL-FDI	4.6485	0.063**

Note: **significance at 10%,*significance at 5%

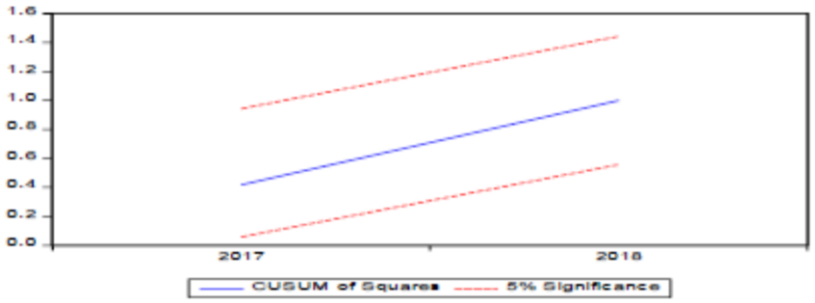


Figure 1: Cumulative sum of Square of recursive residuals

The model's stability was checked by cumulative sum of the recursive residuals stability test figure 1 suggest that model is stable. The findings reveal that foreign direct investment and total greenhouse emission are positively correlated with the logistic operation whereas fossil fuel consumption and GDP per capita have a negative relation.

5. Conclusion

The World Bank reports logistic performance index (LPI) at an interval of every two years. It is the weighted averages of six key dimensions i.e. speed of custom clearance process, quality of infrastructure related to trade and transport, ease of competitive price shipment, quality of logistic service, ability to track and trace consignments and expected delivery time of goods and services to destination. Logistic operation measures the relative ease and efficiency with which products can move inside a country, countries can compete globally by maintaining a good logistic as it would help to reduce the trade cost. The focus of this study is to find out is logistic performance index in India environmentally-economically sustainable during the period 2007-2018.To find the short run and long run relation among the dependent (LPI) and independent variables i.e. GDP per capita, FDI inflow (economic variable) and fossil fuel consumption, greenhouse gas emission (environmental variable) we employed ARDL technique. The findings confirmed that both in the short and long run foreign direct investments (FDI) are significant and are positively co-related with the LPI, which is a positive sign for the growth of an economy. Whereas Total greenhouse gas emission (TGHGE) is also significant and has a positive relation with the LPI, which is a sign of concern, as they are harmful for humanity and effect environment sustainability negatively, which is not the case in many other country like UK (Khan et al., 2017).GDP per capita being significant and having a negative relation highlights the ignorance from the government side. In the recent report India rank have slipped down from 35^{th} to 44^{th} position in the world LPI ranking were by all the six dimensions have fallen down (World Bank, 2018). This is mainly because of government policies are not primarily focus on logistic operations. With the ongoing expansion of the logistic industry government need to focus on efficient and effective policies in order to curb the pollution and waste generated by the logistic industries as countries like Germany, Sweden and Belgium are leading the list of logistic performance index whereby India lacks behind.

6. References

[1] Aldakhil, A. M., Nassani, A. A., Awan, U., Abro, M. M. Q., & Zaman, K. (2018). Determinants of green logistics in BRICS countries: An integrated supply chain model for green business. *Journal of Cleaner Production*, *195*, 861-868

[2] Abdullahi, H., Reyes-Rubiano, L., Ouelhadj, D., Faulin, J., & Juan, A. A. (2022). Modelling and multi-criteria analysis of the sustainability dimensions for the green vehicle routing problem. *European Journal of Operational Research*, *292*(1), 143-154.

[3] Baah, C., Jin, Z., & Tang, L. (2020). Organizational and regulatory stakeholder pressures friends or foes to green logistics practices and financial performance: investigating corporate reputation as a missing link. *Journal of cleaner production*, *247*, 119125.

[4] Chawla, V. K., Chhabra, D., Gupta, P., & Naaz, S. (2021). Evaluation of green operations management by fuzzy analytical hierarchy process. *Materials Today: Proceedings*, *38*, 274-279.

[5] Caggiani, L., Colovic, A., Prencipe, L. P., & Ottomanelli, M. (2021). A green logistics solution for last-mile deliveries considering e-vans and e-cargo bikes. *Transportation Research Procedia*, *52*, 75-82.

[6] Gupta, A., & Singh, R. K. (2020). Developing a framework for evaluating sustainability index for logistics service providers: graph theory matrix approach. *International Journal of Productivity and Performance Management*.

[7] Guirong, Z., Qing, G., Bo, W., & Dehua, L. (2012, October). Green logistics and Sustainable development. In *2012 International Conference on Information Management, Innovation Management and Industrial Engineering* (Vol. 1, pp. 131-133). IEEE.

[8] Harjanne, A., & Korhonen, J. M. (2019). Abandoning the concept of renewable energy. *Energy policy*, *127*, 330-340.
[9] https://lpi.worldbank.org/international/
[10] https://data.worldbank.org/
[11] Karaman, A. S., Kilic, M., & Uyar, A. (2020). Green logistics performance and sustainability reporting practices of the logistics sector: The moderating effect of corporate governance. *Journal of Cleaner Production*, *258*, 120718.
[12] Kwak, S. Y., Cho, W. S., Seok, G. A., & Yoo, S. G. (2020). Intention to Use Sustainable Green Logistics Platforms. *Sustainability*, *12*(8), 3502.
[13] Khan, S. A. R., Sharif, A., Golpîra, H., & Kumar, A. (2019). A green ideology in Asian emerging economies: From environmental policy and sustainable development. *Sustainable Development*, *27*(6), 1063-1075.
[14] Karaman, A. S., Kilic, M., & Uyar, A. (2020). Green logistics performance and sustainability reporting practices of the logistics sector: The moderating effect of corporate governance. *Journal of Cleaner Production*, *258*, 120718.
[15] Kwak, S. Y., Cho, W. S., Seok, G. A., & Yoo, S. G. (2020). Intention to Use Sustainable Green Logistics Platforms. *Sustainability*, *12*(8), 3502.
[16] Khan, S. A. R., & Qianli, D. (2017). Does national scale economic and environmental indicators spur logistics performance? Evidence from UK. *Environmental Science and Pollution Research*, *24*(34), 26692-26705
[17] Khan, S. A. R., Zhang, Y., Anees, M., Golpîra, H., Lahmar, A., & Qianli, D. (2018). Green supply chain management, economic growth and environment: A GMM based evidence. *Journal of Cleaner Production*, *185*, 588-599.
[18] Larson, P. D. (2021). Relationships between Logistics Performance and Aspects of Sustainability: A Cross-Country Analysis. *Sustainability*, *13*(2), 623.
[19] Li, X., Sohail, S., Majeed, M. T., & Ahmad, W. (2021). Green logistics, economic growth, and environmental quality: evidence from one belt and road initiative economies. *Environmental Science and Pollution Research*, 1-11.
[20] Managi, S., Lindner, R., & Stevens, C. C. (2021). Technology policy for the sustainable development goals: From the global to the local level.
[21] Majid, M. A. (2020). Renewable energy for sustainable development in India: current status, future prospects, challenges, employment, and investment opportunities. *Energy, Sustainability and Society*, *10*(1), 1-36.
[22] Naqvi, S. A. A., Shah, S. A. R., & Mehdi, M. A. (2020). Revealing empirical association among ecological footprints, renewable energy consumption, real income, and financial development: a global perspective. *Environmental Science and Pollution Research*, *27*(34), 42830-42849.
[23] Nair, P. U., & Thankamony, P. (2021). Social Issues in Supply Chain Sustainability–Focus Areas for Energy and Manufacturing sectors in India and USA. *European Journal of Sustainable Development*, *10*(1), 495-495.
[24] Oliveira, C., Coelho, D., & Antunes, C. H. (2016). Coupling input–output analysis with multiobjective linear programming models for the study of economy–energy–environment–social (E3S) trade-offs: a review. *Annals of Operations Research*, *247*(2), 471-502.
[25] Pishvaee, M. S., Torabi, S. A., & Razmi, J. (2012). Credibility-based fuzzy mathematical programming model for green logistics design under uncertainty. *Computers & Industrial Engineering*, *62*(2), 624-632.
[26] Pao, H. T., Chen, H. A., & Li, Y. Y. (2015). Competitive dynamics of energy, environment, and economy in the US. *Energy*, *89*, 449-460.
[27] Richnák, P., & Gubová, K. (2021). Green and Reverse Logistics in Conditions of Sustainable Development in Enterprises in Slovakia. *Sustainability*, *13*(2), 581.
[28] Stekelorum, R., Laguir, I., Gupta, S., & Kumar, S. (2021). Green supply chain management practices and third-party logistics providers' performances: A fuzzy-set approach. *International Journal of Production Economics*, *235*, 108093.
[29] Sarkodie, S. A., Strezov, V., Weldekidan, H., Asamoah, E. F., Owusu, P. A., & Doyi, I. N. Y. (2019). Environmental sustainability assessment using dynamic autoregressive-distributed lag simulations—nexus between greenhouse gas emissions, biomass energy, food and economic growth. *Science of the total environment*, *668*, 318-332.
[30] Wang, S., Song, J., Wang, X. E., & Yang, W. (2019). The spatial and temporal research on the coupling and coordinated relationship between social economy and energy environment in the Belt and Road Initiatives. *Sustainability*, *11*(2), 407.
[31] Yadav, R., & Pathak, G. S. (2016). Young consumers' intention towards buying green products in a developing nation: Extending the theory of planned behavior. *Journal of Cleaner Production*, *135*, 732-

739.
[32] Zhang, Q., Yang, M., & Zhou, X. (2021, April). Sustainable Development Analysis of Urban Green Logistics Based on Subway Transportation. In *2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 601-606). IEEE.
[33] Zheng, L. (2021). Construction of Green Logistics Evaluation System for Agricultural Products Cold Chain Based on Data Analysis. In *2020 International Conference on Data Processing Techniques and Applications for Cyber-Physical Systems* (pp. 431-437). Springer, Singapore.
[34] Zhang, Z. (2020). Energy and environment issues in carbon capture, utilization and storage. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 1-4.
[35] Zhang, Y., Khan, S. A. R., Kumar, A., Golpîra, H., & Sharif, A. (2019). Is tourism really affected by logistical operations and environmental degradation? An empirical study from the perspective of Thailand. *Journal of Cleaner Production*, *227*, 158-166.
[36] Zhang, W., Zhang, M., Zhang, W., Zhou, Q., & Zhang, X. (2020). What influences the effectiveness of green logistics policies? A grounded theory analysis. *Science of the Total Environment*, *714*, 136731.

Isolation Enhancement using CSRR Slot in the Ground for Compact Two-Element Textile MIMO Antenna

Navneet Sharma[1], Anubhav Kumar[1], Asok De[2], and R. K. Jain[1]

[1]Department of Electronics and Communication Engineering
Shobhit Institute of Engineering & Technology (Deemed to be University), Meerut (U.P.) 250110, India
navneet1979@gmail.com, rajput.anubhav@gmail.com, rakesh.jain@shobhituniversity.ac.in

[2]Delhi Technological University, Delhi 110042, India
asok.de@gmail.com

Abstract – A textile integrated, two-element, multiple-input multiple-output (MIMO) antenna is designed for Wi-Fi, wireless local area network (WLAN), and wearable biomedical applications. A CSRR slot and horn-shaped decoupling structure increases impedance matching and reduces the current movement from one element of antenna to another element which increases the isolation to 7.4 dB. The CSRR slot in the ground is accomplished between the elements to mitigate the surface current and enhance the isolation up to 27.4 dB. The 10-dB radiation exists from 4.65 to 5.97 GHz with more than 20-dB isolation in the full operating band where the maximum isolation is found to be 49.45 dB at 4.7 GHz. MIMO parameters are investigated in a two-element antenna, in which acceptable results are obtained. Bending analysis and phantom analysis are performed for wearable applications which demonstrate the acceptable results.

Index Terms – Defected ground structure (DGS), flexible antenna, isolation enhancement, MIMO antenna, textile antenna.

I. INTRODUCTION

There has been an extreme influence of body worn antennas for communication system recently. To satisfy the demands of communication in today's scenario, the antenna has to be robust, flexible, cost effective, compact, and reliable for wearable applications [1]. Flexible and textile antenna could be easily incorporated in the wearable and portable devices, but accomplishing the antenna in the compact space is a challenging task for a researcher. Among several technologies of antenna design, the use of multiple-input multiple-output (MIMO) antenna is a popular choice because in MIMO, same data is sent as multiple signals simultaneously through multiple antennas. This results in high data rate, diversity performance, high channel bandwidth, link reliability, high channel capacity, multiple transmission and reception capability, as well as spatial multiplexing [2, 3]. When multiple antennas are incorporated in small space, the distance between antennas is minimized due to which the mutual coupling increases and affects the MIMO parameters. High isolation between antenna elements could primarily be achieved in MIMO antennas by positioning the radiators at high distance ($> \lambda/2$) from each other so that mutual coupling is neutralized. This results in a large sized antenna that is not suitable for wearable applications. Other methods to escape mutual coupling is to place the radiator either perpendicularly [3] or at orthogonal positions [4, 5]. Extended ground structure [6] or a conductive line between the radiators [7] also considerably reduces the size of antenna by bringing the radiators closer. To further reduce the size, which is compatible to the requirements of wearable devices, several decoupling structures such as meander lines, slots, stubs, as well as metamaterial structures such as EBG and SRR are utilized. Meander line structures are easy to implement on rigid FR-4 substrate [8] and provide good isolation, but antennas have to compromise with flexibility. Meander lines are also implemented on flexible jeans substrate to enhance isolation [2], but this approach increases design and fabrication complexities. Alternatively, T-shaped [9–12] and E-shaped [13] stubs, vertical and horizontal slotted ground [14], funnel-shaped defected ground structure [15], neutralized line between the radiators along with defected partial ground [16], decoupling stubs [17], and combination of vertical and horizontal ground for enhanced isolation [14] are very common, easy to implement, and simpler approach on jeans substrate. Researchers had also utilized dielectric resonators along with annular rings [18] to curb mutual coupling. In [19], four-element CPW-fed MIMO antenna is proposed for UWB applications with CSRR notch,

Submitted On: January 28, 2022
Accepted On: June 14, 2022

1054-4887 © ACES

https://doi.org/10.13052/2022.ACES.J.370503

where distance between antenna and parasitic stub is used to improve the isolation. Researchers have also used metamaterial structures such as H-shaped EBG [20] for isolation enhancement. Orthogonalfeed with stubs [4], defected ground structures, and square patch with chamfered feed [21] are some of the hybrid structures deliberately utilized for improved isolation and reducing size of antenna.

An antenna with flexible textile substrate has the following advantages over their fixed counterparts.

- The antenna can be designed conformal, enabling it to fit to the design of the device. Flexible textile antennas are also shock resistant, especially applicable on fast moving devices or wearables such as shoes or wrist watch.
- Fabric antennas are environment friendly when disposed and are immune to high temperature deformations that may exist in antennas with fixed substrates.

There are several types of textile materials that are used as a substrate to design flexible antennas [22] where jeans material, popularly known as denim material, is an obvious choice for research on wearable antennas because of its rugged and long lasting quality [23], low cost, and time saving fabrication techniques [1]. The fabric is highly researched and the textile characteristics are rigorously studied [24] and dielectric constant is determined [25].

The proposed textile antenna could be used in wearable applications with the following novelty and technical contributions.

1. The proposed antenna design is modest as compared to other structures [2, 9, 16] and is integrated on most popular, rugged, flexible, easily available, and inexpensive jeans textile material.
2. The suggested decoupling structure in this work enhances the port to port isolation significantly up to 49.45 dB which is expressively higher than the other related works [2, 4, 13, 15, 17].
3. The antenna is analyzed on different curvatures and maintains its band for wireless local area network (WLAN), Wi-Fi, and 5.8 GHz (ISM) with minimal effect on impedance matching as well as isolation.
4. To analyze the on-body radiation effect of antenna, specific absorption rate (SAR) is obtained using four layered phantom model. The obtained SAR is 0.2680 W/Kg which is lower than other reported work [2, 13, 16].

The antenna is designed with required Wi-Fi, WLAN, and ISM band applications and fulfilled the compact and flexible requirement for wearable applications. The demonstration of human body effect on the antenna is discussed in the manuscript where antenna is designed with traditional equations, explained in eqn (1)–(3), and CSRR mitigates the current which is explained in Figure 1 (f). The low envelope correlation coefficient (ECC) and high isolation ensures that the antenna can be used for wireless communication applications.

II. MIMO ANTENNA DESIGN AND EVOLUTION STEPS

The optimized, compact two-element MIMO antenna is fabricated on lossy [24, 26] jeans material of 1 mm thickness (dielectric constant, $\varepsilon_r = 1.7$) [2, 6, 24**?**] which is illustrated in Figure 1 (a) with a prototype in hardware as shown in Figure 1 (b). The edge-to-edge dimensions of the antenna are 28 mm (Wb) × 24 mm (Wh). Primarily, we have to design a single element antenna and calculate the length (λ_g) and the width of the feed with help of eqn (1)–(3) [27]. In MIMO antenna design, two elements are incorporated with the conventional edge-to-edge distance.

$$\text{For } \frac{W}{h} > 1,$$

where $W =$ width of the patch, and $h =$ height of the substrate where dielectric constant (ε_r) for jeans (denim) material is 1.7.

The effective dielectric constant is given by

$$\varepsilon_{\text{reff}} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 12\frac{h}{W}\right)^{-\frac{1}{2}}, \tag{1}$$

$$\lambda_g = \frac{300}{f}\sqrt{e_{\text{eff}}}, \tag{2}$$

where $f =$ resonant frequency.

The characteristic impedance Z_0 can be given by

$$Z_0 = \frac{120\pi}{\sqrt{\varepsilon_{\text{reff}}}\left[\frac{w}{h} + 1.393 + 0.667\ln\left(\frac{W}{h} + 1.444\right)\right]}. \tag{3}$$

For MIMO antenna, the two elements are placed minimum at the distance of $\lambda/4$ for considerable isolation. However, this distance could be reduced by using isolation structures such as SRR and EBG. The calculated feed is W1 = W2 = 2.7 mm (width) and W3 = W4 = 13.6 mm (height) with $Z_0 = 52.36\ \Omega$.

Other dimensions of two-element MIMO antenna in mm are: W5 = 3.9, W6 = 23, W7 = 8.5, W8 = 8, W9 = 8.9, r1 = 6.7, r2 = 6.6, W10 = 1.5, and W11 = 1.5. The simulation of flexible and textile MIMO antenna is performed on Ansys HFSS-19 simulation software.

The two-element MIMO is evolved in four major iterations/steps, as is demonstrated in Figure 1 (c). In step-1, the ground height is 3.9 mm with 13.6-mm microstrip feed height where feed width of 2.7 mm is used to achieve approximately 50-Ω impedance matching. The $|S_{11}|$ in dB exists from 4.75 to 5.69 GHz, whereas the averaged isolation is at 10 dB with highest isolation of 13.66 dB as revealed in Figures 1 (d) and

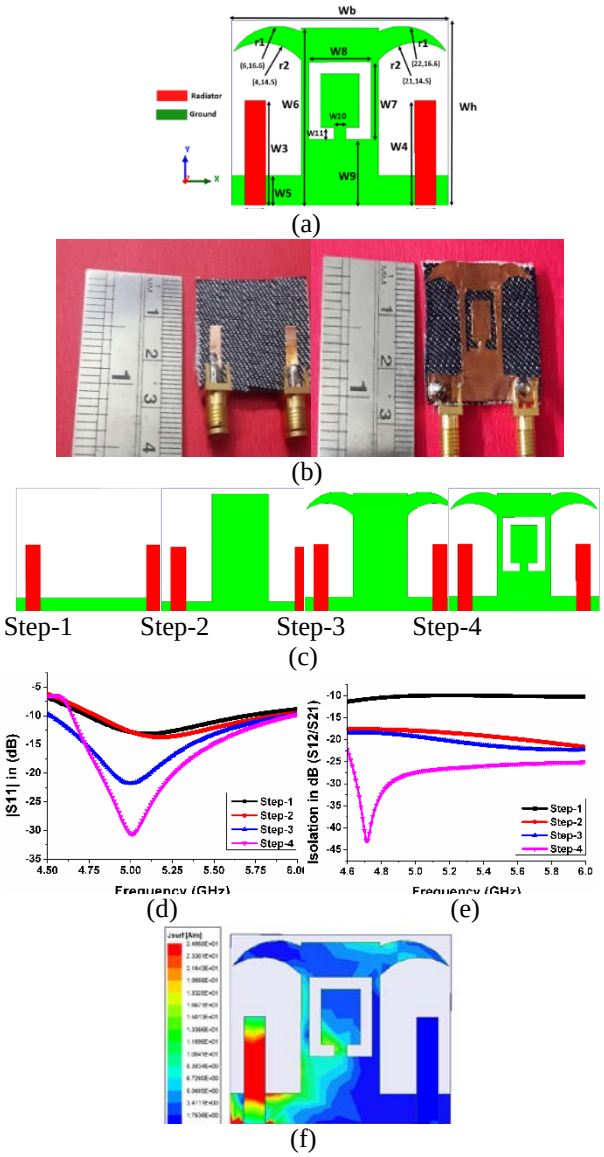


Fig. 1. (a) Proposed MIMO antenna (dimensions). (b) Prototype (fabricated). (c) Steps (evolution). (d) $|S_{11}|$ parameters (steps) in dB. (e) $|S_{12}|$ parameters (steps) in dB. (f) Surface current distribution at 5.8 GHz.

(e), respectively. In step-2, while increasing the ground vertically between two elements, isolation increases $|S_{21}|$ because this modification perturbs the current movement toward the modified ground structure from port one [6]. The $|S_{11}|$ in dB ranges from 4.80 to 5.87 GHz, whereas isolation is improved at this stage and ranges close to 17 dB, as is evident from Figures 1 (d) and (e). In step-3, a horn-shaped structure is implemented in the ground, which curates the impedance matching of antenna and increases the $|S_{11}|$ that varies from 4.53 to 5.96 GHz. The horn-shaped decoupling structure reduces the surface current between antenna elements and improves the isolation by 7 dB. In the last step, CSRR slot is accomplished in the decoupling structure, thereby mitigating the surface current and improving the isolation up to 27 dB. The $|S_{11}|$ characteristics are further improved, which varies from 4.65 to 5.97 GHz with enhancement in isolation. The isolation at this stage has reached a peak of 49.45 dB, which is an improvement of 123.09% as compared to step-3 (without CSRR). Surface current analysis of two-port antenna is analyzed at 5.8 GHz which clearly demonstrates that the CSRR effectively diminishes the surface current and improves the isolation as shown in Figure 1 (f).

III. BENDING ANALYSIS OF ANTENNA

To study the performance of two-port antenna with structural deformations, the antenna is deliberately bent on a polystyrene base cylinder with diameter of 25, 20, 15, and 10 cm using the wrapping tool in HFSS 19 [28]. The cylinder is then removed and simulation is performed. The screenshots are shown in Figure 2 (a), whereas Figure 2 (b) depicts the placement of designed antenna on polystyrene base. The $|S_{11}|$ and $|S_{12}|$ characteristics in dB (simulated) are shown in Figures 2 (c) and (d). Figure 2 (f) shows the variation in radiation efficiency, where small variations are observed. The size of antenna is reducing and impedance matching varies when bending process is applied on two-port antenna; therefore, $|S_{11}|$ and $|S_{12}|$ are shifted toward higher frequency, where small variation is found [36]. The minimum variation is found in gain of antenna in bending process is due to small variations in efficiency as depicted in Figure 2 (e). The bending analysis proves that the proposed antenna is working effectively, without compromising with the radiating bandwidth and isolation.

IV. PHANTOM ANALYSIS

The wearable antennas are used near the human tissues; therefore, certain parameters of radiation have to be studied for wearable applications. The effect of radiation on different layers of human skin is measured in terms of SAR.

Absorption of electromagnetic (EM) energy when exposed to radiation of antenna is calculated in SAR. In other words, it is the radiated power absorbed by human tissue of 1 Kg, averaged at 1 or 10 g, as illustrated in eqn (4)

$$\mathrm{SAR}_{\mathrm{average}}(r,\omega)=\frac{1}{V}\int\frac{\sigma(r,\omega)\,|E(r,\omega)|_2}{2\rho(r)}dr, \qquad (4)$$

where $|E(r,\omega)|$ is electrical field in volts/meter, V = volume in meter cube, r = position of vector, σ is sample density, and (r,ω) is the conductivity in Siemens/meter.

As per internationally approved standards, the value of SAR has to be under 1.6 W/Kg for safe use of antenna

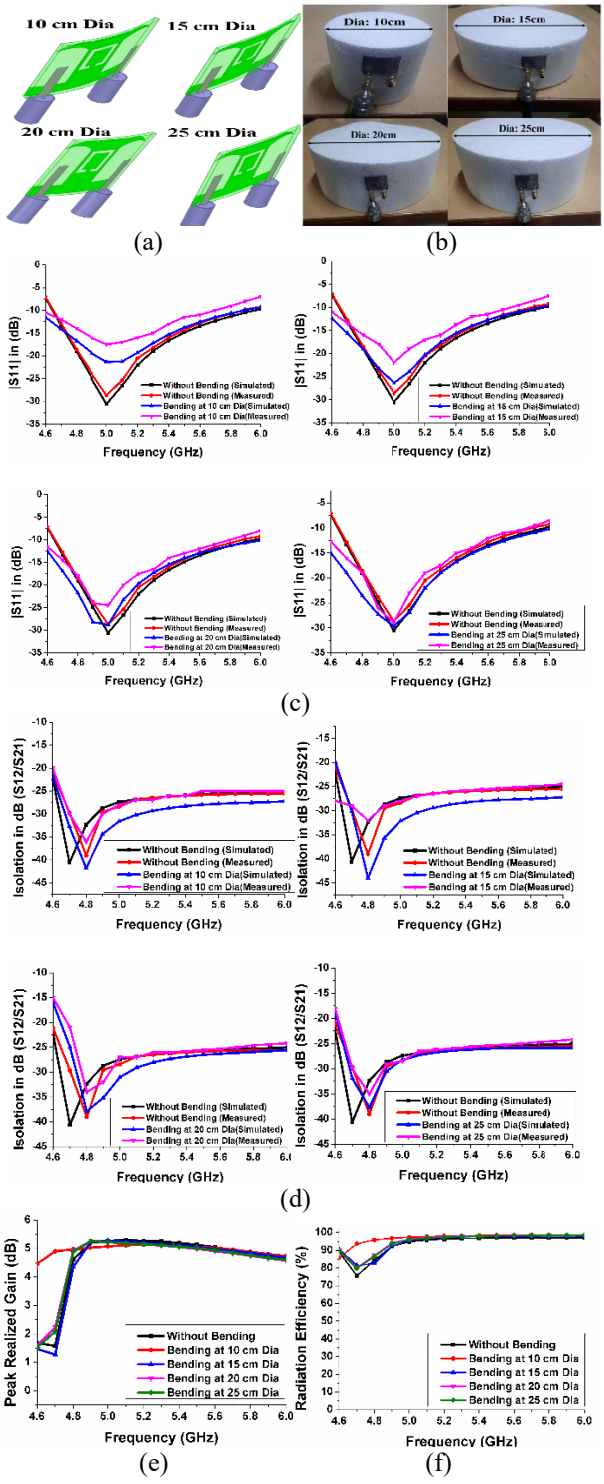


Fig. 2. Bending of antenna at 10, 15, 20, and 25 cm diameter (simulated). (b) Bending of antenna at 10, 15, 20, and 25 mm diameter (measured). (c) $|S_{11}|$ in dB for bending analysis (simulated and measured). (d) $|S_{12/21}|$ in dB for bending analysis (simulated and measured). (e) Peak realized gain variations on bending. (f) Radiation efficiency variations on bending.

Table 1: Characteristic values of human body tissues (5.8 GHz) [31]

Tissue	Conductivity (S/m)	Relative permittivity (ε_r)	Loss tangent (σ)
Skin	3.717	35.114	0.32807
Muscles	4.9616	48.485	0.31715
Fat	0.29313	4.9549	0.18335
Breast fat	0.41974	4.4976	0.28924
Bone	1.1544	9.6744	0.36981

near human body [29]. An analysis of the antenna positioned at 20-mm distance from a phantom model is performed. The phantom model (four layered) is composed of skin layer (2-mm thick), fat layer (3.5-mm thick), muscular layer (10-mm thick), and bone (10-mm thick) as depicted in Figure 3 (a). The tissue characteristics such as conductivity, loss tangent, and relative permittivity at 5.8-GHz frequency are given in Table 1. The $|S_{11}|$ and $|S_{12}|$ parameters of antenna in dB without phantom and with phantom are shown in Figures 3 (c) and (d). The minimum variations are observed in s-parameters when the four-layered phantom model is positioned at 20 mm from antenna, and this could be a feasible distance for wearable applications. The maximum SAR value was evaluated as 0.2680 W/Kg when averaged on 10 g of phantom model, when excited with a source power of 10 mW as is relevant from Figure 3 (b). The radiation efficiency and realized gain, in the absence and presence of phantom model, are depicted in Figures 3 (e) and (f), and a small variation in gain is found as a result of grating lobe reflection from the phantom model.

V. RESULTS AND DISCUSSIONS

The measurements were executed using VNA (Anritsu MS 2025B). The simulated and measured results of $|S_{11}|$ and $|S_{12}|$ in dB are depicted in Figures 4 (a) and (b), where $|S_{11}|$ in dB extends from 4.7 to 5.9 GHz with more than 20-dB isolation and minimum variation is found in the measured results. The anechoic chamber measurement for gain, where the amplitude of the transmitted signal (P_t) from MIMO is measured on the received signal (P_r) by horn antenna at a distance of 3 m. The gain is determined using formulae as given in eqn (5).

$$\frac{P_r}{P_t} = G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2, \tag{5}$$

where G_t and G_r are antenna gains of the MIMO antenna and horn antenna, respectively, and d is the distance between the two antennas and λ is the measured wavelength. The realized gain varies

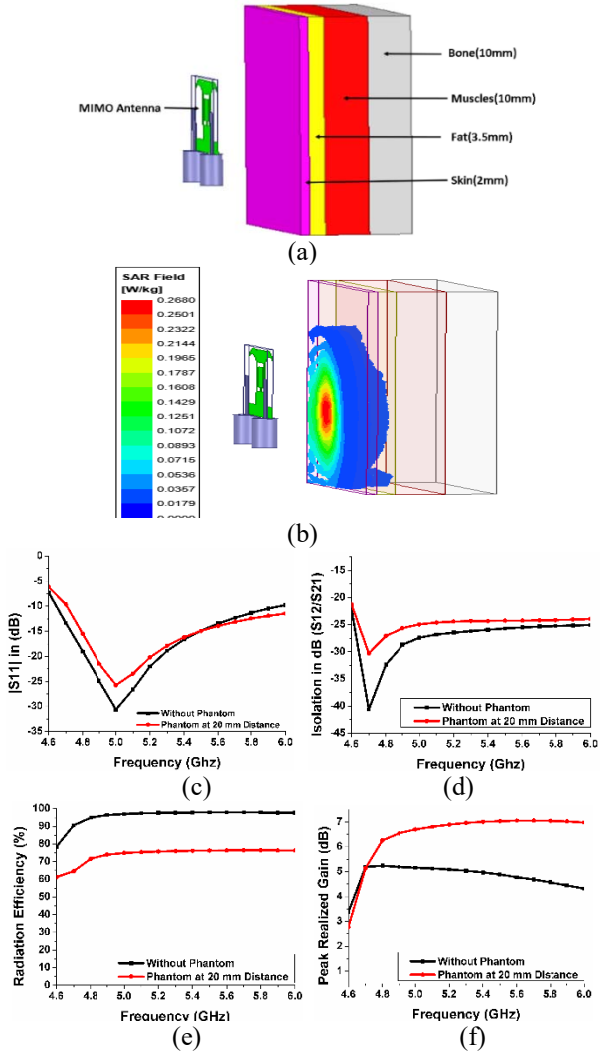


Fig. 3. Antenna positioned on phantom [30]. (b) Analysis of SAR (phantom). (c) $|S_{11}|$ in dB on phantom. (*d*) $|S_{12}|$ on phantom. (e) Radiation efficiency on phantom. (f) Peak realized gain on phantom.

from 3.37 to 5.23dB and less than 0.5 dB variation is observed with measured results due to environment losses associated with high frequency as depicted in Figure 4 (b). Radiation efficiency (simulated) is higher than 90% as is given in Figure 4 (c). The two-port antenna is measured on different body positions such as wrist, thigh, chest, and biceps for "on-body" wearable analysis as depicted in Figure 4 (f). The measured $|S_{11}|$ and $|S_{12}|$ in dB are depicted in Figures 4 (h) and (i). The results show that the antenna is performing satisfactorily when placed on any body part and also the antenna is able to maintain its isolation at less than 20 dB. Further, the antenna is placed on cap, shoes, bag, and riding helmet for portable wearable analysis as depicted in Figure 4 (g), where the antenna has minimum variations in *s*-parameters due to coupling effect, but the experimental results show acceptable performance in terms of $|S_{11}|$ and $|S_{12}|$ in dB as shown in Figures 4 (j) and (k). The measurement of radiation pattern is performed in anechoic chamber and the normalized patterns in *x*-*z* and *y*-*z* planes at 5.8 GHz frequency are shown in Figure 4 (e) and during measurement port 2 is terminated with 50-Ω terminator and port 1 is excited. The difference between co- and cross-plane is observed at more than 20 dB in end-fire direction and stable radiations are observed. Figure 4 (d) represents the simulated and measured VSWR.

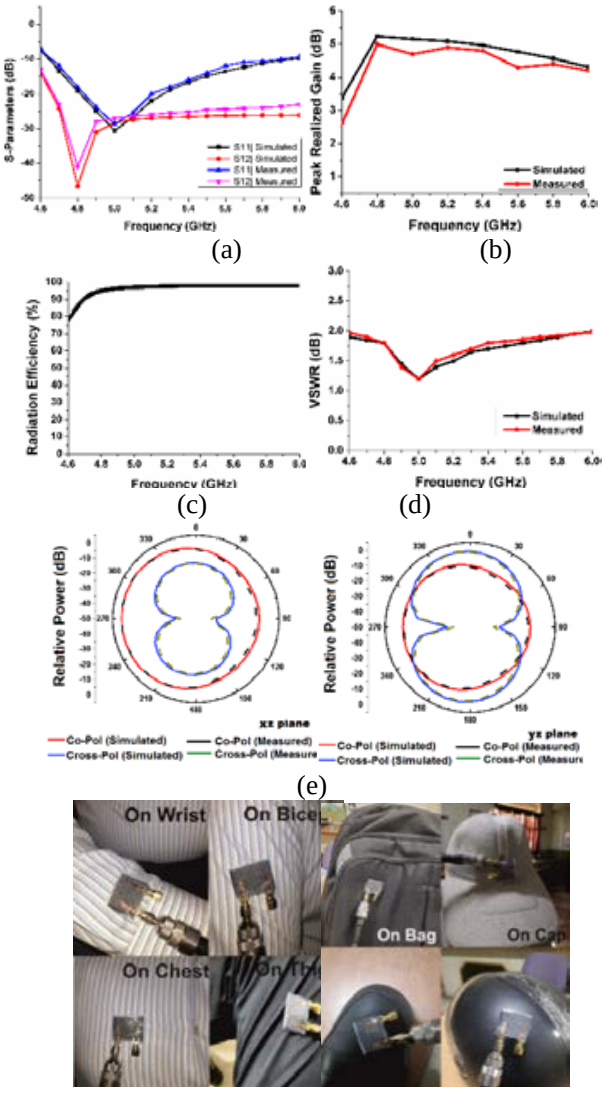
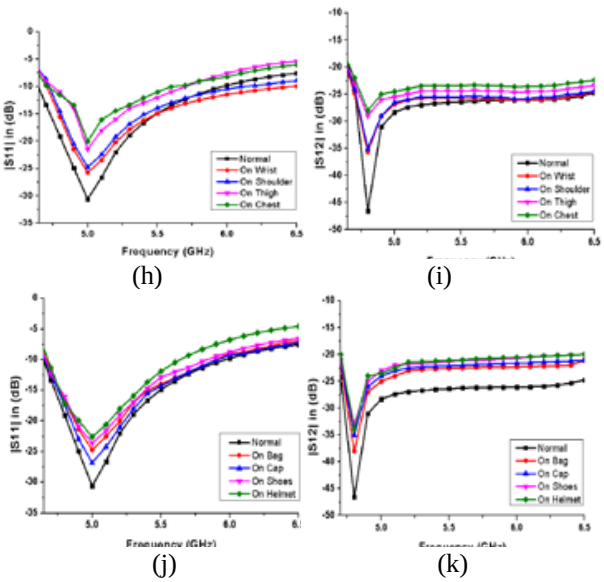


Fig. 4. *Continued.*



(l)

Fig. 4. Results (measured and simulated). (a) *S*-parameters (measured and simulated). (b) Realized gain (dB) (measured and simulated). (c) Radiation efficiency. (d) Simulated and measured VSWR. (e) Normalized radiation patterns in *yz* and *xz* planes at 5.8 GHz (measured and simulated). (f) Antenna placed on various body parts. (g) Antenna placed on different objects (h). $|S_{11}|$ characteristics in dB on different body parts. (i) $|S_{12}|$ characteristics on different body parts. (j) $|S_{11}|$ in dB on different objects. (k) $|S_{12}|$ in dB on different objects. (l) Measurement of antenna in anechoic chamber.

VI. ANALYSIS OF ECC, CCL, TARC, AND DG

A. Formulation

For evaluating MIMO performance in diversity environment, certain parameters such as channel capacity loss (CCL), diversity gain (DG), total active reflection coefficient (TARC), and ECC are taken into consideration. All the parameters are measured using the eqn (6)–(9) [4]

$$\mathrm{ECC} = \frac{\left|S_{11}^{*}S_{12}+S_{21}^{*}S_{22}\right|^{2}}{\left(1-\left|S_{11}\right|^{2}-\left|S_{21}\right|^{2}\right)\left(1-\left|S_{22}\right|^{2}-\left|S_{12}\right|^{2}\right)} \quad (6)$$

$$\mathrm{DG} = 10\sqrt{1-\left|\mathrm{ECC}\right|^{2}}. \quad (7)$$

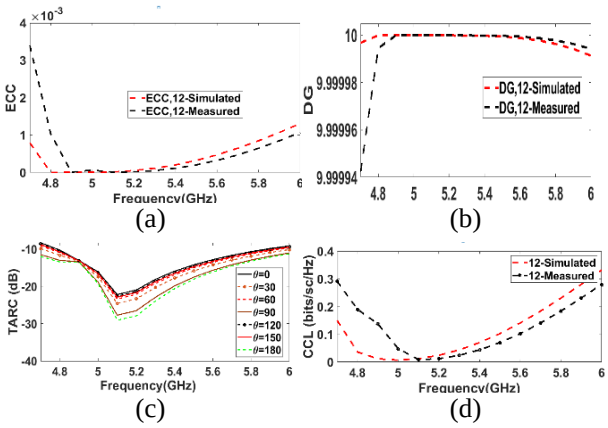


Fig. 5. MIMO parameters. (a) ECC. (b) Diversity gain. (c) TARC. (d) CCL.

$$\mathrm{TARC} = \sqrt{\left|S_{11}+S_{12}e^{j\theta}\right|^{2}+\left|S_{22}+S_{21}e^{j\theta}\right|^{2}}/\sqrt{2}, \quad (8)$$

where θ = phase angle between adjacent/diagonal ports.

$$\mathrm{CCL} = -\log_{2}\det\left(\psi^{\mathrm{R}}\right), \quad (9)$$

and ψ = correlation matrix at the receiving end.

B. Explanation

ECC is an important parameter that represents the correlation between the antenna elements. ECC ranges from 0 to 1 where 0 represents no correlation and antennas are completely decoupled, whereas 1 indicates identical pattern and high coupling between antennas. A truly uncorrelated channel is highly unlikely; therefore, ECC, which is close to 0, is desirable, thereby exhibiting high diversity performance and DG [19]. ECC and DG are calculated from *s*-parameters and are extracted from eqn (6) and (7). The measured and simulated DG and ECC are illustrated in Figures 5 (b) and (a), in which ECC is less than 0.005 and DG is near 10 dB.

Radiating efficiency and bandwidth of MIMO are not accurately characterized by scattering matrix; therefore, TARC is used [20]. Two element antenna scattering parameters are used to extract the TARC from eqn (8). TARC with θ = 0°, 30°, 60°, 90°, 120°, 150°, and 180° is given in Figure 5 (c), where minimum variations are observed.

CCL is a measure of throughput of antenna, where the CCL value should practically be less than 0.4 bit/s/Hz [11]. For the whole operating band, which represents high throughput. The CCL is measured through eqn (9). The measured and simulated CCL is depicted in Figure 5 (d) and is lower than 0.4 bits/s/Hz.

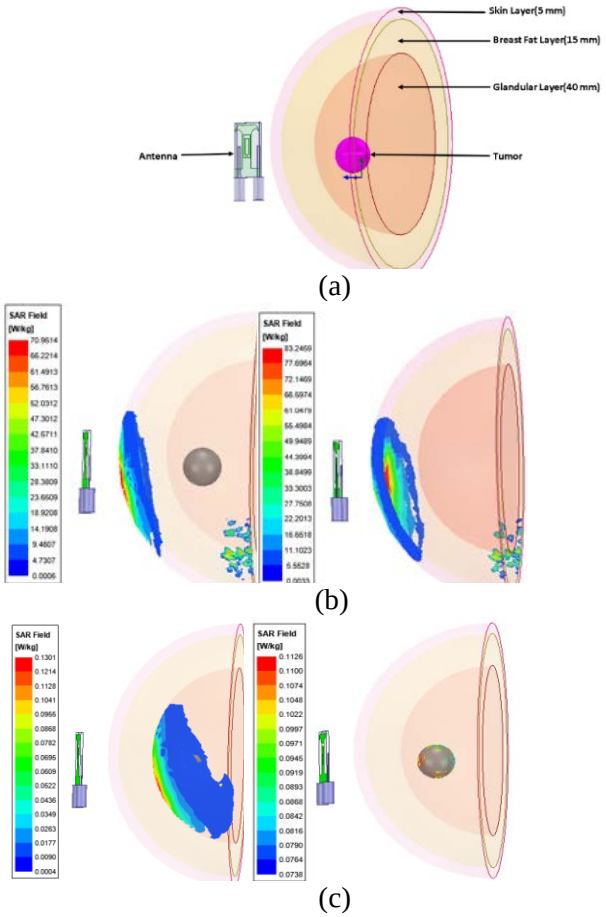


Fig. 6. (a) Breast model. (b) SAR of skin (without and with tumor. (c) SAR of glandular cell (with tumor) and SAR of tumor.

VII. TUMOR ANALYSIS ON FEMALE BREAST PHANTOM MODEL

Tumor in the female breast is very common in the soft glandular tissue, which is more vulnerable to developing malignant cells. In this paper, tumor analysis is investigated on female breast phantom model, with respect to SAR analysis.

The characteristics of tissue such as permittivity and permeability vary on occurrence of malignant (tumor) cell, thereby changing the SAR characteristics of the cell [32, 33]. Therefore, SAR analysis is the direct method for prediction of tumorous cells. A three-dimensional breast model (female) is illustrated in Figure 6 (a) for SAR analysis. The distance between breast model and antenna is 15 mm. The breast model with tumor size of 8 mm radius is composed of glandular tissue (r = 40 mm), breast fat (r = 55 mm), and outermost skin layer (r = 60 mm). The respective thickness of the layers is 40, 15, and 5 mm of glandular, breast fat, and skin layers [30, 32].

Table 2: SAR analysis (with and without tumor)

SAR (W/Kg) analysis at 5.8 GHz				
Tissue		With tumor	Without tumor	SAR diff.
Skin	Max	70.951	83.2459	12.2945
	Min	0.0006	0.0033	0.0027
Breast fat	Max	2.6118	2.6283	0.0165
	Min	0.0004	0.0003	0.0001
Glandular cell	Max	0.1301	0.1240	0.0061
	Min	0.0004	0.0003	0.0001
Tumor cell	Max	0.1126	–	–
	Min	0.0738	–	–

A. Detection of tumor

As the microwave field penetrates the tissue, the power is subsequently rejected through reflections and dissipation. Several tumor detection techniques are conventionally accounted [34, 35]. In the proposed method, tumor detection could be accomplished by using two different hypotheses. In the first hypothesis, the SAR value of the tumor and glandular tissue in the vicinity is analyzed for the detection of tumor cell; here, it is observed that the higher $SAR_{(avg)}$ value of the glandular tissue with tumor varies from 0.1301 to 0.0004 W/Kg, whereas SAR value of tumor tissue in glandular cell varies from 0.1126 to 0.0738 W/Kg, as shown in Figure 6 (c). The SAR difference denotes the presence of tumor, where the huge difference of minimum SAR value is detected in tumor cell and its neighboring glandular cell. In the second hypothesis, the difference in average SAR proves the presence of malignant (tumor) cell. It is observed that SAR values are decreased with the presence of tumor as given in Figure 6 (b). SAR difference (with and without tumor) is depicted in Table 2. In this investigation, the tumor analysis is observed effectively, which can be valuable for cancer detection and biomedical applications. In Table 3, the proposed two-port textile MIMO antenna is compared to the existing MIMO antenna, where proposed antenna is flexible and compact, based on textile material, with high isolation and have acceptable MIMO parameters, which can be applicable for WLAN and Wi-Fi communication in wearable applications.

VIII. CONCLUSION

A two-element MIMO antenna is designed and fabricated on denim textile as a substrate for 5G, WLAN frequency bands: Wi-Fi/WLAN (HIPERLAN, U-NII) (IEEE 802.11 a/h/j) working group as well as biomedical (ISM) band at 5.8 GHz. The isolation is achieved from defected ground decoupling structure such as horn-shape and CSRR. The horn-shaped ground disturbs the surface current and improves the isolation up to 5 dB, whereas

Table 3: Proposed MIMO antenna compared with existing MIMO antenna

Ref	Size (mm^2)	10 dB (GHz)	Isolation (dB)	Gain (dB)	Substrate	Flexibility	Isolation technology used	ECC
[2]	60 × 97	1.5–3.8, 4.1–6.1	15, 20	4	Jeans	Yes	Meander line structures	0.1
[4]	31 × 31	3.3–4.3	24	3.45	Jeans	Yes	Stubs and DGS, and orthogonal feed	0.2
[7]	140 × 70	2.33–2.5	26	4.7	Shieldit Textile	Yes	Conductive line	0.01
[9]	76 × 37	2.00–6.23	29.26	2.88	Felt	Yes	T-shaped stubs, ring-shaped slots, and meander line	0.01
[11]	80 × 170	3.4–3.6 and 3.9–4.5	20	6.8	FR4	No	T-shaped decoupling structure	0.3
[12]	18 × 44	3.3–3.65 and 4.8–5.5	18	5.93	FR4	No	Inverted T-shaped slot	0.002
[13]	55 × 35	2.4–9.0	18	9	Jeans	Yes	E-shaped stubs	0.06
[14]	71.7 × 50	3.24–3.05	19	4.7	FR4	No	Vertical and horizontal slotted ground	0.0035
[15]	35 × 30	3.11–5.15, 4.81–7.39	19, 21	NA	Jeans	Yes	Funnel-shaped DGS	0.18
[16]	30 × 50	3.14–9.73	32	2.7	Jeans	Yes	Neutralized ground	0.12
[17]	70 × 40	1.83–8	22	4	Jeans	Yes	I-shaped stubs	0.01
[21]	93 × 93	2.16–2.75	15	4.02	FR4	No	Square patch with circular chamfered corners	0.03
Proposed	24 × 28	4.65–5.97	20	4.57	Jeans	Yes	Horn-shaped defected coupling structure with CSRR	0.005

CSRR-shaped slot effectively diminishes the surface current and improves the isolation up to 27 dB. Overall, 20-dB isolation is achieved on the 10-dB impedance bandwidth of the proposed antenna that exists from 4.65 to 5.97 GHz; the antenna has simple decoupling structure and flexible substrate. The two-port MIMO antenna parameters are under the acceptable limit and antennas have minimum effect in *s*-parameter when analysis on bending was performed; this shows that the antenna is a suitable choice for wearable applications. Wi-Fi antennas are updated with MIMO technologies that allow group of devices to connect simultaneously to create efficient network and enhance speed and coverage. The MIMO antennas that are compact and integrated on fabrics could invariably be used for such applications in near future. The future technology holds easy-to-use and easy-to-fabricate antennas on other wearable textile substrates for 5G and 6G technologies that would be used in future. In future, more than two antenna elements would be designed for Internet of Things (IoT), sports, entertainment, and wearable technologies.

REFERENCES

[1] C. L. G. Monti and L. Tarricone, "Wearable antennas: Nontextile versus fully textile solutions," *IEEE Antennas and Propagation Magazine,* vol. 61, no. 2, pp. 71-83, 2019.

[2] R. Sourav, G. Soumendu, S. P. Soumya, and C. Ujjal, "Dual-polarized textile-based two/four element MIMO antenna with improved isolation for dual wideband application," *International Journal of RF and Microwave Computer-Aided Engineering,* vol. 30, no. 9, pp. e22292, 2020.

[3] A. A. Ibrahim, J. Machac, and R. M. Shubair, "UWB MIMO antenna for high speed wireless applications," *Applied Computational Electromagnetics Society (ACES) Journal,* pp. 1294-1299, 2019.

[4] K. A. Asok De and R. K. Jain, "Circular polarized two-element textile antenna with high isolation and polarization diversity for wearable applications," *International Journal of Microwave and Wireless Technologies,* pp. 1-9, 2022.

[5] D. Chengzhu, Y. Zhipeng, J. Gaoya, and Z. Shunshi, "Design of a co-planar waveguide-fed flexible ultra-wideband-multiple-input multiple-output antenna with dual band-notched characteristics for wireless body area network," *International Journal of RF and Microwave Computer-Aided Engineering,* vol. 32, no. 3, pp. e22997, 2022.

[6] J. Pankaj, K. Anubhav, D. Asok, and K. J. Rakesh, "Flexible and textile two-port compact antenna for WLAN and wearable applications," *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE*, pp. 308-311, 2021.

[7] A. Ismahayati, R. K. Muhammad, H. R. Ali, H. Norshakila, A. R. Hasliza, A. W. M. Wan Zuki, I. M. Arif, J. Muzammil, and M. Y. Mohd Najib, "Investigation on wearable antenna under different bending conditions for wireless body area network (WBAN) applications," *International Journal of Antennas and Propagation,* 2021.

[8] J. Deng, J. Li, L. Zhao, and L. Guo "A dual-band inverted-F MIMO antenna with enhanced isolation for WLAN applications," *IEEE Antennas and Wireless Propagation Letters,* vol. 16, pp. 2270-2273, 2017.

[9] K. K. Binod, K. Ashwani, S. Kunal, and K. Sachin, "Wideband textile multiple-input-multiple-output antenna for industrial, scientific and medical (ISM)/wearable applications," *International Journal of RF and Microwave Computer-Aided Engineering,* vol. 30, no. 12, e22451, 2020.

[10] K. A. Asok De and R. K. Jain, "Size miniaturization and isolation enhancement of two-element antenna for sub-6 GHz applications," *IETE Journal of Research,* pp. 1-8, 2021.

[11] J. Pankaj, K. Anubhav, D. Asok, and K. J. Rakesh, "Modified CSRR based dual-band four-element MIMO antenna for 5G smartphone communication," *Progress In Electromagnetics Research Letters,* vol. 101, 2021.

[12] A. Chatterjee, M. Midya, L. Mishra, and M. Mitra, "Dual-element Multiple-input-multiple-output system for sub-6 GHz (5G) and WLAN applications with enhanced isolation," *Progress In Electromagnetics Research M,* vol. 103, pp. 197-208, 2021.

[13] S. Rekha and G. Shine Let, "Design and SAR analysis of wearable UWB MIMO antenna with enhanced isolation using a parasitic structure," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering,* pp. 1-11, 2022.

[14] A. Kumar, N. K. Narayaswamy, H. Venkatesh Kumar, B. Mishra, S. A. Siddique, and A. K. Dwivedi, "High-isolated WiFi-2.4 GHz/LTE MIMO antenna for RF-energy harvesting applications," *AEU-International Journal of Electronics and Communications,* vol. 141, pp. 153964, 2021.

[15] B. Ashim Kumar and U. Chakraborty, "Reconfigurable wide band wearable multiple input multiple output antenna with hanging resonator," *Microwave and Optical Technology Letters,* vol. 62, no. 3, pp. 1352-1359, 2020.

[16] B. Ashim Kumar and U. Chakraborty, "Investigation on decoupling of wide band wearable multiple-input multiple-output antenna elements using microstrip neutralization line," *International Journal of RF and Microwave Computer-Aided Engineering,* vol. 29, no. 7, pp. e21723, 2019.

[17] B. Ashim Kumar and U. Chakraborty, "A compact wide band textile MIMO antenna with very low mutual coupling for wearable applications," *International Journal of RF and Microwave Computer-Aided Engineering,* vol. 29, no. 8, pp. e21769, 2019.

[18] D. Gourab, A. Sharma, and R. K. Gangwar, "Dielectric resonator-based two-element MIMO antenna system with dual band characteristics," *IET Microwaves, Antennas & Propagation,* vol. 12, no. 5, pp. 734-741, 2018.

[19] K. Anubhav, "Compact 4x4 CPW-Fed MIMO antenna with Wi-Fi and WLAN notch for UWB applications," *Radioelectronics and Communications Systems,* vol. 64, no. 2, pp. 92-98, 2021.

[20] K. A. Asok De and R. K. Jain, "Novel H-shaped EBG in E-plane for isolation enhancement of compact CPW-fed Two-port UWB MIMO antenna," *IETE Journal of Research,* pp. 1-7, 2020.

[21] A. D. Chaudhari and K. P. Ray, "A compact five-element printed MIMO antenna with pattern and polarization diversity for 2.45 GHz WLAN applications," *Journal of Electromagnetic Waves and Applications,* pp. 1-13, 2022.

[22] M. Sarmad Nozad, J. I. Asnor, S. Tale, A. Hussein, A. Sameer, I. Alyani, and C. S. Azura, "Recent advances in wearable antenna technologies: a review," *Progress in Electromagnetics Research B*, vol. 89, pp. 1-27, 2020.

[23] K. N. Paracha, S. K. Abdul Rahim, P. J. Soh, and M. Khalily "Wearable antennas: A review of materials, structures, and innovative features for autonomous communication and sensing," *IEEE Access,* vol. 7, pp. 56694-56712, 2019.

[24] A. Mohamed Ismail, M. F. Ahmed, and A. H. A. Shaalan, "Novel electro-textile patch antenna on jeans substrate for wearable applications," *Progress In Electromagnetics Research C,* vol. 83, pp. 255-265, 2018.

[25] S. Sankaralingam and B. Gupta, "Determination of dielectric constant of fabric materials and their use as substrates for design and development of antennas for wearable applications," *IEEE Transactions on Instrumentation and Measurement,* vol. 59, no. 12, pp. 3122-3130, 2010.

[26] L. Seung Yoon, C. Moogoong, J. Sohyeon, and H. Wonbi "Optically transparent nano-patterned antennas: A review and future directions," *Applied Sciences* vol. 8, no. 6, pp. 901, 2018.

[27] C. A. Balanis, *Antenna Theory Analysis and Design.* John Wiley & Sons, 2015.

[28] B. Prudhvi Nadh, B. T. P. Madhav, M. Siva Kumar, T. Anil Kumar, M. Venkateswara Rao, and S. S. Mohan Reddy, "MEMS-based reconfigurable and flexible antenna for body-centric wearable applications," *Journal of Electromagnetic Waves and Applications,* pp. 1-15, 2022.

[29] A. Fatemeh, M. M. Paulides, and G. C. Van Rhoon, "SAR thresholds for electromagnetic exposure using functional thermal dose limits," *International Journal of Hyperthermia,* vol. 34, no. 8, pp. 1248-1254, 2018.

[30] S. Navneet, K. Anubhav, A. De, and R. K. Jain, "Design of compact hexagonal shaped multiband antenna for wearable and tumor detection applications," *Progress In Electromagnetics Research M,* vol. 105, pp. 205-217, 2021.

[31] Inst. of Appl. Phys., Italian Nat. Res. Council, "Calculation of the dielectric properties of body tissues in the frequency range 10 Hz-100 GHz," Florence, Italy. [Online]. Available: http://niremf.ifac.cnr.it/tissprop.

[32] S. Navneet, K. Anubhav, A. De, and R. K. Jain, "Compact circular polarized CPW antenna for WLAN and biomedical applications," *Frequenz*, vol. 76, 2021.

[33] S. Subramanian, B. Sundarambal, and D. Nirmal, "Investigation on simulation-based specific absorption rate in ultra-wideband antenna for breast cancer detection," *IEEE Sensors Journal,* vol. 18, no. 24, pp. 10002-10009, 2018.

[34] N. K. Nikolova, "Microwave imaging for breast cancer," *IEEE Microwave Magazine,* vol. 12, no. 7, pp. 78-94, 2011.

[35] B. Ria, A. A. Thathamkulam, and P. Mythili, "An overview of microwave imaging for breast tumor detection," *Progress In Electromagnetics Research B* , vol. 87, pp. 61-91, 2020.

[36] M. U. A. Khan, R. Raad, F. Tubbal, and P. I. Theoharis, "The impact of bending on radiation characteristics of polymer-based flexible antennas for general IoT applications," *Applied Sciences,* vol. 11, no. 19, pp. 9044, 2021.

Navneet Sharma is having an excellent academic record. He received the bachelor's degree in engineering from the Prestigious Delhi College of Engineering (presently, Delhi Technological University), New Delhi, India, in the field of electronics and communication engineering. He received the master's degree in technology from Dr. Abdul Kalam Technical University, Lucknow, India. Currently, he is working toward the Ph.D. degree from the Shobhit Institute of Engineering and Technology (deemed to be university) Meerut, Uttar Pradesh, India, in the field of microwave engineering, where his specialization is wearable and biomedical antennas. He has published his research in SCI/ESCI as well as Scopus indexed journals and also presented papers in scores of national and international conferences. He has 14 years of teaching experience.

Anubhav Kumar received B.Tech. and M.Tech. degrees from Uttar Pradesh Technical University (now Dr. APJ AKTU), Lucknow, India, in electronics and communication engineering (ECE). He has more than 10 years of teaching experience. He is currently working toward the Ph.D. from the Shobhit Institute of Engineering and Technology (deemed to be university), Meerut, Uttar Pradesh, India. He has published many SCIE/ESCI and Scopus indexed research papers in International Journals. His research interests include microstrip antenna, metamaterial, FSS, MIMO, wearable, EBG antenna and Image processing.

Asok De received the B.Tech. and M.Tech. degrees from Jadavpur University, Kolkata, India, and the Ph.D. degree from the Indian Institute of Technology, Kharagpur, India. He served as Faculty with the University of Delhi, University of Kolkata, India, from 1984 to 1997. He joined as a Professor of Electronics and Communication Engineering with the Delhi College of Engineering (at present, Delhi Technological University) in 1997. He was the Founder Principal of Ambedkar Institute of Advanced Communication Technology and Research (2005–2012). He served National Institute of Technology Patna as Director from 2012 to 2017. He also served the National Institute of Technology Durgapur as Director (Additional Charge) from 2015 to 2017.

Prof. De has published more than 200 research papers in international journals and international conferences. He supervised 16 Ph.D. scholars. At present, he is an Emeritus Professor with Delhi Technological University.

R. K. Jain received the Ph.D. degree from Banaras Hindu University, Varanasi, India, in 1990. He is a Professor and Associate Dean with the School of Engineering & Technology, Shobhit Institute of Engineering and Technology (deemed to be university), Meerut, India, where he has been teaching and doing research with the collaboration of BHU, Varanasi and BARC, Mumbai. He has also received Postdoctoral fellowship with the Department of Physics, North Carolina State University, Raleigh, NC, USA from May 2002 to June 2003 and worked on low-intensity and low-energy proton beams for the development of low-energy proton detectors (needed to detect protons emitted in neutron beta decay). He has nearly 23 years of teaching experience and 30 years of research after Ph.D.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1746-8809.htm

M&A performance in emerging markets: do they behave in unison or otherwise?

Sakshi Kukreja, Girish C. Maheshwari and Archana Singh
Delhi School of Management, Delhi Technological University, Delhi, India

Received 7 December 2021
Revised 4 June 2022
25 August 2022
6 October 2022
Accepted 17 October 2022

Abstract

Purpose – The study aims to evaluate and compare the mergers and acquisitions (M&As) performance utilising a sample of deals originating from Brazil, Russia, India, China and South Africa (BRICS). In addition to nation-wise performance analysis, a further sub-sample analysis is conducted based on the target location (domestic and cross-border), development status (developed and emerging) and the acquired ownership stakes (majority and minority).

Design/methodology/approach – The final sample of the study includes 7,105 deals announced between 2000 and 2019. M&A performance is proxied by the abnormal returns earned over the select event windows. Multiple parametric and non-parametric tests are employed for testing the robustness.

Findings – The results indicate significant performance differences across BRICS markets, with the highest and lowest abnormal returns reported for Chinese and Russian acquirers, respectively. The disaggregated analysis also affirms the performance differences for the select sub-samples.

Research limitations/implications – The study highlights the need for acknowledging and expounding the differences in M&As across emerging markets. Further, the results of the study provide a possible explanation of the disagreement over the M&A performance results reported in the previous literature.

Practical implications – Acknowledging and understanding the potential performance differences based on location, ownership strategies and development status can aid executives in sharpening decision-making and also help general investors.

Originality/value – The study contributes by examining a comprehensive sample of deals across five major emerging economies, as against the majority of previous studies which have their results based on either single nation samples or have utilised only a sub-sample of domestic or foreign acquisitions.

Keywords BRICS, Emerging markets, Domestic vs cross-border, Target location, Ownership, Performance

Paper type Research paper

1. Introduction

Globalisation and liberalisation combined with the rapid technological advancements have made the firms ever more vulnerable to market competitive forces. As a mode of inorganic growth, mergers and acquisitions (M&As) are used as a preferred strategy for achieving competitive advantage, enhancing accessibility and strategic flexibility. Given the scope and importance of the M&As, this area has received considerable attention in academic work. Yet, unanimity does not exist in the literature about wealth creation for acquirer shareholders, popularly taken as a measure of deal performance. Previous empirical evidences have reported that expected synergies often fail to materialise, with negligible or rather negative average abnormal acquirer returns around the deal announcements (Weber *et al.*, 2011). Yet, the popularity of M&As as a mode of corporate expansion continues to flourish (Alkaraan, 2019), presenting a paradox.

Traditionally dominated by the developed markets, the contemporary figures validate the increasingly prominent role of emerging market firms (EMF) in the global economy (Caiazza and Volpe, 2015; Reddy *et al.*, 2022). Cross-border M&As constitute the majority of the foreign direct investments from emerging market nations. In the previous decade, EMFs have proved

Conflict of interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

International Journal of Emerging Markets
© Emerald Publishing Limited
1746-8809
DOI 10.1108/IJOEM-12-2021-1827

to be aggressive acquirers, not hesitating to enter the lands of western developed countries (Haasis *et al.*, 2018; Zheng, 2016). Brazil, Russia, India, China and South Africa (BRICS) represent a leading group of emerging markets and have contributed towards the shift in global economic power away from developed economies. Together, BRICS cover more than 40% of the world's population, contribute to about a quarter of global gross domestic product and constitutes around 40% of the world's total foreign reserves (Phiri, 2018). BRICS represent a heterogeneous group of countries, spread across four continents, with different dominant firm features. Concentrated ownership structure with family ownership being a common feature of Indian firms tends to be less affected by agency concerns. Indian economy has witnessed a high propensity for M&A deals in the past two decades. For Chinese firms, State-ownership remains a standard feature. China's 1999 "go-global" policy has caused a surge in the number of deals initiated by Chinese firms, acquiring resources worldwide. China and India remain the leading countries among BRICS in terms of deal count as well as the value of announced deals. A comprehensive emerging market study of M&As performance is warranted for resolving the performance paradox.

The study contributes to the existing knowledge base in several ways. Firstly, with the shift in the trend, the world is now witnessing more and more deals originating from emerging markets, in tandem with their increasing role in the global economy (Dikova *et al.*, 2019; Zhang *et al.*, 2020). Given the differences in the motivations, success factors, institutional differences, corporate governance practices and strategic flexibility of the EMFs against their developed market counterparts, it becomes necessary to investigate the points of divergence in their behaviour. Furthermore, much of the extant research has focussed on developed western markets. Our knowledge of M&A performance in emerging markets lags compared to the extensively studied developed markets. The recent surge in the number of big-ticket deals involving EMFs acquiring developed market targets (like Tata-Corus, Mittal-Arcelor and Anbang Insurance Group Co Ltd- Strategic Hotels & Resorts Inc.) have caught significant media and scholarly attention. Of late, the focus of studies has widened to include emerging markets, but the results remain far from being conclusive. Further, majority of the previous emerging market studies have utilised small samples, focussing on only on few select countries (like India and China) and deal type (focussing either domestic or cross-border) (Figueira *et al.*, 2020; Jain *et al.*, 2021). Empirical evidence on countries like Brazil, South Africa and Russia remain scanty (Ermolaeva, 2019; Ferreira *et al.*, 2019; Opoku-mensah *et al.*, 2019). Few studies have also compared the M&A performance for emerging and developed countries, but comparative research among emerging markets is lacking. It is important to understand the differences before possible generalisations among emerging markets. Prior research have often reported different performance results for deals involving domestic target as against cross-border deals (Adel and Alkaraan, 2019; Du and Kwabi, 2021; Kohli and Mann, 2012). Hence, a further investigation is warranted to gauge the performance differences based on target location. Further for cross-border deals, the host country development status has a potential bearing on the deal performance by influencing the level of acquirer resource requirements, transactional costs and legitimacy issues (Barnard, 2021; Chan *et al.*, 2008; Usman *et al.*, 2021; Zhang, 2022). Finally, the level of acquired target ownership is often guided by diverse underlying deal motivations and rather represent as distinct strategic choices with their peculiar implications (Contractor *et al.*, 2014; Kim *et al.*, 2020). It would be reasonable to investigate and compare the deals resulting in majority ownership as against minority ownership distinctively. To address these gaps, the study seeks to answer two principal research questions:

RQ1. To what extent do emerging markets react differently to M&A announcements?

RQ2. Do performance differences exist for cross-border vs domestic deals, target country development status and the acquired ownership structure? Answering these research questions, the present study provides emerging market evidence on M&A

performance utilising a comprehensive data set of deals, both domestic and cross-border, involving BRICS market acquirers.

The paper is organised as follows. Section two reviews the literature relevant to the theme of the paper highlighting the research gap and formulating the hypotheses. Section three describes the methodology, sample of the study and data sources. Section four presents the discussion of the results, followed by conclusion and future direction of research.

2. Literature review

2.1 M&A performance: the case of emerging markets

Since 2000, emerging market have witnessed a surge in the M&A deals, pushing it to unprecedented levels (UNCTAD, 2015). Emerging market multinational enterprises theoretically differ from their traditional counterparts from developed market (Guillen and Garcia-Canal, 2009; Luo and Tung, 2018). Internationalisation of EMFs is guided by asset-seeking motive against asset-exploitation motive for developed markets firms (Buckley *et al.*, 2016a; Liang *et al.*, 2021; Yang *et al.*, 2014). EMFs "springboard" into mature markets to overcome their latecomer disadvantage and home institutional constraints (Luo and Tung, 2007, 2018). Whilst entering into cross-border deals, EMFs tend to pay higher premiums, especially whilst acquiring developed market targets, on account of informational disadvantage (Zhu *et al.*, 2014) and national pride (Hope *et al.*, 2011).

Although the impact of M&As on shareholder wealth has received considerable academic interest in developed market contexts, less attention has been paid to the emerging markets, and its knowledge remains far from being either conclusive or comprehensive (Ning *et al.*, 2014). The existing emerging market studies have presented mixed, and often contradictory evidence, ranging from positive (Bhagat *et al.*, 2011; Ding *et al.*, 2021; Nicholson and Salaber, 2013; Ning *et al.*, 2014) to negative (Aybar and Ficici, 2009; Bertrand and Betschinger, 2012) average abnormal returns around the deal announcement. Sehgal *et al.* (2012) reports different post-announcement acquirer returns for emerging markets, with strong positive returns for South Africa and significantly negative returns for India, China and South Korea. The majority of the emerging market studies have investigated the acquirer returns for cross-border deals and have remained concentrated on a few particular countries like India and China. There is a need to comprehensively examine M&A as a strategic decision, including domestic and cross-border, for a broader group of countries to understand inter-country variations in deal returns before any possible performance generalisations for emerging markets. Hence, the primary research question, if the emerging markets react differently to M&A announcements forms our first hypothesis and is investigated to expound behaviour for the different groups of deals.

H1. There exist no significant M&A performance differences across emerging markets.

2.2 Domestic vs cross-border

Cross-border offers geographic diversification, allowing firms to expand across national boundaries, access foreign resources and overcome home institutional limitations. Cross-border deals offer the firm a spectrum of opportunities over domestic targets, albeit not without its challenges (Chidambaran *et al.*, 2017). For example, home-host geographic or cultural distance often add to integrational difficulties (Dutta *et al.*, 2013). Also, whilst entering foreign lands, certain challenges like information asymmetry and liability of foreignness are rather magnified (Kukreja *et al.*, 2021). Overcoming these challenges to realise expected synergies and returns demands experience and expertise. There exists a debate on the comparative performance of domestic vs cross-border M&A deals as the previous studies have reported different and often contradictory results. For example, in a prior study utilising

a sample of 268 Indian acquisitions announced between January 1, 1997 and March 31, 2008, cross-border deals realised significantly higher wealth gains as against the domestic deals (Kohli and Mann, 2012). Similar results were also reported for UK (Danbolt and Maciver, 2012; Gregory and O'Donohoe, 2014), Canada (Dutta *et al.*, 2013), India (Rani *et al.*, 2014). In contrast, there also have been studies reporting better performance of domestic M&A deals relative to cross-border deals in case of Europe (Mateev, 2016), USA (Moeller and Schlingemann, 2005), China (Black *et al.*, 2015; Du and Kwabi, 2021; Kling and Weitzel, 2011) and UK (Adel and Alkaraan, 2019; Conn *et al.*, 2005)]. Bertrand and Betschinger (2012) report that Russian acquisitions, both domestic and foreign, result in reduced acquirer performance against non-acquiring firms owing to low M&A capabilities and experience. It is propounded that the idiosyncratic country-level factors may explain the difference in results, impacting the deal returns, especially in foreign located targets. Hence, there exists a need for disaggregated country-wise investigation for each of the above deal types. Taken together, we hypothesise:

H2. There exist no significant performance differences between domestic and cross-border M&A deals.

2.3 Target location: emerging vs developed market

Whilst studies comparing M&A performance based on acquirer country development status have been common, but comparative evidence based on target location has been scanty. Recent studies have highlighted the importance of locational choices in cross-border deals (Lee *et al.*, 2014; Papageorgiadis *et al.*, 2020; Saikia, 2020; Tunyi and Ntim, 2016). Target country's development status plays a crucial role in shaping the firms' internationalisation decisions, influencing the offered opportunities and challenges and also the resource requirements (Barnard, 2021). The risks associated with emerging markets tend to be significantly higher owing to the weaker corporate governance structures, informational and institutional asymmetries (Amewu and Paul Alagidede, 2021; Buckley *et al.*, 2016b; Buckley and Tian, 2017; Chari *et al.*, 2010; Kim *et al.*, 2021). Kim *et al.* (2021) reported listing effect of M&A to be more pronounced for targets based in emerging economies, attributing it to the target country economic growth uncertainty. Studies examining the impact of target location on performance have presented contrasting results. Chalencon and Mayrhofer (2018) assert that acquirers entering developed markets, as against emerging markets, realise higher gains and suggest future studies to use differentiation based on target location. Alternatively, studies have also reported higher bidder returns for deals involving emerging market targets. Chari *et al.* (2010) reported significant value creation for developed market acquirers whilst entering an emerging market, whereas such returns were not realised in case of targets located in developed markets. For a sample of EMFs, Narayan and Thenmozhi (2014) found significant value destruction upon acquisition of developed market targets. Whereas, the vice-versa was not found to be true. Considering the reported significance, a disaggregated analysis based on target location becomes necessary. Thus,

H3. There exists significantly different and higher shareholder returns for deals involving developed market targets as against emerging markets.

2.4 Ownership structure: majority vs minority

In line with the transactional cost economics (Williamson, 1981) and eclectic paradigm (Dunning, 1988), decisions involving acquired ownership levels are viewed as a means of managing the desired levels of target control, resource commitment, risks and synergy realisation. Minority deals, resulting in the acquisition of less than 50% of the target equity stakes, represent a distinct strategic choice against majority or full ownership (Ouimet, 2013). Higher equity stakes in the target reduce the risk of partner opportunism (Lien and Filatotchev, 2015) and allow acquirers to

exercise greater control over target assets, assisting in better knowledge transfer (Gaffney *et al.*, 2016). However, it requires higher financial and managerial commitments, elevating the acquirer risk levels and adding to the liability of foreignness in case of cross-border deals (Chari and Chang, 2009). The motivations behind the choice of minority acquisitions have been reported to differ (Liao, 2014). Information asymmetry, lack of experience, and keeping target managerial incentives intact motivate lower ownership levels (Ouimet, 2013). It supports risk-sharing and promotes gains from the local partner's legitimacy and expertise (Makino and Delios, 1996). Studies have reported various external factors like institutional distance, institutional quality, cultural characteristics and industry relatedness also impact the ownership choices (Pinto *et al.*, 2017; Reddy *et al.*, 2022; Wang and Larimo, 2020; Yang *et al.*, 2022). This decision constitutes a crucial factor determining the future trajectory of a deal (Malhotra and Morgan, 2018). For developed market acquirers entering emerging markets, higher equity stakes held post-acquisition are reported to positively impact the deal performance (Chari *et al.*, 2010; Narayan and Thenmozhi, 2014). Although, the effect of ownership levels on the emerging market acquirer returns remains understudied. Two recent emerging studies report high ownership participation to be positively rewarded by the market (Jain *et al.*, 2019; Yang, 2015). Furthermore, preferences for ownership levels tend to vary for different countries. For example, a recent study comparing Chinese and Indian MNFs confirmed significant differences in ownership preferences whilst engaging in knowledge-intensive foreign acquisitions (Scalera *et al.*, 2020). Chinese MNFs were found to favour lower equity control against their Indian counterparts. Following this line of argument, we propose to examine if the market reaction to M&A deals involving majority and minority acquired ownership stakes would differ for different markets.

H4. There exist significantly different and higher shareholder returns for deals resulting in majority ownership as against minority ownership.

None of the studies was found to present inter-country comparative evidence for emerging market M&A performance, utilising a comprehensive as well as disaggregated deal data set for a broad group of countries. The present study compares the M&A performance for the group of BRICS countries utilising the widely accepted event study methodology, gauging the short-term impact on shareholders' wealth around the deal announcement.

3. Methods and data

3.1 Methodology– event study as performance measure

For studying the shareholders' wealth creation around M&A announcement, the event study methodology is employed. This methodology has been widely accepted for measuring the impact of an event on the shareholder wealth and deal performance (Brown and Warner, 1980; Fama *et al.*, 1969; Roll, 1986). To calculate the abnormal returns owning to an event, expected returns are removed from the actual stock returns for each day over the event window. Expected returns, i.e. the return expected to be earned if the focal event had not occurred, are calculated using the ordinary least squares market model over an estimation window of 180 days, ranging from 31st to 210th day before an event (Brown and Warner, 1985). For market return, a broad-based market index for each country is taken to represent the risk-return conditions for each market better. "Brazilian BOVESPA", "MOEX Russia index", "BSE 500", "Shanghai S.E. Composite index" and "FTSE/JSE All Share index" were used, respectively, for BRICS.

Abnormal returns are calculated as per equation (1):

$$AR_{i,t} = R_{i,t} - E(R)_{i,t} \tag{1}$$

Where, for a firm "i" on day "t", AR = abnormal return, R = Actual Return, E(R) = Expected return calculated by equation (2):

$$E(R)_{i,t} = \alpha + \beta(Rm_t) \tag{2}$$

R_m = Market Return, α and individual risk (β) of each stock are calculated over the estimation window period of 150 days (−180, −31 days).

For the study, average abnormal return (AAR) and cumulative average abnormal returns (CAAR) over multiple event window periods [(−1, +1), (−3, +3), (−5, +5)] around the date of announcement (Day 0) are analysed. The different event windows are utilised acknowledging the trade-off between the shorter as against the longer event window. Whilst a longer event window is equipped to capture any information leakages and delayed market reactions, it poses a critical challenge to ensure a clean event window avoiding any contamination of information reflected in the stock price movements. The AAR and CAAR are calculated as per equations (3) and (4), respectively.

$$AAR_{(t)} = \frac{1}{N} * \sum_{i=1}^{N} AR_{i,t} \tag{3}$$

$$CAAR_{(T_1,T_2)} = \frac{1}{N} * \sum_{i=1}^{N} CAR_{i,(T_1,T_2)} \tag{4}$$

Where in, $AAR_{(t)}$ = Average abnormal return on day t, N = Number of deals, $CAAR_{(T_1,T_2)}$ = Cumulative Average Abnormal Return over the select event window starting from the day T_1 to T_2.

The statistical significance of country-wise AAR and CAAR results are tested by employing a set of two parametric tests, namely the cross-sectional standard deviation (CSS) test (Brown and Warner, 1985) and crude-dependence adjustment (CDA) test (Brown and Warner, 1980). The CSS test statistic is computed as:

$$\text{CSS}\left(\text{AAR}_t\right) = \sqrt{N}\,\frac{\text{AAR}_t}{S.D._{AAR_t}} \tag{5}$$

$$\text{CSS}\left(\text{CAAR}_{(T_1,T_2)}\right) = \sqrt{N}\,\frac{CAAR_{(T_1,T_2)}}{S.D._{CAAR_{(T_1,T_2)}}} \tag{6}$$

CSS holds an underlying assumption of equal variances and no cross-correlation across firms, violation of which might create significant biases in results (Collins and Dent, 1984). To handle such dependence across cross-sectional returns, the procedure of crude-dependence adjustment (Brown and Warner, 1980) is adopted as follows:

$$CDA\,(AAR) = \frac{AAR_t}{S.D._{\,AAR_{(-180,-31)}}} \tag{7}$$

$$CDA\left(CAAR_{(T_1,T_2)}\right) = \frac{CAAR_{(T_1,T_2)}}{\sqrt{150}\left(S.D._{\,AAR_{(-180,-31)}}\right)} \tag{8}$$

Further for testing the robustness of the results, the results are supplemented with two non-parametric tests, viz. generalised sign test (Cowan, 1992) and Corrado's rank test (Corrado Charles, 1989), as parametric tests often pose limitations owing to their underlying assumptions and sensitivity to outliers.

Test statistic for generalised sign test is calculated as follows:

$$Zrank\,(AAR_t) = \frac{\frac{1}{N}\sum_{i=1}^{N} R_{i,t} - \widehat{R}}{S.D._{R_{(-180,\,T_2)}}} \tag{9}$$

$$Zrank\left(CAAR_{(T_1,T_2)}\right) = (T_2 - T_1 + 1)^{1/2} \frac{\overline{R}_{(T_1,T_2)} - \widehat{R}}{\left(\sum_{t=1}^{(150+(T_2-T_1+1))} \left(\overline{R}_t - \widehat{R}\right)^2 \Big/ (150+(T_2-T_1+1))\right)^{1/2}} \tag{10}$$

Where, $\widehat{R}$ is the average rank calculated over the estimation and event window abnormal return ranks. $\overline{R}_t$ is the average rank for "N" deals on day "t".

The generalised sign test is specified in equation (11).

$$Zg\,(AAR_t) = \frac{w - \left(\mathrm{N}*\widehat{\mathrm{p}}\right)}{\left(N*\widehat{\mathrm{p}}\left(1-\widehat{\mathrm{p}}\right)\right)^{1/2}} \tag{11}$$

Where,

w is the number of deals with positive abnormal return on tth day in the event window (or with positive cumulative abnormal return over the event window for $Zg(CAAR_{(T_1,T_2)})$) and $\widehat{\mathrm{p}}$ calculated over the estimation window as follows:

$$\widehat{\mathrm{p}} = \frac{1}{n}\sum_{i=1}^{N}\frac{1}{150}\sum_{t=1}^{150} S_{i,t} \tag{12}$$

Herein, the value of $S_{i,t}$ is equal to one if $AR_{i,t}$ is positive and zero otherwise.

3.2 Data

The initial sample of the study comprised of M&A deals announced and completed between January 1, 2000 and December 31, 2019, avoiding the period of COVID-19 impact on the world markets. The initial population included all deals involving acquirers from BRICS. The deal data were extracted from the Thompson Securities Data Company (SDC) Platinum Database, which is internationally accepted and widely used. Deals involving more than 10% of stakes were only included in the sample to exclude any possible cases of portfolio investments. Financial sector deals were excluded to avoid any bias due to peculiar regulations governing the sector.

As presented in Table 1, the initial population of deals is filtered for fulfilling the requirements of event study methodology adopted in the study. To be retained in the final sample, availability of stock price data over the estimation and event window was a prerequisite for calculating the abnormal returns. Deals with confounding events 30 days around the event day were excluded to ensure a clean event period avoiding any contamination of information influencing the stock returns. Deals with non-synchronous data were also removed to avoid misinterpretation. The final analysis is performed on a sample of 7,105 deals. In line with the Chinese leadership in the actual number of deals announced, our final sample is skewed towards China, accounting for 70% of the deals in the final sample.

Table 1.
Sample selection

Country	Initial sample			Deals dropped due to			Final sample		
	D (1)	CB (2)	Combined (1) + (2) = (3)	Confounding events (4)	Data unavailable (5)	Non-synchronous (6)	D (7)	CB (8)	Combined (9) = (7) +(8) = (3)-(4)-(5)-(6)
Brazil	3,501	381	3,882	3,291	262	81	189	59	248
Russia	8,352	674	9,026	2,732	4,831	803	542	119	661
India	4,228	1,547	5,775	4,771	171	122	411	300	711
China	14,712	2,625	17,337	4,450	6,557	1,349	4,532	449	4,981
South Africa	1,034	701	1,725	318	670	233	282	222	504
TOTAL	31,827	5,928	37,745	15,562	12,491	2,588	5,956	1,149	*7,105*

Note(s): D = Domestic; CB = Cross-Border
Source(s): Compiled by authors from SDC platinum database

Also, the problem of missing data has added to the variations in the number of deals from each country in the final sample. Nevertheless, the performance measures utilised in the study are calculated as sample averages and hence minimises the risk in analysing the results for unequal country samples.

4. Empirical results

The sample summary of the various deal-groups utilised for the study are presented in Table 2. Initially, Deals are classified based on acquirer country for comparing the performance among chosen markets. Further, sub-samples are created based on the deal characteristics to investigate how emerging markets react for different categories of deals. Panel 1 reports the number of deals as bifurcated between cross-border (16%) and domestic (84%). Then, Panel 2 presents cross-border deals as divided by the development status of target countries, viz., developed vs emerging markets, based on its Organisation for Economic Cooperation and Development (OECD) membership. Panel 3 reports the sub-sample of deals resulting in majority ownership (51% and above) and minority ownership (between 10 and 50%). The 50% cut-off level of target equity stake owned after the transaction is utilised for creating the binary ownership classification (Contractor *et al.*, 2014; Dang *et al.*, 2018; Makino and Beamish, 1998). A perusal of the target countries reveal that U.S. is common to most of sample countries except Russia where Ukraine, Cyprus and Germany dominate the scene. In case of China the notable engagement is of Hong Kong obviously because of transfer of lease to China. The results and analysis for these samples are discussed in the following sections.

4.1 BRICS performance: country-wise analysis

Table 3 presents the country-wise AAR and CAAR results for the acquiring firms around the event date. The inter-country differences in the market reaction to the announcement are evident in Figure 1(a) and (b). The results suggest a positive market perception of the M&A announcement with significant and positive zero-day AAR for each sample country. For the announcement day, the highest positive reaction can be observed in the Chinese market (0.91%), followed by South Africa (0.55%) and is the least for Russia (0.03%). The results confirm the significant differences across the BRICS. Hence, H1 stands rejected. For South African firms, market reaction can be observed starting two days before the deal announcement, indicating information leakages and alleged insider trading. Cumulating the abnormal returns around the event window, the results for Brazil, India, China and South Africa confirm that the acquirer shareholders experience addition to their wealth around the M&A announcement with positive CAAR values over the multiple event windows. Whereas for the deals involving Russian acquirers, the market seems to be cold towards the M&A announcements with only a slightly positive abnormal return observed on the day of announcement. Furthermore, the returns present a declining trend with negative AAR in the days following the event date contributing towards negative CAAR values over the multiple event windows $(-1, +1)$, $(-3, +3)$, $(-5, +5)$. The deal information seems to be only gradually absorbed in the market and reflected in the share prices causing a delay, attributable to market inefficiencies. This seems to be a case of lead and lag relationship. The results are largely in line with the previous literature, showing that emerging market acquirers earn a positive abnormal return on M&A announcements (Jain *et al.*, 2019).

4.2 Domestic vs cross-border

Cross-border M&A represent a distinct strategic choice against its domestic counterpart, with its own set of opportunities, costs and risks. For EMFs, cross-border deals provide a means of accessing foreign resources, albeit not without its challenges.

Table 2.
Sample distribution

	Brazil		Russia		India		China		South Africa		Total	
	No. of deals	%	No. of deals	%	No. of deals	%	No. of deals	%	No. of deals	%	No. of deals	%
Full sample	248	100	661	100	711	100	4,981	100	504	100	7,105	100
PANEL 1: Type of deal												
Cross-border	59	23.79	119	18.00	300	42.19	449	9.01	222	44.05	1,149	16.17
Domestic	189	76.21	542	82.00	411	57.81	4,532	90.99	282	55.95	5,956	83.83
PANEL 2: Target country development status												
Developed market	30	50.85	45	37.82	232	77.33	337	75.06	139	62.61	783	68.15
Emerging market	29	49.15	74	62.18	68	22.67	112	24.94	83	37.39	366	31.85
PANEL 3: Target ownership level												
Majority	213	85.89	531	80.33	613	86.22	4,022	80.75	399	79.17	5,778	81.32
Minority	35	14.11	130	19.67	98	13.78	959	19.25	105	20.83	1,327	18.68
Top 3 target countries	Argentina, United States, Canada, Mexico		Ukraine, Cyprus, Germany		United States, United Kingdom, Germany		United States, Germany, Hong Kong		Australia, United Kingdom, United States, Canada			

Source(s): Compiled by authors from SDC platinum database

Acquirer nation	Event window [day]	No. of days in event window	AAR/ CAAR	CSS	CDA	Z-rank	Zg
Brazil	0	1	0.28	1.24	2.18**	−0.01	0.32
	(−1, +1)	3	0.46	1.45	2.06**	2**	1.6
	(−3, +3)	7	0.21	0.52	0.62	−0.01	1.21
	(−5, +5)	11	0.08	0.16	0.18	0.81	1.09
Russia	0	1	0.12	0.89	0.99	0.81	0.53
	(−1, +1)	3	−0.01	−0.03	−0.03	0.45	0.68
	(−3, +3)	7	−0.26	−0.76	−0.84	−1.33	0.92
	(−5, +5)	11	−0.56	−1.49	−1.44	−2.72***	−0.1
India	0	1	0.26	1.09	2.33**	3.31***	3.73***
	(−1, +1)	3	0.47	2.41**	1.12	2.74***	3.88***
	(−3, +3)	7	0.51	1.75*	0.8	2.15**	3.20***
	(−5, +5)	11	0.09	0.23	0.12	0.44	1.09
China	0	1	0.91	14.71***	16.34***	10.17***	12.05***
	(−1, +1)	3	1.43	14.12***	14.87***	9.91***	12.70***
	(−3, +3)	7	1.75	11.65***	11.89***	6.97***	10.29***
	(−5, +5)	11	1.85	10.03***	10.06***	5.84***	9.24***
South Africa	0	1	0.55	3.34***	4.52***	3.79***	3.68***
	(−1, +1)	3	1.14	4.09***	5.38***	7.16***	3.95***
	(−3, +3)	7	1.32	3.65***	4.06***	6.20***	3.15***
	(−5, +5)	11	1.3	3.06***	3.21***	4.85***	2.26

Note(s): *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$

Table 3.
Country-wise average abnormal returns (AAR) and cumulative average abnormal return (CAAR)

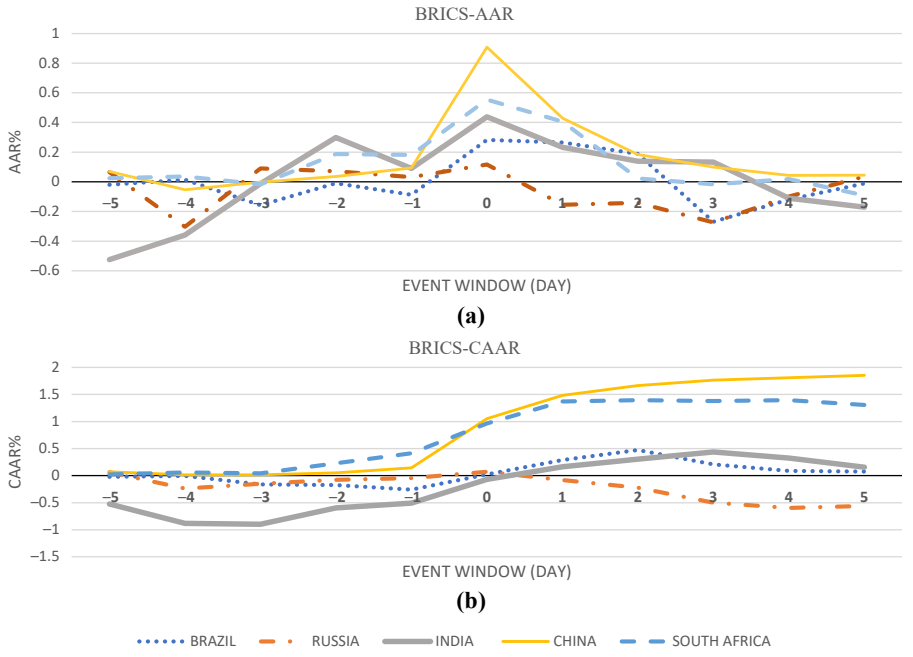


Figure 1.
Average abnormal returns (AAR) and cumulative average abnormal returns (CAAR) during 11-day event-window

Table 4 report the AAR and CAAR values along with respective test statistics for multiple event windows classified for domestic and cross-border deals. The graphical representation of the results for the country-wise domestic and cross-border deals is presented in Figure 2 (a) and (b), respectively. Other than Russian and Indian domestic M&A deals, the zero-day AAR prove to be statistically significant for all the sub-samples (see Table 4). Except for Indian domestic (−0.03%) and Brazilian cross-border deals (−0.6%), the announcement day AAR for all the countries, for both domestic and cross-border deals, are found to be positive. Despite a positive announcement day abnormal return (0.41%), a steep fall is observed in the abnormal return for Russian cross-border deals following the event day contributing towards statistically insignificant yet negative cumulative abnormal returns over the 5-day [−2, +2] and above event windows. H2 is rejected since there is found to be a contrast in the pattern of market responses for the BRICS countries based on the target location. For the sample of acquirers based in Brazil, China and South Africa the market appears to yield higher returns for domestic deal announcements as against the cross-border deals. Whereas, in the case of Indian and Russian acquirers, cross-borders are more positively rewarded by the market. The highest difference (1.16%) can be observed in the case of Brazilian deals with a statistically significant zero-day abnormal return of 0.56% for domestic and −0.6% for cross-border deals.

Analysing the cumulative abnormal returns over the event windows ranging from 3-day to 11-days around the event, stark differences can be observed both in terms of inter-country returns and for intra-country returns based on the target locations. Chinese and South African markets view M&A positively regardless of the target location. However, domestic deals are favoured over cross-border deals by the Brazilian and Chinese markets as evident from the consistently higher cumulative returns over the analysed event windows. In contrast, an opposite pattern is observed for the Indian market with a better perception of the cross-border deals. Brazilian cross-borders deals and Russian domestic deals are rather evident to be eating into the shareholder wealth with statistically insignificant but negative cumulative returns around the event.

4.3 Target country development status: emerging vs developed

In this section, cross-border deals are further analysed based on target country development status. Figure 3(a) and (b) plots the AAR surrounding the announcement for the sub-samples of targets based out of developed and emerging markets, respectively. Table 5 presents the calculated CAAR values over different event window periods, spread over 11-days around the event. The abnormal returns for the cross-border deals involving targets located in emerging markets against those from developed markets are compared.

As evident from the AAR and CAAR results, Indian and South African acquirers add more to their shareholders' wealth for deals involving developed market targets against those located in emerging markets, supporting H3. Given the negative zero-day abnormal returns, deals involving emerging market targets are instead causing a wealth deterioration for the South African (−0.02%) and Brazilian (−0.32%) acquirers. Furthermore, against H3, emerging market targets have been more rewarding for the Chinese acquirers. Likewise, CAAR results for Russia indicate wealth creation in case of targets located in emerging markets against wealth destruction for deals involving developed market targets.

In case of Brazil, M&A announcements prove to be ill-perceived by the market, with negative event day abnormal returns regardless of the target country development status. Chinese firms present higher abnormal returns on the day following the deal announcement as against day zero for deals including both emerging and developed market targets. Hence, H3 is partially supported.

4.4 Ownership structure

Table 6 summarises the results of the country-wise CAAR calculated for the different levels of acquired ownership. The acquired ownership structure is an influential factor that directly

Acquirer nation	Event window [day]	Domestic					Cross-border				
		AAR/ CAAR	CSS	CDA	Z-rank	Zg	AAR/ CAAR	CSS	CDA	Z-rank	Zg
Brazil	0	0.56	2.01**	3.56***	0.93	0.85	−0.6	−1.78*	−2.53**	0.31	−1.64
	(−1, +1)	0.85	2.28**	3.13***	3.34***	2.02**	−0.78	−1.34	−1.91	−2.25	−1.11
	(−3, +3)	0.64	1.37	1.55	1.47	1.14	−1.17	−1.48	−1.86	−3.55	−0.33
	(−5, +5)	0.58	1.06	1.11	2.31**	1	−1.53	−1.46	−1.95	−3.21	−0.33
Russia	0	0.05	0.36	0.41	0.49	0.45	0.41	1.35	1.7*	0.91	0.11
	(−1, +1)	−0.03	−0.11	−0.12	2.08**	1.39	0.08	0.16	0.2	−4.87	−1.36
	(−3, +3)	−0.26	−0.66	−0.75	−0.49	1.05	−0.28	−0.41	−0.44	−3.36	−0.07
	(−5, +5)	−0.56	−1.31	−1.3	−2.07	0.36	−0.59	−0.73	−0.74	−3.52	−0.99
India	0	−0.03	−0.19	−0.07	0.09	0.21	0.66	3.85***	4.47***	4.77***	5.48***
	(−1, +1)	0.12	0.44	0.17	1.5	−6.23	0.94	3.77***	3.66***	7.01***	3.86***
	(−3, +3)	0.11	0.26	0.1	2.25**	−6.23	1.06	2.93***	2.7***	4.62***	2.93***
	(−5, +5)	−0.24	−0.39	−0.17	1.14	−7.13	0.56	1.29	1.13	0.97	1.66*
China	0	0.97	15.12***	19.64***	10.25***	11.55***	0.28	1.29	0.85	2.89***	3.44***
	(−1, +1)	1.5	14.14***	17.54***	10.18***	12.44***	0.74	2.15**	1.26	3.89***	−7.26
	(−3, +3)	1.84	11.66***	14.1***	7.14***	9.88***	0.79	1.68*	0.89	2.81***	−7.07
	(−5, +5)	1.95	10.04***	11.93***	5.82***	8.87***	0.85	1.44	0.76	3.46***	−7.07
South Africa	0	0.67	3.3***	4.32***	2.67***	2.47**	0.41	1.49	2.13**	2.61***	2.77***
	(−1, +1)	1.12	3.17***	4.17***	5.45***	3.18***	1.17	2.61***	3.53***	6.6***	2.36**
	(−3, +3)	1.52	2.96***	3.71***	4.82***	2.23**	1.07	2.13**	2.1**	5.6***	2.23**
	(−5, +5)	1.3	2.2**	2.53***	3.32***	0.92	1.32	2.14**	2.07**	4.92***	2.36**

Note(s): *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$

Table 4.
Country-wise average abnormal returns (AAR) and cumulative average abnormal return (CAAR) for domestic and cross-border deals

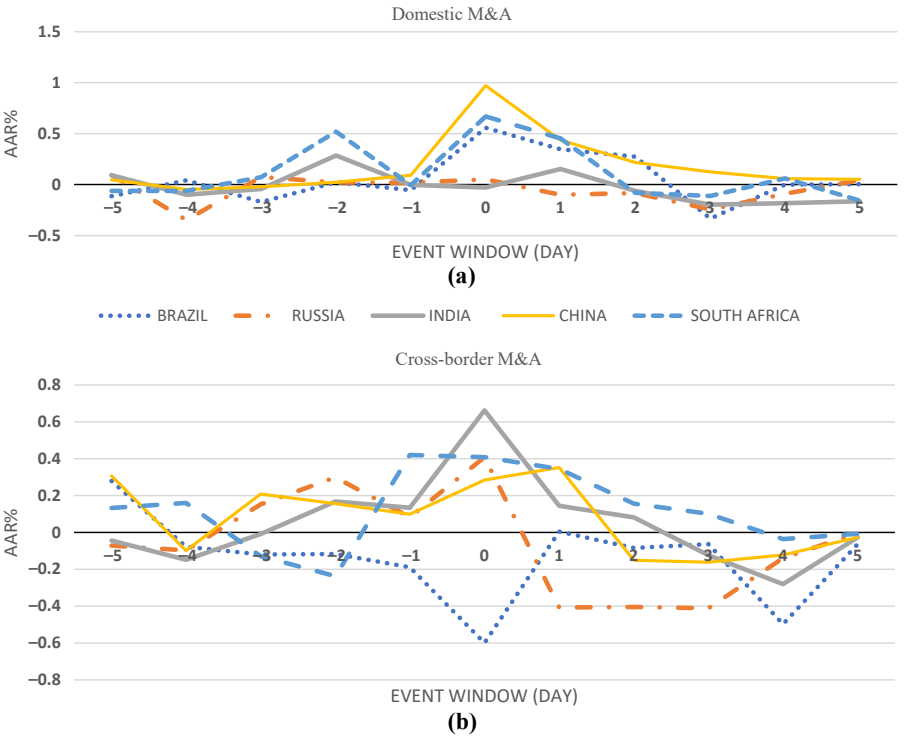


Figure 2.
Average abnormal returns (AAR) for cross-border and domestic M&A

impacts flexibility and decision-making, ultimately impacting the synergy realisations. A majority stakes in the target aids towards shaping board decision-making in the target firm, whereas a minority stake often helps in sustained cooperative acquirer-target partnership but does not provide decision-making authority to the acquirers (Tao *et al.*, 2017). The country samples are disaggregated based on the level of ownership stakes held by the acquirer post the M&A deal, creating a binary classification of minority vs majority ownerships representing distinct strategic choices.

Supporting H4, distinguished market reactions can be observed for deals leading to majority ownership as against minority ownership. For Indian and Chinese acquirers, majority ownership deals are rewarded higher by the market as against minority. As evident from Figure 4, Chinese majority ownership deals top the wealth creation ranking with a statistically significant and positive zero-day abnormal return of 1.03%. The results reveal a positive zero-day return for the majority ownership deals in the case of all the five emerging markets. In comparison, an insignificant yet negative abnormal return can be observed for Indian (−0.05%) and Brazilian (−0.31%) minority ownership deals on announcement day.

5. Discussion

This paper evaluated and compared the M&A performance across BRICS countries utilising a sample of 7,105 deals announced between 2000 and2019. The results reveal that BRICS acquirers do not experience similar returns around deal announcements. There does exist a significant difference across the pairs of countries. Chinese acquirers document the highest positive market returns against the lowest by Russian acquirers. Whilst contributing to the

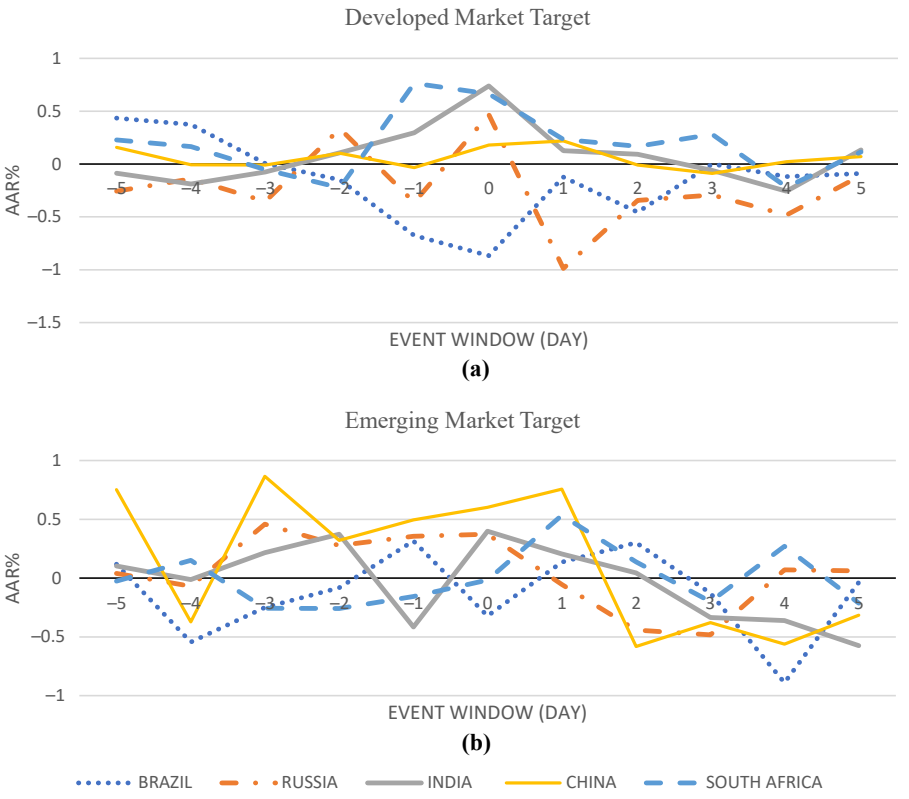


Figure 3.
Average abnormal returns (AAR) for developed and emerging market targets

boom in the number and value of deals, the "Go-out" policy adopted by the Chinese government has also contributed positively to the market sentiment towards deal announcements. Consistent with the resource-based view, Chinese firms tend to target natural resources which cannot be replicated in their home country and can further strengthen their home advantage of labour abundance (Yang and Deng, 2017). The positive market reaction to the Chinese deal announcements is in congruence with the previous research evidence (Jain *et al.*, 2021; Lin *et al.*, 2020; Tao *et al.*, 2017). Interestingly, Russian deal returns are performing differently as against other emerging market counterparts. It recorded a positive abnormal return (0.11%, lowest among BRICS) on the announcement day, which is not sustained in the days following the event. This contributes to a negative CAAR, causing shareholder wealth destruction in the short-run. This result is consistent with previous study reporting negative performance results for Russian M&A deals owing to the lack of experience and capabilities (Bertrand and Betschinger, 2012). Further, an early market reaction was captured for the South African deals, starting two days before the deal announcement. This seems to be the case of information leakages. This is supported by Amewu and Alagidede (2018).

Not just limiting to the broad inter-country comparisons, the study further examines the performance variations based on select deal characteristics. First, the disaggregated samples of domestic and cross-border deals have been investigated based on the target location. For cross-border deals, Brazil demonstrated a distinct performance pattern with a negative zero-day return of 0.6% against statistically significant positive returns for all other markets

Table 5. Country-wise average abnormal returns (AAR) and cumulative average abnormal return (CAAR) for emerging and developed market targets

Acquirer nation	Event window [day]	AAR/CAAR	Emerging CSS	Emerging CDA	Emerging Z-rank	Emerging Zg	Developed CAAR	Developed CSS	Developed CDA	Developed Z-rank	Developed Zg
Brazil	*0*	−0.32	−1.44	−0.94	−1.24	−1.78*	−0.87	−1.38	−2.64***	−1.1	−0.54
	(−1, +1)	0.13	0.23	0.22	1.28	−0.67	−1.67	−1.67*	−2.93***	−5***	−0.91
	(−3, +3)	−0.03	−0.03	−0.04	0.14	0.08	−2.27	−1.86*	−2.61***	−6.02***	−0.54
	(−5, +5)	−1.38	−1.1	−1.22	−3.42***	−0.67	−1.67	−0.99	−1.53	−1.94**	0.19
Russia	0	0.37	0.84	1.17	0.39	0.37	0.47	1.38	1.25	1.02	−0.29
	(−1, +1)	0.67	0.85	1.22	−1.48	−0.1	−0.88	−2.34**	−1.36	−8.11***	−2.08**
	(−3, +3)	0.48	0.48	0.57	0.47	0.83	−1.54	−2.07**	−1.55	−7.77***	−1.19
	(−5, +5)	0.58	0.53	0.56	0.19	0.13	−2.53	−2.22**	−2.03**	−7.72***	−1.78*
India	0	0.4	0.9	1.41	2.42**	2.64	0.74	4.08***	4.13***	4.34***	0.52
	(−1, +1)	0.19	0.31	0.39	0.43	1.43	1.16	4.32***	3.74***	8.43***	1.8*
	(−3, +3)	0.49	0.67	0.65	1.36	0.45	1.22	2.94***	2.58***	5.14***	1.42
	(−5, +5)	−0.36	−0.37	−0.39	−3.19***	−0.52	0.83	1.72*	1.39	2.42**	1.29
China	0	0.6	1.16	2.05**	1.43	1.83	0.18	0.75	0.4	2.54**	2.91***
	(−1, +1)	1.85	2.13**	3.64***	5.59***	1.83*	0.36	1.04	0.47	2.41**	2.14**
	(−3, +3)	2.08	2.15**	2.67***	3.8***	1.08	0.36	0.67	0.3	1.85*	2.69***
	(−5, +5)	1.58	1.51	1.62	3.41***	1.83*	0.6	0.86	0.41	2.83***	1.93*
South Africa	0	−0.02	−0.06	−0.06	1.69*	2.63	0.66	1.64	2.55**	1.97**	1.47
	(−1, +1)	0.36	0.87	0.72	7.09***	1.53	1.66	2.47**	3.68***	4.5***	77***
	(−3, +3)	−0.22	−0.37	−0.29	3.02***	−0.01	1.83	2.56**	2.66***	5.92***	83***
	(−5, +5)	−0.03	−0.04	−0.04	3.83***	1.31	2.12	2.49**	2.46**	4.38***	78***

Note(s): *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$

Table 6.
Country-wise average abnormal returns (AAR) and cumulative average abnormal return (CAAR) for majority and minority acquired ownership stakes

			Majority					Minority			
Acquirer nation	Event window [day]	AAR/CAAR	CSS	CDA	Z-rank	Zg	CAAR	CSS	CDA	Z-rank	Zg
Brazil	*0*	0.38	1.49	2.68***	0.09	0.05	−0.31	−0.72	−1.06	−0.23	−0.28
	(−1, +1)	0.49	1.41	2.01**	1.12	0.74	0.25	0.35	0.5	4.25***	1.41
	(−3, +3)	0.01	0.02	0.02	−1.87	0.33	1.45	1.78	1.88*	6.96***	1.41
	(−5, +5)	−0.23	−0.43	−0.49	−0.88	0.46	1.93	1.73	1.99**	6.7***	0.74
Russia	0	0.04	0.27	0.31	0.43	0.17	0.43	1.72*	1.82*	1.08	0.68
	(−1, +1)	−0.08	−0.35	−0.37	0.19	0.43	0.31	0.84	0.77	1.06	0.68
	(−3, +3)	−0.35	−0.86	−1.01	−1.24	1.04	0.12	0.26	0.19	−0.97	−0.02
	(−5, +5)	−0.73	−1.65*	−1.67	−2.75	−0.27	0.13	0.21	0.17	−1.3	0.33
India	0	0.31	2.53**	1.11	3.53***	3.82***	−0.05	−0.2	−0.22	0.36	0.48
	(−1, +1)	0.53	2.48**	1.09	4.98***	3.82***	0.06	0.14	0.13	1.57	0.89
	(−3, +3)	0.61	1.87**	0.81	4.21***	3.09***	−0.1	−0.16	−0.15	1.04	0.89
	(−5, +5)	0.22	0.47	0.23	1.54	1.22	−0.68	−0.94	−0.83	−0.42	−0.13
China	0	1.03	14.63***	16.75***	9.83***	11.43***	0.38	3.14***	2.99***	3.86***	4.06***
	(−1, +1)	1.64	14.24***	15.38***	10.57***	12.44***	0.53	2.62***	2.43**	3.48***	3.48***
	(−3, +3)	2.06	12.17***	12.64***	7.71***	9.59***	0.42	1.33	1.25	1.46	3.8***
	(−5, +5)	2.19	10.54***	10.7***	6.17***	8.23***	0.44	1.11	1.05	2.21**	4.19***
South Africa	0	0.58	3***	4.13***	3.46***	3.17***	0.46	1.5	2.01**	1.83*	1.89*
	(−1, +1)	1.08	3.29***	4.43***	5.69***	3.07***	1.39	2.76***	3.51***	8.09***	2.67***
	(−3, +3)	1.36	3.22***	3.68***	6.38***	3.07***	1.14	1.74*	1.89*	3.06***	0.91
	(−5, +5)	1.3	2.61***	2.79***	4.29***	1.76*	1.34	1.7*	1.77*	4.33***	1.5

Note(s): *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$

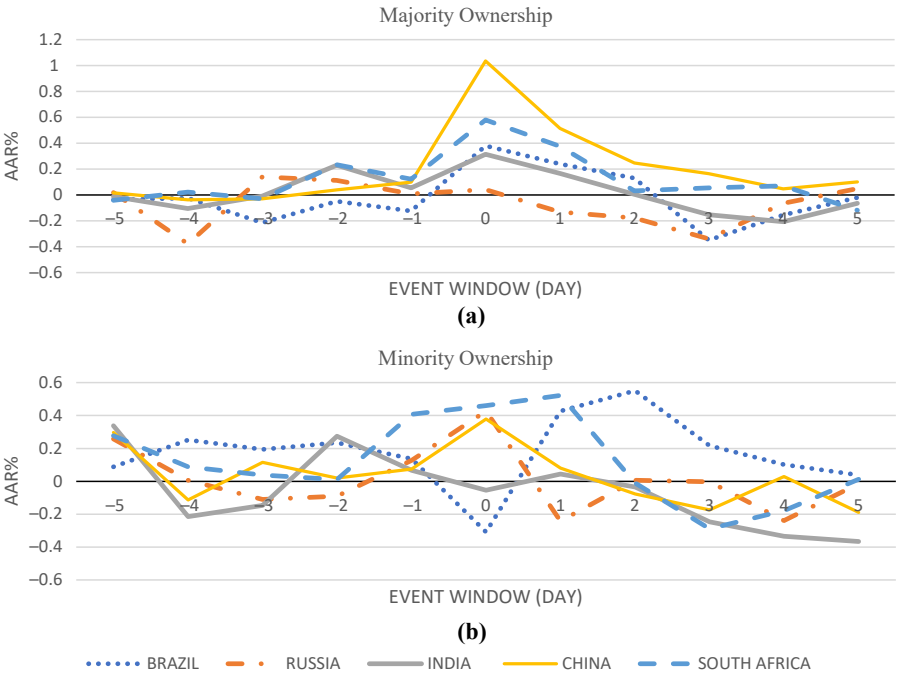


Figure 4.
Average abnormal returns (AAR) for majority and minority acquired ownership stakes

ranging between 0.28% and 0.66%. In the case of domestic deals, all countries recorded a positive market reaction to deal announcement, except for India which had an insignificant yet negative zero-day return of 0.03%. Indian and Brazilian acquirers report an opposite market response, where Indian acquirers are rewarded for cross-border deals and penalised for domestic deal announcements by the market. UK acquirers are rewarded higher for domestic targets than foreign targets (Adel and Alkaraan, 2019). For South Africa, evidence for insider trading is found in both, domestic and cross-border deals.

Second, inter-country deal performances are compared by disaggregating the cross-border deals based on the development status of the target country. Developed and emerging target market offers a distinct package of opportunities and risk. Due to the supportive home country government policies, international aspirations and opening up of the world economies creating flatter competitive global fields, EMFs are now more confident and aggressively penetrating developed markets. In both the sub-samples, the zero-day returns are found to be minimum for Brazilian deals. The market does not seem to have confidence in the synergy realisation capabilities of Brazilian cross-border acquirers. Thus, it is argued that Brazilian cross-border deals are penalised by the market irrespective of the target country development status, causing shareholder wealth destruction.

Third, country-wise deal performances are examined based on the target ownership stakes as a proxy for the acquired level of control. Majority ownership stakes often provide operational control and power to steer board decisions. The results reveal a positive zero-day return (ranging between a minimum of 0.04% for Russia to maximum of 1.03% for China) for the majority ownership deals in the case of all five emerging markets. Whereas for deals involving minority ownership, an insignificant yet negative announcement day abnormal return is observed for India (−0.05%) and Brazil (−0.31%) against positive returns for other

countries. The negative market response indicates market suspicion of the ability to realise the intended synergies through minority acquisitions. Although positive on event day, the abnormal returns for Russian firms are not sustained in the days succeeding announcement irrespective of the acquired ownership level.

5.1 Theoretical contribution and practical implications

The results of study also offer significant theoretical and practical implications. First, the study broadens the literature by providing comparative evidence on emerging market M&A performance. The findings indicate that there does exist differences in deal performance for BRICS countries as evident from the comprehensive country-wise results as well as in the investigated sub-samples. The disagreement on the M&A performance in the previous literature can be partly attributed to their limited samples. Generalising emerging market M&A performance based on single country results can be misleading and calls for a more cautious approach in sample selection of future studies.

Secondly, acknowledging the potential performance differences across markets and deal characteristics shall aid executives in sharpening decision-making and help general investors to better formulate and evaluate their investment strategies. Higher returns can be expected in the case of deals resulting in majority stake ownership, granting greater control to the acquirers, reducing the risk of partner opportunism and better knowledge sharing. Further, the study offers public policy implications as well. The governments can take a cue from these results for an enhanced understanding of the market temperament for M&A deals and to better manage their investment flows.

5.2 Limitations and future research directions

The limitations of the study offer relevant avenues for future research. First, the present study utilises a market-based measure for gauging deal performance. A further comparative evaluation of market-based measures against the accounting measures of performance can provide insightful results. Second, the study compared the M&A performance for BRICS markets based on target location and acquired stakes. To further add to this line of research, other factors can be analysed and compared to gauge their differing impact. Qualitative analysis of emerging market deals can aid in highlighting novel factors contributing to the potential performance differences. Comparing M&A performances across markets based on home institutional quality can also provide as another interesting research direction. Finally, building upon the idea from this study, a time-series analysis of how the M&A performance has progressed in different markets can contribute to a better understanding of the contemporary situation and estimating future evolution.

References

Adel, N. and Alkaraan, F. (2019), "Strategic investment acquisitions performance in UK firms: the impact of managerial overconfidence", *Journal of Financial Reporting and Accounting*, Vol. 17 No. 1, pp. 24-41.

Alkaraan, F. (2019), "Making M&A less risky: the influence of due diligence processes on strategic investment decision making", *Advances in Mergers and Acquisitions*, Emerald Publishing, Vol. 18, pp. 99-110.

Amewu, G. and Alagidede, P. (2018), "Do mergers and acquisitions announcements create value for acquirer shareholders in Africa", *International Journal of Finance and Economics*, Vol. 23 No. 4, pp. 606-627.

Amewu, G. and Paul Alagidede, I. (2021), "Mergers, executive compensation and firm performance: the case of Africa", *Managerial and Decision Economics*, Vol. 42 No. 2, pp. 407-436.

Aybar, B. and Ficici, A. (2009), "Cross-border acquisitions and firm value: an analysis of emerging-market multinationals", *Journal of International Business Studies*, Vol. 40 No. 8, pp. 1317-1338.

Barnard, H. (2021), "Host countries' level of development and internationalization from emerging markets: a typology of firm strategies", *Journal of International Management*, Vol. 27 No. 3, 100828.

Bertrand, O. and Betschinger, M.A. (2012), "Performance of domestic and cross-border acquisitions: empirical evidence from Russian acquirers", *Journal of Comparative Economics*, Vol. 40 No. 3, pp. 413-437.

Bhagat, S., Malhotra, S. and Zhu, P.C. (2011), "Emerging country cross-border acquisitions: characteristics, acquirer returns and cross-sectional determinants", *Emerging Markets Review*, Elsevier B.V., Vol. 12 No. 3, pp. 250-271.

Black, E.L., Doukas, A.J., Xing, X. and Guo, J.M. (2015), "Gains to Chinese bidder firms: domestic vs foreign acquisitions", *European Financial Management*, Vol. 21 No. 5, pp. 905-935.

Brown, S.J. and Warner, J.B. (1980), "Measuring security price performance", *Journal of Financial Economics*, Vol. 8 No. 3, pp. 205-258.

Brown, S.J. and Warner, J.B. (1985), "Using daily stock returns. The case of event studies", *Journal of Financial Economics*, Vol. 14 No. 1, pp. 3-31.

Buckley, P.J. and Tian, X. (2017), "Internalization theory and the performance of emerging-market multinational enterprises", *International Business Review*, Elsevier, Vol. 26 No. 5, pp. 976-990.

Buckley, P.J., Munjal, S., Enderwick, P. and Forsans, N. (2016a), "Cross-border acquisitions by Indian multinationals: asset exploitation or asset augmentation", *International Business Review*, Vol. 25 No. 4, pp. 986-996.

Buckley, P.J., Munjal, S., Enderwick, P. and Forsans, N. (2016b), "The role of experiential and non-experiential knowledge in cross-border acquisitions: the case of Indian multinational enterprises", *Journal of World Business*, Vol. 51 No. 5, pp. 675-685.

Caiazza, R. and Volpe, T. (2015), "M&A Process: a literature review and research agenda", *Business Process Management*, Vol. 21 No. 1, pp. 205-220.

Chalencon, L. and Mayrhofer, U. (2018), "Do cross-border mergers-acquisitions in mature and emerging markets create similar value", *Journal of Organizational Change Management*, Vol. 31 No. 4, pp. 944-958.

Chan, C.M., Isobe, T. and Makino, S. (2008), "Which country matters? Institutional development and foreign affiliate performance", *Strategic Management Journal*, Vol. 29, March, pp. 1179-1205.

Chari, M.D. and Chang, K. (2009), "Determinants of the share of equity sought in cross-border acquisitions", *Journal of International Business Studies*, Vol. 40 No. 8, pp. 1277-1297.

Chari, A., Ouimet, P.P. and Tesar, L.L. (2010), "The value of control in emerging markets", *Review of Financial Studies*, Vol. 23 No. 4, pp. 1741-1770.

Chidambaran, N.K., Krishnakumar, D. and Sethi, M. (2017), "Cross-border vs Domestic acquisitions: evidence from India", *Journal of Economics and Business*, Elsevier, Vol. 95, February 2018, pp. 3-25.

Collins, D.W. and Dent, W.T. (1984), "A comparison of alternative testing methodologies used in capital market research", *Journal of Accounting Research*, Vol. 22 No. 1, pp. 48-84.

Conn, R.L., Cosh, A., Guest, P.M. and Hughes, A. (2005), "The impact on UK acquirers of domestic, cross-border, public and private acquisitions", *Journal of Business Finance and Accounting*, Vol. 32 Nos (5) and (6), pp. 815-870.

Contractor, F.J., Lahiri, S., Elango, B. and Kundu, S.K. (2014), "Institutional, cultural and industry related determinants of ownership choices in emerging market FDI acquisitions", *International Business Review*, Elsevier, Vol. 23 No. 5, pp. 931-941.

Corrado Charles, J. (1989), "A non-parametric test for abnormal security-price performance in event studies", *Journal of Financial Economics*, Vol. 23 No. 2, pp. 385-395.

Cowan, A.R. (1992), "Nonparametric event study tests", *Review of Quantitative Finance and Accounting*, Vol. 2 No. 4, pp. 343-358.

Danbolt, J. and Maciver, G. (2012), "Cross-border versus domestic acquisitions and the impact on shareholder wealth", *Journal of Business Finance and Accounting*, Vol. 39 Nos (7) and (8), pp. 1028-1067.

Dang, M., Henry, D., Nguyen, M.T. and Hoang, V.A. (2018), "Cross-country determinants of ownership choices in cross-border acquisitions: evidence from emerging markets", *Journal of Multinational Financial Management*, Elsevier B.V., Vol. 44, pp. 14-35.

Dikova, D., Panibratov, A. and Veselova, A. (2019), "Investment motives, ownership advantages and institutional distance: an examination of Russian cross-border acquisitions", *International Business Review*, Vol. 28 No. 4, pp. 625-637.

Ding, Y., Zhang, X. and Liu, Z. (2021), "Differences in returns to cross-border M&A in the short and long run: evidence from Chinese listed firms", *Journal of Asian Economics*, Elsevier, Vol. 74, p. 101299.

Du, M. and Kwabi, F. (2021), "State ownership, prior experience and performance: a comparative analysis of Chinese domestic and cross-border acquisitions", *International Journal of Accounting and Information Management*, Vol. 29 No. 3, pp. 472-491.

Dunning, J.H. (1988), "The eclectic paradigm of international production: a restatement and some possible extensions", *Journal of International Business Studies*, Vol. 19 No. 1, pp. 1-31.

Dutta, S., Saadi, S. and Zhu, P.C. (2013), "Does payment method matter in cross-border acquisitions?", *International Review of Economics and Finance*, Elsevier, Vol. 25, pp. 91-107.

Ermolaeva, L. (2019), "M&A deals completion and abandonment by Russian MNE", *International Journal of Emerging Markets*, Vol. 14 No. 3, pp. 475-494.

Fama, E.F., Fisher, L., Jensen, M.C. and Roll, R.W. (1969), "The adjustment of stock prices to new information", *International Economic Review*, Vol. 10, February, pp. 1-21.

Ferreira, S.J., Mohlamme, S., Van Vuuren, G. and Dickason (Koekemoer), Z. (2019), "The influence of corporate financial events on selected JSE-listed companies", *Cogent Economics and Finance*, Cogent, Vol. 7 No. 1, 1597665.

Figueira, S., Torres de Oliveira, R., Rottig, D. and Spigarelli, F. (2020), "Lessons on a novel integration approach of emerging market acquisitions in developed countries", *International Journal of Emerging Markets*, Vol. 16 No. 4, pp. 645-673.

Gaffney, N., Karst, R. and Clampit, J. (2016), "Emerging market MNE cross-border acquisition equity participation: the role of economic and knowledge distance", *International Business Review*, Elsevier, Vol. 25 No. 1, pp. 267-275.

Gregory, A. and O'Donohoe, S. (2014), "Do cross border and domestic acquisitions differ ? Evidence from the acquisition of UK targets", *International Review of Financial Analysis*, Vol. 31, pp. 61-69.

Guillen, M.F. and Garcia-Canal, E. (2009), "The American model of the multinational firm and the 'new' multinationals from emerging economies", *Academy of Management Perspectives*, Vol. 23 No. 2, pp. 23-35.

Haasis, T.I., Liefner, I. and Garg, R. (2018), "The organization of knowledge transfer in the context of Chinese cross-border acquisitions in developed economies", *Asian Business and Management*, Palgrave Macmillan, Vol. 17 No. 4, pp. 286-311.

Hope, O.K., Thomas, W. and Vyas, D. (2011), "The cost of pride: why do firms from developing countries bid higher", *Journal of International Business Studies*, Nature Publishing Group, Vol. 42 No. 1, pp. 128-151.

Jain, S., Kashiramka, S. and Jain, P.K. (2019), "Wealth effects on cross-border acquisition firms from emerging economies", *Emerging Markets Review*, Elsevier, Vol. 40, p. 100621.

Jain, S., Kashiramka, S. and Jain, P.K. (2021), "Determinants of stock market reaction to cross-border acquisitions: evidence from emerging economies", *International Journal of Emerging Markets*, Vol. ahead-of-print No. ahead-of-print.

Kim, H.G., Gaur, A.S. and Mukherjee, D. (2020), "Added cultural distance and ownership in cross-border acquisitions", *Cross Cultural and Strategic Management*, Vol. 27 No. 3, pp. 487-510.

Kim, B., Jung, J. and Cho, S. (2021), "Listing effect in acquirer returns and economic growth uncertainty in the target country: the case of cross-border M&A from emerging economies listing effect in acquirer returns and economic growth", *Emerging Markets Finance and Trade*, Routledge, Vol. 57 No. 2, pp. 427-443.

Kling, G. and Weitzel, U. (2011), "The internationalization of Chinese companies: firm characteristics, industry effects and corporate governance", *Research in International Business and Finance*, Vol. 25 No. 3, pp. 357-372.

Kohli, R. and Mann, B.J.S. (2012), "Analyzing determinants of value creation in domestic and cross border acquisitions in India", *International Business Review*, Elsevier, Vol. 21 No. 6, pp. 998-1016.

Kukreja, S., Maheshwari, G.C. and Singh, A. (2021), "Institutional distance in cross-border M&As: Indian evidence", in *Advances in Manufacturing and Industrial Engineering*, Springer, Singapore, pp. 877-888.

Lee, Y., Hemmert, M. and Kim, J. (2014), "What drives the international ownership strategies of Chinese firms? The role of distance and home-country institutional factors in outward acquisitions", *Asian Business and Management*, Nature Publishing Group, Vol. 13 No. 3, pp. 197-225.

Liang, Y., Giroud, A. and Rygh, A. (2021), "Emerging multinationals' strategic asset-seeking M&As: a systematic review", *International Journal of Emerging Markets*, Emerald Publishing, Vol. 16 No. 7, pp. 1348-1372.

Liao, R.C. (2014), "What drives corporate minority acquisitions around the world? The case for financial constraints", *Journal of Corporate Finance*, Elsevier B.V., Vol. 26, pp. 78-95.

Lien, Y.C. and Filatotchev, I. (2015), "Ownership characteristics as determinants of FDI location decisions in emerging economies", *Journal of World Business*, Elsevier, Vol. 50 No. 4, pp. 637-650.

Lin, X., Li, Y., Wan, X. and Wei, J. (2020), "Market reaction to the international acquisitions by Chinese firms: the role of potential intelligence sourcing and preannouncement", *Chinese Management Studies*, Vol. 14 No. 4, pp. 915-934.

Luo, Y. and Tung, R.L. (2007), "International expansion of emerging market enterprises: a springboard perspective", *Journal of International Business Studies*, Vol. 38 No. 4, pp. 481-498.

Luo, Y. and Tung, R.L. (2018), "A general theory of springboard MNEs", *Journal of International Business Studies*, Palgrave Macmillan, Vol. 49 No. 2, pp. 129-152.

Makino, S. and Beamish, P.W. (1998), "Performance ventures and survival of joint with ownership structures", *Journal of International Business Studies*, Vol. 29 No. 4, pp. 797-818.

Makino, S. and Delios, A. (1996), "Local knowledge transfer and performance: implications for alliance formation in Asia", *Journal of International Business Studies*, Vol. 27 No. 5, pp. 905-927.

Malhotra, S. and Morgan, H.M. (2018), "Sticky decisions: anchoring and equity stakes in international acquisitions", *Journal of Management*, Vol. 44 No. 8, pp. 3200-3230.

Mateev, M. (2016), "Do cross-border and domestic bidding firms perform differently ? New evidence from continental Europe and the UK", *Research in International Business and Finance*, Vol. 37, pp. 327-349.

Moeller, S.B. and Schlingemann, F.P. (2005), "Global diversification and bidder gains: a comparison between cross-border and domestic acquisitions", *Journal of Banking and Finance*, Vol. 29, pp. 533-564.

Narayan, P.C. and Thenmozhi, M. (2014), "Do cross-border acquisitions involving emerging market firms create value: impact of deal characteristics", *Management Decision*, Vol. 52 No. 8, pp. 1451-1473.

Nicholson, R.R. and Salaber, J. (2013), "The motives and performance of cross-border acquirers from emerging economies: comparison between Chinese and Indian firms", *International Business Review*, Vol. 22 No. 6, pp. 963-980.

Ning, L., Kuo, J.M., Strange, R. and Wang, B. (2014), "International investors' reactions to cross-border acquisitions by emerging market multinationals", *International Business Review*, Elsevier, Vol. 23 No. 4, pp. 811-823.

Opoku-mensah, E., Yin, Y., Sandra, A.A. and Tuffour, P. (2019), "Mergers and acquisitions antecedents in BRICS", *Global Journal of Emerging Market Economies*, Vol. 11 No. 3, pp. 202-214.

Ouimet, P.P. (2013), "What motivates minority acquisitions? The trade-offs between a partial equity stake and complete integration", *Review of Financial Studies*, Vol. 26 No. 4, pp. 1021-1047.

Papageorgiadis, N., McDonald, F., Wang, C. and Konara, P. (2020), "The characteristics of intellectual property rights regimes: how formal and informal institutions affect outward FDI location", *International Business Review*, Elsevier, Vol. 29 No. 1, p. 101620.

Phiri, A. (2018), "Robust analysis of convergence in per capita GDP in BRICS economies", *Munich Personal RePEc Archive*, p. 86936, available at: https://mpra.ub.uni-muenchen.de/86936/

Pinto, C.F., Ferreira, M.P., Falaster, C., Fleury, M.T.L. and Fleury, A. (2017), "Ownership in cross-border acquisitions and the role of government support", *Journal of World Business*, Elsevier, Vol. 52 No. 4, pp. 533-545.

Rani, N., Yadav, S.S. and Jain, P.K. (2014), "Impact of domestic and cross-border acquisitions on acquirer shareholders' wealth: empirical evidence from Indian corporate", *International Journal of Business and Management*, Vol. 9 No. 3, pp. 88-110.

Reddy, R.K., Fabian, F. and Park, S.-J. (2022), "Whether, how and why home country environments influence emerging market firm acquisition behavior", *International Journal of Emerging Markets*, Vol. ahead-of-print No. ahead-of-print.

Roll, R. (1986), "The hubris hypothesis of corporate takeovers", *The Journal of Business*, Vol. 59 No. 2, pp. 197-216.

Saikia, M. (2020), "Location determinants of Indian multinationals: a multilevel analysis", *Global Business Review*, Vol. 21 No. 5, pp. 1200-1217.

Scalera, V.G., Mukherjee, D. and Piscitello, L. (2020), "Ownership strategies in knowledge-intensive cross-border acquisitions: comparing Chinese and Indian MNEs", *Asia Pacific Journal of Management*, Vol. 37 No. 1, pp. 155-185.

Sehgal, S., Banerjee, S. and Deisting, F. (2012), "The impact of M&A announcement and financing strategy on stock returns: evidence from BRICKS markets", *International Journal of Economics and Finance*, Vol. 4 No. 11, pp. 76-90.

Tao, F., Liu, X., Gao, L. and Xia, E. (2017), "Do cross-border mergers and acquisitions increase short-term market performance? The case of Chinese firms", *International Business Review*, Vol. 26 No. 1, pp. 189-202.

Tunyi, A.A. and Ntim, C.G. (2016), "Location advantages, governance quality, stock market development and firm characteristics as antecedents of African M&As", *Journal of International Management*, Vol. 22 No. 2, pp. 147-167.

UNCTAD (2015), "World investment report 2015: reforming international investment governance".

Usman, M., Shabbir, R., Ahmad, I. and Zubair, A. (2021), "Host countries' institutional environment and multinational enterprises: does home-host developmental status matter?", *Journal of the Knowledge Economy*, Springer.

Wang, Y. and Larimo, J. (2020), "Survival of full versus partial acquisitions: the moderating role of firm's internationalization experience, cultural distance, and host country context characteristics", *International Business Review*, Elsevier, Vol. 29 No. 1, p. 101605.

Weber, Y., Tarba, S.Y. and Bachar, Z.R. (2011), "Mergers and acquisitions performance paradox: the mediating role of integration approach", *European Journal of International Management*, Vol. 5 No. 4, p. 373.

Williamson, O.E. (1981), "The economics of organization: the transaction cost approach", *The American Journal of Sociology*, Vol. 87 No. 3, pp. 548-577.

Yang, M. (2015), "Ownership participation of cross-border mergers and acquisitions by emerging market firms: antecedents and performance", *Management Decision*, Vol. 53 No. 1, pp. 221-246.

Yang, M. and Deng, P. (2017), "Cross-border M&As by Chinese companies in advanced countries: antecedents and implications", *Thunderbird International Business Review*, Vol. 59 No. 3, pp. 263-280.

Yang, Y., Yang, X., Chen, R.R. and Allen, J.P. (2014), "What drives emerging-economy outbound FDI decisions to obtain strategic assets?", *Asian Business and Management*, Nature Publishing Group, Vol. 13 No. 5, pp. 379-410.

Yang, Y., Xu, J., Allen, J.P. and Yang, X. (2022), "Strategic asset-seeking foreign direct investments by emerging market firms: the role of institutional distance", *International Journal of Emerging Markets*, Vol. ahead-of-print No. ahead-of-print.

Zhang, J. (2022), "Liability of emergingness and EMNEs' cross-border acquisition completion: a legitimacy perspective", *International Business Review*, Vol. 31 No. 2, p. 101951.

Zhang, Z., Lyles, M.A. and Wu, C. (2020), "The stock market performance of exploration-oriented and exploitation-oriented cross-border mergers and acquisitions: evidence from emerging market enterprises", *International Business Review*, Elsevier, Vol. 29 No. 4, p. 101707.

Zheng, Y. (2016), "Aggressive acquirers, laidback owners? Organisational dynamics of subsidiary integration in Chinese MNCs", *Asian Business and Management*, Palgrave Macmillan, Vol. 15 No. 4, pp. 317-342.

Zhu, P.C., Jog, V. and Otchere, I. (2014), "Idiosyncratic volatility and mergers and acquisitions in emerging markets", *Emerging Markets Review*, Elsevier B.V., Vol. 19, pp. 18-48.

Corresponding author
Sakshi Kukreja can be contacted at: kukrejasakshi09@gmail.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

arXiv:2211.12385v1 [cs.SI] 22 Nov 2022

MCD: A Modified Community Diversity Approach for Detecting Influential Nodes in Social Networks

Aaryan Gupta*
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
aryan227227@gmail.com

Inder Khatri*
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
inderkhatri999@gmail.com

Arjun Choudhry
Biometric Research Laboratory
Delhi Technological University
New Delhi, India
choudhry.arjun@gmail.com

Sanjay Kumar
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
sanjay.kumar@dtu.ac.in

November 23, 2022

Abstract

Over the last couple of decades, Social Networks have connected people on the web from across the globe and have become a crucial part of our daily life. These networks have also rapidly grown as platforms for propagating products, ideas, and opinions to target a wider audience. This calls for the need to find influential nodes in a network for a variety of reasons, including the curb of misinformation being spread across the networks, advertising products efficiently, finding prominent protein structures in biological networks, etc. In this paper, we propose Modified Community Diversity (MCD), a novel method for finding influential nodes in a network by exploiting community detection and a modified community diversity approach. We extend the concept of community diversity to a two-hop scenario. This helps us evaluate a node's possible influence over a network more accurately and also avoids the selection of seed nodes with an overlapping scope of influence. Experimental results verify that MCD outperforms various other state-of-the-art approaches on eight datasets cumulatively across three performance metrics.

Keywords Community Diversity · Community Structure · Influential nodes · Independent Cascade Model · Social Networks · Viral Marketing

1 Introduction

Nowadays, Social networks have become an integral part of our daily lives and serve as an excellent medium for people to interact and propagate their views and ideas. Due to the large number of users available on various social networks across the globe, they act as expedient channels for advertisements and promotion of the products (Mislove, Marcon, Gummadi, Druschel and Bhattacharjee, 2007; Garton, Haythornthwaite and Wellman, 1997; Kumar, Novak and Tomkins, 2006). Over the years, advertisements over social networks, also known as "Viral Marketing" have seen a significant boost due to their cost-effective nature and ability to propagate faster and wider. This approach, also known as "word-of-mouth advertising," exploits people's trust in a known or influential person's suggestions over other traditional means of advertisement (Ferguson, 2008; Zhou, Fan, Wang, Wang and Li, 2019). Further, social networks have been used for various other uses, including building collaborations, friendships, and other social or professional relations. Hence, the problem of finding nodes with wide influence over a network is a prominent area of research in

*Equal Contribution

network analysis. This is further required for various tasks like curbing rumors spread in social networks and identifying prominent proteins in protein graphs and other graph-based networks.

The problem of finding influential nodes in a network can be delineated as the task of identifying a set of k influential nodes in a network whose influence over the whole network is maximum across any set of k nodes, where k is a small positive integral constant (Kempe, Kleinberg and Tardos, 2003a; Liu, Qiu and Sun, 2022). Each individual entity acts as a node in the network, whereas the relationship between two nodes (friendship, follower-followee, co-authorship between users, links between proteins, etc.) helps form edges in the network to connect the nodes. For example, in social networks, the task entails maximizing the spread of an opinion by communicating it across the maximum number of nodes possible. Formally, the problem of finding the influential nodes in a network, modelled as a graph $G(V, E)$ where V is the set of all nodes in the network, and E is the set of edges connecting the nodes, can be defined as shortlisting a small number of nodes (k) such that the spread of influence can be maximized in the network. Mathematically, we can represent it as follows.

$$S = argmax\ \sigma(S)\ \ where\ S \in V,\ \|S\| = k, \|V\| \gg k \tag{1}$$

Here, argmax $sigma(S)$ denotes the overall influence spread generated by a set S of k number of nodes, and k is a small and positive constant with respect to the total number of nodes in the network. In general, selecting influential nodes in a network consists of two main components: choosing a set of k spreaders or seed nodes and then computing the spread of influence across the network originating from the seed nodes utilizing a suitable information propagation model. While selecting the set of seed nodes, we first ensure that the number of seed nodes is significantly lesser than the number of nodes in the network to use the resources available optimally. We also need to consider the influence or propagating capability of each node in the network based on its locality and links within the network. Owing to the presence of a community structuring system in social networks, involving community-dependent features can be used to evaluate better the network structure, as well as aid in accelerating the spread of the propaganda.

In this work, we propose the concept of Modified Community Diversity (MCD) for finding influential nodes in a network. Our approach extends the idea of the previously proposed Community-based Spreaders Ranking algorithm (CSR) (Kumar, Gupta and Khatri, 2022a) for two-hop neighborhood information. The proposed work intends to enable better utilization of available seed nodes for information spread. We further propose the use of Extended Community Diversity for gathering information about the neighborhood based upon a two-hop mechanism. This information is further used to avoid the overlapping influence of influential nodes in a network, thus improving the efficacy of finding influential nodes in a network and providing an optimized spread of information or influence. For further experimental analysis, we compare the performance of our proposed approach with other state-of-the-art approaches for influential node detection on eight datasets, seven of which are real-world datasets. We use the Independent Cascade (IC) model (Goldenberg, Libai and Muller, 2001) to simulate the influence propagation on all datasets. Our experimental evaluations over three performance metrics validate the superiority of our proposed approach over the other approaches across different domains and act as evidence of its efficacy and widespread applicability. The main contributions of our paper can be summarised as follows.

- We introduce Modified Community Diversity (MCD) method using community structure and extended community diversity to find influential nodes in social networks.
- The proposed work utilizes a two-hop concept which enables us to gather information about the neighborhood of a node better to judge its influence appropriately. This helps in avoiding the selection of influential nodes with overlapping influence.
- We verify the results of the proposed approach on eight network datasets against seven other state-of-the-art methods. The obtained results reveal that the proposed method outperforms different approaches cumulatively across three metrics. Statistical tests verify the superiority of MCD over competing approaches.

The breakdown of the remaining paper is as follows. Section 2 details the related works and the existing research gaps. Section 3 elaborates on our proposed methodology and algorithms. Section 4 consists of information about the various experimental details used in our paper, including the datasets used, the Information Propagation models, the performance metrics, and the implementation environment. Section 5 contains our experimental results, their analysis, and broad outcomes. Section 6 delineates the statistical tests we have done to verify the efficacy of MCD for influence maximization. Section 7 concludes our paper.

2 Related Works

Due to the advent of the internet and the rise in the global usage of social networks, the task of finding influential spreaders in a network as a research problem has seen significant contributions from researchers. Researchers have proposed various approaches to find the most influential nodes in a network, to maximize the spread of information in a

network (Li, Fan, Wang and Tan, 2018; Banerjee, Jenamani and Pratihar, 2020). These approaches can be predominantly categorized as diffusion process-based approaches, descriptive approaches, topological measures-based approaches, statistical and stochastic approaches, data mining-based approaches, and machine learning-based (Mendonça, da Motta Salles Barreto and Ziviani, 2022) approaches. Experimental evaluations have found topological measures-based approaches to be the most effective in identifying suitable spreader nodes (Kumar, Lohia, Pratap, Krishna and Panda, 2022b). Topological measures-based approaches utilize predefined centrality measures for ranking nodes based on their efficacy in spreading information across a network. These centrality measures aid in selecting suitable spreader nodes with high influence on the other nodes in the network based on their specific topological properties. Centrality measures-based approaches generally consist of two main steps: generating a score for each node based on a defined centrality measure and selecting the top k nodes based on the score obtained for the measure.

One of the first approaches using centrality scores for finding influential nodes in a network was proposed by Freeman (1978). The approach used the Degree Centrality measure, a local centrality measure that considers the number of neighbors of a node to generate its score. Degree Centrality has linear time complexity and was effective and efficient for small networks. However, due to its reliance on the local immediate neighbors for a given node, the effectiveness of Degree Centrality drops for nodes with few but influential neighbors. To counter the issues faced with local centrality measures, especially Degree Centrality, researchers proposed semi-local centrality measures like LC (Chen, Lü, Shang, Zhang and Zhou, 2012) and LSC (Gao, Ma, Chen, Wang and Changming, 2014), which go beyond the immediate scope of a node, but up to a limited range. While LC ranks the nodes using one-hop and two-hop nearest neighbors, LSC also considers the clustering coefficient between immediate neighbors.

These local centrality-based approaches do not consider a node's global location and only use the information corresponding to a certain locality, thus reducing their efficacy in real-life large networks. To counter this, researchers have also used global centrality measures like Betweenness Centrality, Closeness Centrality, and K-shell Decomposition, among others, to Influence Maximization. Freeman (1977) defined Betweenness for a specific node as the ratio of the shortest path which passes through the node and the total number of possible shortest paths in the network. Okamoto, Chen and Li (2008) proposed Closeness for a node as the inverse of the sum of the shortest path to all other nodes, thus determining the average closeness of a node to the rest of the graph. Betweenness and Closeness suffer from high time complexity, as they require the calculation of the shortest path between every node pair of the network.

Some other global centrality measures successfully used for Influence Maximization are Eigenvector Centrality (Bonacich, 2007) and Pagerank Centrality measure (Brin and Page, 1998). They aid in ranking nodes based on their importance among the neighborhood nodes. According to these measures, the presence of important nodes around the central node makes it a viable spreader node. Another global centrality-based measure, GLR (Salavati, Abdollahpouri and Manbari, 2019), divides the graph into communities and then selects each community's local cores and gateway nodes. The nodes are ranked as influential spreaders based on their shortest paths to the local cores and the gateway nodes. Researchers have also proposed greedy search-based approaches (Kempe, Kleinberg and Tardos, 2003b; Leskovec, Krause, Guestrin, Faloutsos, VanBriesen and Glance, 2007c) for ranking spreader nodes. These approaches use the sub-modularity and monotony of the Independent Cascade (IC) model along with Monte Carlo simulations to enhance the overall performance of the approach. However, greedy approaches are generally inefficient due to many simulations, and they require a lot of time to compute.

Some recent approaches also consider the bridge nodes, which connect two different groups of nodes in a network, as influential nodes. Liu, Xiong, Shi, Shi and Wang (2016) proposed the DIL centrality measure, where the importance of lines and the degree value are used to highlight a bridge node. DIL also considers the number of connected triangles when determining the relevance of a link, with the bridge node taking the primary role in controlling information propagation. Berahmand, Bouyer and Samadi (2019) introduced the DCL centrality measure, which finds the bridge nodes using a node's degree, clustering coefficient, and the relationship between its one-hop neighbors. Wen and Deng (2019) introduced LID, an approach where prominent influencers in complex networks are extracted using quasi-local information, i.e., local structural properties.

Recently, researchers have proposed various community structure-based approaches (Kumar, Panda and Aggarwal, 2021a; Chen, Zheng, Li, Liu, Zhao and Tang, 2022) for spreader ranking, which were found to be more effective in identifying suitable spreader nodes in a network (Huang, Shen, Meng, Chang and He, 2019; Kumar, Singhla, Jindal, Grover and Panda, 2021b; Chen, Deng, Zhao, Zhou and Zheng, 2021). Some recent community-based algorithms used for Influence Maximization are IM-ELPR (Kumar et al., 2021b), CFIN (Khomami, Rezvanian, Meybodi and Bagheri, 2021), and CAOM (He, Wang, Mao, Lv, Cai, Huang and Xu, 2020). CAOM divides the whole network into communities and selects only the most significant ones. Certain candidate nodes are then selected from each community using the one-hop measure. Hereafter, seed nodes are selected from candidate nodes by eliminating those with influence overlap. CFIN follows a similar approach, but the candidate nodes are selected using degree centrality, and the seed nodes are selected using the clustering coefficient measure. IM-ELPR uses node seeding, community detection, and

label propagation for influence maximization. The seed nodes are first identified using an extended h-index, and then the communities are detected using label propagation. The smaller, related communities are further merged into larger communities, and then the top k influential nodes are identified.

The approaches mentioned above utilized local structures to identify the bridge nodes. However, unlike these approaches, the recently proposed Community-based Spreaders Ranking algorithm (CSR) (Kumar et al., 2022a) for influence maximization selects global bridge nodes by considering the community bridge nodes as the most influential spreaders in the network. The CSR algorithm utilized three measures to identify bridge nodes, i.e., community diversity, community modularity, and community density. However, CSR didn't consider the overlapping influence of the influential nodes in a network. Our proposed approach (MCD) considers this factor and thus enables better utilization of available seed nodes. Further, CSR considers only one-hop neighbors to detect influential nodes. In contrast, MCD uses a two-hop mechanism, which provides a better overview of the node's neighborhood provided to the algorithm. This leads to markedly superior results by our approach as compared to CSR.

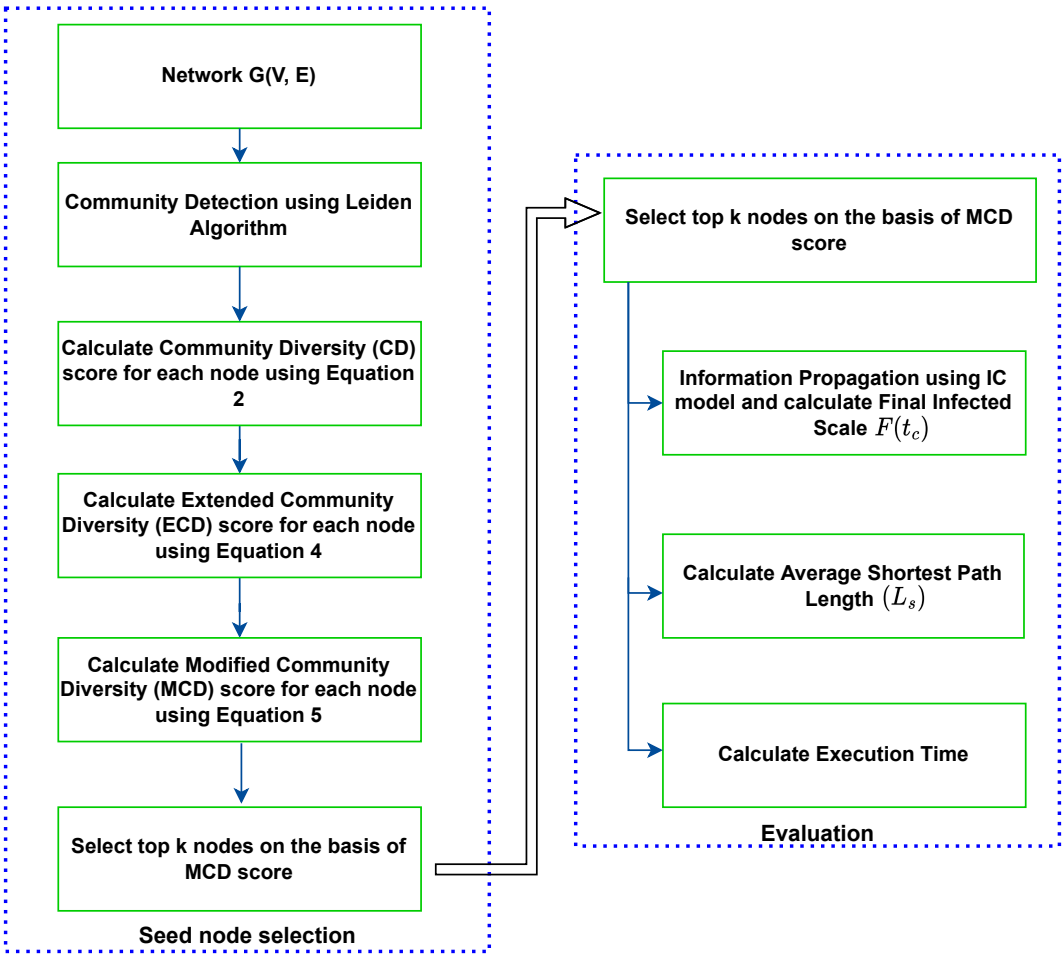


Figure 1: Flowchart representation of the steps involved for the calculation of MCD for each node.

3 Proposed Methodology

In this section, we present our proposed methodology for finding influential nodes in social networks using community detection and a modified community diversity approach. First, we uncover the community structure in the input network using Leiden's algorithm. We consider the bridge nodes in the network to play a vital role in spreading the information, and thus, we identify the bridge nodes in the network using the Community Diversity (CD) score. To incorporate the two-hop mechanism for gathering more information regarding a node's surroundings, we calculate the Extended Community Diversity (ECD) score. Then we calculate the Modified Community Diversity score using ECD and CD

to ensure the proper distribution of the information. We rank the nodes based on the MCD score and select the top k nodes, where k signifies the constraint on the number of nodes that act as influencers. Finally, we evaluate the influence of the selected k seed nodes as influential nodes over the entire network by using various evaluation parameters. The overall proposed work can be understood using Fig. 1, which depicts the process flow for calculating the MCD score for each node. The various phases of the proposed work are discussed below.

3.1 Community Detection

A community can be illustrated as a subset of a graph with nodes representing each member of the community (Yang, Cheng, Su, Zhang, Zhao and Chen, 2021; Chouchani and Abed, 2020; Ma, Chiew, Huang and He, 2017). The nodes have similar characteristics and are connected to some of the other nodes by edges. These nodes can represent any entity in the world, and their meaning is dependent on the scope of the community and, to a major extent, the scope of the graph. Nodes present specifically in a community are expected to have meaningful and substantial connections with each other in comparison to the rest of the graph, thus ensuring a better spread of information within the community as compared to outside the community. In our proposed algorithm, we use Leiden's algorithm (Traag, Waltman and van Eck, 2019) to identify the communities in a network. Leiden's algorithm is based on greedy optimization techniques, which results in efficient processing and quicker convergence. Leiden's algorithm consists of three steps:

- the local moving of the nodes.
- refining the partition between communities.
- accumulation of the nodes in the network on the basis of the refined partition, with the non-refined partition creating an initial partition for the accumulated network.

3.2 Community Diversity Score

It is often observed in real-world networks that influential nodes are usually linked to multiple communities in the network. Thus, they have a significant influence over the state of the network, as they can simultaneously disseminate their influence to multiple communities. We consider these high-degree nodes, which are connected to multiple communities, as influential nodes by applying the concept of Community Diversity (Kumar et al., 2022a), as they act as bridge nodes between communities. The community diversity score of a node can be formally defined as the measure of the probability of finding a certain community in the neighborhood of a node. Community diversity gives a high score to a node that is connected to multiple communities, as well as considers the number of communities to which the node transmits information. The concept of community diversity revolves around the concept of Shannon's entropy (Lin, 1991) over the term of the probability of finding the v^{th} node in the c^{th} community in the neighborhood of a certain node. The Community Diversity score for a node is computed as shown in Equation 2, where CDv_i is the community diversity value of the i^{th} node, and CDv_{ic} is the contribution of c^{th} community in Community Diversity of i^{th} node (Kumar et al., 2022a). The neighborhood contribution of each community is shown in Equation 3, where η_{in} represents the number of nodes in the c^{th} community in the neighborhood of node i, and η_i represents the number of nodes in the neighborhood of node i.

$$CD_i = -\sum_{c=1}^{comm} CDv_{ic} = -\sum_{c=1}^{comm} P_{ic} * log(P_{ic}) \tag{2}$$

$$P_{ic} = \frac{\eta_{ic}}{\eta_i} \tag{3}$$

Algorithm 1 represents the procedure used to compute the Community Diversity (CD) score for a graph G. It can be summarized as follows:

- We divide the graph into communities using Leiden's algorithm.
- For each node, we calculate the number of neighbours. Further, for each community, the number of neighbours of a node that are part of the respective community is computed.
- We evaluate the ratio of the total number of neighbours of the node and the neighbour of the node in the respective community. This is further used to calculate the entropy score, i.e., the Community Diversity score.

If a node is connected to multiple communities and the connections are found to be uniformly distributed, then the community diversity score of that node is observed to be higher. However, if a node has connections to only a few

Algorithm 1 : Community Diversity Score (Kumar et al., 2022a)

Input: $G = (V, E)$, where $n = \|V\|, e = \|E\|$
Output: $Community_Diversity$

$Initialization$
$Communities$ {c1, c2....cm} : $partitioned\ using\ Leiden's\ Algorithm$
$value \leftarrow 0$
$CD \leftarrow \{\}$
for $node \in \{Nodes\ of\ Graph\ G\}$ **do**
    $sum \leftarrow 0$
    $neigh_len \leftarrow number\ of\ neighbours\ of\ node$
    for $c \in$ {Communities} **do**
        $neigh_com_len \leftarrow number\ of\ neighbours\ of\ node\ belonging\ to\ c$
        $ratio \leftarrow \frac{neigh_com_len}{neigh_len}$
        $value \leftarrow -ratio * log(ratio)$
        $sum \leftarrow sum + value$
    end for
    $CD(node) \leftarrow sum$
end for
return CD

communities or the distribution of connections is not uniform, then the community diversity score is observed to be low. Hence, the concept of community diversity can be utilized to find out the nodes capable of efficiently spreading information to multiple parts of the network.

3.3 Extended Community Diversity Score

We formally propose the Extended Community Diversity score as the ratio of the community diversity score of a node to that of its neighborhood. The concept of Extended Community Diversity revolves around considering a node along with its neighborhood for the prediction of its influence in a locality. The neighbors of a node contribute to helping the algorithm understand what relative position a node holds with respect to its neighborhood, which ultimately tells how much a node will benefit from being made the seed node. Extended Community Diversity gains information about the neighborhood of a node by taking two hops, i.e., one hop from the node to its neighbors, and the second hop from its neighbors to farther nodes. This provides more information for the proposed algorithm to process and deduces which node is most influential. The provided information is further used to avoid having many influential nodes in the same region and hence avoid the wastage of useful resources. Extended Community Diversity score for a node is computed as shown in Equation 4 where ECD_v is the Extended Community diversity value of v^{th} node, $neigh$ is the neighbours of node v and CD_v is the community diversity score for node v as computed in Equation 2.

$$ECD_v = CD_v + \sum_{n=1}^{neigh} CD_n \tag{4}$$

Algorithm 2 represents the algorithmic procedure used to compute the Extended Community Diversity (ECD) score for graph G. The summary of the algorithm is as follows:

- We calculate the previously defined Community Diversity score (CD) in Algorithm 1.
- Then, for each node, we calculate the summation of its Community Diversity score, and its neighbor's Community Diversity Score, which gives the Extended Community Diversity score.

3.4 Modified Community Diversity Score

We formally define the Modified Community Diversity (MCD) score as $MCD = -P * logP$, where P is the ratio of the Community Diversity of a node to its Extended Community Diversity. Modified Community Diversity uses the limited available resources judiciously by ensuring that no two influential nodes are in the same locality, which, if not taken care of, would result in the wastage of valuable spreading power and overlapping of the region of influence of influential nodes in the same locality. The problem of several influential nodes existing in the same neighborhood is known as the Rich Club effect. Our proposed approach tackles this problem by scoring the nodes based on their strength

Algorithm 2 : Extended Community Diversity Score

Input: $G = (V, E)$, where $n = \|V\|,\ e = \|E\|$
Output: $Extended_Community_Diversity$

$Initialization$
$ECD \leftarrow \{\}$
$CD \leftarrow Community\ Diversity\ Score\ using\ algorithm\ 1$
$value \leftarrow 0$
for $v \in \{Nodes\ of\ Graph\ G\}$ **do**
    $value \leftarrow CD(v)$
    for $nb \in \{Neighbours\ of\ node\ v\}$ **do**
        $value \leftarrow value + CD(nb)$
    end for
    $ECD(v) \leftarrow value$
end for
return ECD

relative to their neighboring nodes, thus ensuring that the influential nodes are well spread throughout the network. The ratio of the results of the previously calculated CD score and ECD score is used for analyzing the MCD score. The concept of MCD maximizes when the distribution of spreading power is uniform, which makes sense here since the algorithm not only wants the respective node to be influential but also the neighborhood to be sufficiently influential in carrying on the spread of information further. This enables the node to maximize the spread of information in that locality and beyond it. This results in the respective node being somewhat equivalent to the whole neighborhood in terms of spreading power, thus making it the most influential node in the locality. This mitigates the problem of the rich club effect significantly by considering the overlap in the influence and handling it in a balanced way to avoid wastage of useful spreading strength. This results in an optimized spread of information with assured maximum throughput. Modified Community diversity score for a node is computed as shown in Equation 6, where P_v is as shown in Equation 6. In Equation 6, CD_v is the community diversity score of v^{th} node, as shown in Equation 2, and ECD_v is the Extended Community diversity value of v^{th} node, as calculated in Equation 4.

$$MCD_v = -P_v * log(P_v) \tag{5}$$

$$P_v = \frac{CD_v}{ECD_v} \tag{6}$$

Algorithm 3 : Modified Community Diversity Score

Input: $G = (V, E)$, where $n = \|V\|, e = \|E\|$
Output: $Modified_Community_Diversity$

$Initialization$
CD ← Community Diversity Score using algorithm 1
ECD ← Extended Community Diversity Score using algorithm 2
$MCD \leftarrow \{\}$
$value \leftarrow 0$
$p \leftarrow 0$
for $v \in \{Nodes\ of\ Graph\ G\}$ **do**
    $p \leftarrow \frac{CD(v)}{ECD(v)}$
    $value \leftarrow -p * log(p)$
    $MCD(v) \leftarrow value$
end for
return MCD

Algorithm 3 represents the procedure used to calculate the Modified Community Diversity (MCD) score for graph G. The summary of the algorithm is as follows:

- The community diversity score (CD) is calculated using Algorithm 1, and the extended community diversity score (ECD) is calculated using Algorithm 2.

- Then, for each node, we calculate the ratio of the Community Diversity score and Extended Community Diversity Score, which gives us the MCD score for the respective node.

Figure 1 illustrates the process flow for calculating the MCD score for each node. We first calculate the CD and ECD scores for each node. Using the generated values of CD and ECD scores, we compute the MCD scores for each node. We further rank the nodes on the basis of the MCD score and select the top k nodes, where k signifies the constraint on the number of nodes that act as influencers. We further evaluate the influence of the selected k nodes over the entire network in Section 5.

4 Datasets, Information Propagation Model, & Performance Metrics

4.1 Dataset and Baseline Models

We evaluate our proposed approach on seven real-life datasets and one synthetic dataset of varying sizes and types. Table 1 describes the different datasets used during our evaluation. Further, we do a comparative analysis of the proposed approach with seven other state-of-the-art approaches, namely CSR (Kumar et al., 2022a), GLR (Salavati et al., 2019), DCL (Berahmand et al., 2019), LID (Wen and Deng, 2019), DIL (Liu et al., 2016), H-Index (HI) (Lü, Zhou, Zhang and Stanley, 2016) and PageRank (PR) (Brin and Page, 1998), over three distinct evaluation criteria. For the optimal performance for the baseline approaches, we took the input parameters as reported in the original papers throughout all our experiments.

- Community Spreaders Ranking (CSR) (Kumar et al., 2022a): CSR measures a node's spreading ability with respect to different communities on the basis of its connection to the communities.
- Gateway Local Rank (GLR) (Salavati et al., 2019): GLR extends over the closeness centrality measure and simplifies it by reducing the search set to the local and gateway nodes. To get the optimal performance, we take parameters α_1 and α_2 used for weighing the effect of core and gateway nodes to be 1 as reported by (Salavati et al., 2019).
- Degree and Clustering Coefficient and Location (DCL) (Berahmand et al., 2019): DCL measures the spreading ability of nodes on the basis of a node's degree, its neighbour, and the clustering coefficient.
- Local Information Dimensionality (LID) (Wen and Deng, 2019): LID measures the spreading ability of a node by considering the quasilocal structure of a node.
- Degree and Importance of Lines (DIL) (Liu et al., 2016): DIL ranks the nodes on the basis of the degree and the importance of lines.
- H-index (HI) (Lü et al., 2016): H-index is a node centrality measure that considers the neighborhood measures, the quality, as well as the quantity of the neighboring nodes simultaneously.
- PageRank (PR) (Brin and Page, 1998): PageRank measures the significance of a given node in the network based on its outgoing degree and damping factor. Similar to the work Brin and Page (1998) and various other works, we take the value of d to be 0.85 in all experiments to get optimal performance.

Table 1: The different datasets used for experimentation along with the various characteristics such as Number of Nodes, Edges, and Communities by Leiden Algorithm

S.No.	Dataset	Nodes (n)	Edges (m)	Communities
1	BA	2,000	9,974	14
2	PGP	10,638	24,301	104
3	Jazz Musicians	198	2,742	4
4	Ca-hepth	7,126	35,324	481
5	GR-QC	5,242	14,496	394
6	Dolphins	62	159	5
7	Erdos992	6,100	7,500	50
8	p2p-Gnutella25	22,687	54,705	59

A brief description of different datasets is given below:

- BA (Barabási and Albert, 1999): BA is a random graph generated using Barabási-Albert preferential attachment model. A network of *n* nodes is generated by adding new nodes, each having *m* edges that are preferentially coupled to existing nodes with a high degree.

- PGP (Boguñá, Pastor-Satorras, Diaz-Guilera and Arenas, 2004): Pretty Good Privacy (PGP) is an encrypted communication network. PGP algorithm is used for secure information interchange. Here, the users are nodes, and the edges of this huge network are the connectivity between users.
- Jazz Musicians (GLEISER and DANON, 2003): The Jazz musicians dataset is a real-world social network dataset, where each node represents a musician, and each edge between two nodes indicates that the two musicians have played together in a band.
- Ca-hepth (Leskovec, Kleinberg and Faloutsos, 2007a); It represents a collaboration network of authors who have collaborated on papers submitted to the High Energy Physics - Theory Category.
- General Relativity and Quantum Cosmology (GR-QC) (Leskovec, Kleinberg and Faloutsos, 2007b): GR-QC dataset is a real-world social network dataset from arXiv and covers scientific contributions by authors in General Relativity and Quantum Cosmology. The authors are nodes, and if they have collaborated with another author on a paper, then there is an undirected edge between them.
- Dolphins dataset (Lusseau, Schneider, Boisseau, Haase, Slooten and Dawson, 2003): The dolphins dataset is a social network of bottlenose dolphins, where a node represents a dolphin, while an edge represents frequent associations between dolphins.
- Erdos992 (Batagelj and Mrvar, 2000): This dataset is about the collaborations of Paul Erdős, a Hungarian Mathematician, with his co-authors. The co-authors are nodes while collaborations between the nodes signify the edges.
- p2p-Gnutella25 (Ripeanu, Foster and Iamnitchi, 2002): This dataset is one of the snapshots of the Gnutella peer-to-peer file sharing network from August 2002. Nodes represent hosts in the Gnutella network, while edges represent connections between the Gnutella hosts.

4.2 Information Propagation Model

Information Propagation (IP) is described as the task of circulating information in complex networks (Chen, Lakshmanan and Castillo, 2013; Guille, Hacid, Favre and Zighed, 2013). IP can be used to assess various spreader-ranking approaches. Generally, epidemic models are used to model information propagation in a complex network. In epidemic models, the population of users is represented as nodes, and the requisite information is regarded as an infectious disease, which propagates through the population upon the interaction between nodes.

While various epidemic models can be used for information propagation tasks, like SIR (Hethcote, 2000), IC (Goldenberg et al., 2001) and LT (Granovetter, 1978), we employ the Independent Cascade (IC) model to assess the performance of our proposed approach and evaluate it against other state-of-the-art approaches. The nodes in an IC model can be allocated one out of two states: infected node or susceptible node. The propagation of information in the network begins with setting the state of a few nodes as infected. These nodes further interact with their neighboring nodes and try to infect them. Further, the newly infected nodes further interact with their neighbors to further propagate the information to their neighbouring nodes until there are no new nodes in the network that are infected. A probability (P) is associated with these interactions between nodes, which elucidates the possibility of a susceptible node getting infected upon interaction with an infected node. The total influence of a seed node is calculated by taking the sum of all nodes which get infected by the end of information propagation.

4.3 Performance Metrics

We use three performance metrics for comparing our approach with other state-of-the-art approaches. They are as follows:

(i) Final Infected Scale: It is defined as the ratio of activated or infected nodes in the network to the total number of nodes in the network at the end of information propagation. The nodes only get activated when the activation probability for a node is favourable for activation. The fraction of the total nodes taken as the initial spreaders is called the spreader fraction, and it plays a significant role in the final number of nodes that get infected. Generally, a high spreader fraction leads to a high percentage of the total nodes getting infected.

(ii) Average Shortest Path Length (ASPL): ASPL can be defined as the average of the shortest paths between two influential nodes for all sets of influential nodes. It enables us to evaluate the efficacy of information propagation in a complex social network. The higher the value of ASPL, the more even the spread of the influential nodes in the network. A high value of ASPL further indicates that influential nodes cover a wider part of the network instead of being concentrated in a small region, leading to the poor spread of information in the network.

(iii) Execution Time: Execution Time: We evaluate our approach against other approaches based on the time taken to rank nodes based on their ability to propagate information in a network. The time consumption for all centralities was assessed in a Google Colaboratory environment. The code was run on an Intel Xeon CPU with one core, with 13GB of RAM available for use.

5 Experimental Results and Analysis

This section presents the experimental results and analysis performed on the eight datasets using three performance metrics as mentioned in Section 4. We conducted a comparative analysis of the proposed approach with seven other state-of-the-art approaches for finding influential nodes, namely, H-Index (HI) (Lü et al., 2016), PageRank (PR) (Brin and Page, 1998), Community-based Spreader Ranking (CSR) (Kumar et al., 2022a), Gateway Local Ranking (GLR) (Salavati et al., 2019), Degree and Clustering Coefficient and Location (DCL) (Berahmand et al., 2019), Local Information Dimensionality (LID) (Wen and Deng, 2019), and Degree and Importance of Lines (DIL) (Liu et al., 2016). We utilized the Information Cascade (IC) model to simulate information propagation. Our analysis of the results obtained over three metrics, illustrated in Figures 2, 3, and 4, is elaborated below.

5.1 Final Infected State

The Final Infected Scale illustrates the final degree of propagation or *infection* in the network at the end of the simulation, for varying values of fraction of nodes taken as initial spreaders. We evaluated our proposed approach against other competing approaches using the IC simulation model for 100 test runs and plotted the average results in Figure 2. For small datasets, i.e., datasets with the number of nodes less than 2000, we take the values of the initial spreader fraction from the set 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, while for larger datasets with the number of nodes greater than 2000 nodes, we take the values of the initial spreader fraction from the set 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04. We have taken larger values of the initial spreader fraction for smaller datasets, as the absolute number of initial spreaders will be small for smaller values of the initial spreader fraction in smaller datasets. We have taken the *infection probability*, i.e., the probability of influencing a susceptible node as 0.1 for all combinations of datasets and approaches.

The graphs for Final Infected Scale with variations in Spreader Fraction for all approaches and datasets are illustrated in Figure 2. We observed that the final infected scale generally increased with an increase in spreader fraction, irrespective of the dataset. We observed that our proposed approach, MCD, outperformed competing approaches significantly across all datasets, except the Dolphins dataset, where PR outperforms our approach. We found DCL and PR to perform reasonably well across the majority of the datasets too. HI was consistently among the worst-performing approaches of the lot for most datasets. Thus, we conclude that our approach handily outperforms other approaches in maximizing information propagation in a network for a wide set of values of the initial spreader fraction.

5.2 Average Distance between Spreaders

It is desirable for the initial spreaders to have the least overlapping area of influence so that a wider portion of the network can be activated more efficiently. Thus, it is preferable that the average distance between spreaders is maximized. To compare our proposed MCD approach against other approaches, we calculated the average shortest distance between the initial spreaders (L_s) for a wide set of values of the spreader fraction.

The graphs for the Average Distance between Spreaders vs. Initial Spreader Fraction for all approaches and datasets are illustrated in Figure 3. We observed that MCD significantly outperforms other approaches across datasets for almost all values of the initial spreader fraction. This ensures that MCD selects nodes that are far apart, thus avoiding overlapping of influence as stated in the manuscript. It verifies that MCD is able to cater to the rich club effect, which other competing approaches are unable to counter. The significant gap in the figure between our approach and competing approaches indicates that MCD selects nodes that are different from the ones selected by other approaches, and significantly farther apart. This further ensures nodes selected by MCD may not belong to the same region, thus having a wider spread of influence and lesser overlap between the regions of influence for each seed node. Further, we observe that on increasing the spreaders fraction, the difference between seed nodes decreases due to the increased absolute number of seed nodes for the whole network. This is observed in the case of all approaches, as increasing the absolute number of spreaders in a graph is bound to decrease the average distance between the spreaders. However, MCD still has a higher average spreader distance as compared to other approaches, even for greater values of Spreader Fraction, verifying MCD's superiority over competing approaches. Among the other approaches, we found DCL and PR to perform reasonably well. However, there was still a significant gap between these approaches and MCD. We

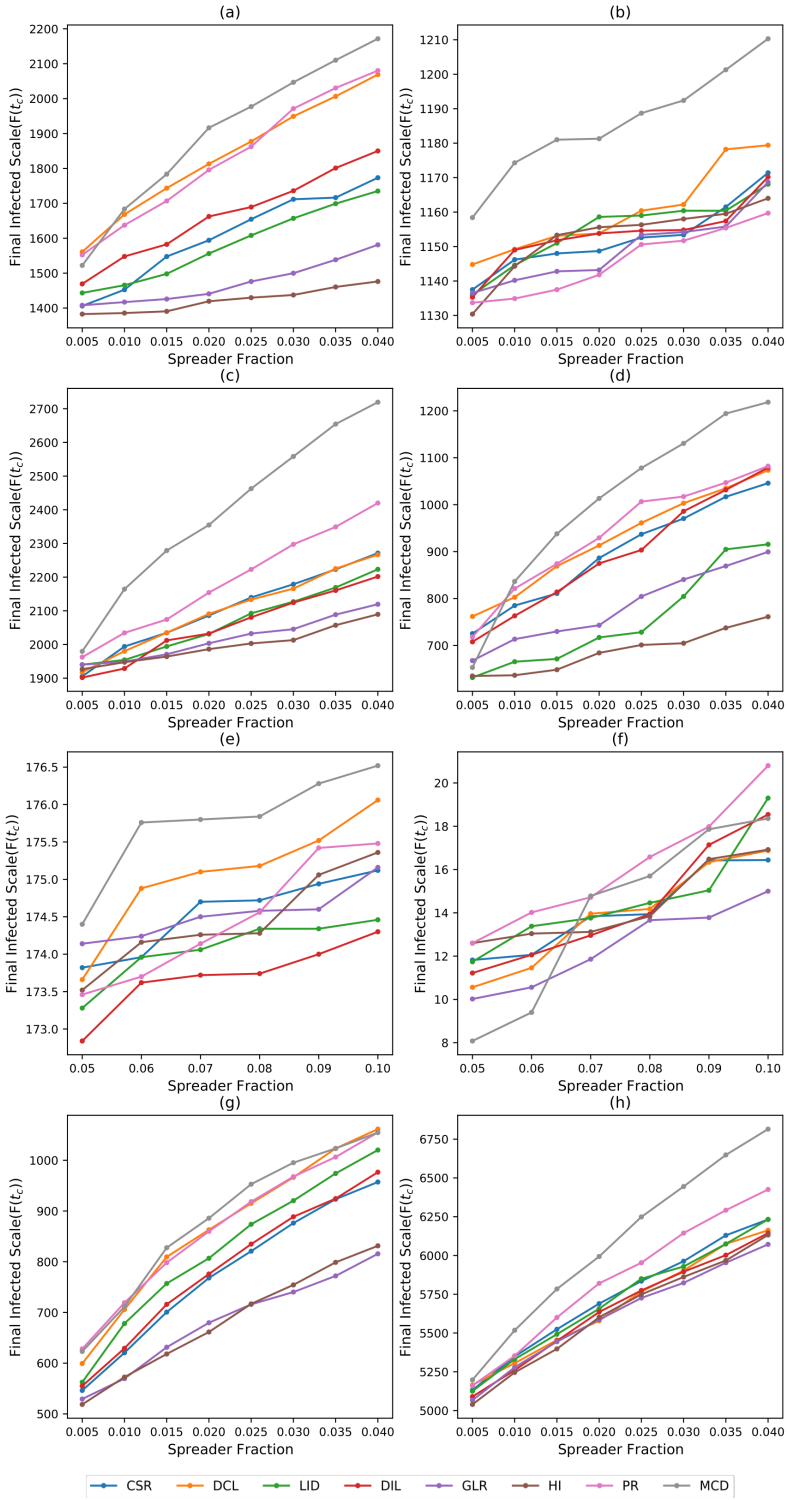


Figure 2: Final infected scale plots for various values for inital spreader fraction on the (a) PGP dataset (b) BA dataset (c) Ca-hepth dataset (d) GR-QC dataset (e) Jazz Musicians dataset (f) Dolphins dataset (g) Erdos992 dataset (h) p2p-Gnutella25 dataset. The results are averaged for 100 independent simulations of the Information Cascade model with an activation probability (P) equal to 0.1.

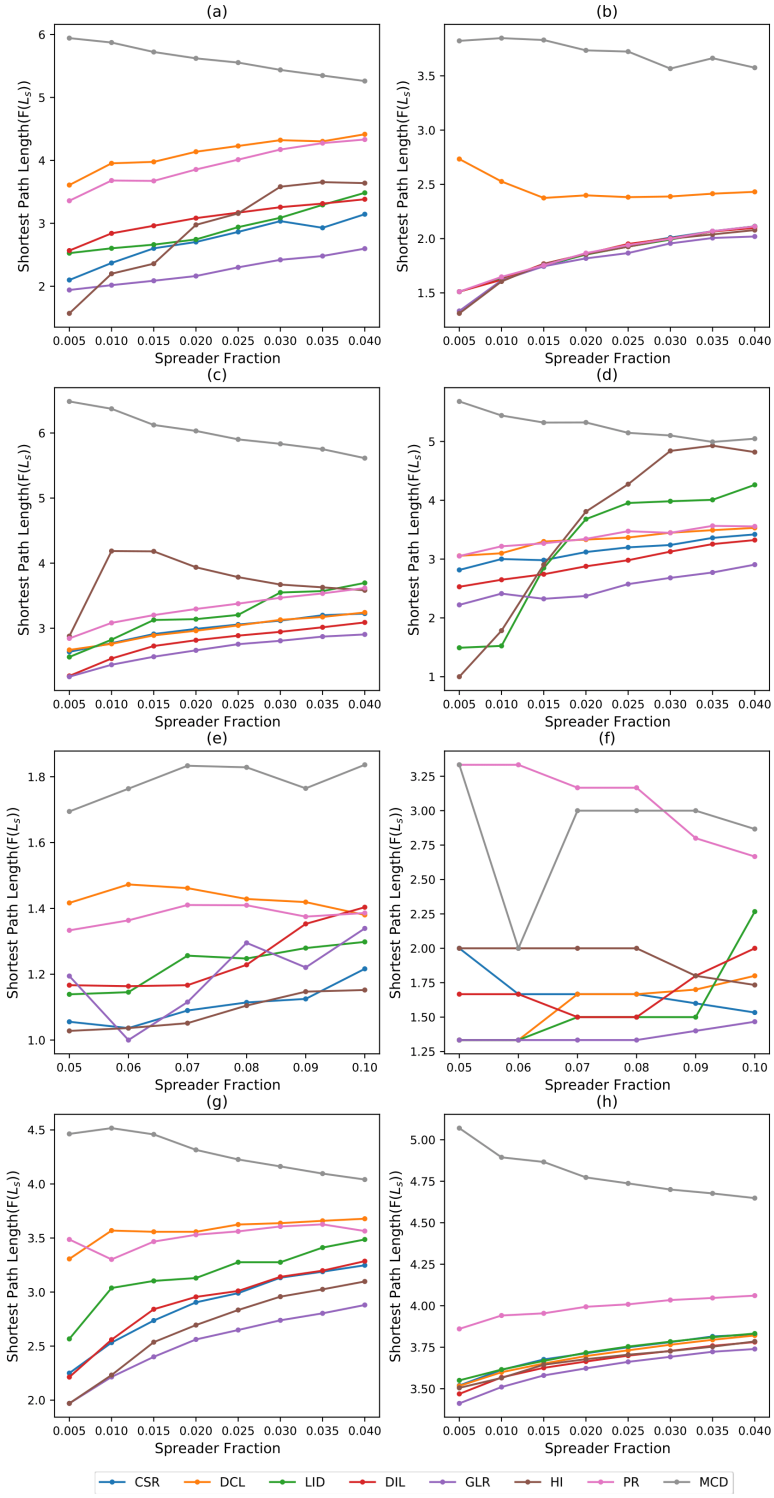


Figure 3: Average distance between spreaders with respect to different initial spreaders fraction values for (a) PGP dataset (b) BA dataset (c) Ca-hepth dataset (d) GR-QC dataset (e) Jazz Musicians dataset (f) Dolphins dataset (g) Erdos992 dataset (h) p2p-Gnutella25 dataset. The results are averaged for 100 independent simulations of the Information Cascade model with an activation probability (P) equal to 0.1.

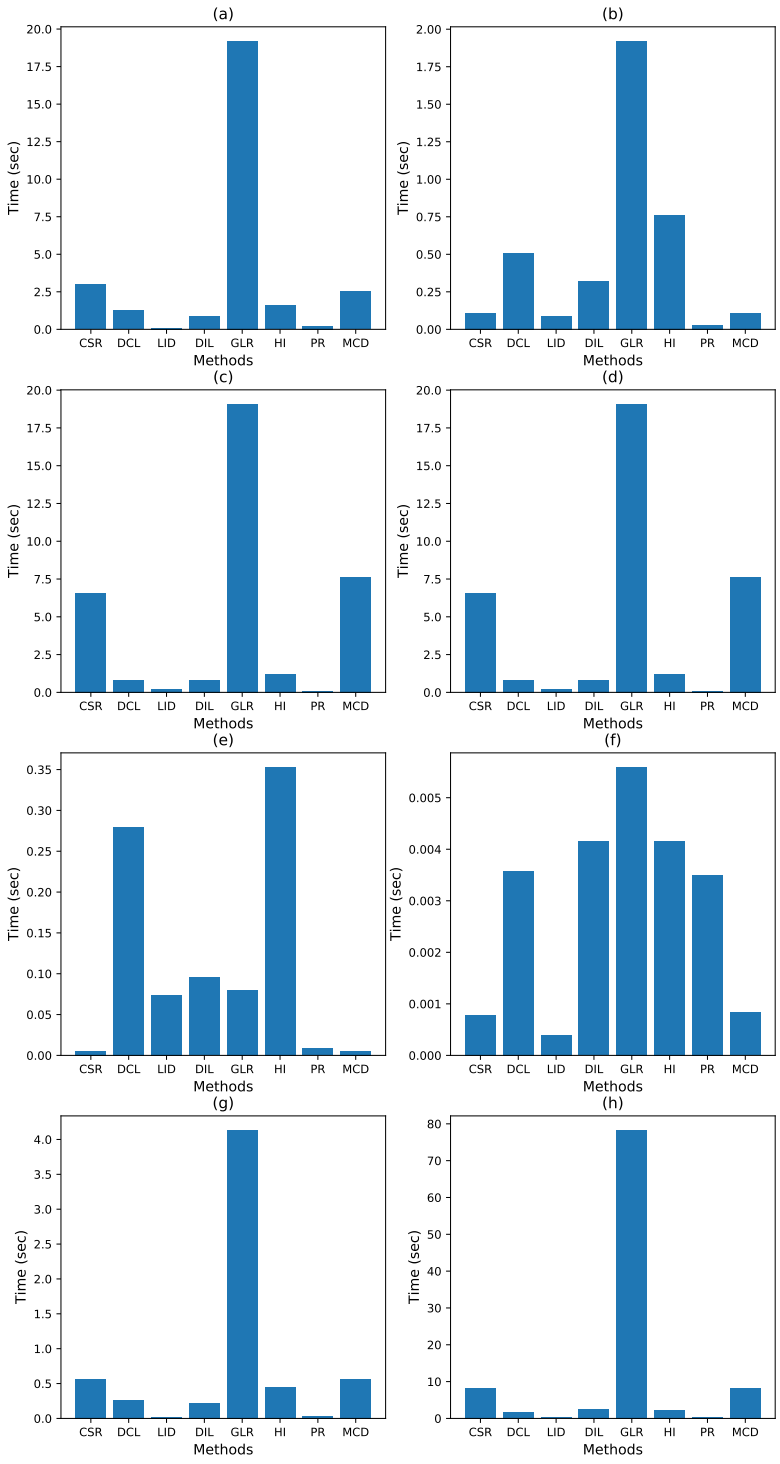


Figure 4: The execution time plots for different methods for ranking the spreaders for (a) PGP dataset (b) BA dataset (c) Ca-hepth dataset (d) GR-QC dataset (e) Jazz Musicians dataset (f) Dolphins dataset (g) Erdos992 dataset (h) p2p-Gnutella25 dataset.

found GLR consistently performs the worst in this metric across almost all datasets. Thus, we conclude that MCD suffers from a significantly lesser *rich club* effect and outperforms all other approaches significantly.

5.3 Execution Time

We compared our proposed approach, MCD, against other competing approaches based on the execution time required by each approach for ranking the nodes for all datasets. Figure 4 illustrates the line graph comparing MCD with other approaches based on the execution time for all datasets. DIL performed well on larger datasets but was relatively more computationally expensive for smaller datasets. We found LID and PR to be the least expensive computationally, while GLR was consistently the most expensive. Our proposed approach was computationally the least expensive for smaller datasets, while it was more expensive than local hybrid centrality-based approaches like LID for larger datasets. Thus, we observe that MCD performs well in terms of execution time when compared to other approaches, especially when considering the improved performance in other metrics.

6 Statistical Testing

To further evaluate the performance of MCD using statistical tests, we have conducted the Friedman test (Friedman, 1937), which is a non-parametric statistical test used for multiple comparisons (more than two given methods). We have performed this test to determine whether the performance of our proposed method is significantly different than the other approaches for two or more data sets. The Friedman test attempts to detect significant differences based on the ranking of the methods, rather than their errors. It consists of two hypotheses: the null (H_0) and the alternate hypothesis (H_1). The former states that there are no prominent differences between these algorithms (equality of medians) and the latter insists that the algorithms have significant differences in their median of the population, thus negating the null hypothesis. We have performed the following test on the Final Infected Scale metric as well as the Average Distance between Spreaders metric.

In order to perform the Friedman Statistical Test, we execute the following procedure:

- Gather the generated results for each problem pair.
- Rank the values in ascending order from 1 (best value) to n (worst value) for each problem i for a particular algorithm j.
- For algorithm j, the average rank for each problem i is calculated using the Eq. 7:

$$R_j = \frac{1}{n}\sum_{i=1}^{n}(r_i^j) \tag{7}$$

  where r^j is the rank ($1 < j < k$) and R_j is the average rank.
- Now that all algorithms are ranked according to their priority, compute the Friedman Statistic F_f using the following equation. F_f is based on a Chi-square distribution, with $k-1$ degree of freedom. F_f is computed as shown in Eq. 8.

$$F_f = \frac{12n}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{8}$$

- The previous statistic produced relatively conservative results, which were not desired. Thus, (Iman and Davenport, 1980) introduced a new statistic F_{id}, as shown below, which follows the $F-distribution$ with the degree of freedom as $k-1$ and $(n-1)(k-1)$, given in Eq. 9:

$$F_{id} = \frac{(n-1)\chi_F^2}{n(k-1)-\chi_F^2} \tag{9}$$

Table 2 and Table 3 show the average ranking calculated from the Friedman test for all the approaches used in this paper. MCD clearly has a better average ranking than the other competing approaches.

The unadjusted P-value or Holm P-value, obtained from the Iman-Davenport statistic, corresponding to the performance of MCD, advocates the rejection of the null hypothesis H_0. All of the computed P-values are less than the standard significance level $\alpha = 0.05$. This indicates that there is a significant difference between the performance of the baseline approaches and our proposed approach. Hence, the obtained P-values conclude the negation of null hypothesis H_0 but are not appropriate for comparison with different methods. In order to compare these approaches with each other, we must calculate their respective adjusted P-values keeping MCD as the control algorithm.

Hence, we use adjusted P-values (APVs) for evaluation in order to perform a comparative study between the control method and benchmark methods on a statistical ground. Adjusted P-value provides the correct correlation between these

Table 2: Average ranking of algorithms using the Friedman test based on Final Infected State

	Method	Average Rank
1	MCD	1.567
2	PR	2.983
3	DCL	3.467
4	CSR	4.675
5	LID	5.033
6	DIL	5.225
7	GLR	6.517
8	HI	6.533

Table 3: Average ranking of algorithms using the Friedman test based on Average Distance between Spreaders

	Approach	Average Rank
1	MCD	1.067
2	PR	2.983
3	DCL	3.517
4	LID	4.675
5	HI	5.3
6	CSR	5.367
7	DIL	5.508
8	GLR	7.583

algorithms, taking into account the accumulated family error with respect to the MCD control algorithm. Moreover, the adjusted P-values can be directed compared with the significance level α, which is equivalent to 0.05. In order to calculate the adjusted P-values, a few post-hoc procedures need to be defined. Various post-hoc procedures such as (Dunn, 1961) and (Holland and Copenhaver, 1987) differ as they adjust the value of α to compensate for multiple comparisons for multiple methods. In this paper, we have used the common Holm's procedure (Holm, 1979) to evaluate the respective APVs. The values of Holm P-values and APVs are always sorted in ascending order. The equation for the same is given below, where indices i and j refers to the main hypothesis whose APVs are being computed and different hypothesis in the set respectively. P_j is the P-value for the j^{th} hypothesis. Holm APV is computed as shown in Eq. 10.

$$HolmAPV_i = \min\{v, 1\}, \text{where } v = \max\{(k-j)p_j : 1 \leq j \leq i\} \tag{10}$$

Table 4: Adjusted P-values (APVs) using Holm procedure for Final Infected State

	Approach	Z-score	P-value	APV
1	HI	-11.11	5.88E-29	4.11E-28
2	GLR	-11.06	8.91E-29	5.35E-28
3	DIL	-8.18	1.42E-16	7.08E-16
4	LID	-7.75	4.53E-15	1.81E-14
5	CSR	-6.95	1.82E-12	5.46E-12
6	DCL	-4.24	1.08E-05	2.15E-05
7	PR	-3.16	7.68E-04	7.68E-04

Table 5: Adjusted P-values (APVs) using Holm procedure for Average Distance between Spreaders

	Approach	Z-score	P-value	APV
1	GLR	-14.57	2.13E-48	1.49E-47
2	DIL	-9.93	1.51E-23	9.08E-23
3	CSR	-9.61	3.45E-22	1.73E-21
4	HI	-9.46	1.45E-21	5.81E-21
5	LID	-8.06	3.56E-16	1.07E-15
6	DCL	-5.47	2.15E-08	4.29E-08
7	PR	-4.28	9.10E-06	9.10E-06

The results of the APVs using Holm's procedure are shown in Table 4 and Table 5. The intended APVs are less than the significance level, thus rejecting the null hypothesis. It is clearly observed from the results obtained that MCD is significantly better than the competing approaches on these statistical tests as well.

7 Conclusion

Finding influential nodes in complex networks like social networks play a vital role in achieving maximum information spread across the network. It has found numerous applications in business, marketing, and generating awareness for a social cause. In this work, we proposed a modified Community Diversity (MCD) method for finding influential nodes. First, we introduced the concept of Extended Community Diversity, which represents the community diversity score of the extended neighborhood of a node. Next, we introduced Modified Community Diversity, built upon the concepts of Community Diversity and Extended Community Diversity. To verify the efficacy of our approach in determining the most effective influential nodes in a network, we compared MCD with seven recent state-of-the-art methods over eight datasets. We found that MCD cumulatively outperformed the competing approaches across three metrics. Further, we observed that seed nodes selected by MCD were reasonably close together for extended activation due to a node's influence but far enough to not significantly overlap with each other's scope of influence, thus mitigating the Rich Club effect significantly.

References

Banerjee, S., Jenamani, M., Pratihar, D., 2020. A survey on influence maximization in a social network. Knowledge and Information Systems 62. doi:10.1007/s10115-020-01461-4.

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. science 286, 509–512.

Batagelj, V., Mrvar, A., 2000. Some analyses of erdos collaboration graph. Social Networks 22, 173–186.

Berahmand, K., Bouyer, A., Samadi, N., 2019. A new local and multidimensional ranking measure to detect spreaders in social networks. Computing 101. doi:10.1007/s00607-018-0684-8.

Boguñá, M., Pastor-Satorras, R., Diaz-Guilera, A., Arenas, A., 2004. Models of social networks based on social distance attachment. Physical review. E, Statistical, nonlinear, and soft matter physics 70, 056122. doi:10.1103/PhysRevE.70.056122.

Bonacich, P., 2007. Some unique properties of eigenvector centrality. Social Networks 29, 555–564. doi:10.1016/j.socnet.2007.04.002.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Computer Networks 30, 107–117. URL: http://www-db.stanford.edu/~backrub/google.html.

Chen, D., Lü, L., Shang, M.S., Zhang, Y.C., Zhou, T., 2012. Identifying influential nodes in complex networks. Fuel and Energy Abstracts 391. doi:10.1016/j.physa.2011.09.017.

Chen, L., Zheng, H., Li, Y., Liu, Z., Zhao, L., Tang, H., 2022. Enhanced density peak-based community detection algorithm. Journal of Intelligent Information Systems , 1–22doi:10.1007/s10844-022-00702-y.

Chen, W., Lakshmanan, L., Castillo, C., 2013. Information and influence propagation in social networks. Synthesis Lectures on Data Management 5, 1–177. doi:10.2200/S00527ED1V01Y201308DTM037.

Chen, X., Deng, L., Zhao, Y., Zhou, X., Zheng, K., 2021. Community-based influence maximization in location-based social network. World Wide Web 24, 1903–1928. URL: https://doi.org/10.1007/s11280-021-00935-x, doi:10.1007/s11280-021-00935-x.

Chouchani, N., Abed, M., 2020. Online social network analysis: detection of communities of interest. Journal of Intelligent Information Systems 54, 1–17. doi:10.1007/s10844-018-0522-7.

Dunn, O.J., 1961. Multiple comparisons among means. Journal of the American statistical association 56, 52–64.

Ferguson, R., 2008. Word of mouth and viral marketing: Taking the temperature of the hottest trends in marketing. Journal of Consumer Marketing 25, 179–182. doi:10.1108/07363760810870671.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35–41. URL: http://www.jstor.org/stable/3033543.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Social Networks 1, 215–239.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 32, 675–701. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1937.10503522, doi:10.1080/01621459.1937.10503522, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1937.10503522.

Gao, S., Ma, J., Chen, Z., Wang, G., Changming, X., 2014. Ranking the spreading ability of nodes in complex networks based on local structure. Physica A: Statistical Mechanics and its Applications 403, 130–147. doi:10.1016/j.physa.2014.02.032.

Garton, L., Haythornthwaite, C., Wellman, B., 1997. Studying online social networks. Journal of computer-mediated communication 3, JCMC313.

GLEISER, P.M., DANON, L., 2003. Community structure in jazz. Advances in Complex Systems 06, 565–573. URL: https://doi.org/10.1142/S0219525903001067, doi:10.1142/S0219525903001067, arXiv:https://doi.org/10.1142/S0219525903001067.

Goldenberg, J., Libai, B., Muller, E., 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters 12, 211–223. doi:10.1023/A:1011122126881.

Granovetter, M., 1978. Threshold models of collective behavior. American Journal of Sociology 83, 1420–1443. doi:10.1086/226707.

Guille, A., Hacid, H., Favre, C., Zighed, D.A., 2013. Information diffusion in online social networks: A survey. ACM Sigmod Record 42, 17–28.

He, Q., Wang, X., Mao, F., Lv, J., Cai, Y., Huang, M., Xu, Q., 2020. Caom: A community-based approach to tackle opinion maximization for social networks. Inf. Sci. 513, 252–269. URL: https://doi.org/10.1016/j.ins.2019.10.064, doi:10.1016/j.ins.2019.10.064.

Hethcote, H.W., 2000. The mathematics of infectious diseases. SIAM Review 42, 599–653. URL: https://doi.org/10.1137/S0036144500371907, doi:10.1137/S0036144500371907, arXiv:https://doi.org/10.1137/S0036144500371907.

Holland, B.S., Copenhaver, M.D., 1987. An improved sequentially rejective bonferroni test procedure. Biometrics , 417–423.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65–70. URL: http://www.jstor.org/stable/4615733.

Huang, H., Shen, H., Meng, Z., Chang, H., He, H., 2019. Community-based influence maximization for viral marketing. Applied Intelligence 49, 2137–2150. URL: https://doi.org/10.1007/s10489-018-1387-8, doi:10.1007/s10489-018-1387-8.

Iman, R.L., Davenport, J.M., 1980. Approximations of the critical region of the fbietkan statistic. Communications in Statistics - Theory and Methods 9, 571–595. URL: https://doi.org/10.1080/03610928008827904, doi:10.1080/03610928008827904, arXiv:https://doi.org/10.1080/03610928008827904.

Kempe, D., Kleinberg, J., Tardos, E., 2003a. Maximizing the spread of influence through a social network. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 137-146. doi:10.1145/956750.956769.

Kempe, D., Kleinberg, J., Tardos, E., 2003b. Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 137–146. URL: https://doi.org/10.1145/956750.956769, doi:10.1145/956750.956769.

Khomami, M.M.D., Rezvanian, A., Meybodi, M.R., Bagheri, A., 2021. Cfin: A community-based algorithm for finding influential nodes in complex social networks. J. Supercomput. 77, 2207–2236. URL: https://doi.org/10.1007/s11227-020-03355-2, doi:10.1007/s11227-020-03355-2.

Kumar, R., Novak, J., Tomkins, A., 2006. Structure and evolution of online social networks, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 611–617. doi:10.1145/1150402.1150476.

Kumar, S., Gupta, A., Khatri, I., 2022a. Csr: A community based spreaders ranking algorithm for influence maximization in social networks. World Wide Web doi:10.1007/s11280-021-00996-y.

Kumar, S., Lohia, D., Pratap, D., Krishna, A., Panda, B.S., 2022b. Mder: Modified degree with exclusion ratio algorithm for influence maximisation in social networks. Computing 104, 359–382. URL: https://doi.org/10.1007/s00607-021-00960-8, doi:10.1007/s00607-021-00960-8.

Kumar, S., Panda, B.S., Aggarwal, D., 2021a. Community detection in complex networks using network embedding and gravitational search algorithm. Journal of Intelligent Information Systems 57, 1–22. doi:10.1007/s10844-020-00625-6.

Kumar, S., Singhla, L., Jindal, K., Grover, K., Panda, B.S., 2021b. Im-elpr: Influence maximization in social networks using label propagation based community structure. Applied Intelligence 51. doi:10.1007/s10489-021-02266-w.

Leskovec, J., Kleinberg, J., Faloutsos, C., 2007a. Graph evolution: Densification and shrinking diameters. ACM Trans. Knowl. Discov. Data 1, 2–es. URL: https://doi.org/10.1145/1217299.1217301, doi:10.1145/1217299.1217301.

Leskovec, J., Kleinberg, J., Faloutsos, C., 2007b. Graph evolution: Densification and shrinking diameters. ACM Trans. Knowl. Discov. Data 1, 2–es. URL: https://doi.org/10.1145/1217299.1217301, doi:10.1145/1217299.1217301.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N., 2007c. Cost-effective outbreak detection in networks, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 420–429. URL: https://doi.org/10.1145/1281192.1281239, doi:10.1145/1281192.1281239.

Li, Y., Fan, J., Wang, Y., Tan, K.L., 2018. Influence maximization on social graphs: A survey. IEEE Transactions on Knowledge and Data Engineering 30, 1852–1872. doi:10.1109/TKDE.2018.2807843.

Lin, J., 1991. Divergence measures based on the shannon entropy. IEEE Transactions on Information theory 37, 145–151.

Liu, J., Xiong, Q., Shi, W., Shi, X., Wang, K., 2016. Evaluating the importance of nodes in complex networks. Physica A: Statistical Mechanics and its Applications 452. doi:10.1016/j.physa.2016.02.049.

Liu, Y., Qiu, L., Sun, C., 2022. An influence maximization algorithm based on low-dimensional representation learning. Applied Intelligence , 1–18doi:10.1007/s10489-022-03178-z.

Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M., 2003. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Behavioral Ecology and Sociobiology 54, 396–405.

Lü, L., Zhou, T., Zhang, Q.M., Stanley, H., 2016. The h-index of a network node and its relation to degree and coreness. Nature Communications 7, 10168. doi:10.1038/ncomms10168.

Ma, L., Chiew, K., Huang, H., He, Q., 2017. Evaluation of local community metrics: from an experimental perspective. Journal of Intelligent Information Systems 51, 1–22.

Mendonça, M.R.F., da Motta Salles Barreto, A., Ziviani, A., 2022. Efficient information diffusion in time-varying graphs through deep reinforcement learning. World Wide Web doi:10.1007/s11280-021-00998-w.

Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B., 2007. Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pp. 29–42.

Okamoto, K., Chen, W., Li, X.Y., 2008. Ranking of closeness centrality for large-scale social networks, in: Frontiers in Algorithmics, pp. 186–195. doi:10.1007/978-3-540-69311-6_21.

Ripeanu, M., Foster, I., Iamnitchi, A., 2002. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. IEEE Internet Computing Journal 6.

Salavati, C., Abdollahpouri, A., Manbari, Z., 2019. Ranking nodes in complex networks based on local structure and improving closeness centrality. Neurocomput. 336, 36–45. URL: https://doi.org/10.1016/j.neucom.2018.04.086, doi:10.1016/j.neucom.2018.04.086.

Traag, V.A., Waltman, L., van Eck, N.J., 2019. From louvain to leiden: guaranteeing well-connected communities. Scientific Reports 9. URL: https://doi.org/10.1038%2Fs41598-019-41695-z, doi:10.1038/s41598-019-41695-z.

Wen, T., Deng, Y., 2019. Identification of influencers in complex networks by local information dimensionality. Information Sciences 512. doi:10.1016/j.ins.2019.10.003.

Yang, H., Cheng, J., Su, X., Zhang, W., Zhao, S., Chen, X., 2021. A spiderweb model for community detection in dynamic networks. Applied Intelligence 51, 5157–5188. URL: https://doi.org/10.1007/s10489-020-02059-7, doi:10.1007/s10489-020-02059-7.

Zhou, J., Fan, J., Wang, J., Wang, X., Li, L., 2019. Cost-efficient viral marketing in online social networks. World Wide Web 22, 1–24. doi:10.1007/s11280-018-0651-5.

arXiv:2211.05853v1 [cs.CL] 10 Nov 2022

Measuring Reliability of Large Language Models through Semantic Consistency

Harsh Raj
Delhi Technological University
Delhi, India
harsh777111raj@gmail.com

Domenic Rosati
scite.ai
Brooklyn, NY, USA
dom@scite.ai

Subhabrata Majumdar
Trustworthy ML Initiative
Seattle, WA, USA
zoom.subha@gmail.com

Abstract

While large pretrained language models (PLMs) demonstrate incredible fluency and performance on many natural language tasks, recent work has shown that well-performing PLMs are very sensitive to what prompts are feed into them. Even when prompts are semantically identical, language models may give very different answers. When considering safe and trustworthy deployments of PLMs we would like their outputs to be consistent under prompts that mean the same thing or convey the same intent. While some work has looked into how state-of-the-art PLMs address this need, they have been limited to only evaluating lexical equality of single- or multi-word answers and do not address consistency of generative text sequences. In order to understand consistency of PLMs under text generation settings, we develop a measure of semantic consistency that allows the comparison of open-ended text outputs. We implement several versions of this consistency metric to evaluate the performance of a number of PLMs on paraphrased versions of questions in the TruthfulQA dataset, we find that our proposed metrics are considerably more consistent than traditional metrics embodying lexical consistency, and also correlate with human evaluation of output consistency to a higher degree.

1 Introduction

As large pretrained language models (PLMs) get adopted more and more into next-generation automated workflows for natural language-based tasks, concerns about the safety and reliability of such models have also come in prominence [21]. In Natural Language Generation (NLG), it is reasonable to expect a reliable PLM that is prompted with semantically equivalent prompts to return semantically equivalent outputs. This property is called *consistency*. Without this property, *we cannot ensure the safety of PLMs* since we won't be certain about the outputs of models under inputs that semantically equivalent to seen examples. However, the degree to which current approaches to NLG are consistent and how to measure their consistency are still unclear. Addressing this gap, we introduce a novel framework for measuring consistency of NLG that generalizes previous approaches measuring consistency evaluation of single-token outputs (such as [5]), to measuring consistency of generated text sequences.

Consistency measurement was introduced in the LAMA probe to understand PLMs as knowledge bases [18]. Elazar et al. [6] extended this and developed the ParaRel dataset to evaluate the consistency of masked language models by understanding the tokens they would predict for masked tuples. Taking this direction of research forward, Fierro and Søgaard [7] extended their methods to a multilingual, multi-token setting. Jang et al. [11] developed a novel framework for understanding consistency on fine-tuned models on sentence similarity tasks. Zhou et al. [26] developed an approach based on using multiple prompts to specify single tasks to improve consistency metrics by more than 10% across diverse data and task settings. Finally, Newman et al. [16] developed a robust

ML Safety Workshop, 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

method to accurately extract factual information from PLMs as measured by a consistency measure they introduce.

In this paper, we present a novel measure of consistency for NLG that goes beyond the simple lexical measures proposed in past work, encoding realistic variation across prompts that semantically carry the same meaning but are lexically dissimilar. We empirically evaluate the effectiveness of this metric for consistency evaluation using multiple measures of semantic equivalence, PLM architectures, and answer generation/decoding methods. We find that (1) PLMs tend to generate *less accurate* answers as their size increases (the so-called inverse scaling phenomenon, see Lin et al. [15]), (2) PLMs tend to become *more consistent* with size. (3), consistency and accuracy are not correlated and seem to reflect orthogonal properties, and (4) *the biggest contributor to inconsistency is the decoding method*. Greedy approaches to decoding are much more consistent than a sampling-based approach.

2 Method

Elazar et al. [5] first formalized consistency analysis of PLMs. The goal of consistency in PLMs is that prompts which are paraphrases of each other should produce the equivalent outputs to be considered consistent. Given a set of n semantically similar prompts: $X = \{x_1, \ldots, x_n\}$, with generated outputs $Y = \{y_1, \ldots, y_n\}$, the consistency measure of Elazar et al. [5] is defined as

$$\text{Cons}_{lex}(Y) = \frac{1}{n(n-1)} \sum_{i,j=1,i\neq j}^{n} \mathbb{I}(y_i = y_j). \tag{1}$$

A key limitation of the work of Elazar et al. [5] and its follow-up studies [7, 12, 16, 26] is that their experimental consistency evaluations of PLMs were limited to settings where outputs are considered consistent if they are *exactly equal* to each other, or have a certain amount of token overlap. This means *we cannot use the same measure to evaluate freeform natural language outputs of language models*. Furthermore, the constraint of exact equality (or multi-token equality [13]) is not realistic where we can have consistent answers that are not lexically identical but are in fact semantically identical.

We extend the consistency measure in Eq. (1) by relaxing the exact equality constraint and replacing it with an agreement function that can identify semantic agreement between two outputs such as whether they are paraphrases or whether there is an entailment relationship between them.

$$\text{Cons}_{sem}(Y) = \frac{1}{n(n-1)} \sum_{i,j=1,i\neq j}^{n} f(y_i, y_j). \tag{2}$$

Eq. (2) describes a generalization of the consistency metric in Eq. (1) originally proposed in Elazar et al. [5], replacing token equality of outputs with a *semantic equivalence function* $f(\cdot)$ to produce a consistency score in the range $[0, 1]$. Note that we can easily recover the original metric of Elazar et al. [5] by setting the token equality indicator as $f(\cdot)$: $f(y_i, y_j) \equiv \mathbb{I}(y_i = y_j)$. 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.

In this paper, we explore four such agreement functions: semantic similarity using BERTScore [24, BERTs], semantic equivalence using paraphrase detection (PP, using the approach in Appendix A), and entailment (Entail) and contradiction (Contra) using a natural language inference model (DeBERTa v2 xxlarge trained on mnli). For comparison purposes, we also implement rouge1 (R1-C) and named-entity overlap (NER) as heuristic measures of token overlap.

3 Experimental Setup

To evaluate semantic consistency under the framework proposed above, we take the generated text responses of PLMs on the TruthfulQA benchmark dataset [15], enriched with several paraphrases

Table 1: Accuracy and consistency of OPT and GPT-3 models on TruthfulQA

Metric type	Acc		Consis						
Model	R1-A	BLEURT	PP	PP+acc	BERTs	Entail	Contra	R1-C	NER
Greedy									
OPT-125M	40.0	50.0	27.7	21.4	90.7	26.1	25.7	12.0	12.2
OPT-350M	41.6	50.3	17.1	17.7	88.8	18.0	23.1	12.6	13.5
OPT-1.3B	36.5	49.2	30.6	25.2	89.5	23.6	29.4	12.7	11.9
OPT-2.7B	35.2	43.6	37.6	32.6	90.3	27.4	28.6	11.6	11.1
GPT-3	**59.0**	**62.6**	**62.2**	**71.5**	**92.3**	**42.5**	**11.3**	**30.4**	**19.8**
Sampled									
OPT-125M	41.9	50.7	6.2	2.5	86.1	5.2	36.7	0.2	4.7
OPT-350M	43.2	50.7	4.6	2.8	85.6	4.4	31.0	0.2	3.9
OPT-1.3B	40.1	50.4	11.3	6.0	85.6	5.7	36.0	0.3	3.7
OPT-2.7B	40.0	48.9	14.2	9.2	85.8	6.6	36.7	0.3	3.2
GPT-3	**56.3**	**60.5**	**54.0**	**66.0**	**90.8**	**32.0**	**14.1**	**19.9**	**14.2**

for each of the questions. We chose TruthfulQA because it has a series of metrics and baselines for evaluating freeform text generation. In order to have access to a diverse set of paraphrases for the questions in TruthfulQA, we utilize three methods of question paraphrase generation: DocT5Query [3], QC [1], and GPT-3 [2] with few-shot prompts for generating paraphrases (using the method found in Chowdhury et al. [4]). This results in a total of 8,956 paraphrased questions.

Since having consistent answers requires that the question paraphrases actually ask the same thing, we train a paraphrase detection model on PAWS [25] (see Appendix A for details), and filter out all generated paraphrases that fail to be detected as paraphrase after passing through this model using a two-step process. First, we rank paraphrases by their output probabilities from the PAWS model of being a paraphrase, and select the top 6 paraphrases. Second, we manually remove question paraphrases which we do not consider to be actual paraphrases of the original question. These steps result in a final set of 3,962 paraphrased questions covering 817 original questions (see Appendix B for results on unfiltered questions and further analysis).

To understand the effects of model parameter size on consistency, we prompt a series of OPT models [23] (from 125M to 2.7B parameters) with the question paraphrases and an answer prompt resulting in a set of answers for each question. We generate answers using greedy decoding and nucleus sampling [10] procedures in order to understand how sampling-based decoding might effect consistency (inspired by Lee et al. [14] who found a dramatic effect of decoding method on factuality). Finally we run these set of answers through both the accuracy metrics for freeform text generation outlined in TruthfulQA [15] and the semantic consistency metrics proposed above. We further evaluate a random sample of 100 questions with GPT-3 text-davinci-002 [2, 17] to understand how the computed consistency values compare with those of a model that has a much larger number (175B) of parameters.

To validate our semantic consistency measurement framework and chosen agreement functions, we ran a human study with three volunteer participants who labeled a random sample of 100 questions with 903 answer pairs resulting from paraphrases of the 100 sampled questions. Participants were instructed to label answer pairs as consistent if they considered the two answers as semantically equivalent and inconsistent otherwise. The inter-annotator agreement has a Fleiss' κ value of 0.84, indicating strong agreement between participants. We use these human scores and perform a correlation study with our proposed semantic consistency scores and the TruthfulQA accuracy scores to understand what metrics might be most reflective of what humans understand as consistency. Since we do not have access to the Judge model or human evaluation scores from the TruthfulQA paper, we did not compare those scores with our approach.

Table 2: Correlation of consistency scores (Spearman ρ) with accuracy scores as well as human consistency annotations. Scored on outputs from OPT-2.7B with greedy decoding.

	Human	R1-C	NER	BERTs	PP	Entail	Contra	R1-A	BLEURT
Human	1.00	0.32	0.15	0.54	0.52	0.70	-0.28	-0.10	-0.10
R1-C		1.00	0.27	0.23	-0.10	0.15	-0.08	0.10	-0.11
NER			1.00	-0.05	-0.07	0.01	-0.05	-0.05	-0.06
BERTs				1.00	0.66	0.78	-0.29	-0.13	-0.22
PP					1.00	0.76	-0.48	-0.15	-0.09
Entail						1.00	-0.49	-0.25	-0.22
Contra							1.00	0.05	0.11
R1-A								1.00	0.48
BLEURT									1.00

4 Results

Table 1 outlines a number of key findings. **First**, we are able to reproduce the finding that as OPT models get larger in parameter size they tend to do worse at providing truthful answers in TruthfulQA [15], as measured by rouge1 (R1-A) and BLEURT [19]. **Second**, there is a general trend of models became more consistent with parameter size especially in the case of paraphrase and entailment which correlated most with human judgements of consistency. This is also the case for answers that are both consistent under paraphrase and accurate (PP+acc). **Third**, while nucleus sampling is more accurate than greedy sampling for the OPT models, greedy sampling tends to be much more consistent on all consistency measures. Finally lexical measures of consistency (R1-C, NER) and BERTScore are relatively uninformative measures.

Table 2 outlines results from the correlation analysis between each consistency metric and human scores. Another important finding here is that *accuracy measures do not correlate with consistency measures*. The two accuracy measures show moderate positive correlation with one another ($\rho = 0.48$). However, consistency measures tend to not correlate with one another, except in the case of PP, entailment, and BERTscore measures where there are moderate positive correlations. Human scores show the strongest correlation with entailment (ρ=0.70), PP (ρ=0.52), and BERTScore (ρ=0.54), but have weak positive correlation with NER (ρ=0.15) and R1-c (ρ=0.32)These observations indicate that our semantic consistency scores do indeed measure the property of consistency as understood by humans, and do a better job than lexical measures based on token similarity.

Given the above results, we recommend that entailment and paraphrase-based agreement functions should be further investigated as consistency measures. Lexical approaches do not correlate well with human judgement. While BERTScore does have a similar correlation with human judgement as paraphrase, this is due to the fact that what BERTScore really measures is semantic *similarity*. Two outputs can still be semantically *similar* while being inconsistent. This is not the case with the paraphrase-based approach. Additionally, Table 1 shows that BERTScore is relatively uninformative for analyzing models as they scale.

5 Conclusion

In this work, we presented a novel framework for evaluating consistency of NLG that correlates well with human judgements when using paraphrase or entailment as agreement functions. While we have demonstrated that our semantic consistency framework works well for measuring consistency in NLG, future work should continue to validate our proposal across other types of text generation tasks such as dialogue or table-to-text generation. Additionally, a key limitation of our work is that any error in the agreement function will be reflected as error in the consistency score. Finally, since Lin et al. [15] has shown an inverse scaling on TruthfulQA, future work should follow up with datasets where scaling laws hold to understand how consistency looks on those.

Many of the above findings are intuitive or in line with previous results. Lin et al. [15] found similar inverse scaling on TruthfulQA. While Webson and Pavlick [20] has shown that PLMs have poor understanding of prompts, and prompts are better understood by larger models. If larger models

understand their prompts more, it makes sense that they would be more consistent than smaller ones. Lee et al. [14] found sampling-based decoding was much less factual then greedy decoding with the explanation that sampling introduces randomness causing factual errors. Along these lines it is plausible that sampling-based decoding methods could be introducing inconsistency by randomness during decoding. This is worth further investigation in future studies.

Less intuitive is that consistency and accuracy do 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 PLMs. Future work could extend and validate our framework by using it as the basis for measurable improvements of consistency in NLG settings.

References

[1] Elron Bandel, Ranit Aharonov, Michal Shmueli-Scheuer, et al. Quality controlled paraphrase generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 596–609, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.45. URL https://aclanthology.org/2022.acl-long.45.

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners, 2020. URL https://arxiv.org/abs/2005.14165.

[3] David R. Cheriton. From doc2query to doctttttquery. 2019.

[4] Jishnu Ray Chowdhury, Yong Zhuang, and Shuyi Wang. Novelty controlled paraphrase generation with retrieval augmented conditional prompt tuning, 2022. URL https://arxiv.org/abs/2202.00535.

[5] Yanai Elazar, Nora Kassner, Shauli Ravfogel, et al. Measuring and Improving Consistency in Pretrained Language Models. *Transactions of the Association for Computational Linguistics*, 2021. doi: 10.1162/tacl_a_00410.

[6] Yanai Elazar, Hongming Zhang, Yoav Goldberg, and Dan Roth. Back to Square One: Artifact Detection, Training and Commonsense Disentanglement in the Winograd Schema, October 2021. URL http://arxiv.org/abs/2104.08161. arXiv:2104.08161 [cs].

[7] Constanza Fierro and Anders Søgaard. Factual Consistency of Multilingual Pretrained Language Models. In *FINDINGS*, 2022. doi: 10.48550/arXiv.2203.11552.

[8] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. Deberta: Decoding-enhanced bert with disentangled attention, 2020. URL https://arxiv.org/abs/2006.03654.

[9] Pengcheng He, Jianfeng Gao, and Weizhu Chen. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, 2021. URL https://arxiv.org/abs/2111.09543.

[10] Ari Holtzman, Jan Buys, Li Du, et al. The curious case of neural text degeneration, 2019. URL https://arxiv.org/abs/1904.09751.

[11] Myeongjun Jang, Deuk Sin Kwon, and Thomas Lukasiewicz. Accurate, yet inconsistent? consistency analysis on language understanding models. *CoRR*, abs/2108.06665, 2021. URL https://arxiv.org/abs/2108.06665.

[12] Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. How Can We Know What Language Models Know?, May 2020. URL http://arxiv.org/abs/1911.12543. arXiv:1911.12543 [cs].

[13] Nora Kassner, Philipp Dufter, and Hinrich Schütze. Multilingual LAMA: Investigating knowledge in multilingual pretrained language models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3250–3258, Online, April 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.eacl-main.284. URL https://aclanthology.org/2021.eacl-main.284.

[14] Nayeon Lee, Wei Ping, Peng Xu, et al. Factuality enhanced language models for open-ended text generation, 2022. URL https://arxiv.org/abs/2206.04624.

[15] Stephanie Lin, Jacob Hilton, and Owain Evans. TruthfulQA: Measuring how models mimic human falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3214–3252, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.229. URL https://aclanthology.org/2022.acl-long.229.

[16] Benjamin Newman, Prafulla Kumar Choubey, and Nazneen Rajani. P-Adapters: Robustly Extracting Factual Information from Language Models with Diverse Prompts, April 2022. URL http://arxiv.org/abs/2110.07280. arXiv:2110.07280 [cs].

[17] Long Ouyang, Jeff Wu, Xu Jiang, et al. Training language models to follow instructions with human feedback, 2022. URL https://arxiv.org/abs/2203.02155.

[18] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, et al. Language models as knowledge bases? In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *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*, pages 2463–2473. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1250. URL https://doi.org/10.18653/v1/D19-1250.

[19] Thibault Sellam, Dipanjan Das, and Ankur P Parikh. Bleurt: Learning robust metrics for text generation. In *Proceedings of ACL*, 2020.

[20] Albert Webson and Ellie Pavlick. Do Prompt-Based Models Really Understand the Meaning of Their Prompts? *NAACL*, 2022. doi: 10.18653/v1/2022.naacl-main.167.

[21] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, et al. Taxonomy of risks posed by language models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, page 214–229, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393522. doi: 10.1145/3531146.3533088. URL https://doi.org/10.1145/3531146.3533088.

[22] Adina Williams, Nikita Nangia, and Samuel R. Bowman. A broad-coverage challenge corpus for sentence understanding through inference, 2017. URL https://arxiv.org/abs/1704.05426.

[23] Susan Zhang, Stephen Roller, Naman Goyal, et al. Opt: Open pre-trained transformer language models, 2022. URL https://arxiv.org/abs/2205.01068.

[24] Tianyi Zhang, Varsha Kishore, Felix Wu, et al. Bertscore: Evaluating text generation with bert, 2019. URL https://arxiv.org/abs/1904.09675.

[25] Yuan Zhang, Jason Baldridge, and Luheng He. 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)*, pages 1298–1308, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1131. URL https://aclanthology.org/N19-1131.

[26] Chunting Zhou, Junxian He, Xuezhe Ma, Taylor Berg-Kirkpatrick, and Graham Neubig. Prompt consistency for zero-shot task generalization, 2022.

Appendix

A Paraphrase model details

We finetuned a DeBERTa v3 [9] large model on PAWS [25] 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.

We used a pretrained DeBERTa base model [8] trained on MNLI [22] for entailment and contradiction measures.

Table 3: Accuracy and consistency of OPT and GPT-3 models on TruthfulQA on unfiltered questions

model	Acc R1-A	BLEURT	Consis PP	PP+acc	BERTScore	R1-C	NER
greedy							
opt-125m	40.0	50.0	26.2	19.7	90.2	9.9	10.5
opt-350m	41.6	50.3	16.1	14.3	88.4	10.7	12.0
opt-1.3b	36.5	49.2	28.8	22.3	89.0	10.5	10.8
opt-2.7b	35.2	43.6	35.0	28.6	89.6	9.5	9.7
sampled							
opt-125m	41.9	50.7	6.0	2.5	86.0	0.2	3.8
opt-350m	43.2	50.7	4.4	2.7	85.4	0.2	3.3
opt-1.3b	40.1	50.4	11.2	5.4	85.5	0.2	3.1
opt-2.7b	40.0	48.9	14.4	9.2	85.7	0.3	2.8

B Question consistency and answer consistency

We can use our evaluation framework to understand the consistency of the prompts we used before and after human filtering. Both question sets were surprisingly consistent when using paraphrase as an agreement function with 93.34% consistency of question paraphrases without human filtering and 92.42% consistency of question paraphrases after automatic and human filtering. Since we would expect human filtering to have improved consistency, future work should follow up on what the discrepancy might be between filtering inconsistent results and measuring them.

To show what the accuracy and consistency results were when using the unfiltered set we present Table 3 below that summarizes results on the unfiltered set. Table 3 shows that they are almost the same results as Table 1 presented before with a downward accuracy trend, improved paraphrase consistency, and worse consistency with sampling-based decoding methods.

Silicon
https://doi.org/10.1007/s12633-022-02168-7

ORIGINAL PAPER

Mechanical and Tribological Properties of AA6061/SiC/Aloe Vera Powder Hybrid Al Composites Fabricated by Stir Casting

N. Yuvaraj[1] · Yashwant Koli[2] · G. Vedabouriswaran[3] · Plash Issar[1]

Received: 25 July 2022 / Accepted: 10 October 2022
© Springer Nature B.V. 2022

Abstract
Aluminium Matrix Composites (AMC) are fabricated through the stir casting route is more beneficial in comparison with other techniques since the methodology is inexpensive, easier in fabrication with mass production capabilities. Two types of AA6061 composites were produced by stir casting route namely AA6061 alloy reinforced with SiC particles and mixture of SiC & natural reinforcement of aloe vera particles (hybrid). FESEM and XRD analyses were conducted for the estimation of the presence of reinforcement particles in the composites. Microhardness, tensile properties and wear properties of composites were evaluated, results establish that composite samples are superior than cast AA6061 alloy. The lowest wear rate of 0.1874 10^{-3} g N^{-1} m^{-1} with a lower coefficient of friction (COF) 0.242 were found in hybrid AA6061 composite. The hard SiC ceramic particles acted as load-bearing elements while aloe vera particles enhanced the tribological behavior of composite by forming thin lubricating film during wear test. Results showed that aloe vera as a natural reinforcement is compatible with SiC to fabricate aluminium composites with impressive tensile, microhardness, and wear characteristics.

Keywords Aluminum matrix composites · Stir casting · Al6061 alloy · Silicon carbide · Aloe vera · Wear

1 Introduction

Nowadays, metal matrix composites (MMC) have generated lot of interest and attention among researchers and is evident by the ongoing research on various Aluminium alloys. Aluminium based MMC is one of the most common metals and its use is increasing day by day in various industries, most significantly in the aerospace, automobile, and defense industry which is purely attributed to its comparatively lower weight, higher specific strength, and shows good mechanical & better tribological properties [1, 2]. Various ceramic-based reinforcements such as SiC [3], B_4C [4], ZrO_2 [5], TiB_2 [6], Al_2O_3 [7], Y_2O_3 [8], TiO_2 [9] had already been used to produce aluminium composites through stir casting. Aluminium composites reinforced with SiC particles find their extensive usage in high-temperature applications of automobile parts such as pistons, connecting rods, gears, propeller shafts, bearing surfaces, camshafts, brake components, engine blocks, etc. [3, 10]. Due to the occurrence of hard ceramic particles in the MMC, Uthayakumar et al. [11] worked on stir casting machine to develop A1100/SiC + B_4C hybrid composite samples with higher wear resistance than base Al alloy. Priyaranjan Samal et al. [12] had fabricated Al5154 composite with different volume % (0, 5, 10 & 15) of SiC particles through stir casting methodology. The higher volume weight % of SiC particles enhances the hardness, tensile strength and wear resistance of the fabricated sample. The impact strength of the MMC, on the other hand, is reduced due to thermal incongruity between the matrix and the reinforcements. Addition of SiC and Mo particles to an Al hybrid composite improves mechanical properties and wear resistance because the particles good bonding inhibits dislocation in the Al matrix [13]. M. Saravana Kumar et al.

✉ Yashwant Koli
yashwantkoli5@gmail.com

N. Yuvaraj
yuvraj@dce.ac.in

G. Vedabouriswaran
veda@aiactr.ac.in

Plash Issar
plash.issar@yahoo.in

[1] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

[2] Department of Mechanical Engineering, IIT Delhi, Delhi 110016, India

[3] Department of Mechanical Engineering, NSUT (East Campus), Delhi 110031, India

Published online: 29 October 2022

[14] produced Al/SiC composite by glycerol-water method and studied the viscosity (1.04–1.24 mPa.s) of Aluminum melt in the composite. The homogeneous dispersal of SiC particles strengthens the matrix at 1.13 mPa.s viscosity, increasing the composite's tensile strength and wear resistance by 4% and 21%, respectively. I. Manivannan et al. [15] reported that Al/1.2SiC/0.5Gr nano-composite fabricated through ultrasonic assisted stir casting have better hardness and wear resistance than base alloy. The adequate bonding of nanoparticles of SiC in the MMC opposes the elimination of material from the surface during the wear test. In addition, graphite particles act as a lubrication effect for obtaining a lower wear rate and COF of the composite. Sonagiri Suresh et al. [16] added up (1–4% weight) of nano SiC particles in Al7075 alloy for fabrication of composites. With a decreased weight % of reinforcements, the tribological characteristics of the composite decrease linearly. SiC nanoparticles carry the load and reduce the plastic deformation with matrix alloy. Shahin Soltani et al. [17] concluded that appropriate stirring temperature (650 °C) and time (4 min) required for better incorporation of SiC microparticles with uniform distribution into molten pure aluminium which consequences in augmentation of tensile strength of the fabricated composite. Excessive stirring temperature (850 °C) leads to the formation of Al_4C_3 and shrinkage porosity in the matrix/ceramic interface, which outcomes in the decrease of mechanical properties of the fabricated composite. The researcher's attention has recently shifted to the use of eco-friendly agro (natural) derivatives as reinforcements for the fabrication of aluminium based MMC, intending to complement SiC or ceramic reinforcements. Agriculture and industrial waste are finding prominent use in the composite to achieve a reduction in cost as well as their ability to show satisfactory enhancement in properties [18]. Eggshell, coconut shell, melon shell, rice husk, bamboo leaf, and other agro derivatives had been treated into ashes with suitableness used as reinforcing material in the fabrication of AMC has been investigated by several researchers [19–30]. Aluminium with ceramic and natural reinforcement combined composites fabricated by stir casting and their mechanical & wear properties examined by numerous researchers are exhibited in Table 1.

Aloe vera is from the Asphodelaceae family which has shown resistance to drought. Aloe vera can be described as lance-shaped leaves having jagged edges and a sharp point. The Aloe Vera plant is commonly grown for its gel, which is mostly used in the cosmetic and health care industries [31]. Aloe Vera composed with aluminium has shown good strength to weight ratio and exceptional tensile properties. The fabrication of agriculture particulate aluminium composite is eco-friendly i.e. powder form of aloe vera in place of fly-ash in the form of ash [18]. SiC reinforcement particles have superb properties like high strength, hardness, good wear resistance, shock resistance, thermal stability, wettability and good distribution in the Aluminum matrix. Better wetting and good interfacial bonding take place with matrix help in fabrication through stir casting route efficiently and economically [3]. It is accomplished from the literature review that the natural reinforcements with ceramic particles give a better option to produce hybrid composites. In this paper, SiC particles and Aloe vera powder were used as reinforcements for the fabrication of AA6061 hybrid composite through the stir casting methodology. The consequence of reinforcements within the composite, microhardness, tensile strength and wear properties of the MMC samples were studied.

2 Experiment Procedure

2.1 Materials and Methods

Al6061 is opted as the base material. The SiC particles size of 40 µm and Aloe vera (natural reinforcement) particles size of 40–50 µm are used as reinforcements. Aloe vera particles were extracted naturally from aloe vera leaves, dried and made in powder form in micron particles. The chemical composition of Al6061, SiC and aloe vera are specified in Table 2 and FESEM images of reinforcement particles and EDS plots are shown in Fig. 1. For the fabrication of composite samples, the stir casting technique is used and the set-up is shown in Fig. 2(a). Cast iron die was used to fabricate composite samples which is an exhibited in Fig. 2(b). The AA6061 is brought to a liquid state at a temperature of 750 °C. Upon the formation of the vortex, the reinforcements are added. To attain uniform mixing of reinforcement particles in the Al alloy matrix, the mechanical stirrer was rotated at 125 rpm for 5 min. Different combinations of these reinforcements are added into the melt as shown in Table 3. The casting parameters are chosen after extensive literature survey [16, 17] and exhaustive trial tests which include casting of sample and checking the uniform distribution of reinforcement particles under optical microscope.

Prior to adding reinforcements in the furnace, the SiC particles are preheated at 200 °C for 1 h. The hardness, tensile and wear test specimens were obtained from the fabricated samples as shown in Fig. 2(c), (d) & (e). Tensile test samples were prepared as per ASTM E8/E8M-011 standard and tensile specification is shown in Fig. 3. For checking the repeatability of each composite sample three tensile and wear specimens were prepared. The tensile properties of samples are examined using UTM at a speed of 1 mm/min. The microhardness of the fabricated MMC was estimated using a microhardness tester (FISCHERSCOPE HM2000 S) at room temperature applying 100 g of load for 10 s of dwell time of 10 s. Wear test were

Table 1 Aluminium with ceramic and natural reinforcement MMC fabricated by stir casting

S.No	Base Material & reinforcements	Hardness	Tensile properties	Wear properties	Remarks	References
1	Al-12% Si and Melon shell ash (MSA) in the range of (5–20%wt.)	Increase in hardness to 103.1 HRC (15% wt. MSA) from 95 HRC (0% wt. MSA)	Highest tensile strength of 206.5 N/mm^2 for 15% wt. MSA and 120 N/mm^2 for 0% wt. MSA	Decrease in wear rate of 0.0091 g/mm (Al-12%Si 20%wMSA) from 0.022 g/mm (Al-12%Si-20 wt.%0 MSA)	MSA may be a good replacement for ceramic reinforcements due to cheap and environmentally friendly The interface between matrix and MSA particles dislocation density increases for enhancement of composite properties	Suleiman et al. [19]
2	Al6063 + Al_2O_3 + Rice husk ash (RHA) + Graphite	The hardness decreases with an increase in RHA and graphite particles weight ratio in the composite	Higher tensile strength observed in 0.5 wt.% graphite and 50% RHA composites	Wear resistance is improved with graphite particles composite	The presence of a larger quantity of RHA and graphite particles in the matrix reduces the load transmitting from the matrix to the reinforcements	Alaneme et al. [20]
3	Al-Si–Fe alloy + Bread Fruit seed hull ash (BFA)	The highest hardness value of 85 HRB was obtained for Al/10%SiC/ 10% Fly ash composite. Base alloy hardness = 65 HRB	The highest tensile strength of 215 ± 2 MPa was obtained for 12% BLA reinforced composite. Base alloy tensile strength = 172 ± MPa		An increase in BFA particles in the composites enhances the mechanical properties owing to expansion in dislocation density in base alloy interface	Atuanya et al. [29]
4	ADC12 aluminium alloy + B_4C + ground nut shell ash (GSA)	Highest hardness value of 115.3 BHN obtained for Al + 2.5% GSA + 7.5% B_4C. Base alloy hardness = 98.5 BHN	Highest tensile strength of 348.5 MPa obtained for Al + 2.5% GSA + 7.5% B_4C Base alloy tensile strengths = 322.5 MPa		Economic and environmentally friendly GSA ash is a probable reinforcement to improve mechanical properties of the composite	Venkatesh et al. [30]
5	Al6063 + nano fly ash (NFA) stir casting and further Friction stir welding (FSW) of composite	The hardness of composite sample FSW is enhanced with an increase in NFA weight %	Tensile strength of weld joint is improved with 2 wt% of NFA		NFA provides grain refinement and acts as a load-bearing element	Gosangi et al. [21]
6	Al6061 + Red mud powder(casting and ageing of composite)	Optimum micro harness value of 87 HV observed for 20% weight of Red mud powder, 74 μm particle size and 24 h of ageing time			Macro and microhardness of composite are increasing with the increase in % of reinforcement particles. An increase in particle size causes a reversal of the trend	Panwar et al. [22]
7	Al-Cu-Mg + Egg shell	Increase in hardness from 59.12 HRF for 0 wt.% to 74.17 HRF at 12 wt.% for eggshell particles	Increase in tensile strength from 98.28 MPa for 0 wt.% to 112.84 MPa for 12% wt.% eggshell particles	Dilo	Enhancement of mechanical properties of composites due to decrease in a grain of the composite and creation of an excessive dislocation density in the matrix	Hassan et al. [23]

Table 1 (continued)

S.No	Base Material & reinforcements	Hardness	Tensile properties	Wear properties	Remarks	References
8	Al 6063 + ceramic mortar ash (MA) + nano fibrillated composite (NFC) + industrial waste met coke ash (MCA) + agro waste straw ash (SA)	Highest hardness value of 62.63 kg/mm^2 obtained with Al + 5%wt. MA composite Al 6063 hardness value = 50.36 kg/mm^2	Higher tensile strength of 179 MPa was obtained for all combined matrix composites Al 6063 Tensile strength = 130 MPa	Al6063 + 5 wt.% NFC composite has better wear resistance than other composite samples	Good bonding of reinforcements and better refinement in the composite is responsible for better mechanical and wear properties	Rozhbiany et al. [24]
9	Al-4Cu + Bamboo leaf ash (BLA)	Increase in microhardness of 24.43% for 4% BLA reinforced composite with base alloy. Base alloy microhardness = 83.9 HV	The highest tensile strength of 177.3 MPa was obtained for 4% BLA reinforced composite. Base alloy tensile strength = 150.5 MPa		Decrease in grain size and no interfacial reaction takes place between matrix and reinforcements	Kumar et al. [25]
10	A356 + Rice husk ash (RHA) + Fly ash	Highest hardness value of 96 HV obtained for hybrid composite contains 10% RHA and 10% Fly ash particles			Reinforcement particles formed more interfaces and good mixing of particles in the matrix	Vinod et al. [26]
11	Al6061 + SiC + Jute ash	The highest hardness value of 66 HV was obtained for the Al + 3.4%SiC composite. 57HV hardness value for hybrid composite	The highest tensile strength of 123 MPa was obtained for the Al + 3.4%SiC composite. Marginal decrease in tensile strength observed in hybrid composite	Better wear properties were obtained for SiC reinforced mono composite	SiC act as a load-bearing element in the matrix for the enhancement of mechanical properties. Jute ash may be improper adherence takes place in the matrix for the decrease in properties	Coyal et al. [27]
12	Al7075 + SiC + Fly ash	The highest hardness value of 142HV was obtained for Al/10%SiC/ 10% Fly ash composite. Base alloy hardness is 102HV	Higher tensile strength was observed for 10% SiC & 10% Fly ash composite reinforced composite		An increase in SiC and Fly ash particles enhances the mechanical properties of the composite	Kanth et al. [28]

Table 2 Chemical composition of AA6061 and reinforcements (% weight)

Base material									
Composition	Al	Si	Fe	Cu	Mn	Mg	Zn	Cr	Ti
AA6061-T6	97.75	0.665	0.253	0.153	0.0676	0.840	0.0226	0.178	0.0225
Reinforcement									
Composition	SiC	Si	SiO_2	Fe	Al	C			
SiC	98.71	0.3	0.5	0.08	0.1	0.3			
Composition	Ca	Mg	Na^+	K	P	Fe	Cu	Zn	
Aloe Vera	3.58	1.22	3.66	4.06	0.02	0.1	0.06	0.02	

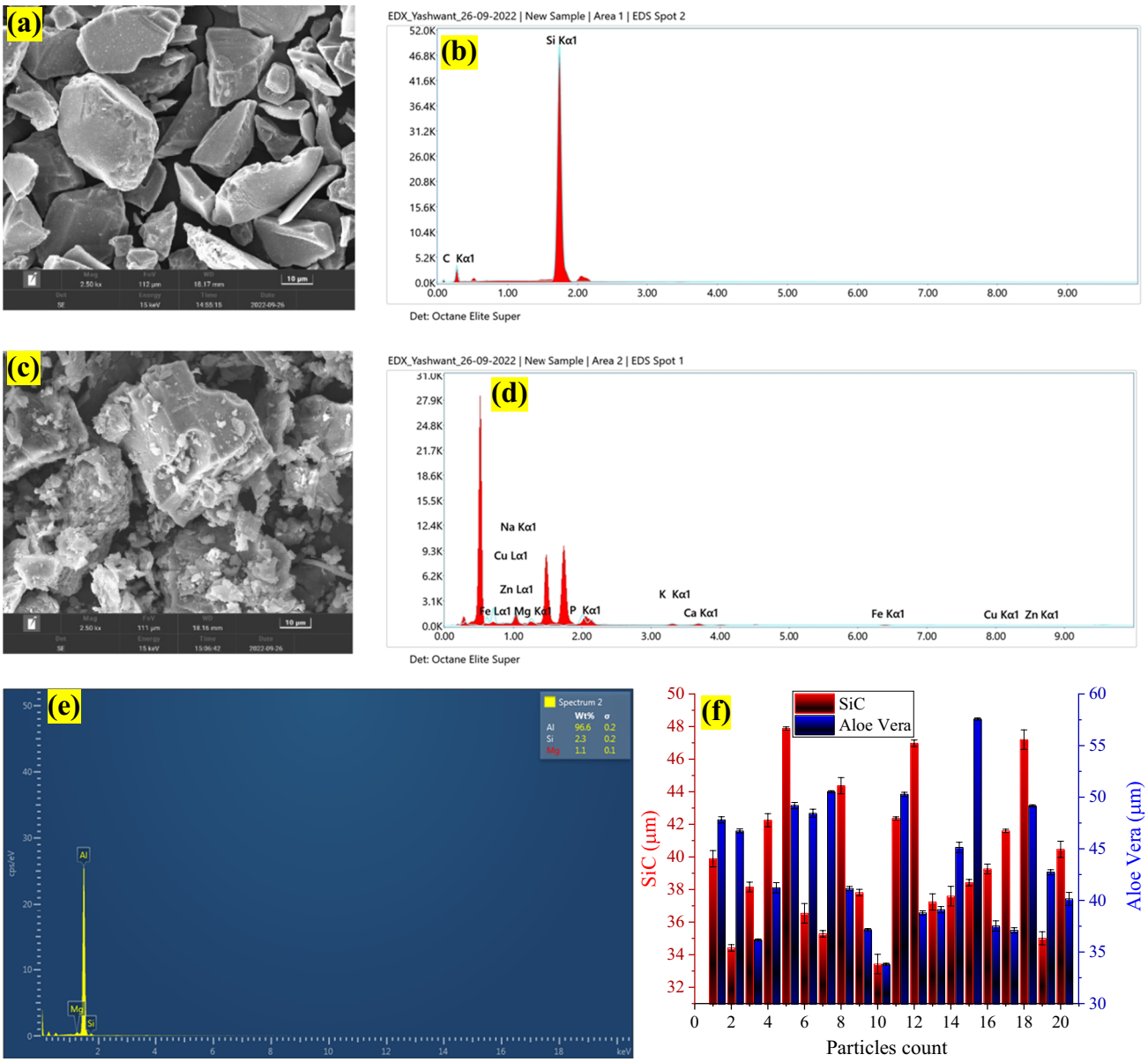


Fig. 1 (**a**) SiC particles (**b**) EDS of SiC (**c**) Aloe Vera particles (**d**) EDS of Aloe Vera (**e**) EDS of AA6061 (**f**) Particle size analysis

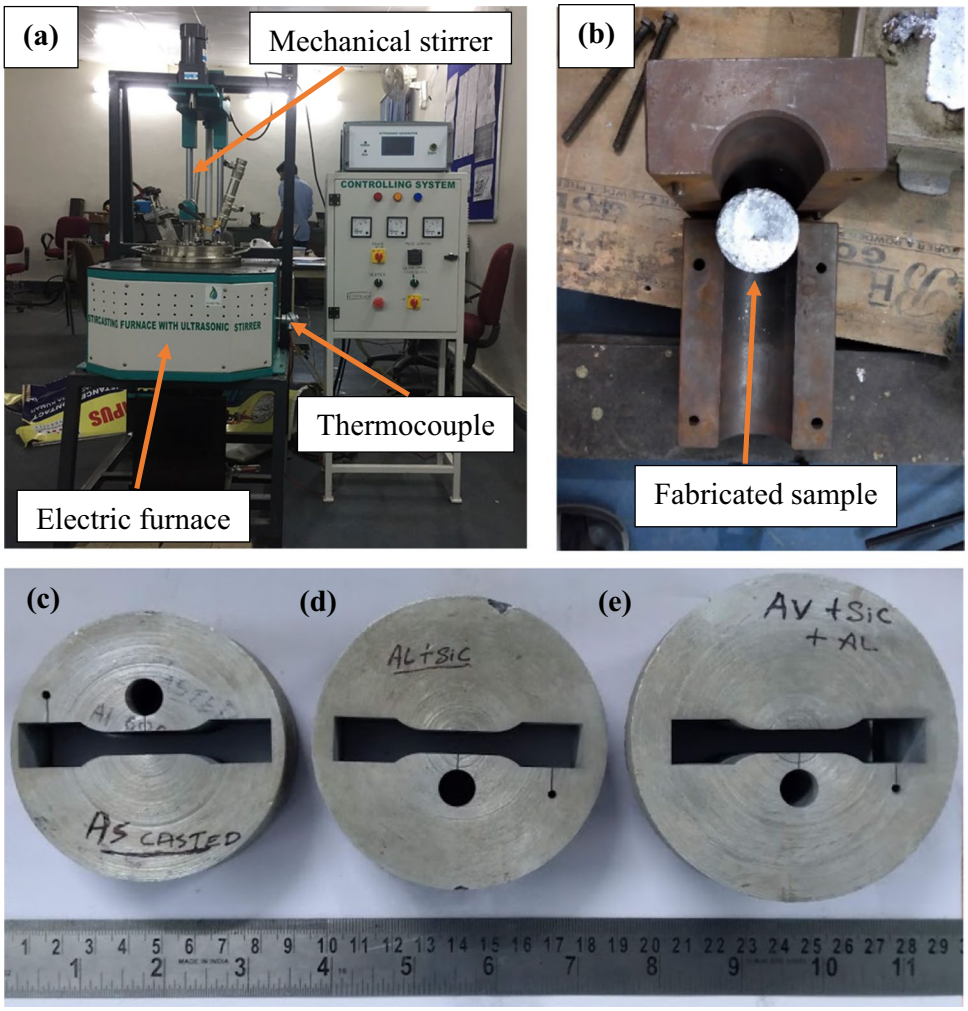


Fig. 2 (**a**) Stir casting setup (**b**) Cast iron die used for fabrication of AA composites (**c**) As casted 6061 (**d**) AA6061/SiC (**e**) AA6061/SiC/Aloe vera

Table 3 Fabricated samples

Sample Identification	AA6061 (wt.%)	SiC (wt.%)	Aloe Vera (wt.%)
AC	100	0	0
SiC-AA	95	5	0
AV-AA	95	2.5	2.5

AC = as cast AA6061; SiC-AA = AA6061/SiC; AV-AA = AA6061/SiC/aloe vera

accomplished using pin-on-disc High-Temperature Rotary Tribometer (DUCOM) as per ASTM G99 standards. The fabricated samples were used to create wear test specimens with a diameter of 10 mm. The counter-face is a hardened EN31 steel disc having a hardness of 55–58 HRC is used. The pin-on-disc setup is shown in Fig. 4. Sliding speed and load are kept constant at 1 m/s and 20 N respectively. Sliding distance ranged between 500 and 3000 m with an increment of 500 m. Prior to the wear test, the wear pin contact surface was grounded with 1600 grade emery paper and cleaned using acetone. Wear test samples were weighed (before and after the test) using an electronic precision weighing balance at an accuracy of 0.01 mg. To ensure repeatability all the wear test samples was performed three times. The COF was recorded between the hardened disc and the wear pin.

Porosity analysis is also conducted on the fabricated samples as per Eq. 1 [32, 33].

$$\text{Porosity} = \frac{\text{Volume of pores}}{\text{Volume of specimen}} = \frac{\text{Weight of water absorbed}}{\text{Density of water}} \quad (1)$$

2.2 Preparation of Aloe vera powder (Natural Reinforcement)

Aloe vera powder is made by chopping the aloe vera leaves into small pieces of 5 mm in length. The leaves are washed using hot water (80–100˚C) and dehydrated in a hot oven (150˚C for approximately 8–10 h) to eliminate moisture

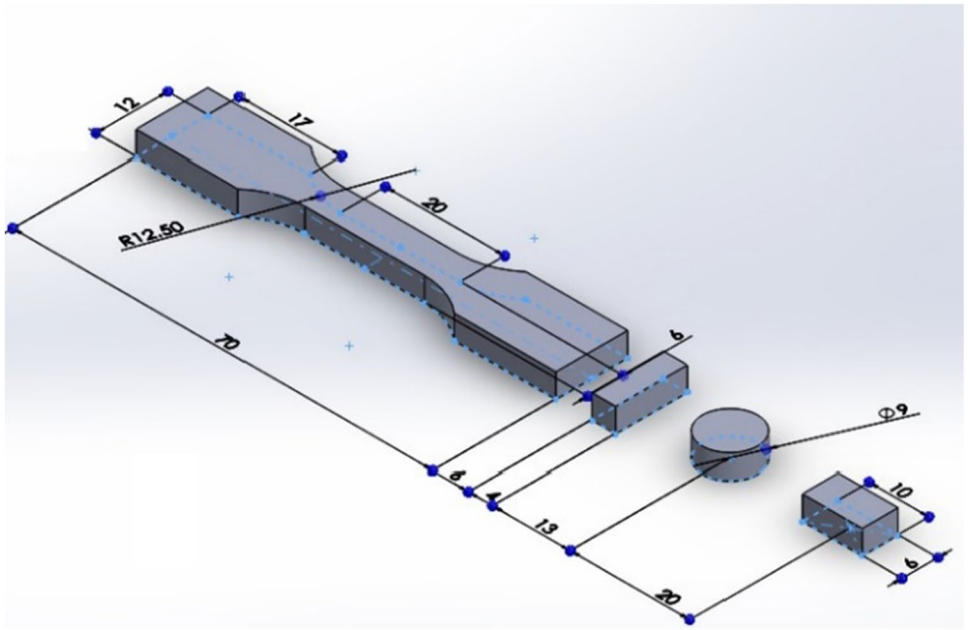


Fig. 3 Tensile specimen dimensions

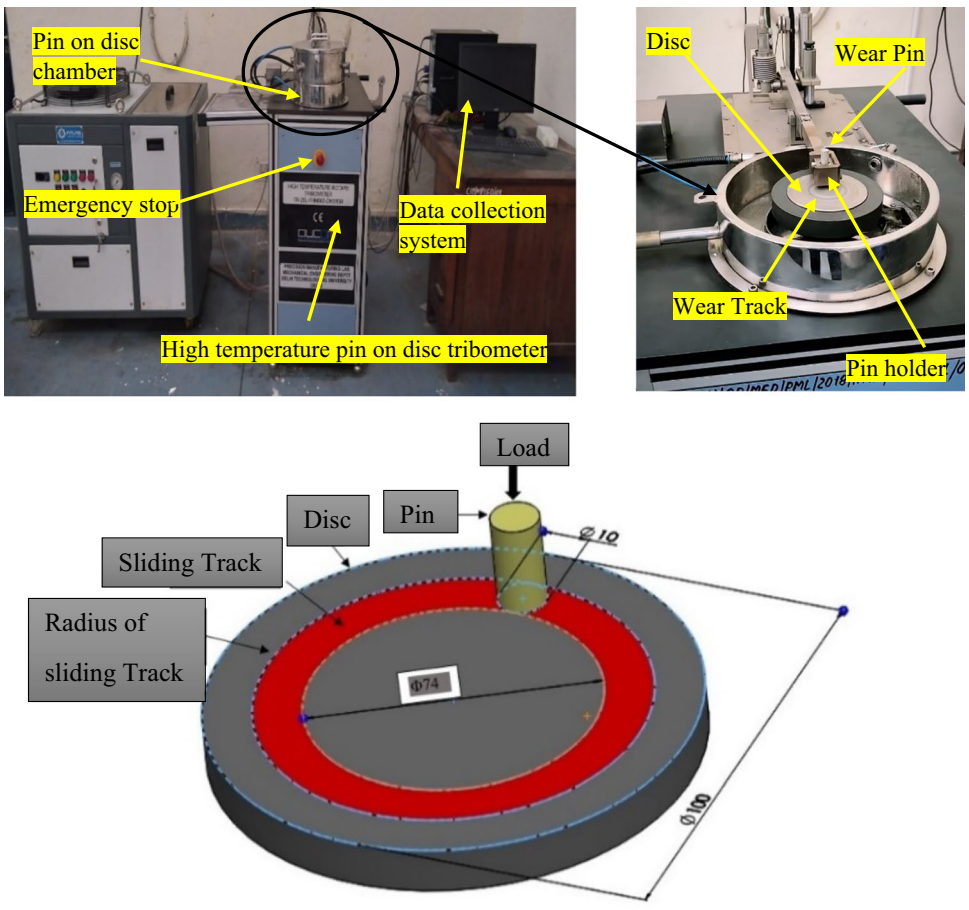


Fig. 4 (**a**) High-temperature rotary Tribometer (**b**) Pin on disc diagram

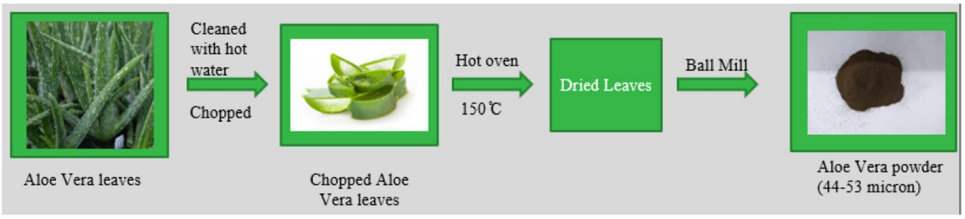


Fig. 5 Methodology of preparing Aloe vera powder

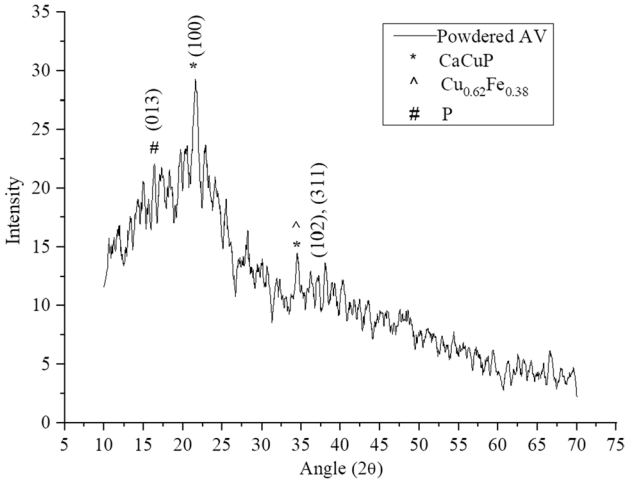


Fig. 6 XRD plot of Aloe vera powder

content. Using a ball mill, these dried leaves were grounded into approximately 45 µm size powder. Figure 5 shows the methodology of preparing aloe vera powder and it was analyzed using XRD. It shows the presence of elements such as calcium, iron, phosphorous having a higher intensity which justifies its chemical composition (Fig. 6).

3 Results and Discussion

3.1 Intermetallic Compounds (IMC) Analysis

The developed composites were studied with X-ray diffraction for the classification of the formation of IMC. XRD results establish a fact that the additives and the matrix retained their identity without forming any intermetallic thus

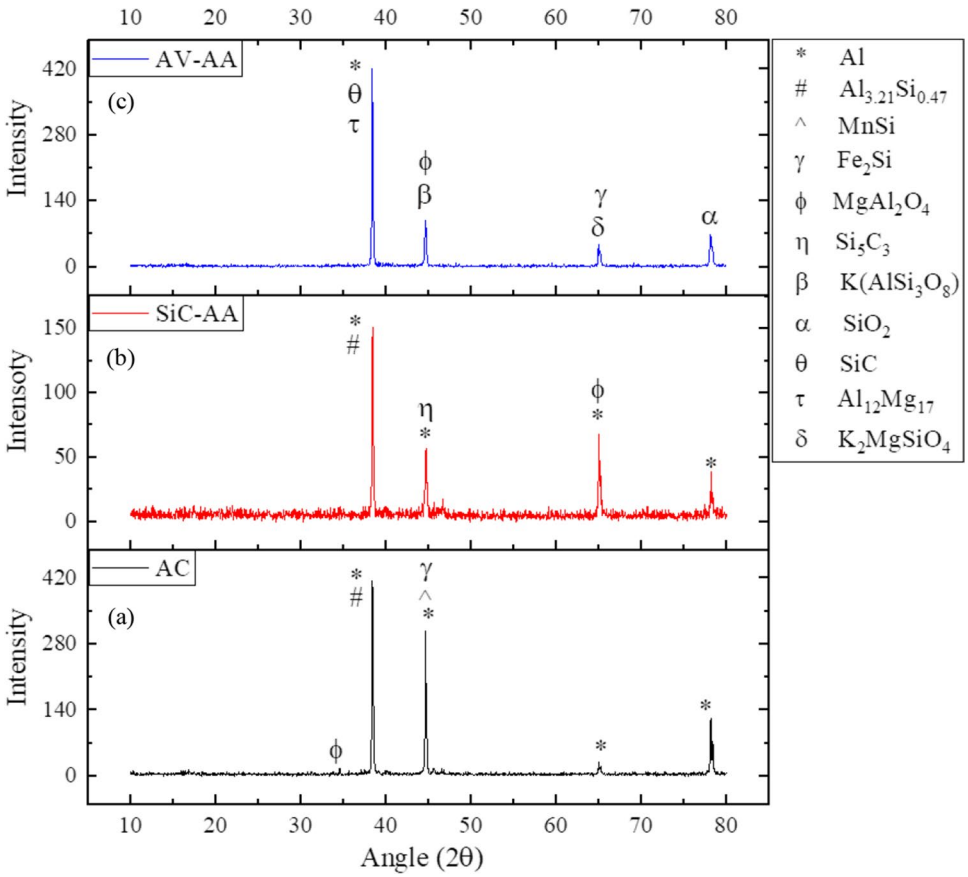


Fig. 7 XRD plots (**a**) as-cast AA6061 (**b**) SiC-AA composite (**c**) AV-AA hybrid composite

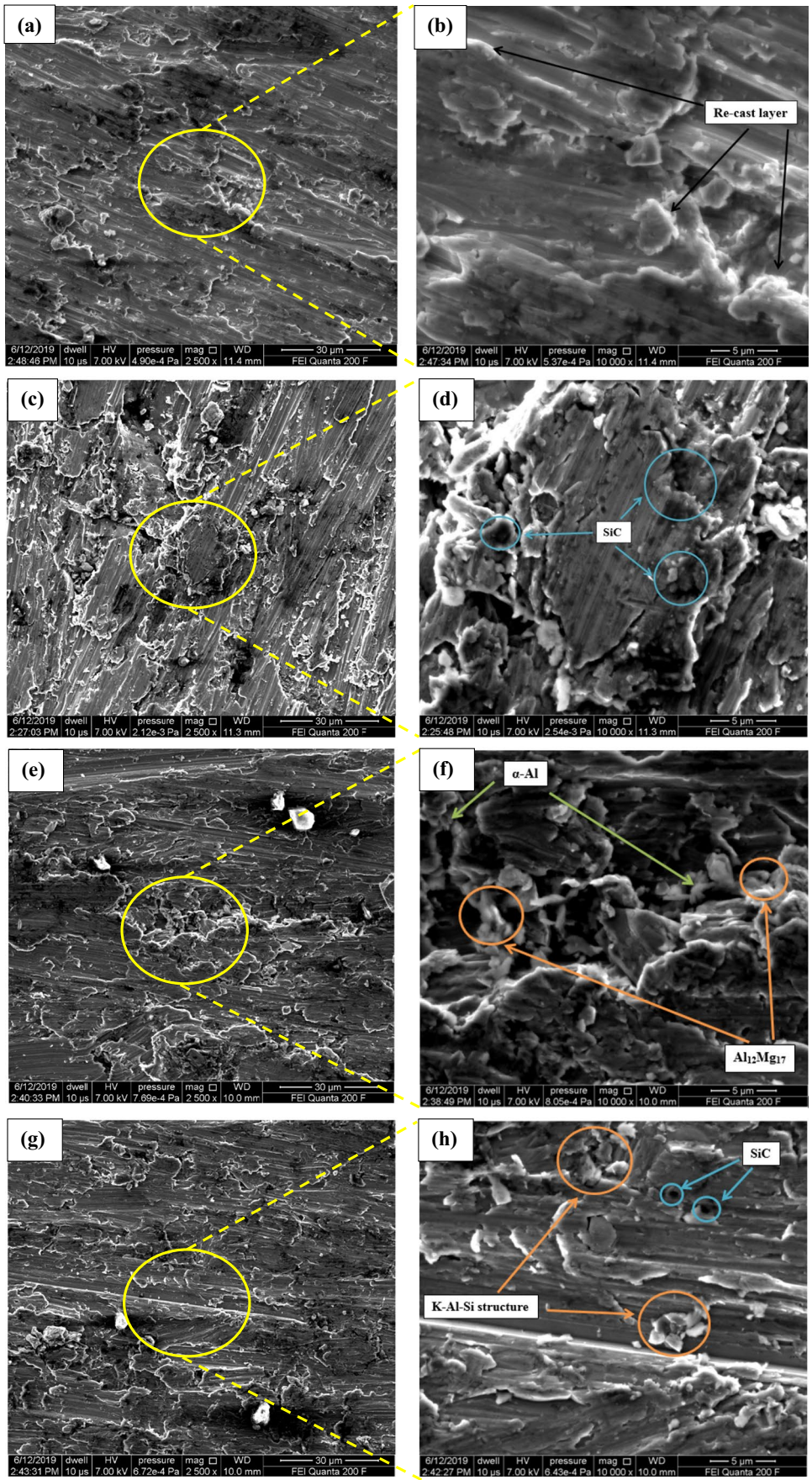


Fig. 8 FESEM images (**a**) & (**b**) AC (**c**) & (**d**) SiC-AA composite (**e**), (**f**), (**g**) & (**h**) AV-AA composite

producing a composite as desired [34]. The XRD analysis of as-cast Al6061 and MMC samples are shown in Fig. 7.

In the as-cast sample, the major peaks correspond to aluminium and other alloying elements such as Si, Mg and Fe are identifiable. Higher intermetallic compounds (IMC) are forming in the hybrid composite as compared with SiC-AA composite sample. Higher brittle IMC results in higher hardness and improved wear resistance compared with the as-casted sample. FESEM images of the fabricated samples are exhibited in Fig. 8. When aluminium is solidified, there is the evolution of aluminium oxide on the surface of the solidified casting which is visible in Fig. 8(a) & (b) as a recast layer. K-Al-Si structure, $Al_{12}Mg_{17}$ and α-Al particles are seen in the hybrid composite. This justifies the formation of a high amount of IMC in the hybrid composite. Porosity analysis is calculated with the help of Eq. 1. The amount of porosity in the as casted sample is around $0.8 \pm 0.01\%$ whereas for SiC-AA and AV-AA is $1 \pm 0.02\%$ and $1.2 \pm 0.01\%$ respectively.

3.2 Microhardness

For repeatability, the microhardness was estimated at ten different points in each specimen and the mean value was taken. Table 4 gives the microhardness test result of fabricated composites. The microhardness value of as-cast AA6061 is 87.12 HV. The highest hardness of 116.74 HV was witnessed in the AA/SiC composite. In hybrid and AA/SiC composite samples hardness are improved by 25% and 33% respectively as compared to as-cast alloy. The SiC ceramic reinforcement particles enhance the hardness of the composite sample. In addition, the formation of intermetallic particles during casting also contributes to the improvement of hardness. which is seen from the XRD plots of the composite samples. A similar type of enhancement of hardness was observed by other researchers while adding reinforcements in the composite [35, 36].

3.3 Tensile Test

The tensile specimen of the fabricated samples before and after the testing is shown in Fig. 9. The stress vs strain plot for tensile tested samples is exhibited in Fig. 10. The strength of the composite samples is greater than the as-cast AA6061 alloy with a marginal decline in the elongation. The ceramic reinforcements addition may have in brittle nature for the decrease in elongation. SiC reinforced composite sample has shown maximum tensile strength of 181 MPa and the hybrid composite also recorded 175 MPa while the as casted sample has only 139 MPa of tensile strength. The increase in UTS can be credited to the pass on of applied tension load to the heavily bonded SiC reinforcements in the aluminium matrix. Similar results are achieved by synthetic and natural reinforcements in aluminium composites [19–21, 23–25, 27–30] An escalation in dislocation density at the metal matrix-reinforcement interface, as well as differential in coefficient of thermal expansion among Al alloy and reinforcements, induce the strengthening effect [37, 38].

3.4 Wear Properties

Under a sliding load of 20 N, the impact of reinforcing particles on microhardness and wear rate of casted samples is shown in Fig. 11. SiC-AA composite has the highest microhardness when compared to hybrid composite and the as-cast AA6061 material. The microhardness of the hybrid fabricated composite (AV-AA sample) is slightly lesser than the SiC-AA composite however, the hybrid composite has enhanced wear resistance than the SiC-AA matrix. The enhanced wear properties of the composite depend upon several attributes such as size and shape of reinforcement, and their distribution in the MMC. The presence of hard particles, behave like load-bearing members and results in the enhancement of wear resistance. The harder particles result in offering greater resistance to wear and hence reduced metal removal and eventually lower wear rate. As per Archard's wear relation, a higher hardness sample has a lower wear rate [39]. In a hybrid sample, the wear resistance is improved in spite of lower hardness than the SIC-AA sample due to detachment of $Al_{12}Mg_{17}$ and α-Al intermetallic particles during wear test protect the matrix material. Similar trending of wear test results is reported for fabrication of Al composite by various researchers [40–42].

Figure 12 gives the wear rate of the fabricated samples over sliding distance. With the inclusion of the reinforcement particles, the composite sample's wear resistance has greatly increased. Similar results are achieved by synthetic and natural

Table 4 Mechanical and wear properties of fabricated samples

Sample	Average microhardness (HV)	Tensile (MPa)	Yield strength (MPa)	Elongation (%)	Wear Rate x (10^{-3} g N^{-1} m^{-1})	Coefficient of Friction
AC	87.12 ± 0.99	139 ± 2	93 ± 2	7.33 ± 0.09	0.2295 ± 0.003	0.302 ± 0.008
SiC-AA	116.74 ± 1.95	181 ± 3	122 ± 1	6.45 ± 0.05	0.2021 ± 0.005	0.284 ± 0.006
AV-AA	109.31 ± 0.97	175 ± 2	118 ± 2	6.37 ± 0.08	0.1874 ± 0.004	0.242 ± 0.007

AC = as cast AA6061; SiC-AA = AA6061/SiC; AV-AA = AA6061/SiC/aloe vera

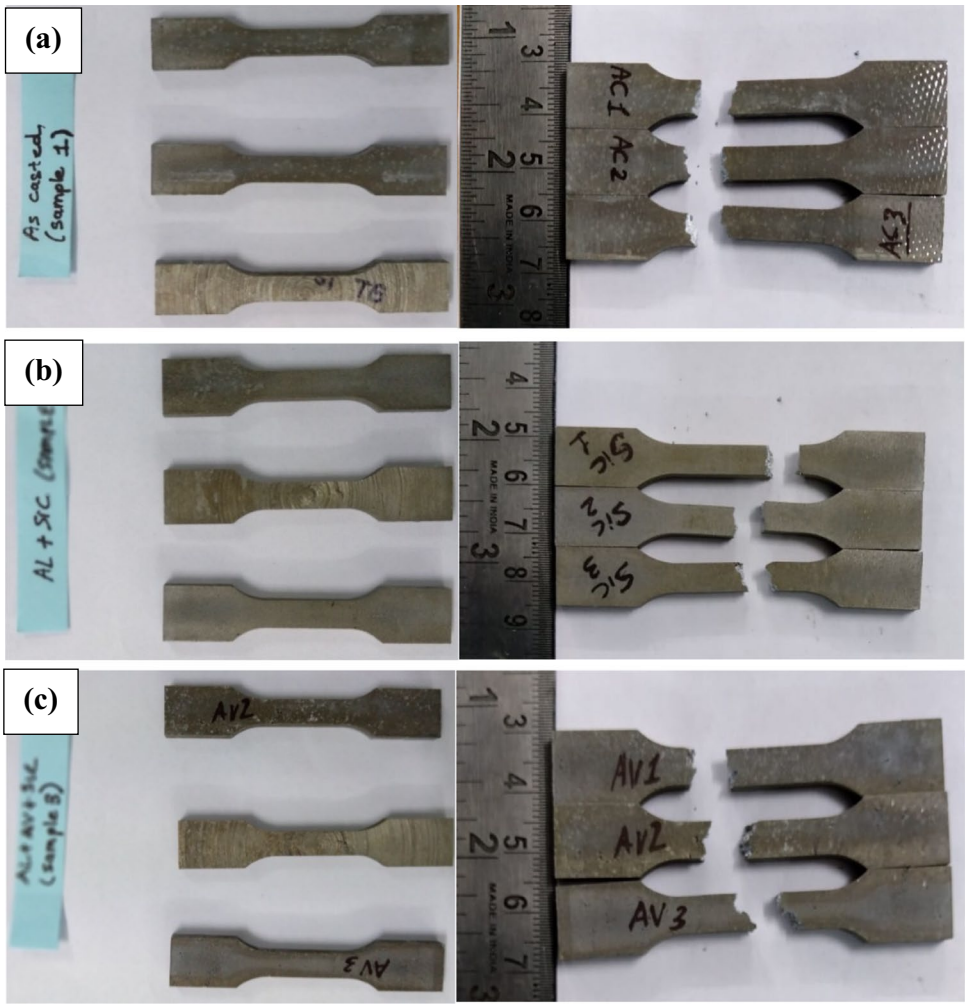


Fig. 9 Tensile test specimen (**a**) AC (**b**) SiC-AA composite (**c**) AV-AA composite

reinforcements in aluminium composites [11, 19, 20, 24, 27]. The variance in wear rate of the specimen is more clearly visible as presented in Table 4. The aloe vera reinforcement forms a thin tribo-layer between hard disc and wear pin which results in the lessening of wear loss of the metal from the wear pin.

Figure 13 depicts the fluctuation in COF as a function of sliding distance. Composites samples have a lower COF than base cast alloy. The average COF for the base metal sample is 0.302 which is greater than mono and hybrid fabricated composite samples. The lowest COF of 0.242 was witnessed in the hybrid composite sample with lower fluctuation in comparison with mono composite sample due to hard SiC particles acting as a load-bearing element whereas Aloe vera particles may act as a lubricating element, which forms a thin solid lubricating layer during wear test. The incorporation of secondary reinforcements acts as a solid lubricant in hybrid composites, which results in a reduction of COF [35]. FESEM studies were performed to study the wear mechanism of worn-out samples. FESEM micrographs are shown in Fig. 14. Abrasive groove marks and oxide layers exhibit parallel to the sliding direction in all fabricated samples. Deeper groves with shallow pits are observed in casted samples which results in a higher wear rate. On other hand, micro groves were observed in composite samples. In a hybrid composite sample, aloe vera powder reduces the detachment of material from wear specimen due to lesser COF. Abrasive wear and delamination wear is observed in as-cast and SiC fabricated sample while the AV-AA sample experiences lighter grooves with small wear debris.

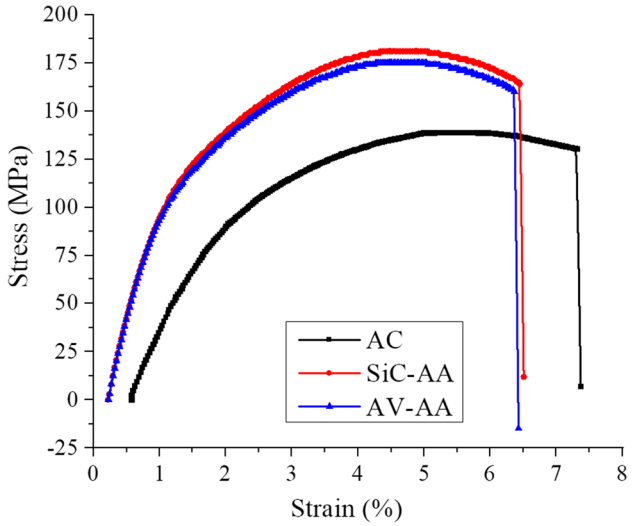


Fig. 10 Typical stress vs strain graphs of fabricated samples

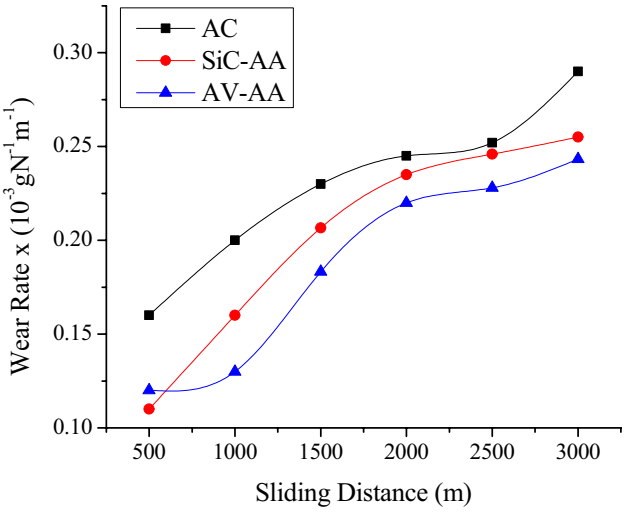


Fig. 12 Wear rate of fabricated samples

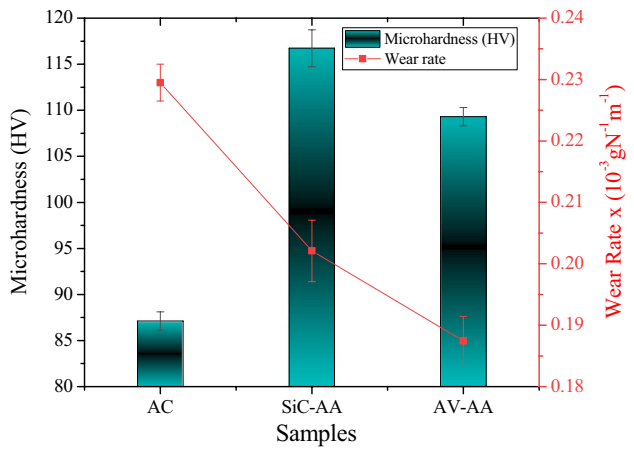


Fig. 11 Microhardness and wear rate of composite samples

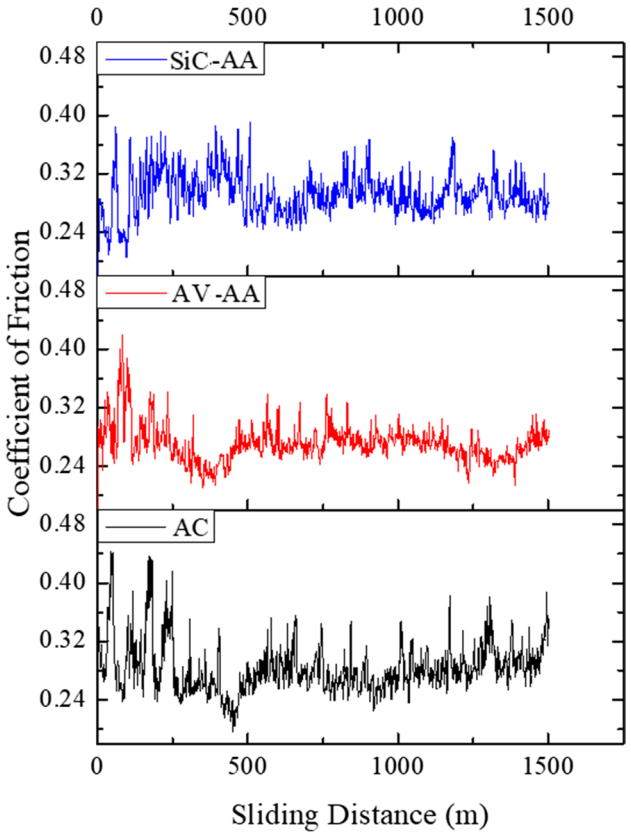


Fig. 13 COF graph of wear tested samples at 20 N load and 1 m/s sliding velocity

4 Conclusions

AA6061 with SiC particles and aloe vera particles composite was successfully prepared by stir casting process. The tensile, microhardness and wear properties of the samples were studied.

1. Aloe vera as a natural reinforcement is proven to have good compatibility with SiC reinforcement to fabricate aluminium composites.
2. FESEM and XRD results verify the existence of more intermetallic compounds and reinforcements in the composite samples.
3. Al/SiC and Al/SiC/Aloe-vera composite have an increase in microhardness value of 30% and 26% respec-

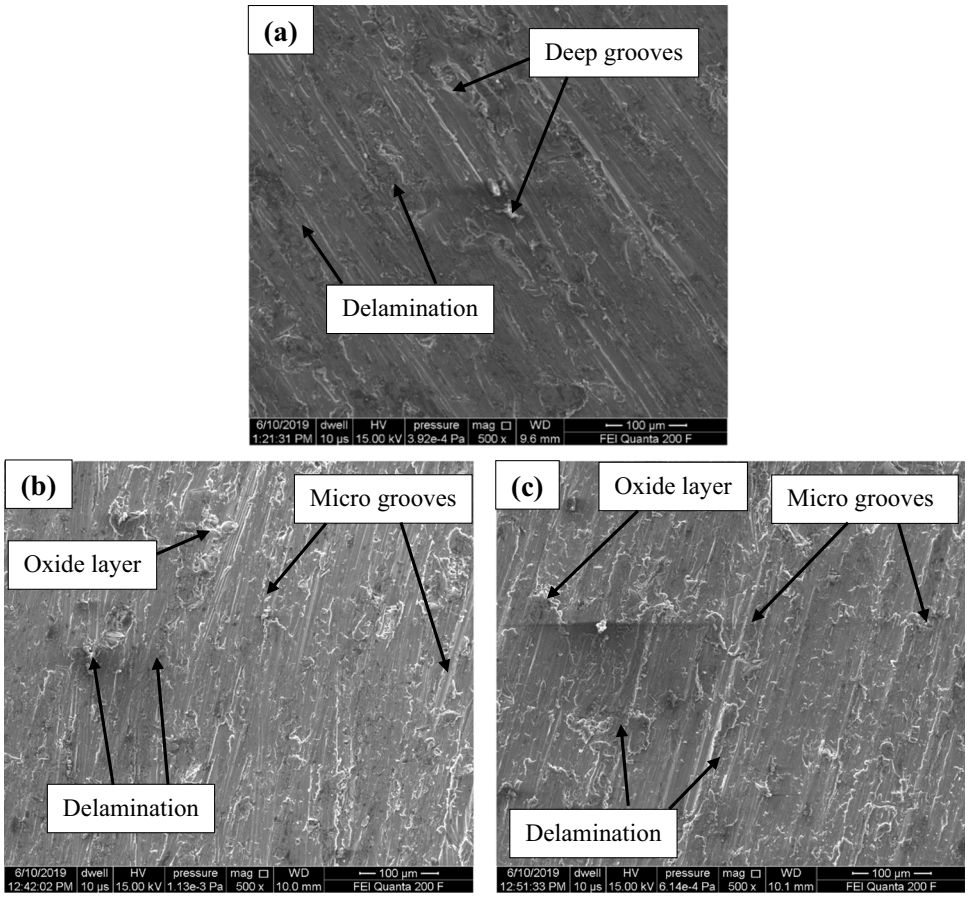


Fig. 14 FESEM images of wear samples worn surface of (**a**) AC (**b**) SiC-AA (**c**) AV-AA

tively from the Al cast alloy due to the existence of hard reinforcement particles in the fabricated composite.

4. Highest tensile strength of 175 MPa was obtained in Al/SiC composite sample with a marginal decrease in elongation compare with Al cast alloy.
5. Hybrid composite (Al/SiC/Aloe-vera) and Al/SiC composite have shown a decrease in wear rate of 11.2% and 18.3% respectively in comparison with cast samples.

Author Contributions Conception and design of study: N. Yuvaraj and Plash Issar, Acquisition of data: Yashwant Koli and Plash Issar, Analysis and/or interpretation of data: Yashwant Koli, Drafting the manuscript: N. Yuvaraj and Yashwant Koli, Revising the manuscript: G. Vedabouriswaran, N. Yuvaraj and Yashwant Koli.

Funding The author(s) received no financial support for the research, authorship and/or publication of the article.

Data Availability The authors confirm that the data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects.

Competing Interests The authors declare no competing interests.

Conflict of Interest The authors declare that there is no conflict of interest.
The authors declared that the manuscript ethics is approved as per the journal.

Consent to Participate Not applicable.

Consent for Publication The authors give full consent for the publication of this research work.

References

1. Ramanathan A, Krishnan PK, Muraliraja R (2019) A review on the production of metal matrix composites through stir casting – Furnace design, properties, challenges, and research opportunities. J Manuf Process 42:213–245. https://doi.org/10.1016/j.jmapro.2019.04.017

2. Prasad Reddy A, Vamsi Krishna P, Narasimha Rao R, Murthy NV (2017) Silicon carbide reinforced aluminium metal matrix nano composites-a review. Mater Today Proc 4:3959–3971. https://doi.org/10.1016/j.matpr.2017.02.296
3. Singh J, Jawalkar CS, Belokar RM (2020) Analysis of mechanical properties of AMC fabricated by vacuum stir casting process. SILICON 12:2433–2443. https://doi.org/10.1007/s12633-019-00338-8
4. Li Y, LI Q lin, LI D, et al (2016) Fabrication and characterization of stir casting AA6061—31%B4C composite. Transactions of Nonferrous Metals Society of China (English Edition) 26:2304–2312. https://doi.org/10.1016/S1003-6326(16)64322-4
5. Khalili V, Heidarzadeh A, Moslemi S, Fathyunes L (2020) Production of Al6061 matrix composites with ZrO2 ceramic reinforcement using a low-cost stir casting technique: Microstructure, mechanical properties, and electrochemical behavior. J Market Res 9:15072–15086. https://doi.org/10.1016/j.jmrt.2020.10.095
6. Manoj Kumar G (2021) Characterization of Al-6063/TiB2/Gr hybrid composite fabricated by stir casting process. Met Powder Rep 76:S30–S38. https://doi.org/10.1016/j.mprp.2020.06.063
7. Sajjadi SA, Ezatpour HR, Beygi H (2011) Microstructure and mechanical properties of Al-Al2O3 micro and nano composites fabricated by stir casting. Mater Sci Eng, A 528:8765–8771. https://doi.org/10.1016/j.msea.2011.08.052
8. Butola R, Yuvaraj N, Singh RP et al (2022) Evaluation of microhardness and wear properties of Al 6063 composite reinforced with yttrium oxide using stir casting process. World Journal of Engineering 19:361–367. https://doi.org/10.1108/WJE-12-2020-0645
9. Shayan M, Eghbali B, Niroumand B (2019) Synthesis of AA2024-(SiO2np+TiO2np) hybrid nanocomposite via stir casting process. Mater Sci Eng, A 756:484–491. https://doi.org/10.1016/j.msea.2019.04.089
10. Prasad SV, Asthana R (2004) Aluminum metal-matrix composites for automotive applications: Tribological considerations. Tribol Lett 17:445–453. https://doi.org/10.1023/B:TRIL.0000044492.91991.f3
11. Uthayakumar M, Aravindan S, Rajkumar K (2013) Wear performance of Al-SiC-B4C hybrid composites under dry sliding conditions. Mater Des 47:456–464. https://doi.org/10.1016/j.matdes.2012.11.059
12. Samal P, Surekha B, Vundavilli PR (2022) Experimental investigations on microstructure, mechanical behavior and tribological analysis of AA5154/SiC composites by stir casting. Silicon 14:3317–3328. https://doi.org/10.1007/s12633-021-01115-2
13. Kumar J, Singh D, Kalsi NS et al (2021) Investigation on the mechanical, tribological, morphological and machinability behavior of stir-casted Al/SiC/Mo reinforced MMCs. J Market Res 12:930–946. https://doi.org/10.1016/j.jmrt.2021.03.034
14. Kumar MS, Begum SR, Pruncu CI, Shahedi Asl M (2021) Role of homogeneous distribution of SiC reinforcement on the characteristics of stir casted Al–SiC composites. J Alloys Compd 869:159250. https://doi.org/10.1016/j.jallcom.2021.159250
15. Manivannan I, Ranganathan S, Gopalakannan S, Suresh S (2018) Mechanical properties and tribological behavior of Al6061–SiC–Gr self-lubricating hybrid nanocomposites. Trans Indian Inst Met 71:1897–1911. https://doi.org/10.1007/s12666-018-1321-0
16. Suresh S, Gowd GH, Deva Kumar MLS (2019) Tribological behavior of Al 7075/SiC metal matrix nano-composite by stir casting method. Journal of The Institution of Engineers (India): Series D 100:97–103. https://doi.org/10.1007/s40033-018-0167-2
17. Soltani S, Azari Khosroshahi R, Taherzadeh Mousavian R et al (2017) Stir casting process for manufacture of Al–SiC composites. Rare Met 36:581–590. https://doi.org/10.1007/s12598-015-0565-7
18. Hima Gireesh C, Durga Prasad KG, Ramji K, Vinay PV (2018) Mechanical characterization of aluminium metal matrix composite reinforced with aloe vera powder. Mater Today Proc 5:3289–3297. https://doi.org/10.1016/j.matpr.2017.11.571
19. Suleiman IY, Salihu SA, Mohammed TA (2018) Investigation of mechanical, microstructure, and wear behaviors of Al-12%Si/reinforced with melon shell ash particulates. Int J Adv Manuf Technol 97:4137–4144. https://doi.org/10.1007/s00170-018-2157-9
20. Alaneme KK, Sanusi KO (2015) Microstructural characteristics, mechanical and wear behaviour of aluminium matrix hybrid composites reinforced with alumina, rice husk ash and graphite. Engineering Science and Technology, an International Journal 18:416–422. https://doi.org/10.1016/j.jestch.2015.02.003
21. Gosangi NR, Dumpala L (2019) Characterization of friction stir welded AL6063/nano fly ash composite. Mater Res Express 6:. https://doi.org/10.1088/2053-1591/ab179b
22. Panwar N, Chauhan A (2019) Parametric behaviour optimisation of macro and micro hardness for heat treated Al 6061-red mud composite. J Market Res 8:660–669. https://doi.org/10.1016/j.jmrt.2018.04.018
23. Hassan SB, Aigbodion VS (2015) Effects of eggshell on the microstructures and properties of Al–Cu–Mg/eggshell particulate composites. Journal of King Saud University - Engineering Sciences 27:49–56. https://doi.org/10.1016/j.jksues.2013.03.001
24. Rozhbiany FAR, Jalal SR (2019) Influence of reinforcement and processing on aluminum matrix composites modified by stir casting route. Adv Compos Lett 28:1–8. https://doi.org/10.1177/2633366X19896584
25. Kumar BP, Birru AK (2017) Microstructure and mechanical properties of aluminium metal matrix composites with addition of bamboo leaf ash by stir casting method. Transactions of Nonferrous Metals Society of China (English Edition) 27:2555–2572. https://doi.org/10.1016/S1003-6326(17)60284-X
26. Vinod B, Ramanathan S, Ananthi V, Selvakumar N (2019) Fabrication and characterization of organic and in-organic reinforced A356 aluminium matrix hybrid composite by improved double-stir casting. Silicon 11:817–829. https://doi.org/10.1007/s12633-018-9881-5
27. Coyal A, Yuvaraj N, Butola R, Tyagi L (2020) An experimental analysis of tensile, hardness and wear properties of aluminium metal matrix composite through stir casting process. SN Appl Sci 2:1–10. https://doi.org/10.1007/s42452-020-2657-8
28. Kanth UR, Rao PS, Krishna MG (2019) Mechanical behaviour of fly ash/SiC particles reinforced Al-Zn alloy-based metal matrix composites fabricated by stir casting method. J Market Res 8:737–744. https://doi.org/10.1016/j.jmrt.2018.06.003
29. Atuanya CU, Ibhadode AOA, Dagwa IM (2012) Effects of breadfruit seed hull ash on the microstructures and properties of Al-Si-Fe alloy/breadfruit seed hull ash particulate composites. Results Phys 2:142–149. https://doi.org/10.1016/j.rinp.2012.09.003
30. Venkatesh L, Arjunan TV, Ravikumar K (2019) Microstructural characteristics and mechanical behaviour of aluminium hybrid composites reinforced with groundnut shell ash and B4C. J Braz Soc Mech Sci Eng 41:1–13. https://doi.org/10.1007/s40430-019-1800-1
31. Chaitanya S, Singh I (2018) Ecofriendly treatment of aloe vera fibers for PLA based green composites. International Journal of Precision Engineering and Manufacturing - Green Technology 5:143–150. https://doi.org/10.1007/s40684-018-0015-8
32. Alam MT, Azhar M, Rafat Y (2022) Physical, mechanical and morphological characterization of A356/Si3N4 nanoparticles stir casting composites. Journal of Engineering Research. https://doi.org/10.36909/jer.16205
33. Tanwir Alam M (2021) Physical, corrosion and microstructural analysis of A356/SiC nanocomposites fabricated through stir casting process. Mater Sci Forum 1034:73–86. https://doi.org/10.4028/www.scientific.net/MSF.1034.73

34. Vedabouriswaran G, Aravindan S (2019) Wear characteristics of friction stir processed magnesium RZ 5 composites. J Tribol 141:. https://doi.org/10.1115/1.4042039
35. Arora G, Sharma S (2017) A review on monolithic and hybrid metal–matrix composites reinforced with industrial-agro wastes. J Braz Soc Mech Sci Eng 39:4819–4835. https://doi.org/10.1007/s40430-017-0910-x
36. Verma AS, Cheema MS, Kant S, Suri NM (2019) Porosity study of developed Al–Mg–Si/Bauxite residue metal matrix composite using advanced stir casting process. Arab J Sci Eng 44:1543–1552. https://doi.org/10.1007/s13369-018-3613-4
37. Hynes NRJ, Sankaranarayanan R, Tharmaraj R et al (2019) A comparative study of the mechanical and tribological behaviours of different aluminium matrix–ceramic composites. J Braz Soc Mech Sci Eng 41:1–12. https://doi.org/10.1007/s40430-019-1831-7
38. Madhukar P, Selvaraj N, Rao CSP, Veeresh Kumar GB (2019) Tribological behavior of ultrasonic assisted double stir casted novel nano-composite material (AA7150-hBN) using Taguchi technique. Compos B Eng 175:107136. https://doi.org/10.1016/j.compositesb.2019.107136
39. Karthik A, Srinivasan SA, Karunanithi R et al (2021) Influence of CeO2 reinforcement on microstructure, mechanical and wear behaviour of AA2219 squeeze cast composites. J Market Res 14:797–807. https://doi.org/10.1016/j.jmrt.2021.06.056
40. Prasad Reddy A, Vamsi Krishna P, Rao RN (2019) Tribological Behaviour of Al6061–2SiC-xGr Hybrid Metal Matrix Nanocomposites Fabricated through Ultrasonically Assisted Stir Casting Technique. SILICON 11:2853–2871. https://doi.org/10.1007/s12633-019-0072-9
41. Sardar S, Karmakar SK, Das D (2020) Experimental investigation on two-body abrasion of cast aluminum-alumina composites: influence of abrasive size and reinforcement content. J Tribol 142:1–21. https://doi.org/10.1115/1.4045378
42. Jojith R, Radhika N (2021) Reciprocal dry sliding wear of SiCp/Al–7Si-0.3 Mg functionally graded composites: Influence of T6 treatment and process parameters. Ceram Int 47:30459–30470. https://doi.org/10.1016/j.ceramint.2021.07.225

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.

Materials Research Express

ACCEPTED MANUSCRIPT • OPEN ACCESS

Mechanical properties of high strength concrete incorporating Chopped Basalt fibers: Experimental and Analytical Study.

To cite this article before publication: MOHIT GUPTA *et al* 2022 *Mater. Res. Express* in press https://doi.org/10.1088/2053-1591/aca644

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 **© 2022 The Author(s). Published by IOP Publishing Ltd**.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by/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 and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated 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 30/11/2022 at 04:10

Mechanical properties of high strength concrete incorporating Chopped Basalt fibers: Experimental and Analytical Study

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Mohit Gupta [1*], Dr. Ritu Raj [2], Dr. Anil Kumar Sahu [3]

[1*]Research Scholar, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India (mohit_2k17phdcept029@dtu.ac.in)

[2] Assistant Professor, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India (rituraj@dtu.ac.in)

[3] Professor, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India (sahuanilkr@yahoo.co.in)

Correspondence should be addressed to Mohit Gupta; mohit_2k17phdcept029@dtu.ac.in

Abstract

Many civil engineering structures are constructed in terms of aspect like economy, strength and serviceability requirement. The fibres play a prominent role in bridging the micro cracks at the early stage of crack propagation and makes the structure ductile. The impact of chopped basalt fibre on high-strength reinforced concrete composites is the focus of this research. The chopped basalt fibre hardened property of concrete was investigated and compared it with controlled concrete. Moreover, physical property of concrete i.e slump cone and compaction factor incorporating chopped basalt fibre volume fraction in variation of 0%,0.75%,1.5%,2.25%,3% to the total volume of concrete mix was also investigated. To analyse the hardened properties of chopped basalt fibre high strength concrete, 135 test samples were made and cured for 7, 14, and 28 days. The test findings show that increasing the volume percentage of fibres reduces slump from 135mm to 132mm for BFHS0.75 but does not result in a significant increase in concrete compressive strength of 77.1 MPa for BFHS0.75 However, on addition of chopped basalt fibres at 2.25% (BFHS2.25) the percentage increment of flexural strength increases by 72.8% (10.3MPa) with control mix (5.96 MPa) and after that it decreases when 3% fibres were incorporated. Similarly, split-tensile strength increases at all fibre dosage and fibre addition of 3% (BFHS3) increases by 12.28% (5.65MPa) with controlled concrete (5.04MPa). The analytical results and suggested regression model could be used in real-world situations with fiber reinforced high strength concrete related issues. There is a significant relationship between mechanical property and independent variable through ANOVA in SPSS.

Keywords: High strength concrete, flexural strength, chopped basalt fibres, split tensile strength, workability, compressive strength, regression model.

1. Introduction

High strength concrete is the requirement of today's era as lots of multi-storey structures requires advanced materials to inculcate the need of the structure [1]. The use of high strength concrete in modern structures is increasing day by day due to reduced dead load of the structural columns which in turns reduce the cost [2]. High strength concrete possesses high compressive strength greater than 60 MPa as per Indian standard code. The material requirement for the high strength concrete includes silica fume, fly ash as mineral admixture which combines with chemical admixture to attain the required workability with low w/c ratio. Silica Fume (SF) acts as a prominent role in the hydration process of high strength concrete. As the concrete is having high strength which results into brittle behaviour so to reduce the brittleness the fibers are added together with concrete [3]. Different fibers are available like polypropylene, carbon, glass fibres etc. but in this

Accepted Manuscript

study the chopped basalt fibres are taken into consideration due to lack of research on chopped basalt fibres. Basalt fiber can be regarded as the eco-friendly and safe materials to be used. This is not a new material, but the application of it, it is no doubt to be innovative in a wide range of industrial and economic activities of energy-efficiency improvements to the building, from the automotive to the aerospace, aviation, due to the high mechanical, chemical and thermal properties. Therefore, basalt fiber (BF), it has received more and more attention to as a reinforcement of the material, especially when compared with the traditional glass fiber [4]. Basalt fibres are inorganic fibre originated from volcanic rocks which is having characteristics of resistance against acids, good thermal resistance etc [7]. The basalt fibres are having high energy absorption capacity. When the content of basalt fibres was increased, the workability of high strength concrete (HSC) based on slump value dropped [5]. According to Shafiq et al. [50], Basalt Fibers (BF) can replace steel fibre, since they can be used to make lightweight cement compounds. According to Ahmed et al (2016), BF greatly increases the ductility of Fly ash/ Magnesium Phosphate Cement (FA/MPC) composites. The composites with 1% of BF showed the highest compressive strength (CS) at roughly 77.6 MPa at 28 days, which was a relative increase of 25% over the standard control sample, according to Qing et al (2018). Furthermore, they showed that mixtures with more than 1% BF incorporation showed the compressive strength started to decline, which was caused by the increased development of pores within these solid compounds, and splitting tensile strength, Flexural Strength (FS), and Flexural Tension (FT) clearly found to increase with increase of BF contents. 59% & 108% split tensile strength was increased on addition of 0.75% and 1.5% basalt fibers. Additionally, Ahmed and Chen (2018) showed that the Magnesium Phosphate Cement (MPC) matrix 0.5% BF provided the highest compressive and flexural strength at 28 days, with values of 49.42 MPa and 5.31 MPa, respectively. Even yet, they [53] also noted an important point: there was a slight loss of strength for MPC mixtures with 0.75% BF as opposed to 0.5% BF. The answer for this reduced strength containing 0.75% BF was related to the flocculation tendency of BF at higher dosages. This is consistent with Kizilkanat et al. (2015) research work. 12 mm & 24 mm basalt fibres length were investigated and tested for fresh and mechanical properties of concrete and the results were verified with ANOVA (statistical tool) to validate the results and it revealed that on changing the length of basalt fibres the workability, air content and compressive strength of concrete reduced [8]. The basalt fibres and Polyvinyl alcohol (PVA) fibres were taken to determine the mechanical properties of concrete and the results revealed that PVA fibres improved the property as compared to basalt fibre due to bridging ability of PVA [6]. There is a greater change in the post-peak residual strength of flexural performance of basalt fibre concrete as compared to control concrete [9]. Basalt fibre reinforcement for geopolymer concrete was researched by Xu et al (2021). The findings demonstrated that basalt fibre could improve the pores and raised the compressive and flexural strengths of geopolymer concrete when basalt fibre content was increased. The fiber-matrix transition zone benefited from the addition of basalt fibres, according to Saloni et al. (2020) and the right amount of basalt fibres increased the concrete's setting time and strength. Sahar A. Mostafa et al. (2022) studied their work on self-compacting concrete with basalt fiber addition for ultra-high-performance concrete and the behaviour of (Ultra high-performance basalt fiber- self compacting concrete) UHPBF-SCC is also examined in this work in relation to increased temperatures of 300°C and 600°C. Studies were done on the physical characteristics, such as segregation resistance, flowability. They found that when nanoparticles are present, mechanical properties dramatically improve—by more than 18% in CS, 32% in TS, and 28% in FS when compared to the reference mix. Paschal Chimeremeze Chiadighikaobi et al. (2022) worked on basalt fibers with light weight expanded clay composite and found that porosity index, CS and elastic modulus was improved when 1.6 % BF was added to the mix. This paper is written to understand the characterization of chopped basalt fibers composite on high strength concrete in varying proportion by keeping the volume fraction of fibers limited to 3% of total volume of concrete. As other fibers such as steel, polypropylene, carbon fibers were used to study the behaviour of fiber reinforced concrete of normal strength & still there is an ample scope to utilize the chopped basalt fibers in high strength concrete and work on its analytical study. The objective of this study is to determine the mechanical properties of concrete and compare it with M70 grade of conventional concrete with analytical study.

1.1 Motivation of the research

The civil engineering structures are encountered by impact loads during their life span of structures which is very significant from strength criterion. Major contribution of basalt fibre as construction material in the pavement construction for better flexural strength and at the same time pavement requires very high strength concrete which is subjected to impact load of heavy vehicles. This is the prime motivation of doing this research. The application of fibres includes the design of high strength pavement for airport runway, off shore constructions, high rise buildings, bridges etc. The proposed research is limited to design of M70 grade of concrete as per IS 10262:2019 and the chopped basalt fibres are incorporated in variation of 0.75%,1.5%,2.25% & 3% by the volume fraction of total concrete mix.

1.2 Problem Statement

Concrete faults become more noticeable as the service life lengthens, which causes numerous issues in engineering applications. As concrete ages, local failure mechanisms that regulate its ability to bridge are reduced bonding capacity, fibre pull-out, and

Accepted Manuscript

fibre sliding. The linear elastic fracture mechanics form the foundation of the fibre spacing theory. It is believed that the concrete has flaws and microscopic fissures. The crack enlarges as a result of the stress concentration created under the influence of the external force. By using fibre, the crack can be successfully stopped from spreading. This study represents the utilization of basalt fibres which greatly reduces the effect due to shrinkage in construction practices.

2. Experimental program

2.1 Selection of Materials

In this design of high strength concrete with and without chopped basalt fibres with proper material characterization and their tests in the laboratory. Ordinary Portland Cement (OPC) of 53 grade refers to IS 269 was used to design high strength concrete. The cement was taken from Ultratech cement plant, Sahibabad, Ghaziabad. The maximum content of OPC is restricted to 450Kg/m3. The specific gravity of OPC was calculated as 3.15 and was determined from Le-Chatelier's flask. The cement content for high strength was estimated as 427kg/m3. The specific gravity of fine aggregate in SSD condition is 2.65 and FA conforming to IS 383 belongs to grading Zone-II. The FA used to design high strength concrete is 588 kg/m3. Fine aggregate was procured from Shiv building material supplier, Jhandapur, Ghaziabad. The specific gravity of Coarse aggregate in SSD condition is 2.74 and CA conforming to IS 383 and CA % different fraction of 10- 4.75 mm & 20-10 mm is 50 & 50 respectively. Coarse aggregate was procured from Krishna stone crusher and supplier, New Delhi. Chopped Basalt fibres originated from volcanic rock is used in high strength concrete. The chopped basalt fibres are considered in the volume fraction of 0 to 3 % of the volume of concrete with increment of 0.75%. Chopped basalt fibres are purchased from CF Composite, Dilshad Garden, New Delhi. Table 2 shows the property of chopped basalt fibre. Water/ Cementitious ratio is a key parameter in deciding the strength of concrete.

For high strength concrete the water-cement ratio is kept as low as normal strength concrete. Water-cementitious ratio of 0.3 is used in this study. Normal potable water is used throughout the research work. The primarily function of superplasticizer is to reduce the water content requirement and is used for high strength concrete. In this study, Auramix 400 chemical admixture made from Fosroc, chemicals India Pvt. Ltd, Bangalore was used based on polycarboxylic technology having advanced low viscosity with high performance and it is Poly-carboxylate ether based (PCE) type superplasticizer and the optimum dosage was determined from Marsh cone test and it reduces water content to 20%. Table 3 summarize the properties of Auramix 400. Fly ash powder of class F was used which is having great affinity during heat of hydration and it is purchased from Aditya enterprise, Surat, Gujarat. Amorphous silica fume which is a pozzolonic material and is purchased from the supplier Dalton mines and minerals, New Delhi.

2.2 Mix Proportion

Total 5 mix groups were prepared and each mix designated as BFHS-0, BFHS.75, BFHS1.5, BFHS2.25, BFHS3 representing variation of chopped basalt fibres as 0%, 0.75%, 1.5%, 2.25%, 3% respectively by the volume fraction of total mix. The mix proportioning is done in accordance with Indian standards 10262:2019 [10] shown in Table-1. 10mm & 20 mm coarse aggregate were proportioned in the ratio of 50:50. The chopped basalt fibres were added in volume fraction of 0-3% of total volume of concrete. The physical properties of basalt fibre as shown in Table-2.

(a)

(b)

Accepted Manuscript

(c)

Fig.1 (a) Chopped basalt fibre of 12 mm length, (b) 150mm cube under CTM (c) Specimen under curing tank

Table 1. Mix Proportioning and Sample Designation with varying % of Basalt fibres in Kg/m^3

S.No.	Mix ID	Cement	Fly ash	Silica Fume	Water	F.Agg.	C. Agg.	Chemical admixture	w/cm	Vf (%)	BF in (Kg)
1	BFHS-0	427	80.1	26.2	161.6	588	1218	2.66	0.30	-	-
2	BFHS.75	427	80.1	26.2	161.6	588	1218	2.66	0.30	.75	19.7
3	BFHS1.5	427	80.1	26.2	161.6	588	1218	2.66	0.30	1.5	39.4
4	BFHS2.25	427	80.1	26.2	161.6	588	1218	2.66	0.30	2.25	59.1
5	BFHS3	427	80.1	26.2	161.6	588	1218	2.66	0.30	3	78.9

Table 2. Properties of Basalt fibres [11]

Properties	Length in mm	Diameter in (micrometer)	Aspect ratio (L/D)	Specific gravity	Elastic Modulus (GPa)	Tensile Strength (MPa)	Elongation at break (%)
Values	12mm (2mm tolerance)	12 (0.5mm tolerance)	1000	2.63	110	3200	3.50

Table 3. Properties of Auramix 400 Superplasticizer

Appearance	pH	Volumetric mass @ 20 deg C	Cl. content	Alkali content
Light yellow coloured liquid	Min. 6.0	1.09 kg/lit	Nil to IS:456 *	less than 1.5 g Na2O equivalent / litre of admixture.

2.3 Mixing and Specimen Casting

Cement, fine aggregate, coarse aggregate, fly ash, silica fume is weighed and mixed together for dry mixing in concrete pan mixture for 2 to 3 minutes and then chopped basalt fibres were mixed by hand for dry mixing in concrete pan mixture then water is added along with the superplasticizer to obtain the required consistency of the concrete mix. The prepared concrete poured for workability test including slump test and compaction factor test. After that, moulds were properly oiled and concrete is poured in the moulds of cubes, cylinders and

Accepted Manuscript

beams followed by compaction on vibrating machine to remove the voids in the mix and is allowed to dry for next 24 hours. The dried samples were taken to curing tank for required age. It is clearly seen from Table 4. that total number of specimens prepared for determining the hardened properties of high strength concrete were 135. The chopped basalt fibres in varying proportion were added and 45 cubes of size (150mmx150mmx150mm) were prepared to determine the compressive strength at 7, 14 & 28 days of curing. Similarly, 45 beams (150mmx150mmx700mm) & 45 cylinders (150mmx300mm) were prepared to determine the flexural strength & split tensile strength of high strength concrete.

Table 4. Total Number of Samples

Mix ID	45 Cubes (150x150x150) mm for Compressive Strength			45 Beams (150x150x700) mm for Flexural Strength			45 cylinders (150x 300) mm for Split Tensile Strength		
	7 Days	14 Days	28 Days	7 Days	14 Days	28 Days	7 Days	14 Days	28 Days
BFHS-0	3	3	3	3	3	3	3	3	3
BFHS.75	3	3	3	3	3	3	3	3	3
BFHS1.5	3	3	3	3	3	3	3	3	3
BFHS2.25	3	3	3	3	3	3	3	3	3
BFHS3	3	3	3	3	3	3	3	3	3
Total	15	15	15	15	15	15	15	15	15

3 Test Methodologies

3.1 Fresh state property of concrete

The fresh property of chopped basalt fibre reinforced concrete is done in accordance with IS: 1199-1959 for slump cone test and for compaction factor test. The slump value of fresh concrete with chopped basalt fiber was measured 135 mm which is satisfactory for the pumpable concrete and the compaction factor was 0.92 which shows that the workability of chopped basalt fibre composite is medium.

3.2 Compressive strength

Compressive strength of concrete is done in accordance with IS 516: 1959. The samples were taken out after 24 hours from the curing tank and wipe with the cloth for testing the 15cmx15cmx15cm cube in compression testing machine. The average of three samples were taken as the compressive strength of concrete. The load applied on the sample was continuously increasing at the rate of 309 kN/min until the specimen fails.

3.3 Flexural Strength

Hydraulic testing device was utilised to apply the stress on the sample. The load on specimen administered between 0.035 and 0.1 mm per minute. The peak load (P) was calculated using this test procedure and residual loads that account for the L/600 and L/150 deflection. The size of beam specimen taken is 150mm x 150mmx 700mm because of 20mm maximum size of aggregate. The testing procedure adopted for this flexural strength test on universal testing machine with loading rate of 3.924 kN/min until the specimen fails. Equation 2 represents modulus of rupture for determining the strength.

$$f = PL/Bd\text{^}2 \quad (1)$$

Where; 'L' is the length of beam, 'd' is the depth of beam, 'P' is the failure load, 'f' is the flexural strength.

3.4 Split Tensile Strength

According to ASTM C39 requirements, split tensile strength tests were performed on five cylinders for each type of mix. A 250 kN capacity universal testing equipment was used to test these specimens. 12 MPa/min of loading every minute. 45 cylinders of basalt fibers in varying proportion were added in all the samples for testing. The size of the specimen used is 150mm x 300mm. The splitting tensile test was loaded at a rate of 0.7 to 1.4 MPa/min till failure. According to ASTM C496 criteria, the average of the three specimens were taken as the split tensile strength of concrete as per formula given in Indian standard.

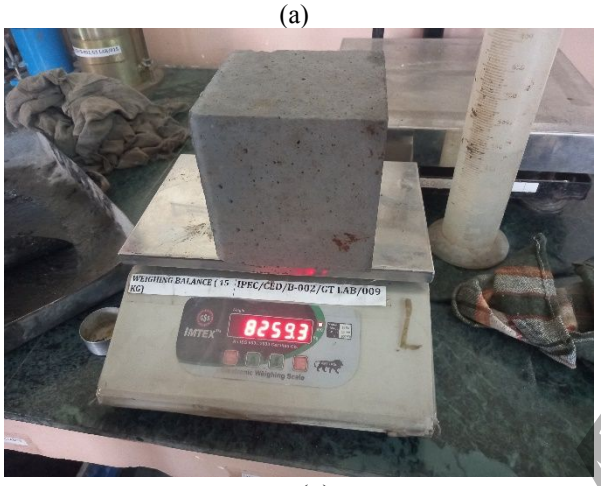
$$fst = 2P/\pi DL \quad (2)$$

Where; L is the length of cylinder, D is the diameter of cylinder, P is the failure load, fst is the Split Tensile strength.

Accepted Manuscript

(a)

(b)



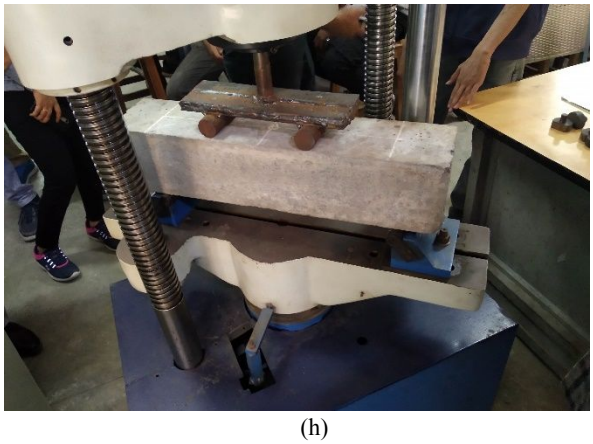
(c)

(d)

(e)

(f)

(g)



(h)

Fig. 2 (a) Cylinder failure under CTM (b) Cube failure under CTM (c) 150mm cube weight after curing (d) demolding of cube specimens (e, f) Compaction factor measurement (g) 150mm x150mm x700mm beam specimen (h) Beam specimen under universal testing machine for flexural strength.

Accepted Manuscript

4. Results and Discussion

The results of workability, mechanical property for high strength concrete with chopped basalt fibres is discussed. Regression models were developed & validated in subsequent section.

4.1 *Effect of chopped basalt fibres on workability.*

The findings reveal that on increasing the volume percentage of fibres it reduces slump. When chopped basalt fibres added in concrete the slump decreases on increasing fibre percentage and this trend is seen throughout the research work. The slump cone of BFHS-0 was found to be 135mm in which no chopped basalt fibres were added in the mixture. For the mix BFHS-0.75 the value of slump cone measured was 132 mm and follows the same trend for subsequent mixes of BFHS-1.5, BFHS-2.25, BFHS-3 with the slump value 128mm, 123mm and 115mm respectively. This reduction is due to accumulation of fibres that changes water- binder ratio and also due to the huge surface area of CBF, when more basalt fibre is added, resulting in a reduction in slump. Green concrete's workability is dependent on its proper placement and the mixture's subsequent performance. Mixing ratio and moisture content are the prominent factors that affects the workability. Jianxun Ma [37] also observed the influence of pre-soaking basalt fibre on concrete quality, as well as the fact that the workability of mixtures containing basalt fibres decreases as the amount and length of basalt fibre increases. Tumadhir [41] also looked at the impact of different basalt fibre contents on reinforced concrete workability and suggest the same pattern of slump value. Similar research Singaravadivelan R [38] and Wang X et al. [39] found that when the fibre content of concrete rises, the workability of freshly formed concrete with varying basalt fibre content diminishes. The research reported [33-36] that flow decreases as the fibre dosage and length of the fiber rises.

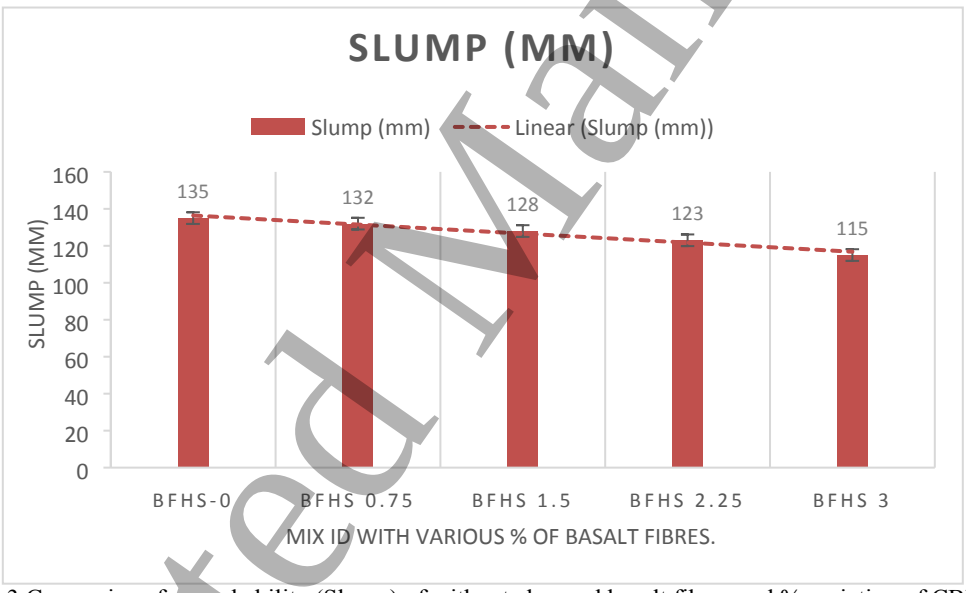


Fig.3 Comparison for workability (Slump) of without chopped basalt fibers and % variation of CBF.

4.2 *Effect of chopped basalt fibres on Compressive strength.*

There is a negligible change on concrete compressive strength on 0.75% fibres and it reduces with increase in fibre dosage up to 3%. It is clear from the Table 4 & Fig.4 that average compressive strength at 28 days for control concrete (BFHS-0) is 76.8MPa. Incorporation of 0.75% chopped basalt fibres the compressive strength remains the same and further on addition of 1.5% the average strength starts decreasing by 2.86% than control concrete. BFHS3 indicating compressive strength decreases by 3.33% than control concrete. Since presence of fibres in compression does not increase appreciably since the failure is in compressive direction and not due to fracture. Another reason is that the cohesiveness among cement paste & aggregate deteriorated due to the basalt fibres and at the same time due to very less diameter of CBF, it starts vanish during mixing and flocculated together and fibres bridging behaviour lacking which in turns reduces the compressive strength and increases the brittleness [18-21]. The increased fibre dosage in cubes shows there is a decrease in the crack width, although compressive strength in terms of load carrying capacity do not meet the requirement. [22-25] Tumadhir and Borhan [41] found similar results on different percentage of fibres and reported that influence of basalt fibres is very less on compressive strength. Wang and Zhang [42] and Revade [43] also support the similar trend during their research. This could be due to increased fibre dosage in the mixture, which produces voids, reducing ITZ between fibers and other constituents.

Accepted Manuscript

Table 4. Average compressive strength for varying proportion of Basalt fibres

S.No.	Mix ID	Avg. strength 7 days	Avg. strength 14 days	Avg. strength 28 days
1	BFHS-0	51.2	69.1	76.8
2	BFHS.75	51.3	69.2	77.1
3	BFHS1.5	49.2	67.1	74.6
4	BFHS2.25	49.8	67.3	74.8
5	BFHS3	49.4	66.8	74.2

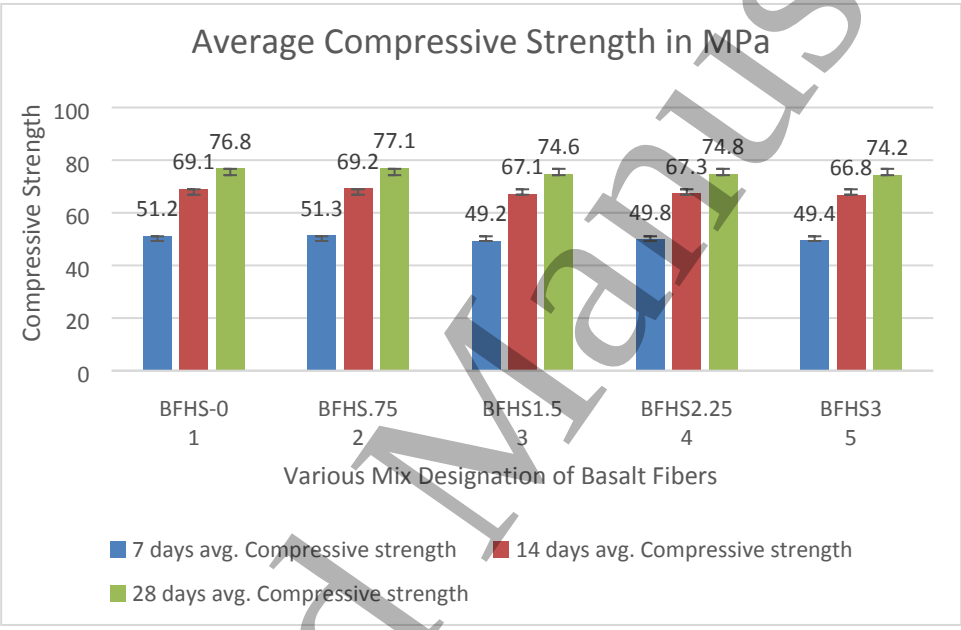


Fig.4 Average compressive strength with various mix designation

(a) (b)

Fig. 5 (a) Failure pattern of BFHS-0 & (b) Failure pattern of BFHS0.75

4.3 Effect of chopped basalt fibres on split tensile strength.

Table 5, Fig. 6 & 7 shows the average split tensile strength, peak load and percentage increment with standard concrete at 28 days. The split tensile strength with 0% fibres was 5.04MPa and it was not changed on increasing 0.75% fibres. Further on addition of 1.5% dosage of CBF the split-tensile strength increases by 5.55% than control concrete and the peak load was 376.04kN. Similarly, on 3% fibre dosage the strength enhances by 12.28% & the peak load notice was 400kN for mix designation BFHS3 and also maximum strength was achieved at 3%. With fibre fraction rises the spalling of concrete increase and also cracking resistance increases which results in improved ductility [26-28]. Revade [43] also found the similar pattern in his research work on adding 1–1.5 percent fibers tensile strength increases

Accepted Manuscript

because of high tensile property of CBF. Similar study was observed by Ketan [44] and Chen Feng [45], adding 1% basalt fibre to concrete reduces tensile strength. The only reason is the increased dosage of CBF and length, and at the same time remarkable balling effect was observed and it reduces fibre bridging in concrete.

Table 5- Average Split-Tensile strength & Peak Load for 28 days of curing

S.No.	Mix ID	Split-Tensile strength (MPa) at 28 Days	Peak Load(kN)	% Increase in strength than Control Mix
1	BFHS-0	5.04	356.25	-
2	BFHS.75	5.04	356.25	0%
3	BFHS1.5	5.32	376.04	5.55%
4	BFHS2.25	5.47	386.65	8.53%
5	BFHS3	5.65	400	12.28%

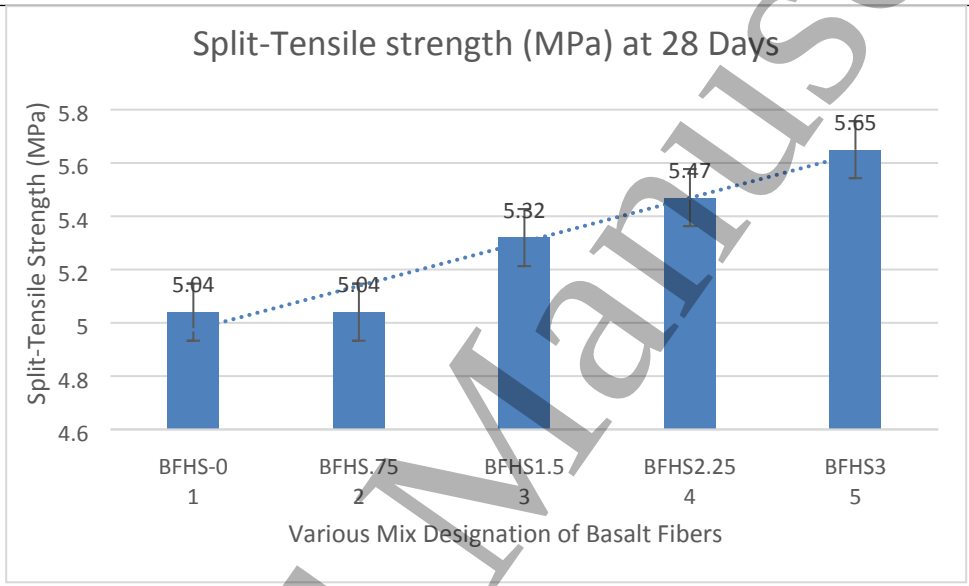


Fig. 6 Average split-tensile strength at 28 days with various mix designation

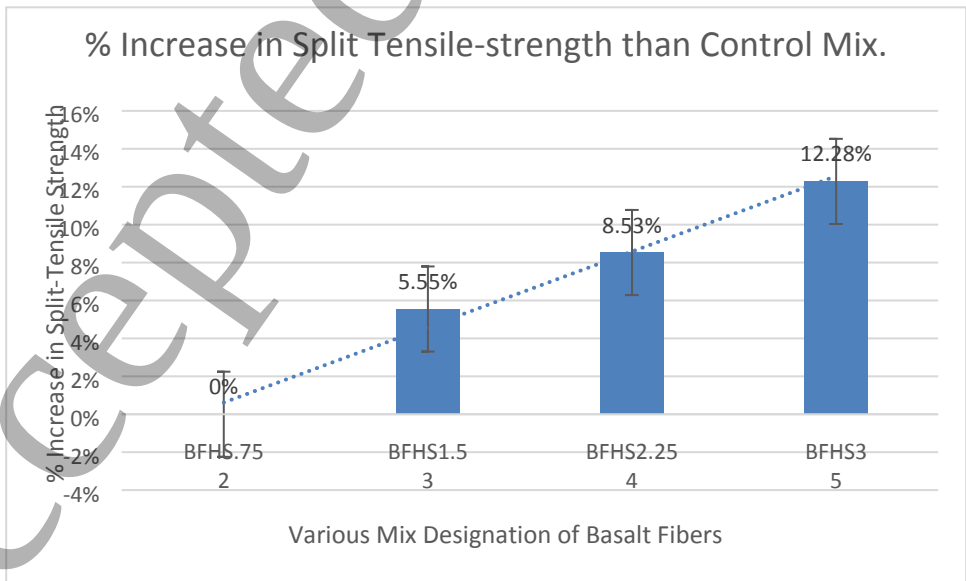


Fig.7 % Increase in split-tensile strength at 28 days with various mix designation.

Accepted Manuscript

(a) (b)

Fig. 8 (a) Failure pattern of BFHS-0 & (b) Failure pattern of BFHS3

4.4 Effect of chopped basalt fibres on flexural strength.

Table 6, Fig. 9 & 10 shows average flexural strength and % increase with control concrete at 28 days. The flexural strength with 0% fibres was 5.96MPa and on increasing 0.75% fibres it increases by 15% for BFHS.75. Further on addition of 1.5% volume fraction of chopped basalt fibres the flexural strength increases by 41.9% than control concrete. Similarly, on increasing dosage of fibres to 2.25% the average flexural strength was 10.3MPa, increases by 72.8% & was maximum for mix designation BFHS2.25 and also 2.25% found the optimum dosage of fibres in split tensile strength. As with fibre addition of 3% the flexural strength starts decreasing. Bridging action of fibres plays an important role in restrict the growth of cracks up to 2.25% fibre fraction and improves ductility [29-31] and the effect of chopped basalt fibres at 3% was insignificant. Basalt fibres were added to the concrete mix and compared to regular concrete. When basalt fibre is added properly, cement paste and aggregates are uniformly distributed; however, when basalt fibre is added excessively, more cement paste bonds to the fibre, weakening the bond between the aggregates and cement paste as a result of the reduced amount of cement paste on the aggregates. The several authors Wang and Zhang [42], Ketan [44], Solikin [46] used prisms to evaluate the flexural strength of numerous specimens and found the same pattern in flexural strength.

Table 6- Average Flexural strength & % Increase of Basalt Fibres at 28 Days

S.No.	Mix ID	Flexural strength (MPa) at 28 Days	% Increase in strength than Control Mix
1	BFHS-0	5.96	-
2	BFHS.75	6.85	15%
3	BFHS1.5	8.46	41.9%
4	BFHS2.25	10.3	72.8%
5	BFHS3	9.1	52.6%

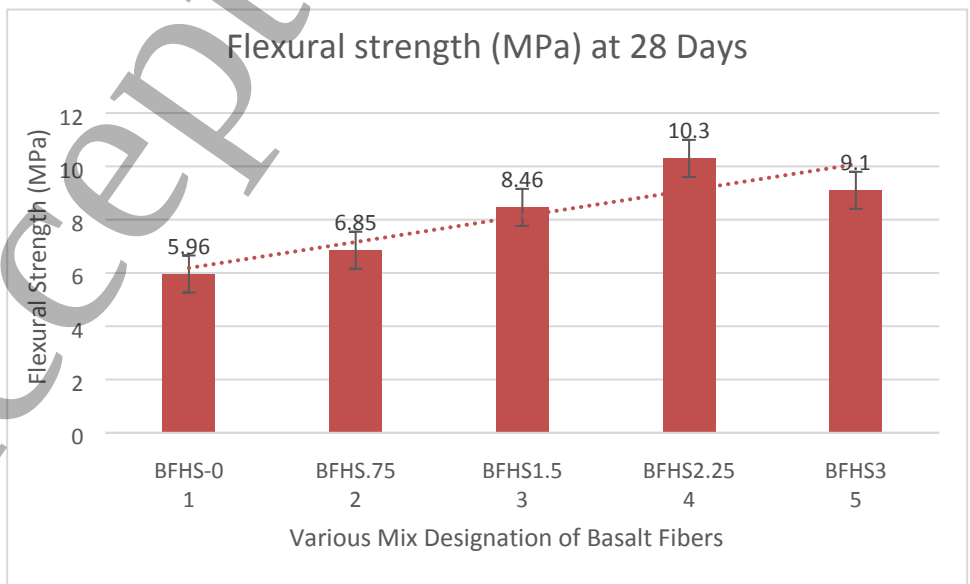


Fig. 9 Average flexural strength at 28 days with various mix designation

Accepted Manuscript

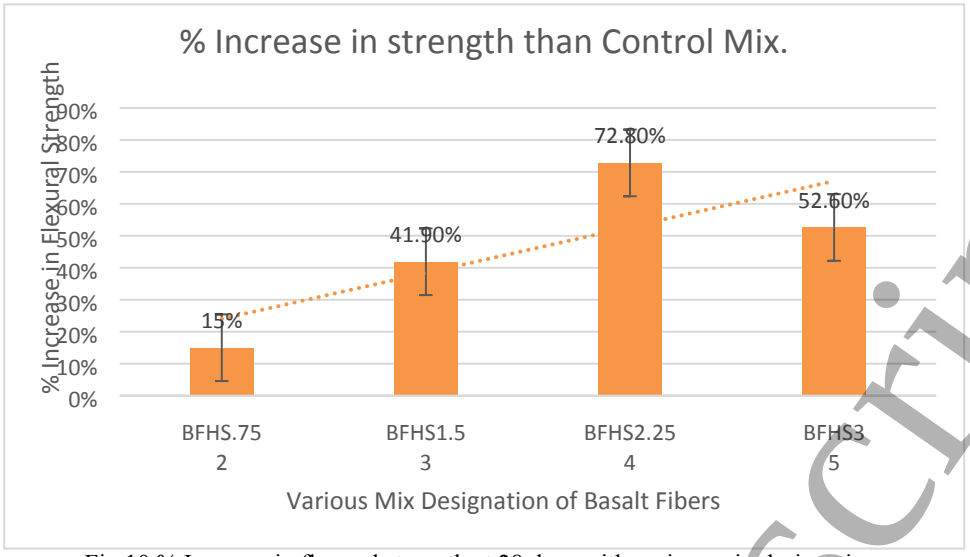


Fig.10 % Increase in flexural strength at 28 days with various mix designation.

(a) (b)

Fig. 11 (a) Cross section view of failure pattern BFHS 3 under CTM & (b) Failure pattern of BFRC 2.25 under flexure.

4.5 Regression Model developed between Split Tensile Strength and Flexural Strength

$$ST = .13337 * FS + 4.22 \quad (3)$$

$$FS = 5.618 * ST - 21.663 \quad (4)$$

Regression model is developed between ST (Split tensile strength) and FS (Flexural Strength) in equation (3) and (4). Fig. 12 shows the experimental vs predicted split tensile strength and Fig. 13 shows the experimental vs predicted flexural strength.

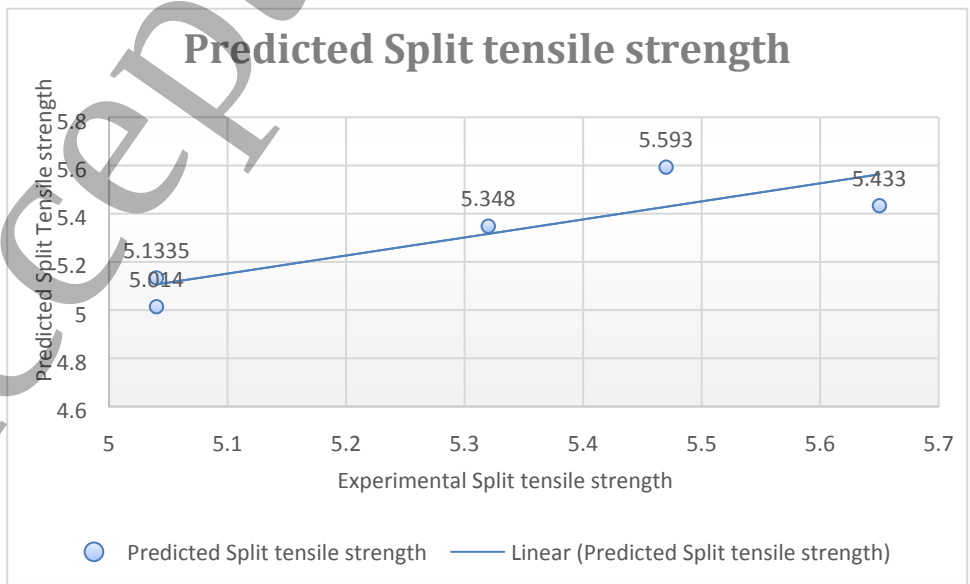


Fig.12 Experimental Split tensile strength vs Predicted Split tensile strength.

Accepted Manuscript

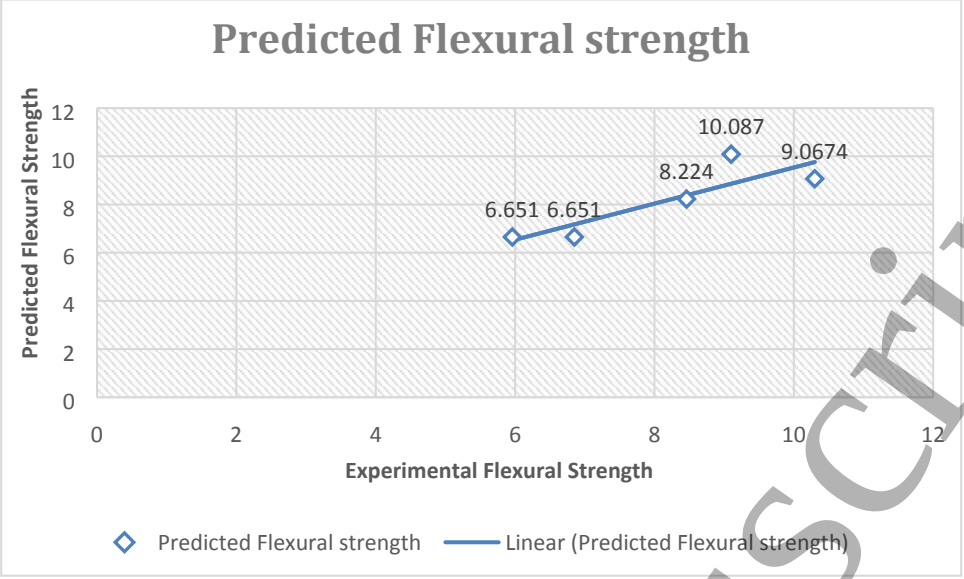


Fig.13 Experimental Flexural strength vs Predicted flexural strength.

4.6 *Regression Model developed between Compressive strength and Split Tensile Strength.*

$$CS = -4.4151 * ST + 98.85 \quad (5)$$

$$ST = -0.19943 * CS + 20.348 \quad (6)$$

Regression model is developed between CS (Compressive strength) and ST (Split tensile strength) in equation (5) and (6). Fig. 14 shows the experimental vs predicted Compressive strength and Fig. 15 shows the experimental vs predicted split tensile strength.

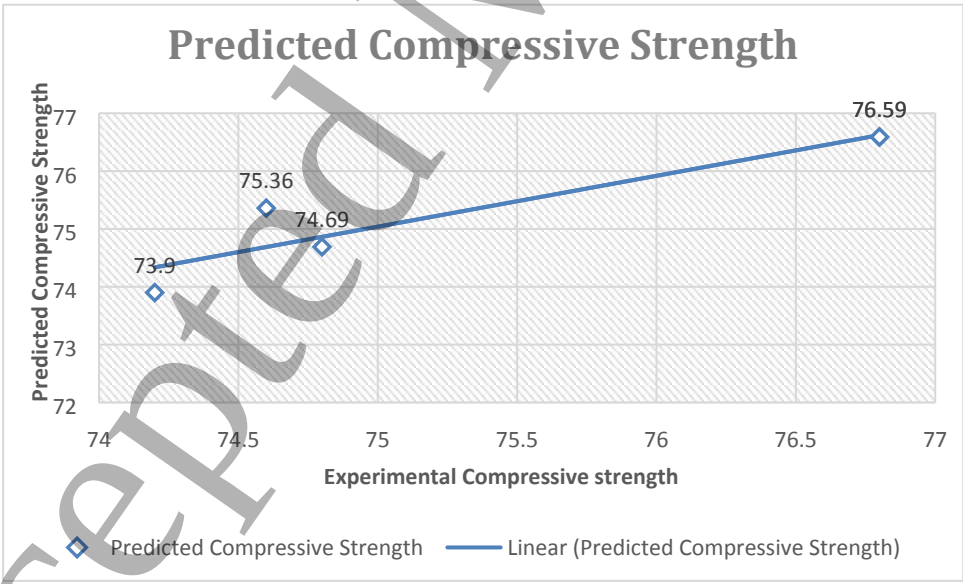


Fig.14 Experimental compressive strength vs Predicted compressive strength.

Accepted Manuscript

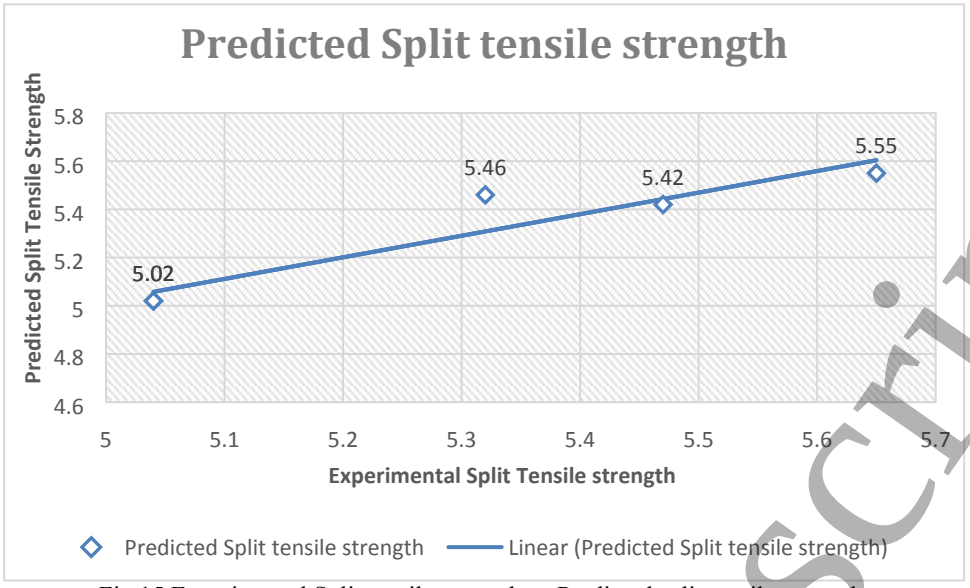


Fig.15 Experimental Split tensile strength vs Predicted split tensile strength.

Table 7. Studies showing regression models published by other authors

S.No.	Published author's work	Regression Model
1	ACI 318R-95[59]	$Fr= 0.62Fc^{0.5}$
2	ACI 363R-92[60]	$Fr= 0.94\ Fc^{0.5}$
3	Ahmad & Shah[61]	$Fr= 0.44\ Fc^{0.5}$
4	Ali Hasanzadeh[62]	$F'_c = 1.6\ (f'_f - 16.5)^2 + 91.8$
5	Mohammad Hematibahar[62]	$F'_t = 0.045\ (f'_f - 16.6)^2 + 5.28$
6	Ali Hasanzadeh[62]	$F'_t = 0.008\ (f'_c - 98.1)^2 + 5.52$

Where; Fr stands for modulus of rupture, Fc & F'_c stands for Compressive strength, F'_t stands for Split-tensile strength, f'_f stands for Flexural strength

4.7 Multiple Regression Analysis on Compressive Strength, Flexural Strength and Split tensile strength.

Multiple regression analysis is analyzed for all the mechanical properties and other variables like, cement, fine aggregate, coarse aggregate, water/ cementitious ratio, fly ash, silica fume, water, chemical admixture, Vf (%), Basalt fibre (BF). The principle of analysis is that the ANOVA test is used to analyze the differences among the means of various groups using certain estimation procedures. ANOVA means analysis of variance. ANOVA test is a statistical significance test that is used to check whether the null hypothesis can be rejected or not during hypothesis testing.

(i) Multiple Regression analysis for Compressive strength.

Null hypothesis (Ho): There is no significant relationship between compressive strength and other variables.

Alternate hypothesis (H1): There is a significant relationship between compressive strength and other variables.

Model Summary of compressive strength

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.903[a]	.816	.755	.62436

a. Predictors: (Constant), X10

Accepted Manuscript

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	5% F-Critical (Table Value).
1	Regression	5.183	1	5.183	13.295	10.13
	Residual	1.169	3	.390		
	Total	6.352	4			

a. Dependent Variable: Compressive Strength

b. Predictors: (Constant), X10

From model summary of compressive strength, it is very clear that R Square value is 0.816 and Adjusted R Square is 0.755 which is close to 1. In addition to this, ANOVA table shows that F-value i.e., 13.295 is greater than F- critical i.e., Table Value (10.13) at 5% level of significance. This signifies that null hypothesis (Ho) is rejected and alternate hypothesis (H1) is accepted. Therefore, there is a positive significant relationship between compressive strength and other variables.

(ii) Multiple Regression analysis for Split tensile strength.

Null hypothesis (Ho): There is no significant relationship between ST strength and other variables.

Alternate hypothesis (H1): There is a significant relationship between Split tensile strength and other variables.

Model Summary of Split-Tensile strength

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.974[a]	.949	.932	.06987

a. Predictors: (Constant), X10

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	5% F-Critical (Table Value).
1	Regression	.272	1	.272	55.781	10.13
	Residual	.015	3	.005		
	Total	.287	4			

a. Dependent Variable: Split tensile strength

b. Predictors: (Constant), X10

From model summary of Split-tensile strength, it is very clear that R Square value is 0.949 and Adjusted R Square is 0.932 which is close to 1. In addition to this, ANOVA table shows that F-value i.e., 55.781 is greater than F-critical i.e., Table Value (10.13) at 5% level of significance. This signifies that null hypothesis (Ho) is rejected and alternate hypothesis (H1) is accepted. Therefore, there is a positive significant relationship between Split-tensile strength and other variables.

Accepted Manuscript

(iii) Multiple Regression analysis for Flexural strength.

Null hypothesis (Ho): There is no significant relationship between Flexural strength and other variables.

Alternate hypothesis (H1): There is a significant relationship between Flexural strength and other variables.

Model Summary of Flexural Strength

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.884[a]	.781	.708	.93956

a. Predictors: (Constant), X10

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	5% F-Critical (Table Value).
1	Regression	9.458	1	9.458	10.714	10.13
	Residual	2.648	3	.883		
	Total	12.106	4			

a. Dependent Variable: Flexural Strength

b. Predictors: (Constant), X10

From model summary of Flexural strength, it is very clear that R Square value is 0.781 and Adjusted R Square is 0.708 which is close to 1. In addition to this, ANOVA table shows that F-value i.e., 10.714 is greater than F- critical i.e., Table Value (10.13) at 5% level of significance. This signifies that null hypothesis (Ho) is rejected and alternate hypothesis (H1) is accepted. Therefore, there is a positive significant relationship between Flexural strength and other variables.

5. Conclusion

- Within the parameters of this investigation, the ideal volume fraction of basalt fibres is between 2.25% and 3%. Concrete's flexural strength and split tensile strength have all greatly increased in this instance. The mechanical characteristics of basalt fibre are diminished when the volume fraction exceeds the ideal volume fraction.
- Presence of fibres in compression does not increase appreciably since the failure is in compressive direction and not due to fracture.
- Bridging action of fibres plays an important role in restrict the growth of cracks up to 2.25% fibre fraction and improves ductility and the effect of chopped basalt fibres at 3% was insignificant in flexure.
- Regression models were developed to predict the compressive strength, flexural strength and split tensile strength of concrete with addition of chopped basalt fibre. The analytical results and suggested regression model could be used in real-world situations with fiber reinforced high strength concrete related issues.
- There is a positive significant relationship between compressive strength and other variables like, cement, fine aggregate, coarse aggregate, water/ cementitious ratio, fly ash, silica fume, water, chemical admixture, Vf (%), Basalt fibre (BF) through ANOVA. Similarly, for split tensile strength and flexural strength.

Data Availability: The data used to support the findings is itself available in this research article and it is author's own experimental work.

Conflicts 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

[1] I. Nováková and L. Bodnárová 2017 High strength concrete with enhanced properties by addition of chopped

Accepted Manuscript

basalt fibres *Mater. Sci. Forum* vol. **908** 164–170 doi: 10.4028/www.scientific.net/MSF.908.164.
[2] Z. Algin and M. Ozen 2018 The properties of chopped basalt fibre reinforced self-compacting concrete *Constr. Build. Mater.* vol. **186** 678–685 doi: 10.1016/j.conbuildmat.2018.07.089.
[3] R. Ralegaonkar, H. Gavali, P. Aswath, and S. Abolmaali 2018 Application of chopped basalt fibers in reinforced mortar: A review *Constr. Build. Mater.*, vol. **164**, 589–602, doi: 10.1016/j.conbuildmat.2017.12.245.
[4] T. Ayub, N. Shafiq, and M. F. Nuruddin, 2014 InCIEC 2013 *InCIEC 2013*, 2014, doi: 10.1007/978-981- 4585-02-6.
[5] C. High, H. M. Seliem, A. El-Safty, and S. H. Rizkalla 2015 Use of basalt fibers for concrete structures *Constr. Build. Mater.* vol. **96** 37–46 doi: 10.1016/j.conbuildmat.2015.07.138.
[6] N. Shafiq, T. Ayub, and S. U. Khan 2016 Investigating the performance of PVA and basalt fibre reinforced beams subjected to flexural action *Compos. Struct.* vol. **153** 30–41 doi: 10.1016/j.compstruct.2016.06.008.
[7] J. Branston, S. Das, S. Y. Kenno, and C. Taylor 2016 Mechanical behaviour of basalt fibre reinforced concrete *Constr. Build. Mater.* vol. **124** 878–886 doi: 10.1016/j.conbuildmat.2016.08.009.
[8] E. S, B. M, and W. G 2016 Mechanical Performance of Reinforced Concrete with Different Proportions and lengths of Basalt Fibres *J. Mater. Sci. Eng.*, vol. **5** no. **5** doi: 10.4172/2169-0022.1000271.
[9] V. Mohanavel and M. Ravichandran, "Experimental investigation on mechanical properties of AA7075- AlN composites," *Mater. Test.*, vol. **61**, no. 6, pp. 554–558, 2019, doi: 10.3139/120.111354.
[10] *Bureau of Indian Standards*, "IS 10262- 2019, Concrete mix proportioning- guidelines," no. January, pp. 1–40, 2019.
[11] S. K. Kirthika and S. K. Singh Experimental Investigations on Basalt Fibre-Reinforced Concrete *J. Inst. Eng. Ser. A*, vol. **99**, no. 4, pp. 661–670, 2018, doi: 10.1007/s40030-018-0325-4.
[12] Nanni Antonio. Fatigue behavior of steel fibre reinforced concrete. *Cem Concr Compos* 1991;**13**(4):239–45.
[13] Deák T, Czigány T. Chemical composition and mechanical properties of basalt and glass fibres: a comparison. Text Res J 2009;**79**(7):645–51.
[14] Lopresto V, Leone C, De Iorio I. Mechanical characterization of basalt fibre reinforced plastic. *Compos B Eng* 2011;**42**(4):717–23.
[15] Borhan TM. Properties of glass concrete reinforced with short basalt fibre. *Mater Des* 2012; **42**:265–71.
[16] Jiang CH, McCarthy TJ, Chen D, Dong QQ 2010 Influence of basalt fibre on performance of cement mortar. *Key Eng Mater* ;426–427:**93–6.**
[17] Wu ZW, Lian HZ. High performance concrete. Beijing: Railway Press of China Beijing; 1999. p. 19–22.
[18] P. Rashiddadash, A. A.Ramezanianpour, and M. Mahdikhani Experimental investigation on flexural toughness of hybrid fiber reinforced concrete (HFRC) containing metakaolin and pumice *Construction and Building Materials*, vol. **51,** pp. 313–320, 2014.
[19] Z. Kan and Y. Li Analysis on mechanical properties and durability of the chopped basalt fiber reinforced concrete *Advanced Materials Research*, vol. **598**, pp. 627–630, 2012.
[20] H. L.Wang and Y. H. Zhong Research status and proposals of basalt fiber reinforced concrete *Advanced Materials Research*, vol. **834**, pp. 730–737, 2014.
[21] T.Ayub,N. Shafiq,M. F.Nuruddin, and S.U.Khan, Mechanical properties of high-strength concrete reinforced with PVA and basalt fibres in *InCIEC 2013*, pp. 567–575, Springer, 2014.
[22] C. Jiang, K. Fan, F.Wu, and D.Chen Experimental study on the mechanical properties and microstructure of chopped basalt fibre reinforced concrete *Materials & Design*, vol. **58**, pp. 187–193, 2014.
[23] Ayub T, Shafiq N, Fadhil Nuruddin M. Mechanical properties of high-performance concrete reinforced with basalt fibers. *Proc Eng* 2014; **77**:131–9.
[24] High C, Seliem HM, Adel El-Safty A, Rizkalla SH. Use of basalt fibers for concrete structures. *Constr Build Mater* 2015; 96:37–46.
[25] Afroughsabet V, Biolzi L, Ozbakkaloglu T. High-performance fiber-reinforced concrete: a review. J Mater Sci 2016; 51:6517–51.
[26] Li W, Xu J. Mechanical properties of basalt fibre reinforced geopolymeric concrete under impact loading. *Mater Sci Eng A Struct* 2009; 505:178–86.
[27] Fiore V, Scalici T, Di Bella G, Valenza A. A review on basalt fibre and its composites. *Compos Part B-Eng* 2015; **74:**74–94.
[28] A.E.A. Elshekh, N. Shafiq, M.F. Nuruddin and A. Fathi: *Evaluation the Effectiveness of Chopped Basalt Fiber on the Properties of High Strength Concrete*. In: Journal of Applied Sciences **14(10)**, ISSN 1812-5654, 2014.
[29] C. Higha, H. M. Seliem, A. El-Saftyc and S. H. Rizkalla: *Use of basalt fibers for concrete structures*. In: Construction and Building Materials, Volume **96,** 2015.
[30] J. Sim, C. Park, D.Y. Moon 2005 Characteristics of basalt fibre as a strengthening material for concrete structures, *Compos. Part B Eng.* **36** 504–512.
[31] S. Paulraj, N. Balasundaram, K. Sates Kumar, M. Dharshna Devi 2017 Experimental studies on strength and SCC characteristics of basalt fiber reinforced concrete, *Int. J. Civil Eng. Technol. (IJCIET)* **8** 704–711
[32] Güneyisi E, Atewi YR, Hasan MF 2019 Fresh and rheological properties of glass fiber reinforced self-compacting concrete with nanosilica and fly ash blended. *Constr Build Mater* **211:**349–362. https://doi.org/10.1016/j.conbuildmat.2019.03.087
[33] Elshafie S, Whittleston GS 2015 A review of the effect of basalt fibre lengths and proportions on the mechanical properties of concrete. *Int J Res Eng Technol* **4(1):** 458–465. https://doi.org/10.15623/ijret.2015.0401069
[34] Jiang C, Fan K, Wu F, Chen D 2014 Experimental study on the mechanical properties and microstructure of chopped basalt fibre reinforced concrete. *Mater Des* **58:**187–193. https://doi.org/10.1016/j.matdes.2014.01.056
[35] De Figueiredo AD, Ceccato MR 2015 Workability analysis of steel fiber reinforced concrete using slump and ve-be test. *Mater Res.* https://doi.org/10.1590/1516-1439.022915
[36] Li MO, Li VC 2013 Rheology, fiber dispersion, and robust properties of engineered cementitious composites. *Mater Struct* **46(3):**405–420.
[37] Jianxun M 2010 Experimental research on the fundamental mechanical properties of presoaked basalt fibre concrete, CICE 2010. In: The 5th International Conference on FRP *Composites in Civil Engineering*, September 27–29, 2010 Beijing, China, pp 1–4. http://www.iifchq.org/proceedings/CICE_2010
[38] Singaravadivelan R 2012 An experimental study has been carried out to evaluate the behavior of concrete elements using basalt chopped strands fibre. *In: International conference on chemical, bio-chemical and environmental sciences (ICBEE'2012)*, pp. 31–34.
[39] Wang X, He J, Mosallam AS, Li C, Xin H 2019 The effects of fibre length and volume on material properties and crack resistance of basalt fibre reinforced concrete (BFRC). *Adv Mater Sci Eng.* https://doi.org/10.1155/2019/7520549
[40] Basalt Fibre Research worldwide 2010 Report of material testing on basalt fibre mortars, testing report,

Accepted Manuscript

basalt fibres products of high quality. Available from www.basaltfibreworld.blogspot.co.uk

[41] Borhan TM 2013 Thermal and mechanical properties of basalt fibre reinforced concrete. *World Acad Sci Eng Technol* 7: 712– 715. www.waset.org

[42] Wang J, Zhang Y 2010 Experimental research on mechanical and working properties of non-dipping chopped basalt fibre reinforced concrete. In: *3rd International Conference on Information Management, Innovation Management and Industrial Engineering*, **4,** pp. 635–637

[43] Revade AB 2017 Compressive and split tensile strength of chopped basalt fibre concrete. *Int Res J Eng Technol* **4(6):**347–351

[44] Ayub T, Shafq N, Nuruddin MF 2014 Mechanical properties of high-performance concrete reinforced with basalt fibers. *Proc Eng* 77:131–139. https://doi.org/10.1016/j.proeng.2014.07.029

[45] Chen F 2013 An experimental study on Mechanical properties of Basalt fibre reinforced concrete. *Prog Ind Civil Eng Appl Mech Mater* 405–408: 2767–2770 https://doi.org/10.4028/www.scientific.net/amm.405-408.2767

[46] Solikin M 2012 High performance concrete with high volume ultra-fine fly ash reinforced with basalt fibre, PhD thesis, School of Civil, Environmental and Chemical Engineering RMIT University Melbourne, Australia.

[47] Jiang C, Fan K, Wu F, Chen D 2014 Experimental study on the mechanical properties and microstructure of chopped basalt fibre reinforced concrete. *Mater Des* **58:**187–193. https://doi.org/10.1016/j.matdes.2014.01.056

[48] Arivalagan S 2012 Study on the compressive and split tensile strength properties of basalt fibre concrete members. *Global J Res Eng Civil Struct Eng* **12(4):** 23–27. www.globaljournal.org

[49] Wang D, Ju Y, Shen H, Xu L 2019 Mechanical properties of high-performance concrete reinforced with basalt fibre and polypropylene fibre. *Construct Build Mater* **197**:464–473. https://doi.org/10.1016/j.conbuildmat.2018.11.181

[50] M.R. Ahmad, B. Chen, J. Yu 2019 A comprehensive study of basalt fiber reinforced magnesium phosphate cement incorporating ultrafine fly ash. *Compos. Part B Eng.* **168:** 204–217, https://doi.org/10.1016/j.compositesb.2018.12.065.

[51] N. Shafiq, T. Ayub, S.U. Khan 2016 Investigating the performance of PVA and basalt fibre reinforced beams subjected to flexural action. *Compos. Struct.* **153**, 30–41, https://doi.org/10.1016/j.compstruct.2016.06.008.

[52] J. Qin, J. Qian, Z. Li, C. You, X. Dai, Y. Yue, Y. Fan 2018 Mechanical properties of basalt fiber reinforced magnesium phosphate cement composites. *Constr. Build. Mater.***188:**946–955 https://doi.org/10.1016/j.conbuildmat.2018.08.044.

[53] M.R. Ahmad, B. Chen 2018 Effect of silica fume and basalt fiber on the mechanical properties and microstructure of magnesium phosphate cement (MPC) mortar. *Constr. Build. Mater.* **190:** 466–478, https://doi.org/10.1016/j.conbuildmat.2018.09.143.

[54] A.B. Kizilkanat, N. Kabay, V. Akyüncü, S. Chowdhury, A.H. Akça 2015 Mechanical properties and fracture behavior of basalt fiber reinforced concrete: An experimental study, *Constr. Build. Mater.* **100:** 218–224, https://doi.org/10.1016/j.conbuildmat.2015.10.006.

[55] Xu, J.; Kang, A.; Wu, Z.; Xiao, P.; Gong, Y. 2021 Effect of high-calcium basalt fiber on the workability, mechanical properties and microstructure of slag-fly ash geopolymer grouting material. *Constr. Build. Mater.* **302**, 124089.

[56] Saloni; Parveen; Pham, T.M. 2020 Enhanced properties of high-silica rice husk ash-based geopolymer paste by incorporating basalt fibers. *Constr. Build. Mater.* **245,** 118422.

[57] Sahar A. Mostafa, Bassam A. Tayeh, Ibrahim Almeshal 2022 Investigation the properties of sustainable ultra-high-performance basalt fibre self-compacting concrete incorporating nano agricultural waste under elevated temperatures. *Case Studies in Construction Materials* Volume **17**, e01453, https://doi.org/10.1016/j.cscm.2022.e01453.

[58] Paschal Chimeremeze Chiadighikaobi, Adedapo Adegoke Muritala, Ibrahim Abu Mahadi Mohamed, Abbas A. Abd Noor, Enoch Mayowa Ibitogbe, Ahmad Milad Niazmand 2022 Mechanical characteristics of hardened basalt fiber expanded clay concrete cylinders. *Case Studies in Construction Materials Volume **17** e01368.*

[59] Shah, S. P.; Ahmad, S. H. J. - Prestress. Concr. Inst. 1985, **30,** 92.

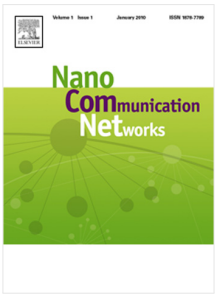
[60] ACI Committee 318 Am. Concr. Inst. 2008, 2007, **191.**

[61] Rahmani, T.; Kiani, B.; Shekarchi, M.; Safari, A. Constr. Build. Mater. 2012, **37**, 360.

[62] Hasanzadeh, A., Vatin, N. I., Hematibahar, M., Kharun, M., & Shooshpasha, I. 2022. Prediction of the Mechanical Properties of Basalt Fiber Reinforced High-Performance Concrete Using Machine Learning Techniques. Materials, 15**(20)**, 7165.

Accepted Manuscript

Journal Pre-proof



Metasurface inspired wideband high isolation THz MIMO antenna for nano communication including 6G applications and liquid sensors

Gaurav Saxena, Sanjay Chintakindi, Mohsin Ahmed Kasim, Praveen Kumar Maduri, Y.K. Awasthi, Sanjay Kumar, Sahil Kansal, Rishabh Jain, Manish Kumar Sharma, Charul Dewan

PII: S1878-7789(22)00024-2
DOI: https://doi.org/10.1016/j.nancom.2022.100421
Reference: NANCOM 100421

To appear in: *Nano Communication Networks*

Received date : 6 June 2022
Revised date : 6 October 2022
Accepted date : 1 November 2022

Please cite this article as: G. Saxena, S. Chintakindi, M.A. Kasim et al., Metasurface inspired wideband high isolation THz MIMO antenna for nano communication including 6G applications and liquid sensors, *Nano Communication Networks* (2022), doi: https://doi.org/10.1016/j.nancom.2022.100421.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Elsevier B.V. All rights reserved.

Journal Pre-proof

Metasurface Inspired Wideband High Isolation THz MIMO Antenna for Nano Communication including 6G Applications and Liquid Sensors

Gaurav Saxena[1], Sanjay Chintakindi[2], Mohsin Ahmed Kasim[2], Praveen Kumar Maduri[1], Y K Awasthi[3*], Sanjay Kumar[4], Sahil Kansal[5], Rishabh Jain[6], Manish Kumar Sharma[6], Charul Dewan[7]

[1]Department of Electronics & Communication Engineering, Galgotia College of Engineering and Technology, Greater Noida, UP India-201306

[2]College of Engineering, King Saud University, Post box 800, Riyadh,11421, Saudi Arabia

[3]Electronics & Communication Engineering, Manav Rachna International Institute of Research and Studies, Faridabad, HR India-121004

[4]Department of Information Technology, Galgotia College of Engineering and Technology, Greater Noida, UP India-201306

[5]Department of Computer Science and Engineering, JIMS Delhi-110053

[6]Department of Computer Science and Engineering, Galgotia College of Engineering and Technology, Greater Noida, UP India-201306

[7]Department of Computer Science and Engineering, Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi, Delhi 110053

Abstract: In this paper, a circular-shaped microstrip feed wideband THz antenna with a small dimension of 480×480×150μm^3 is presented on a gold-plated diffused quartz substrate with a relative permittivity of 3.50. It has an impedance operational bandwidth of 0.51-1.46 THz (80.76%) with a peak gain of 10.16 dBi. Throughout the desired bandwidth, radiation efficiency is more than 70%. This single-element antenna is transformed into a two-element MIMO antenna using a butterfly-shaped decoupling structure that included an electromagnetic coupling structure and a metasurface absorber to increase isolation and diversity characteristics along with impedance bandwidth 0.4-2.0THz. Return loss, gain, radiation efficiency, co-cross E and H-polarization, electric field, magnetic field, current density, SAR, and diversity parameters such as Envelope Correlation Coefficient (ECC), Directive Gain (DG), Total Active Reflection Coefficient (TARC), and Channel Capacity Loss (CCL)

*Correspondence Author: Y K Awasthi: E-mail: yash_ips79@rediffmail.com

Journal Pre-proof

Journal Pre-proof

are all within acceptable limits for Nano wireless applications. The proposed wideband THz MIMO antenna can also be used as a sensor to measure the proportion of crystallized sugar ($C_{12}H_{22}O_{11}$) and salt (NaCl) in water. The fields in which this antenna has applications include 6G, imaging, 3D printing, THz-wave radar, healthcare, liquid sensors with excellent sensitivity, and astronomy radiometric.

Keywords: — Circular complementary split-ring resonator- CCSRR, Rectangular complementary split-ring resonator- RCSRR, Multiband- MB, Wideband- WB, Sixth Generation- 6G, Axial Ratio - AR, Terahertz-THz, MS-Metasurface.

1. Introduction:

The need for high-speed data rates in growing wireless communication doubles every 18 months, reaching Tbps in the next several years. Because the number of users is growing day-by-day, requests for data rates in Tbps are increasing, hence THz bandwidth will be necessary to meet the aforementioned needs. In 2030, 6G technology will be vital in meeting the demands for large channel capacity and high data rate transmission. THz frequency spectrum as shown in Fig.1 is ideally suited for achieving 6G technology criteria in health monitoring systems, internet of nano-things, ultra-high-speed nano-processors, weather forecasting, and other applications [1]-[5].

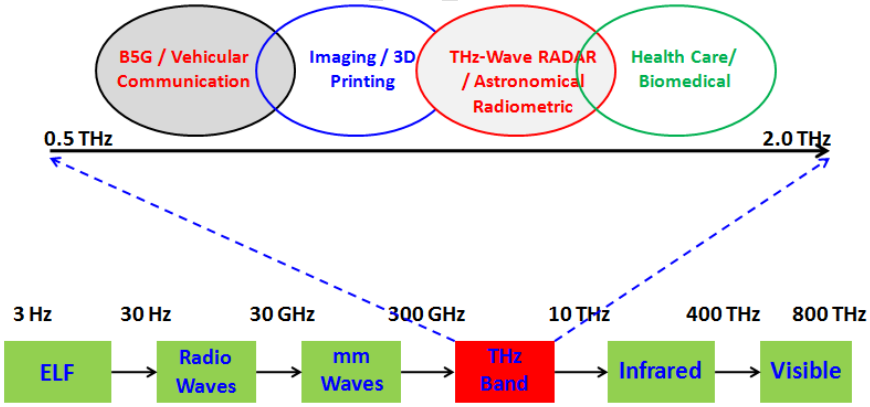


Fig. 1 Frequency spectrum with state-of-art THz Communication.

Many antennas employ a partial ground plane to accomplish the wide bandwidth, which lowers the plate capacitance and quality factor (Q) of the antenna achieving a maximum gain for medical and military applications in THz range [6]-[8].

Furthermore, through a novel decoupling technique, ultra/super-wideband MIMO (Multiple-Input-Multiple-Output) THz antennas have been built for maximum isolation, gain, and efficiency [9]-[12]. Multi-band circularly polarized with polarization diversity THz MIMO antennas are also designed

Journal Pre-proof

using novel materials like tunable metamaterial, graphene, etc. for point-to-point communication, imaging and medical applications [13]-[23].

Today's antennas function as sensors and are extensively utilized because of their easy configuration, low cost, capacity for multimodality sensing, and passive operation. The antenna also functions as a sensor for detecting moisture, temperature, ice and frost, dissolving particles, sugar content, and water quality [24]-[26]. A graphene based 1.3 - 1.6THz antenna is designed for sensing oblique and incidence angle having different polarization [27]-[28]. A monopole/UWB/T-shaped slot antenna is used as a sensor for measuring different concentration of sugar and salt, frost and ice in water in terms of magnitude of return loss and frequency shift [29]. If the temperature rises by 25° to 90°, the printed antenna's 1.2 - 5.8GHz frequency works as a temperature sensor and changes in the s-parameters are more than 3.5 dB [30].

In this article, a circular-shaped antenna on a quartz substrate (ε_r = 3.5) with a height of 150µm is designed with an elliptical-shaped slot etched to prevent charge accumulation on the patch and RCSRR & CCSRR-shaped electromagnetic coupling structures is used in ground plane to convert multiband to wideband operation. CST Microwave Studio simulates all antenna parameters such as return loss, isolation, gain, and radiation efficiency. Antenna design process will be explained in section-II. The S-parameter, gain, efficiency (radiation and total), co-cross polarization, E & H, and current density distributions are studied in section-III. In section-IV and V, design and diversity performance of MIMO antenna is discussed. Section-VI contains details about the proposed design's application as a sensor and conclusion of the article is mentioned in section-VII.

2. Design Procedure and Result Discussion of Proposed Antenna:

The proposed multiband to wideband characteristic antenna is designed on a gold plated (50µm) quartz substrate with a loss tangent (tanδ) of 0.009 as illustrated in Fig. 2. This antenna is constructed up of a circular patch with an oval-shaped etched-slot with different ground structures shown in Fig. 2 with same patch. Antenna-1 has a complete gold laminated conductor ground and is capable of multiband operation at frequencies of 0.57, 0.73, 0.90, 1.07, 1.22, and 1.38 THz, as shown in Fig. 3 (a). Further, modifications in the ground plane have been made to improve the performance of antenna-1, and antenna-2 now has a partial ground plane. The findings still demonstrate multiband functioning at frequencies of 0.65, 0.78, 0.95, 1.25, 1.3, and 1.42 THz as shown in Fig. 3(a). Moreover, antenna-3 has displayed partial ground plane changes as well as electromagnetic coupling structure integrated (RCSRR) in the ground plane to modify the performance of antenna-2 from multiband to wideband from 1.05-1.45THz. Similarly, antenna-4 has displayed partial ground plane changes along with electromagnetic coupling structure integrated (CCSRR) in the ground plane to modify the

Journal Pre-proof

performance of antenna-2 from multiband to wideband from 0.9-1.50THz. Antenna-5 demonstrates the final ground plane modification by combining both electromagnetic coupling structures (RCSRR/CCSRR) within a partial ground plane to change the performance of antenna-2 from multiband to wideband from 0.51-1.46 THz as shown in Fig. 3(a).

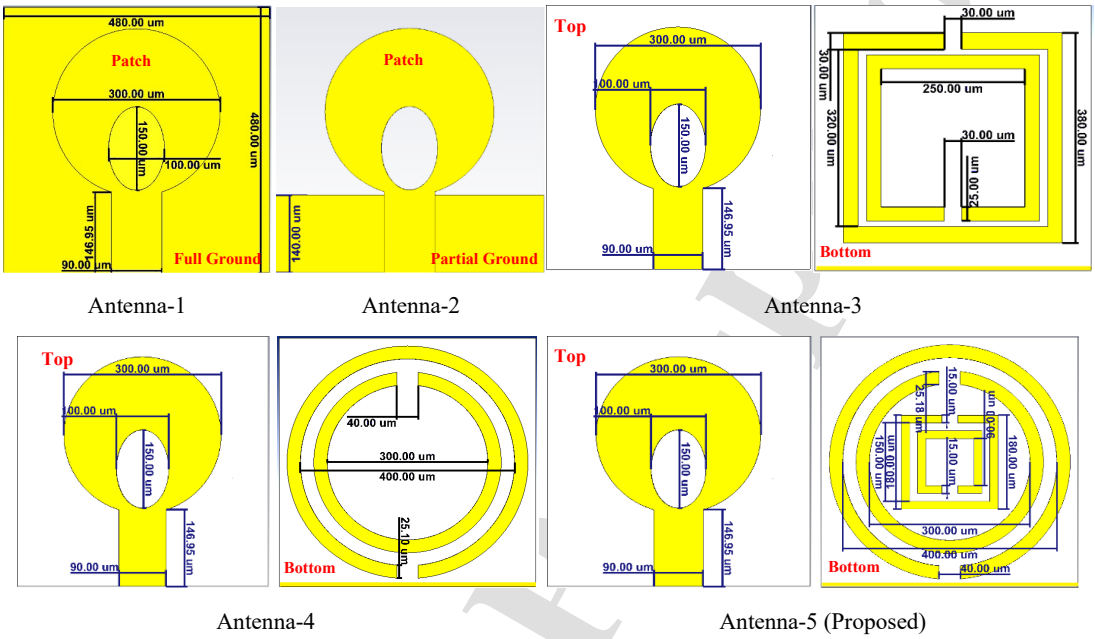


Fig. 2 A planar monopole antenna with different ground structures to convert multiband to wideband operation for THz Application.

In addition, Fig. 3(b) shows that the proposed antenna has a radiation/total efficiency of more than 65% and a maximum gain of 10.3 dBi at 0.95 THz frequency. For SAR analysis human head used for calculation of SAR having some typical parameters used in Table-1. SAR (Specific absorption rate) is a key criterion for determining the mobility of antennas that will be utilized in near-field mobile devices.

Table:1 Parameters of the human head used for calculation of specific absorption rate by the proposed antenna at constant permeability

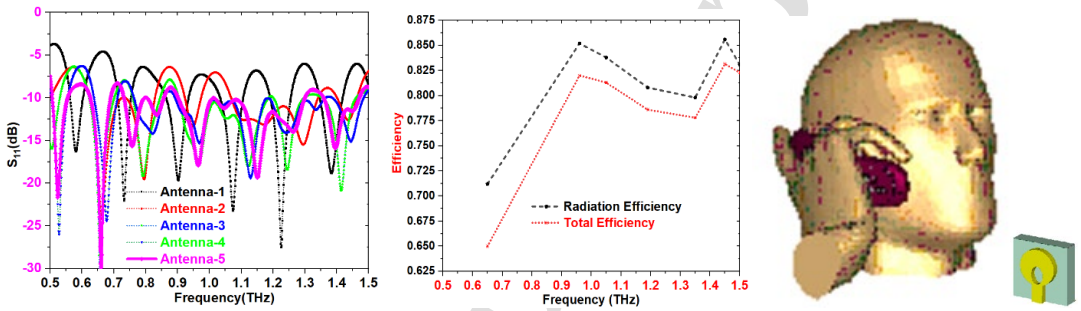
Type	Density (ρ) kg/m^3	Thermal Conductivity (σ) S/m	Heat capacity kJ/K/kg	Blood flow W/K/m^3	Metabolism rate W/m^3
Bio-tissue skin	1100	0.293	3.5	9100	1620
Bio-tissue Bone	1850	0.41	1.3	3400	610

Journal Pre-proof

Bio-tissue Brain	1030	1.13	3.675	40000	7100

The typical value of SAR for 1g and 10g of bio tissue for satisfactory application in 6G mobile communication is less than 1.6 W/Kg and 2.0 W/Kg respectively. The SAR values for 1g and 10g of bio-tissue at 0.95 THz frequency are 0.095 and 0.087 W/Kg, respectively, which are less than 1.4 W/Kg and 2.0 W/Kg of the permitted value given in Fig. 3(c) [17] without any plastic jacket or cover. The SAR of electromagnetic energy is calculated by equation (1) [17]

$$\text{Specific Absorption Rate(SAR)} = \frac{1}{V}\int_0^r \frac{\sigma(r)|E(r)|^2}{\rho(r)}dr \quad (1)$$



(a) (b) (c)

Fig. 3 Simulated results of the proposed antenna (a) S-parameter as per antennas given in Fig. 2 (b) Efficiency, and (c) Set-up for SAR estimation of head voxel model.

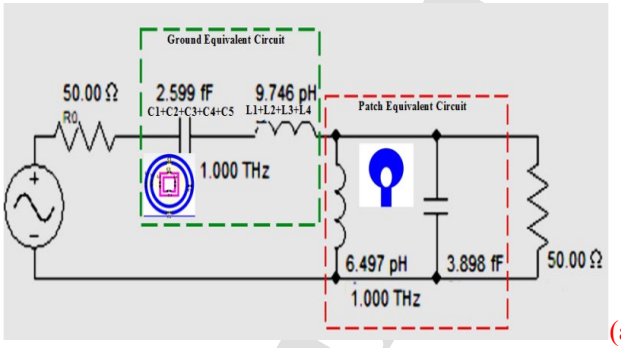
Where $E_{rms}(r)$ is rms electric field associated with input signal in V/m, $\rho(r)$ is bio tissue sample density in kg/m^3, $\sigma(r)$ is thermal conductivity of human voxel model in S/m, and V is the volume of the bio tissue sample in m^3. SAR is directly dependent upon input power used imposed by mobile phone to human head skeleton and 500mW power is used for SAR calculation.

Equivalent Network Analysis: Understanding the electrical behavior of antennas through geometrical adjustments to produce the desired outcomes is extremely challenging in planar antenna technology because planar antennas are the amalgamation of two conductor-dielectric interfaces but it is easy through analogous network analysis. Hence, the equivalent network analysis of the proposed antenna-5 is shown in Fig. 4(a)-(h) corresponding to the results given in Fig. 3(a), where the first interface (patch conductor and dielectric) forms a parallel LC equivalent network and the second interface (Dielectric-ground conductor) forms a series LC equivalent network under the lossless condition, with the values of inductances and capacitances given in Table-2. Where C_1, C_2, C_3, C_4, C_5 and L_1, L_2, L_3, L_4 are gap capacitance and inductance associated with proposed ground structure depicted in Fig. 4(a)

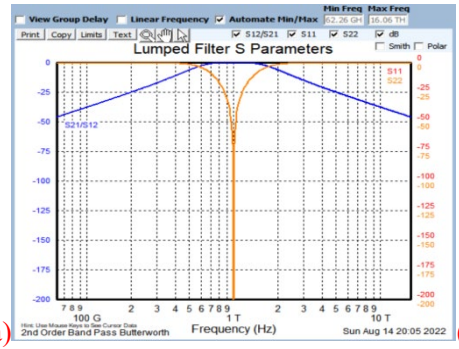
respectively. The value of gap capacitance is to be calculated $C_{gap} = \frac{\varepsilon_o \times \varepsilon_r \times w \times l}{s}$ where ε_o and ε_r are the absolute and relative permittivity, w, l, and s denotes the width, length and separation of transmission line, $C_1 = 4.905\times10^{-16}$F, $C_2 = 5.242\times10^{-16}$F, $C_3 = 6.554\times10^{-16}$F, $C_4 = 7.865\times10^{-16}$F, $C_5 = 1.647\times10^{-16}$F. So, the total capacitance associated them 2.599×10^{-15} F at central frequency 1.0 THz. Loop inductance associated with transmission line can be calculated by $L = 2\times10^{-7}\times\log_e\frac{d}{r\times e^{-0.25}}$ H where, d is the thickness of the transmission line and r denotes the radius associated with transmission line, $L_1 = 1.236$pH, $L_2 = 1.836$pH, $L_3 = 3.985$pH, $L_4 = 2.689$pH so that $L_1+L_2+L_3+L_4 = 9.746$pH.

Table-2: Inductances and capacitances values of the equivalent network of the proposed antenna-5 given in Fig. 1, where port impedance is 50 Ω

Frequency (THz)	Parallel LC (Patch-dielectric)	Series LC (Dielectric-ground)
0.57	L = 3.464pH and C = 22.51fF	L = 56.27pH and C = 1.3860fF
0.73	L = 2.112pH and C = 22.51fF	L = 56.27pH and C = 0.08447fF
0.90	L = 1.389pH and C = 22.51fF	L = 56.27pH and C = 0.5558fF
1.07	L = 1.996pH and C = 11.25fF	L = 28.13pH and C = 0.7864fF
1.22	L = 1.512pH and C = 11.25fF	L = 28.13pH and C = 0.6049fF
1.38	L = 1.182pH and C = 11.25fF	L = 28.13pH and C = 0.4748fF
0.5-1.5(Proposed)	**L = 6.497pH and C = 3.898fF**	**L = 9.746pH and C = 2.599fF**



(a)



(b)

Journal Pre-proof

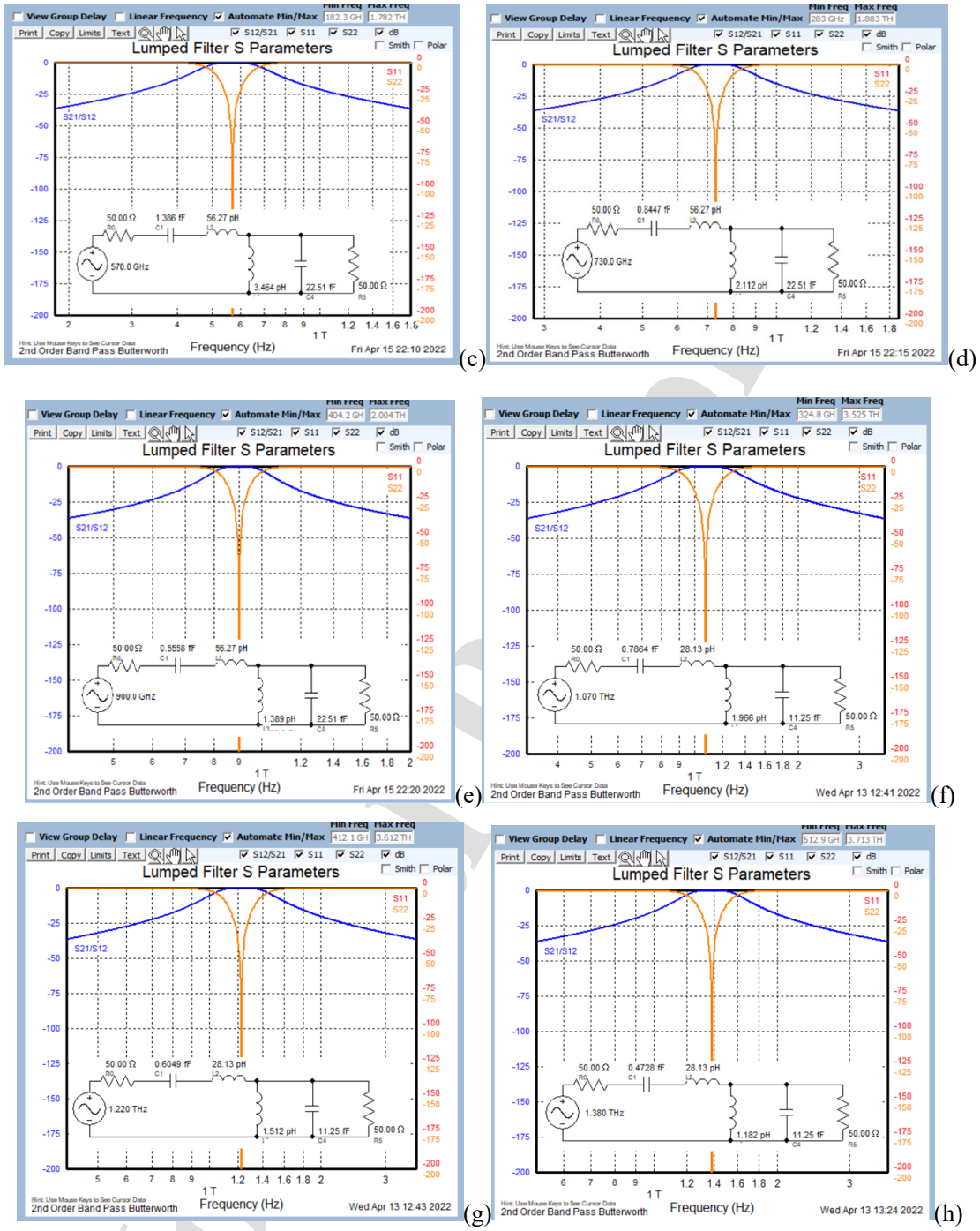
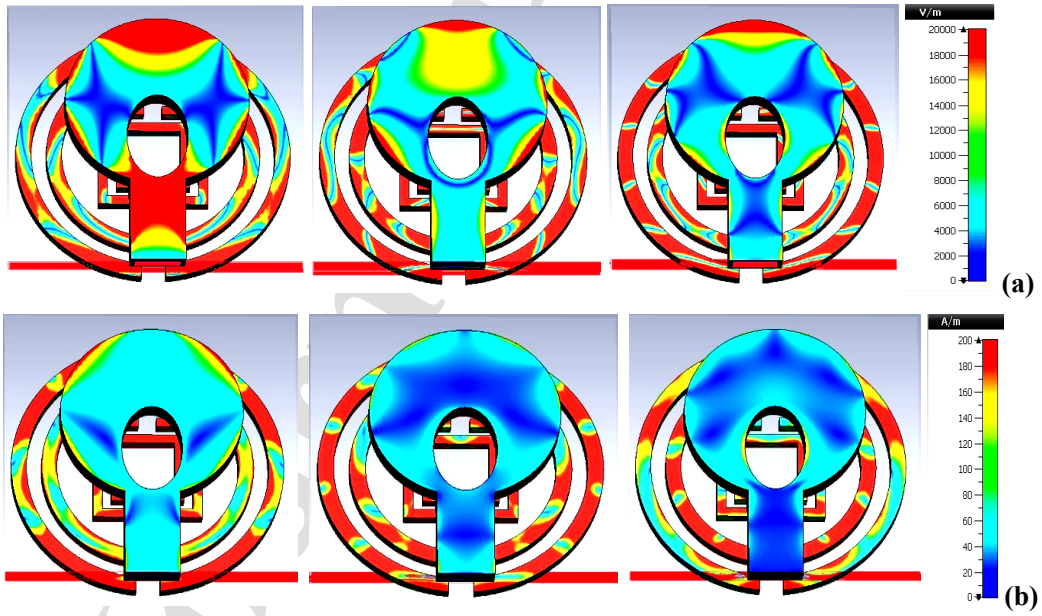


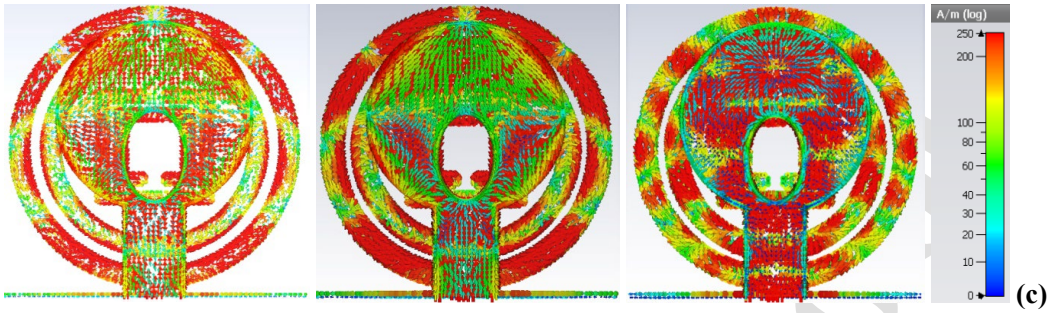
Fig. 4 Equivalent network analysis of the proposed antenna at (a) Electrical equivalent model (0.5-1.5THz) (b) S-parameters from 0.5-1.5THz (c) 0.57 THz (d) 0.73 THz (e) 0.90 THz (f) 1.07 THz (g) 1.22 THz (h) 1.38 THz.

Journal Pre-proof

3. ELECTRIC / MAGNETIC FIELD INTENSITY DISTRIBUTIONS:

The distribution of electric and magnetic field intensity in planar antennas is essential because it shows which modes are propagating across the proposed antenna surface as shown in Fig. 2 (Antenna-5). Therefore, the distribution of E-field intensity propagating TE_{13}, TE_{15}, and TE_{16} modes at 0.65, 0.95, and 1.05THz is shown in Fig. 5(a) respectively. The E-field distribution is only one half-wavelength change in the x-direction and three half-wavelength changes in the y-direction, indicating TE_{13} mode at 0.65 THz, and in similar ways at other frequencies. The current distribution in Fig. 5(b-c) shows that the magnetic field intensity is highest in the ground plane, including electromagnetic coupling structures (RCSRR and CCSRR). This means that the ground radiates the least amount of radiation at the frequencies listed below, while the patch emits the most, indicating high radiation efficiency. The distribution of electromagnetic energy in free space, which might be directed or omini-directional, is represented by the 2D/3D radiation pattern of an antenna. Figure 6(a)-(b) shows simulated far-field radiation patterns in the E- and H-planes at 0.65, 0.95, and 1.05 THz at a distance of 1.0m. The proposed antenna's radiation patterns are directional at all desired frequencies, with peak gains of 7.8, 10.3, and 9.9dBi. The difference between the 2D radiation patterns of antenna at co- (yz-plane) and cross- (xz-plane) polarization is around 80dB, which is quite far apart as seen in Fig. 6(a).



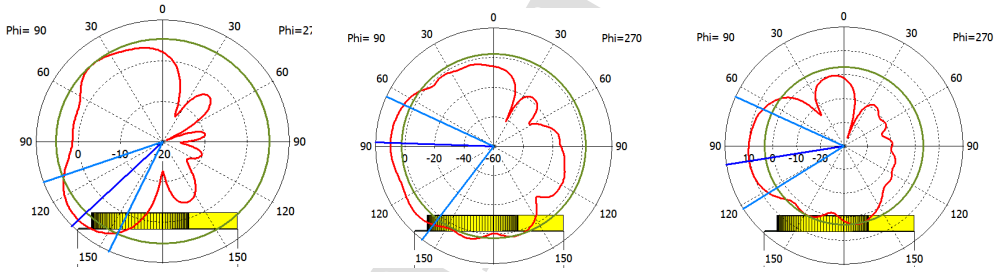


(c)

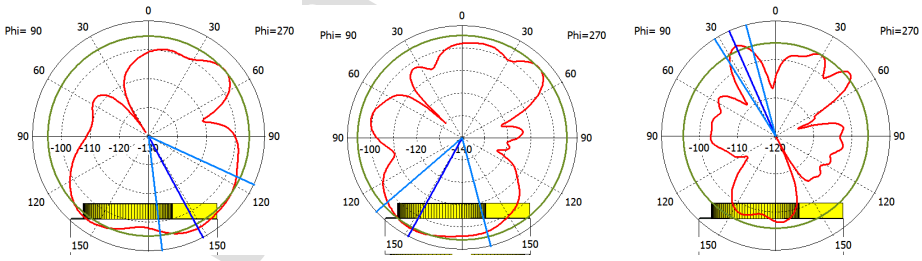
0.65THz **0.95THz** **1.05THz**

Fig. 5 Simulated results at 0.65THz, 0.95THz and 1.05THz (a) Electric field (b) Magnetic field (c) Vectored Current density

Hence, this antenna is appropriate for both multiband and wideband THz applications. To improve the proposed design's practical feasibility, this single-element antenna is transformed into a two-element MIMO antenna using a butterfly-shaped decoupling structure that included an electromagnetic coupling structure and a metasurface absorber to increase isolation and diversity characteristics along with impedance bandwidth 0.4-2.0THz.



YZ-plane



(a)

0.65THz **0.95THz** **1.05THz**

XZ-plane

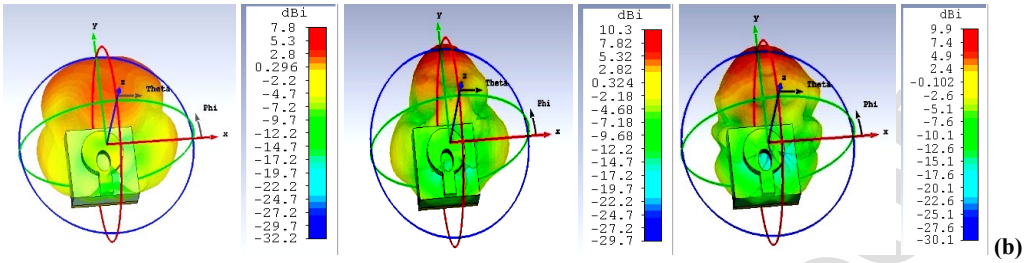


Fig. 6 Simulated far-field radiation patterns at 0.65THz, 0.95THz, and 1.05THz (a) 2D (b) 3D.

4. MIMO ANTENNA DESIGN PROCEDURE AND RESULTS:

MIMO technology is used to enhance channel capacity of antennas due to spatial diversity and multiplexing; the channel capacity of a two-element MIMO system is nearly double that of a SISO (Single-Input-Single-Output) system, making the system more reliable and efficient. The Signal-to-Noise Ratio and correlation between elements determine the quality of signal transmitted/received by MIMO antennas. SNR and correlation are affected by diversity characteristics such as DG (Directive Gain), ECC (Envelope Correlation Coefficient), TARC (Total Active Reflection Coefficient), and CCL (Channel Capacity Loss), among others.

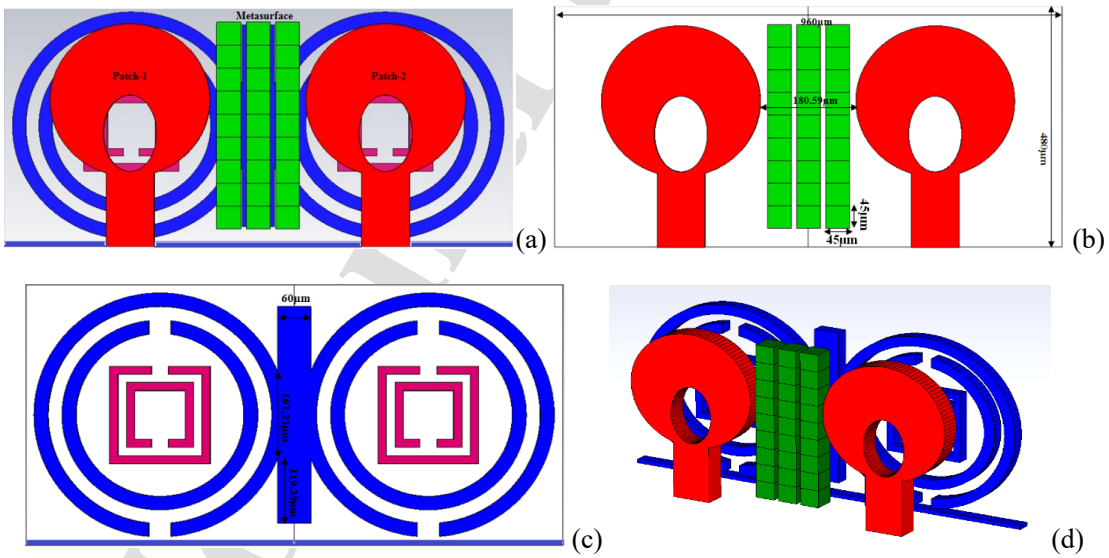


Fig.7 Two-element THz planar MIMO antenna (a) front view with ground (b) front view with dimensions (c) bottom view with dimensions (d) 3D view of the proposed design.

The aforementioned information in section-2 is used to construct a two-element MIMO antenna including a metasurface between the antenna elements to achieve more than 20dB of isolation over

the intended bandwidth of 0.4-2.2THz. Both antenna elements are positioned in the same plane at a distance of less than λ/2 to decrease the size of the MIMO antenna, as illustrated in Fig. 7. (a) - (d).

The proposed MIMO antenna's s-parameters are shown in Fig. 8, with a return loss of less than 10dB and isolation of less than 20dB for the proposed bandwidth. This antenna's metasurface can also be employed for sensing applications such as identifying cancerous tissue in the breast and brain [20].

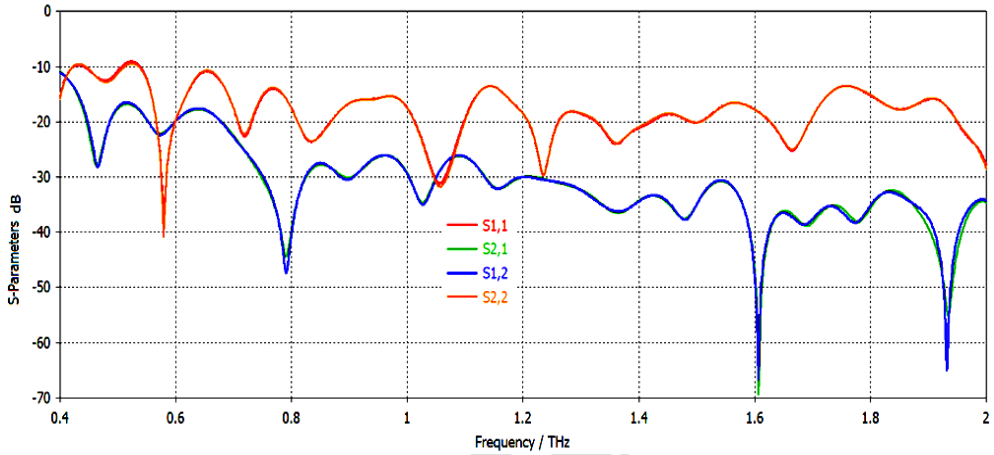
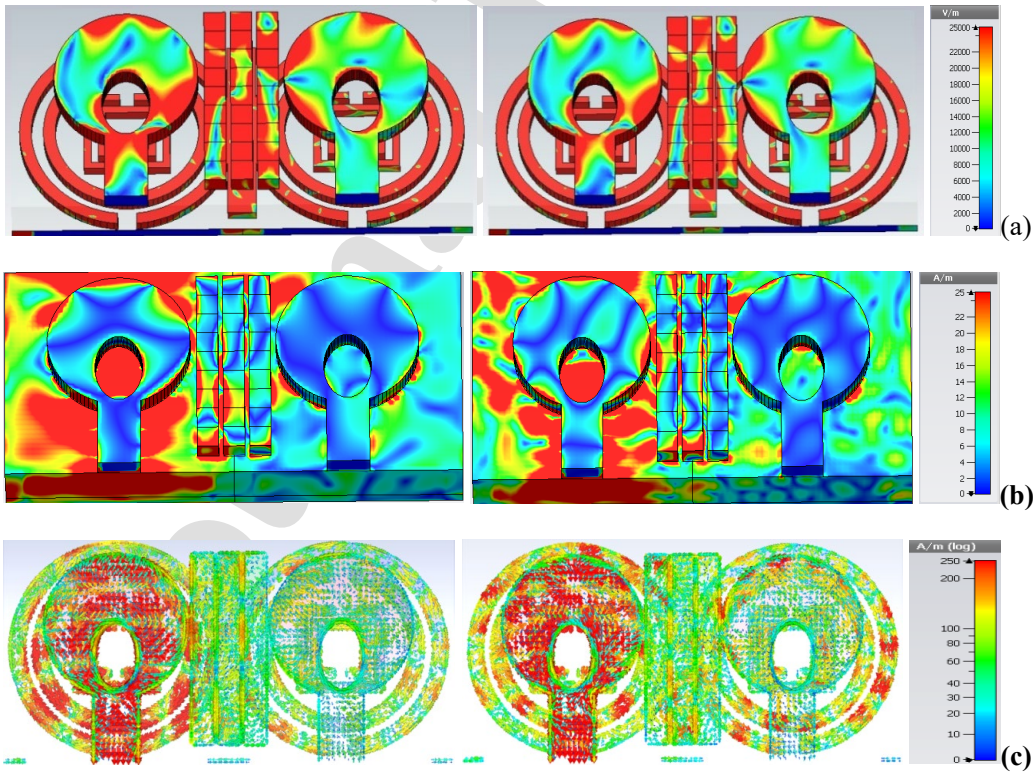


Fig.8 Simulated s-parameters of the proposed MIMO antenna.



Journal Pre-proof

0.8 THz **1.6THz**

Fig.9 Field Intensity of the proposed MIMO antenna at 0.8 and 1.6THz (a) Electric -V/m (b) Magnetic- A/m (c) Vectored current density-A/m.

The distribution of charge accumulation and charge flow over the suggested design is represented in Fig. 9 (a)-(b), which depicts the electric and magnetic field strength at 0.8 and 1.6THz. The maximum electric field is confined in the metasurface and electromagnetic coupling structures in the ground, as shown in Fig. 9(a), and radiates through both elements. Its shows that metasurface and split ring resonators are well designed at proposed band. According to Fig. 9(b), the square-shaped metasurface between the elements plays an essential role in preventing current migration from element-1 to element-2 when element-1 is excited and element-2 is terminated with a 50ohm load. Moreover, TE_{10} and TM_{11} mode is propagating which is helpful in high-speed radar, healthcare and radiometric applications.Fig.9(c) denotes the vectored current density of proposed MIMO antenna which provides that charge has been accumulated on the ground and due to gap discontinuity on the patch this antenna radiates and no current migrated from exited patch to 50 ohm terminated patch.

5. Diversity Parameters of MIMO Antenna:

Envelope Correlation Coefficient: ECC is one of the important diversity parameters to judge the amount of correlation between antenna-elements. The amount of correlation between antennas has an impact on the MIMO system's channel capacity. If the amount of correlation between antennas is low, the channel capacity will be higher, and the data speed will be faster. According to empirical evidence, ECC should be less than 0.1 [8, 17, 18-19], and ECC variation with regard to the intended bandwidth is less than 0.0012, which is suitable for high-data rate wireless transmission as shown in Fig. 10(a) and determined using equation (2) in terms of s-parameters [8,19, 22-23].

$$\rho_{ecc,ij} = \frac{\left[\sum_{n=1}^{N} S_{in}^{*} S_{nj}\right]^{2}}{\left(1-\sum_{n=1}^{N}\left|S_{ni}\right|^{2}\right)\left(1-\sum_{n=1}^{N}\left|S_{nj}\right|^{2}\right)} \quad (2)$$

Diversity Gain: The benefits of MIMO over SISO antennas in terms of signal-to-noise ratio are determined by diversity gain, and the diversity gain of a MIMO antenna in a correlated medium is determined by ECC. In wireless communication, the permissible DG value is more than 9.5dB. As a result, the suggested antenna's diversity gain is 9.9999dB, as illustrated in Fig. 10. (b), and will be determined by equation (3) [8, 19, 22-23].

Journal Pre-proof

$$DG = 10 \times \sqrt{\left(1 - \left(\rho_{ecc,ij}\right)^2\right)} \tag{3}$$

Total Active Reflection Coefficient: TARC is the square root of all reflected power at MIMO input ports divided by the total throughput power available at the same ports [8, 17-22]. It should be larger than 10dB for an applicable MIMO antenna, and it is greater than 25dB over the intended bandwidth in the proposed system employing metasurface as visible in Fig. 10(c). Equation (4) will be used to determine TARC in terms of s-parameters [8, 19, 22-23].

$$TARC = \frac{\sum_{n=1}^{N} |b_i|^2}{\sum_{n=1}^{N} |a_i|^2} = \sqrt{\frac{(S_{ii} + S_{ij})^2 + (S_{jj} + S_{ji})^2}{n}} \tag{4}$$

Where a_i and b_i is an incident (from port to patch) and reflected (patch to port) EM wave respectively and n denotes the number of antenna elements involved in MIMO system at Tx/Rx ends.

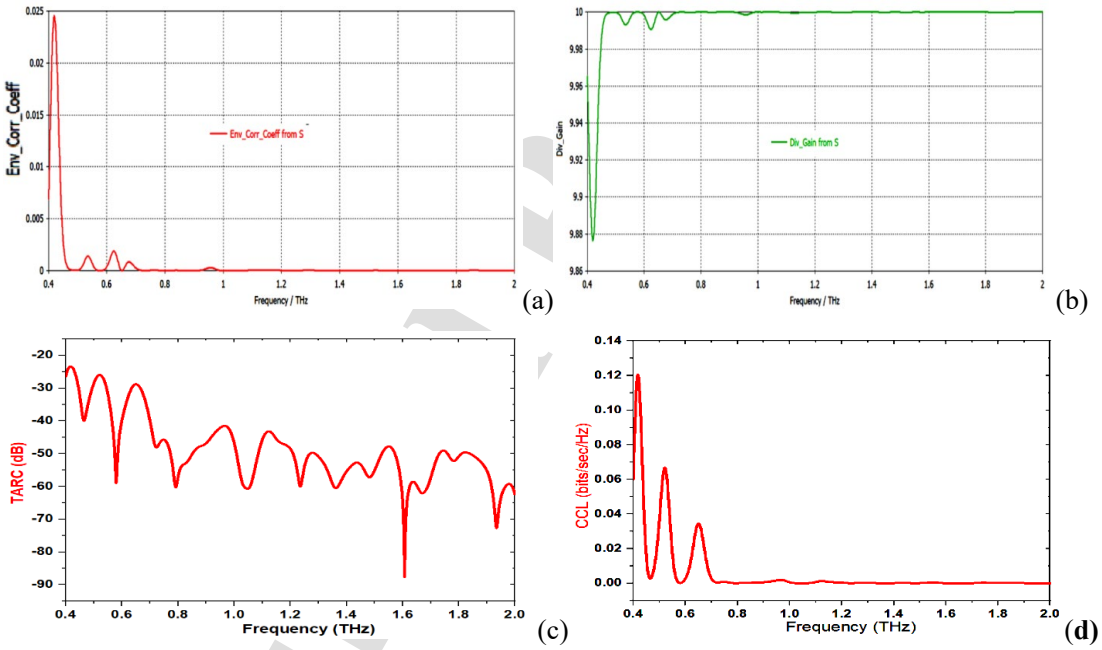


Fig.10. Simulated diversity parameters of the proposed antenna (a) Envelope Correlation Coefficient (b) Diversity Gain (c) Total Active Reflection Coefficient (d) Channel Capacity Loss.

Channel Capacity Loss: The channel capacity of the MIMO system is majorly linearly dependent upon how much antenna elements involved in particular MIMO system and amount of correlation between two antenna elements. Increased antenna elements provide better channel capacity, but when the correlation between antenna elements increases, channel capacity loss also increases, decreasing the data rate. Reduced CCL is a key requirement for assessing how well a MIMO system handles

Journal Pre-proof

actual data rate. For an effective transceiver link, the threshold value for two-port antennas should be less than 0.35 bits/sec/Hz. For the desired frequency band, the CCL of the proposed two-element MIMO antenna is less than 0.12bits/sec/Hz, as shown in Fig. 10(d) [17-23] and it is calculated by equation (5) – (7) [8, 19, 22].

$$CCL = -\log_2(\psi^p) \tag{5}$$

$$\text{where, } \psi^p = \begin{vmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{vmatrix}$$

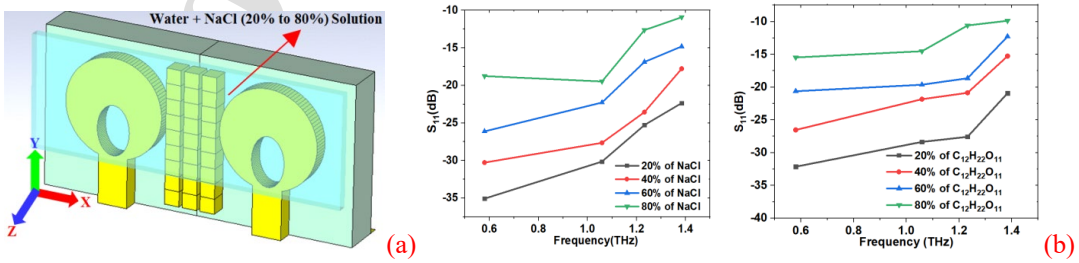
$$\rho_{ii} = 1 - \left(\sum_{n=1}^{N} \left| S_{in}^{*} S_{ni} \right| \right) \tag{6}$$

$$\rho_{ij} = - \left(\sum_{n=1}^{N} \left| S_{in}^{*} S_{nj} \right| \right) \quad \text{where i, j, and n = 1, 2} \tag{7}$$

The antenna is best suited for usage in THz applications in terms of gain, efficiency, and size, according to a comparison of the proposed antenna characteristics with existing antenna literature as illustrated in Table-2.

6. THz MIMO Antenna as a Sensor:

Antenna sensors, which use an antenna as both a sensing platform and a signal transmitter and receiver, are used to detect a wide range of compounds in both liquid and solid form. In this article, when salt and sugar are added in different quantity, the suggested antenna detects the conductivity of the water, demonstrating the sensor's nature. Salt and sugar are crystalline compounds, easy to dissolve in water and the dielectric properties change when salt and sugar added in water. These changes reduced the ionic conductivity of the water and polarization of the water content [24]. When increases the concentration of salt and sugar in water free ions decreases hence ionic conductivity also decreases, these decreases in ionic conductivity will be identified by variations in the return loss of the antenna with respect to the frequencies as indicated in Tables 3 and 4, respectively. For detection of percentage (like 20%, 40%, 60%, 80%) content salt and sugar in water. For investigating antenna as a sensor, the solution has been divided in two categories.



Journal Pre-proof

Fig. 11(a) Set-up for sensing concentration of water with NaCl and $C_{12}H_{22}O_{11}$ (b) S_{11} Vs Frequency graph for sensing concentration of water with NaCl and $C_{12}H_{22}O_{11}$

In first categories four different salt (NaCl) concentration like 20%, 40%, 60%, 80% samples with water have been taken and second categories four different sugar ($C_{12}H_{22}O_{11}$) concentration like 20%, 40%, 60%, 80% samples with water have been taken for MIMO antenna as a sensing application. The return loss variation at the peak resonant frequencies at 0.58, 1.059, 1.232 and 1.384 THz with the percentage change of salt and sugar concentrations in water are shown in Fig. 11(a) – (b). In these figures, identified that reflection coefficient increases with increases the percentage of salt and sugar in water.

Table: 3 Comparison of the proposed single antenna element with existing literature on various aspects

Frequency (THz)	Efficiency	Gain (dBi)	Size (μm^2)	Substrate	References
3.0,4.85,7.02, 8.87,9.43,10.85	0.7	13.92	90×90	polyimide	[2]
0.58-6.6	-	12.5	400×600	-	[3]
5.64-37.56	0.8	15.28	80×100	polyimide	[4]
0.22-0.32	-	9.03	400×500	polyimide	[6]
0.33-10	0.7	19	1000×1400	Roger-5880	[8]
0.52-1.47	**0.7**	**10.3**	**480×480**	**Quartz**	**[P]**

Table-4: Comparison of the proposed two-element MIMO antenna with existing reference on various aspects.

Frequency (THz)	ECC	Isolation (dB)	DG (dB)	TARC (dB)	CCL(Bits/sec/Hz) /References
3.0,4.85,7.02, 8.87,9.43,10.85	0.010	≥ 15	≥9.95	-	-[2]
5.64-37.56	0.110	≥ 20	≥ 9.93	-	- [4]
1.57,2.05,3.46,4.47, 6.82,8.3,8.9,9.6	0.500	≥ 20	≥ 8.50	-	0.25 [5]

Journal Pre-proof

0.33-10	0.0015	≥ 25	≥ 9.99	≥ 10	0.25 [8]
0.4-2.0	**0.025**	**≥ 20**	**≥ 9.95**	**≥ 20**	**0.12 [P]**

[P] Proposed Antenna

7. CONCLUSION:

A microstrip feed wideband THz antenna has been designed with excellent performance listed in Table-3 including stable omnidirectional radiation patterns. This single-element antenna has also been transformed into a two-element MIMO antenna with impedance bandwidth 0.4 - 2.0THz having exotic diversity performance mentioned in Table-4 which is useful in future 6G, High-speed RADAR, Healthcare and astronomical radiometric and liquid sensing applications. CST Microwave studio simulator is used to design and simulate proposed MIMO antenna and its application as liquid sensor.

8. ACKNOWLEDGEMENT:

This Work is funded by deanship of scientific research of King Saud University, Riyadh11451, Saudi Arabia for this work-Research Group (RG 1441-535)

9. REFERENCES:

[1] Hajiyat, Zahraa RM, Alyani Ismail, Aduwati Sali, and Mohd Nizar Hamidon. "Antenna in 6G wireless communication system: Specifications, challenges, and research directions." *Optik* 231 (2021): 166415.

[2] Singh, Mandeep, and Simranjit Singh. "Design and Performance Investigation of Miniaturized Multi-Wideband Patch Antenna for Multiple Terahertz Applications." Photonics and Nanostructures-Fundamentals and Applications 44 (2021): 100900.

[3] Keshwala, Ushaben, Sanyog Rawat, and Kanad Ray. "Design and analysis of eight petal flower shaped fractal antenna for THz applications." *Optik* 241 (2021): 166942.

[4] Keshwala, Ushaben. "Microstrip line fed sinusoidal tapered square shaped MIMO antenna for THz applications." *Optik* 247 (2021): 167905.

[5] Ashvanth, Balu, and Bactavatchalame Partibane. "Multiband characterized high gain MIMO antenna for terahertz applications." *Optical and Quantum Electronics* 53, no. 8 (2021): 1-13.

[6] Keshwala, Ushaben, Sanyog Rawat, and Kanad Ray. "Design and analysis of DNA shaped antenna for terahertz and sub-terahertz applications." *Optik* 232 (2021): 166512.

Journal Pre-proof

Journal Pre-proof

[7] Varshney, Gaurav. "Ultra-wideband antenna using graphite disk resonator for THz applications." *Superlattices and Microstructures* 141 (2020): 106480.

[8] Saxena, Gaurav, Y. K. Awasthi, and Priyanka Jain. "High isolation and high gain super-wideband (0.33-10 THz) MIMO antenna for THz applications." *Optik* 223 (2020): 165335.

[9] Das, Priyanka, Amit Kumar Singh, and Kaushik Mandal. "Metamaterial loaded highly isolated tunable polarisation diversity MIMO antennas for THz applications." *Optical and Quantum Electronics* 54, no. 4 (2022): 1-19.

[10] Upender, P., and Amarjit Kumar. "Quad-band circularly polarized tunable graphene based dielectric resonator antenna for terahertz applications." *Silicon* (2021): 1-14.

[11] Kazemi, Fatemeh. "High Q-factor compact and reconfigurable THz aperture antenna based on graphene loads for detecting breast cancer cells." *Superlattices and Microstructures* 153 (2021): 106865.

[12] Nissiyah, G. Jemima, and M. Ganesh Madhan. "Graphene based microstrip antenna for triple and quad band operation at terahertz frequencies." *Optik* 231 (2021): 166360.

[13] Kiani, Narges, Farzad Tavakkol Hamedani, and Pejman Rezaei. "Polarization controlling idea in graphene-based patch antenna." *Optik* 239 (2021): 166795.

[14] Kiani, Narges, Farzad Tavakkol Hamedani, and Pejman Rezaei. "Polarization controlling plan in graphene-based reconfigurable microstrip patch antenna." *Optik* 244 (2021): 167595.

[15] Kazemi, Fatemeh. "Dual band compact fractal THz antenna based on CRLH-TL and graphene loads." *Optik* 206 (2020): 164369.

[16] Varshney, Gaurav, Astha Verma, V. S. Pandey, R. S. Yaduvanshi, and Rajni Bala. "A proximity coupled wideband graphene antenna with the generation of higher order TM modes for THz applications." *Optical Materials* 85 (2018): 456-463.

[17] Jha, K.R. and Sharma, S.K., 2014, July. Waveguide integrated Microstrip patch antenna at THz frequency. In *2014 IEEE Antennas and Propagation Society International Symposium (APSURSI)* (pp. 1851-1852). IEEE.

[18] Saxena, Gaurav, Y. K. Awasthi, and Priyanka Jain. "Design of metasurface absorber for low RCS and high isolation MIMO antenna for radio location & navigation." *AEU-International Journal of Electronics and Communications* 133 (2021): 153680.

[19] Saxena, Gaurav, Y. K. Awasthi, and Priyanka Jain. "Design of metasurface absorber for low RCS and high isolation MIMO antenna for radio location & navigation." *AEU-International Journal of Electronics and Communications* 133 (2021): 153680.

Journal Pre-proof

Journal Pre-proof

[20] Farooq, Sajid, Diego Rativa, Zafar Said, and Renato E. de Araujo. "Ultra-sensitive narrow-band plasmonic perfect absorber for sensing applications." *Photonics and Nanostructures-Fundamentals and Applications* 50 (2022): 101018.

[21] Benlakehal, Mohamed Elamine, Abdesselam Hocini, Djamel Khedrouche, and Tayeb Ahmed Denidni. "Design and analysis of novel microstrip patch antenna array based on photonic crystal in THz." *Optical and Quantum Electronics* 54, no. 5 (2022): 1-16.

[22] Saxena, Gaurav, Priyanka Jain, and Yogendra Kumar Awasthi. "High diversity gain super-wideband single band-notch MIMO antenna for multiple wireless applications." *IET Microwaves, Antennas & Propagation* 14, no. 1 (2020): 109-119.

[23] Varshney, Gaurav, Shailza Gotra, V. S. Pandey, and Rajveer S. Yaduvanshi. "Proximity-coupled two-port multi-input-multi-output graphene antenna with pattern diversity for THz applications." *Nano Communication Networks* 21 (2019): 100246.

[24] Islam, Mohammad Tariqul, Md Naimur Rahman, Mandeep Singh Jit Singh, and Md Samsuzzaman. "Detection of salt and sugar contents in water on the basis of dielectric properties using microstrip antenna-based sensor." *IEEE Access* 6 (2018): 4118-4126.

[25] Huang, Haiying. "Flexible wireless antenna sensor: A review." IEEE sensors journal 13, no. 10 (2013): 3865-3872.

[26] Esfandiyari, Meisam, Ali Lalbakhsh, Saughar Jarchi, Mohsen Ghaffari-Miab, Hamideh Noori Mahtaj, and Roy BVB Simorangkir. "Tunable terahertz filter/antenna-sensor using graphene-based metamaterials." Materials & Design 220 (2022): 110855.

[27] El Gharbi, Mariam, Marc Martinez-Estrada, Raúl Fernández-García, and Ignacio Gil. "Determination of salinity and sugar concentration by means of a circular-ring monopole textile antenna-based sensor." IEEE Sensors Journal 21, no. 21 (2021): 23751-23760.

[28] Logeswaran, Jeyagobi, and Rajasekar Boopathi Rani. "UWB Antenna as a Sensor for the Analysis of Dissolved Particles and Water Quality." Progress In Electromagnetics Research Letters 106 (2022): 31-39.

[29] Kozak, Ryan, Kasra Khorsand, Telnaz Zarifi, Kevin Golovin, and Mohammad H. Zarifi. "Patch antenna sensor for wireless ice and frost detection." Scientific Reports 11, no. 1 (2021): 1-11.

[30] Bhattacharjee, Mitradip, Fatemeh Nikbakhtnasrabadi, and Ravinder Dahiya. "Printed chipless antenna as flexible temperature sensor." IEEE Internet of Things Journal 8, no. 6 (2021): 5101-5110.

Journal Pre-proof

Journal Pre-proof

Biography

Dr. GAURAV SAXENA
Department of Electronics & Communication Engineering

Dr. Gaurav Saxena received his Bachelor of Technology (B.Tech.) Degree in Electronics and Communication Engineering, from Uttar Pradesh Technical University Lucknow, India, in 2007, M. Tech degree in Microwave Electronics, from University of Delhi, South Campus, Delhi, India and Ph.D. from Delhi Technological University, Delhi, India, In 2021 and 2022 awarded a Commendable Research Award for SCI articles Published in 2019-2020 and 2020-2021 session. He also received a meritorious scholarship from July 2010 to July 2012 given by the University of Delhi. From Dec.2011 to May 2012 he has worked as an internship trainee at CSIR, NPL Delhi where he established traceability of the VNA and received an appreciation certificate from Scientist-F. Now, he is working as an Associate Professor in Electronics & Communication Engineering Department from Galgotias College of Engineering and Technology, Greater Noida, Uttar Pradesh and published one patent and numerous research articles in peer reviewed International Journals and Conferences. His recent research interest includes modelling of passive microwave components like filters Power divider/combiner, Microwave antenna for 5G and Wireless applications, Meta-material absorber, MIMO Antenna receiver/transmitter, THz MIMO antenna, absorbers and LNA etc.

Award Summary and Publication Details

Journal Pre-proof

Journal Pre-proof

Author Agreement

Submission of work requires that the piece to be reviewed has not been previously published. Upon acceptance, the Author assigns to the “Nano Communication Networks” the right to publish and distribute the manuscript in part or in its entirety. The Author's name will always be included with the publication of the manuscript. The Author has the following nonexclusive rights: (1) to use the manuscript in the Author's teaching activities; (2) to publish the manuscript, or permit its publication, as part of any book the Author may write; (3) to include the manuscript in the Author's own personal or departmental (but not institutional) database or on-line site; and (4) to license reprints of the manuscript to third persons for educational photocopying. The Author also agrees to properly credit the Nano Communication Networks as the original place of publication. The Author hereby grants the Results in Optics and exclusive rights to the manuscript, all revisions, and the full copyright. The Nano Communication Networks rights include but are not limited to the following: (1) to reproduce, publish, sell, and distribute copies of the manuscript, selections of the manuscript, and translations and other derivative works based upon the manuscript, in print, audio-visual, electronic, or by any and all media now or hereafter known or devised; (2) to license reprints of the manuscript to third persons for educational photocopying; (3) to license others to create abstracts of the manuscript and to index the manuscript; (4) to license secondary publishers to reproduce the manuscript in print, microform, or any computer readable form, including electronic on-line databases; and (5) to license the manuscript for document delivery. These exclusive rights run the full term of the copyright, and all renewals and extensions thereof. I hereby accept the terms of the above Author Agreement.

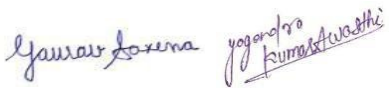
Author: - **Gaurav Saxena, Sanjay Chintakindi, Praveen Kumar Maduri, Y K Awasthi, Sanjay Kumar, Sahil Kansal, Rishabh Jain, Manish Kumar Sharma, Charul Dewan**

Date-06.10.2022

Journal Pre-proof

Declaration of interest statement

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. 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 all of us. 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. Signed by all authors as follows:



Gaurav Saxena
Sanjay Chintakindi, Praveen Kumar Maduri, Y K Awasthi, Sanjay Kumar, Sahil Kansal, Rishabh Jain, Manish Kumar Sharma, Charul Dewan

Date-06.06.2022

Journal Pre-proof

Advanced Production and Industrial Engineering
R.M. Singari and P.K. Kankar (Eds.)
© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE220733

Numerical Analysis of Aerodynamic Drag for the Hyperloop System

Anmol PRABHAKAR[a,1], Arya SHEKHAR[a], Bhuvan JAIN[a], Mohammad ZUNAID[a] and Naushad Ahmad ANSARI[a]

[a] *Department of Mechanical Engineering, Delhi Technological University, Delhi, India*

Abstract. The Hyperloop vehicle is a means of transport in which the pod advances at a fast velocity inside a low-pressure tube to reduce the coefficient of drag. Aerodynamic drag is one of the most significant elements to consider while studying such systems. The design was created on ANSYS SpaceClaim and it consists of a Two-Dimensional model with tapered ends to diminish the drag. Since both ends are tapered, the drag is reduced significantly as compared to the other designs. In this paper, the pod design is analyzed at different speeds, namely 60 ms^{-1}, 120 ms^{-1}, 180 ms^{-1}, 260 ms^{-1}, and 300 ms^{-1} on ANSYS FLUENT and numerical simulations were performed keeping the blockage ratio fixed. Appropriate conditions, equations and k-epsilon turbulence model were applied to produce the desired result. All the drag coefficient values were analyzed, and hence, the graph was plotted. It was observed that the drag coefficient of the pod decreases with an increase in velocity.

Keywords. Hyperloop vehicle; Drag coefficient; Aerodynamic drag

1. Introduction

People's traditional modes of transportation are divided into four categories: road, air, water, and rail. People all across the world are looking for much faster and more efficient ground transportation options as transportation technology has progressed in recent years. The hyperloop is a high-speed land transportation idea that involves passenger pods moving in a partly evacuated tube at transonic speeds [1]. Musk drew the attention of the technical community in 2013 when he proposed the hyperloop concept as a new mode of transportation with performance and comfort benefits for trip distances under 1500 km [2]. Chin et al. introduced the hyperloop concept's general system quantifying the choking risk, which is defined by the throat between the outer tube wall and the pod [3]. The idea is to have a levitating pod that travels inside a low-pressure tube (~100Pa) and at fast speeds (~300 m/s) to minimize the air resistance [4]. Aerodynamic restrictions, such as drag, become major considerations while designing the Hyperloop structure. One of the essential characteristics in the analysis of fast-moving objects, especially within confined tubes (low-pressure), is drag [5][6]. The design of the Hyperloop system is a highly-coupled optimization problem with numerous recursive layout relationships from an engineering viewpoint [7]. Lowering the drag reduces the energy load of the propulsion system.

In this paper, the emphasis is on reducing the drag of the hyperloop structure in comparison with Le et al. [8] by optimizing the design in which both the ends are tapered and omitting the consideration of ideal gas condition and Sutherland model.

2. Methodology

2.1 Designing and details of the model

Previous research looked at the flow attributes of the transonic-speed train entering a tunnel. The two-dimensional axisymmetric model was chosen because it allowed researchers to study a vast parameter space and global patterns [9]. To produce an ideal pod form for two unique goals: increasing lift and lowering drag, a multi-criteria effectivization combined with axisymmetric 2D Computational Fluid Dynamics simulations are applied [10].

As shown in Figure 1, the hyperloop pod was designed aerodynamically on ANSYS SpaceClaim. The dimensions of the pod structure were as follows:

- d_{pod} – 3 m
- d_{tube} – 8 m
- L_{pod} – 55 m
- L_{tube} – 129 m

Here dpod is the diameter of the pod, dtube is the tube's diameter, Lpod is the pod's length, and Ltube represents the length of the tube. Blockage ratio is defined as the projected area of the pod in flow direction to the cross-sectional area of the tube around the pod [11]. The Blockage Ratio is calculated by:

$$B.R = \frac{d_{pod}^2}{d_{tube}^2} \tag{1}$$

The dimensions of the pod were Ø 3m x 55m, and hence, the Blockage Ratio turned out to be 0.140625. The structure shown in Figure 1 was created in 2D and was further used for simulation.

Several research modifying these parameters occur when taking into account the pressure as a variable, the quotient betwixt the cross-sectional area of the capsule and the tunnel (known as Blockage Ratio or BR), and the vehicle speed [12][13].

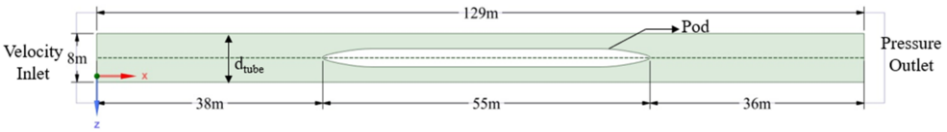


Figure 1. Geometry of hyperloop pod inside a tube

2.2 Meshing and naming of the model

Firstly, all the labelings of the model were done. While generating the mesh, CFD was selected for Physics Preference as the outcome was a very structured and fine mesh that took into account all the relevant edges and curves in the complex geometry. The element size was 0.05, and the smoothing quality was medium. After the mesh was generated, the total number of nodes and elements were 357,083 and 353,263, respectively.

3. Mathematical Model

In the given research, conservation of mass, conservation of momentum (Navier-Stokes), and energy equations are written sequentially [8][14]:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho \mathrm{u_i}) = 0 \tag{2}$$

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_i}\left(\rho u_i u_j\right) = \frac{-\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}[\mu\,(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3}\delta_{ij}\ \frac{\partial u_k}{\partial x_k})\,] + \frac{\partial}{\partial x_j}(-\rho\{u_i' u_j'\}) \tag{3}$$

$$\frac{\partial}{\partial t}(\rho E) + \frac{\partial}{\partial x_j}\left[u_j(\rho E\ +\ P)\right] = \frac{\partial}{\partial x_j}[(k_{eff})\frac{\partial T}{\partial x_j}] + \frac{\partial}{\partial x_j}\left[u_i \mu_{eff}\left(\frac{\partial u_i}{\partial x_j}\right) + \left(\frac{\partial u_j}{\partial x_i}\right) - \frac{2}{3}\delta_{ij}\left(\frac{\partial u_k}{\partial x_k}\right)\right] \tag{4}$$

$$\mathrm{i, j, k = 1, 2, 3}$$

Here, P is the fluid pressure, E is the specific internal energy, μ is the fluid viscosity, ρ is the fluid density, k_{eff} is the effective thermal conductivity, μ_{eff} is the effective dynamic viscosity, and u is the fluid velocity.

The turbulence equation used here is the k-Ɛ model because the k-Ɛ model is the most commonly used. Launder and Spalding (1974) produced the "standard" k-Ɛ model, which is one of the most widely used models for indoor and outdoor airflow modeling. The k-Ɛ model can manage a wide range of fluid flow conditions. It has a straightforward structure, delivers outstanding results, and is generally accepted [15].

4. Validation

The numerical analysis by Le et al. [8] was used for validation to confirm the numerical solution's stability. Figure 2 compares the values of the coefficient of drag between our simulation and the work of Le et al. [8] in graphical form. Our work and their work show a similar trend in coefficient of drag, demonstrating that our numerical method is suitable for simulating moving pod-tube configurations. The maximum percentage error in the simulated result is 4%, indicating that the simulated result was validated successfully.

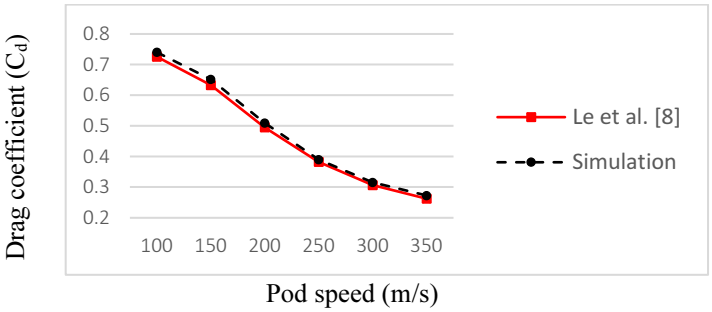


Figure 2. Comparison of drag coefficients between Le et al. [8] and simulation

5. Results

5.1 Parametric Study

The simulation was run on ANSYS Fluent. The units of the pressure were atm, and the solver used here is density-based as it is used for high-speed compressible flows. Gravity is taken into consideration (y = -9.81m/s^2). The energy equation is selected as it represents the conservation of energy of a fluid element. For turbulent flow conditions, the standard k-Ɛ model is commonly used to simulate mean flow characteristics. It is best suited for flow away from the wall. In boundary conditions, five pod speeds were contemplated from 60 ms^{-1} to 300 ms^{-1}. The operating pressure is set at 1/1000 atm, and the operating temperature is 300K. The cross-sectional area of the pod is determined by;

$$A = \frac{\pi d_{pod}^2}{4} \tag{5}$$

and is approximately 7.06858 m^2. The flow behavior upstream is influenced by more than just the inlet boundary condition. The Kantrowitz limit is reached when the inlet is subjected to excessive flow. In reference values, compute from the inlet and insert the area of the pod. Considering the implicit technique can employ a big step size, it is appropriate for solving the calculations that require a lengthy time. Roe FDS was used as a Flux Type, and Least Squares Cell Based was selected as Gradient in Spatial Discretization. Although the first order upwind may get a slighter, less accurate result, it quickly converges and minimizes the computational effort. Pseudo Transient was selected as it helps in stabilizing the case and at the same time gives faster convergence. Afterward, there is a creation of the force report (drag of the vehicle). All the convergence conditions are supposed to be met. Hybrid initialization was used and should be preferred as it accelerates the overall computation. All the contours were kept at the same level in the result.

5.2 Comparative Analysis

Table 1. Drag Coefficients for different velocities of the Hyperloop pod

Velocity (ms^{-1})	**Drag Coefficient (C_d)**
60	0.20603
120	0.18620
180	0.17247
260	0.16414
300	0.16209

The above table (Table 1) records the coefficients of drag for different velocities which came as a result of the simulation. It is evident from the above data that the coefficient of drag decreases with increasing velocity taking area as fixed as mentioned in *5.1*. Here it is observed that at 60 ms^{-1} value of C_d is 0.20603, which further goes down to 0.16209 as velocity equals 300 ms^{-1}. The reason for this phenomenon is that the drag coefficient is inversely proportional to the square of velocity as given in the equation given subsequently:

$$C_d = \frac{D_t}{\left(\frac{1}{2}\right)pv_p^2 A} \tag{6}$$

Compared to open air, the interaction with the tunnel barriers creates extra pressure caused due to the movement of the Hyperloop pod generating additional aerodynamic drag.

Pressure and velocity contours for different velocities are observed. It is evident that the maximum pressure is obtained at the nose of the pod and minimum pressure near the tail. In the case of velocity, the highest magnitude is observed in the spacing between the pod and the tube.

The most accurate result was obtained at a velocity of 300ms^{-1}. The figure given below represents the velocity contour of the same.

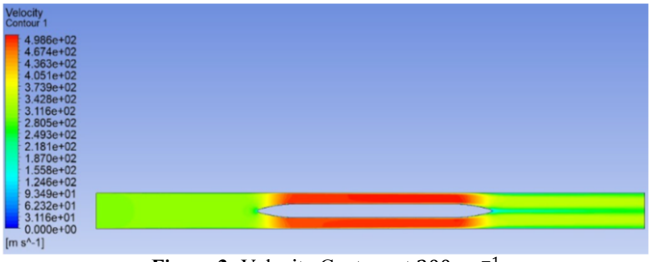


Figure 3: Velocity Contour at $300ms^{-1}$

In Figure 3, the velocity of air is lower at the nose, which eventually becomes greater in the space between the pod and the tube due to the venturi effect, which states that velocity of fluid increases, when passed through constricted and static pressure, decreases similar to the convergence-divergence nozzle.

In their work, Yang et al. [16] found that changing operating speeds and working vacuum pressures had a major impact on the capsule's resistance.

The result obtained by running the simulation is as follows. All the numerical simulations were conducted at different velocities of the pod. Figure 4 indicates that as the drag coefficient decreases, the speed increases.

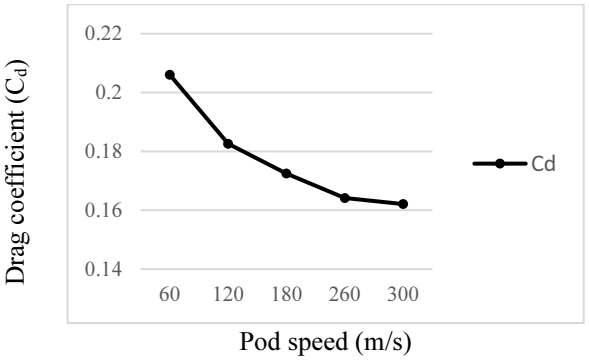


Figure 4. Drag coefficients with respect to velocities of the pod

6. Conclusions

As evident by the results of the present paper and various other research in the domain, the Hyperloop (HL) system emerges as a competitive, powerful, and potential mode of transportation. In the present paper, a Two Dimensional model was created and then analyzed in the hope of optimizing the already existing models. A graph was generated for five different speeds viz. 60 ms^{-1}, 120 ms^{-1}, 180 ms^{-1}, 260 ms^{-1}, and 300 ms^{-1}, and the result indicates that it was possible to converge the coefficient of drag with increasing speed and keeping the blockage ratio constant. The results indicate that when the pod's velocity increases, the coefficient of drag decreases.

Further efforts will target more reduction in drag with minimum dispersal of energy and reduced costs.

7. References

[1] Opgenoord, M.M.J.; Merian, C.; Mayo, J.; Kirschen, P.; O'Rourke.; Izatt, G. MIT Hyperloop Final Project Report; Massachusetts Institute of Technology: Cambridge, MA, USA, 2017.

[2] https://www.tesla.com/sites/default/files/blog_attachments/hyperloop_alpha3.pdf

[3] Opgenoord, M. M. J., Caplan, P. C. (2018). Aerodynamic Design of the Hyperloop Concept. AIAA Journal, 1–10.

[4] Braun, J., Sousa, J., & Pekardn, C. (2017). Aerodynamic Design and Analysis of the Hyperloop. AIAA Journal, 55(12), 4053–4060.

[5] Zhang, Y. (2012). Numerical simulation and analysis of aerodynamic drag on a subsonic train in evacuated tube transportation. Journal of Modern Transportation, 20(1), 44–48.

[6] Pandey, B.K.; Mukherjea, S.K. (2014). Aerodynamic Simulation of Evacuated Tube Transport Trains With Suction at Tail. Volume 12: Transportation Systems.

[7] Kirschen, P.G., Burnell, E. (2021). Hyperloop System Optimization. arXiv: 2104.03907 [cs.CE].

[8] Le, T. T. G., Jang, K. S., Lee, K.-S., & Ryu, J. (2020). Numerical Investigation of Aerodynamic Drag and Pressure Waves in Hyperloop Systems. Mathematics, 8(11).

[9] Li, R., Mei, Y. Study on the Aerodynamic Performance of the High-Speed Train Head with Symmetrical and Asymmetric Nose Shape. In DEStech Transactions on Engineering and Technology Research.

[10] Bizzozero, M., Sato, Y., Sayed, M. A. Aerodynamic study of a Hyperloop Pod equipped with compressor to overcome the Kantrowitz Limit. Journal of Wind Engineering and Industrial Aerodynamics, Volume 218, November 2021.

[11] Lluesma-Rodríguez, F., González, T., & Hoyas, S. (2021). CFD simulation of a hyperloop capsule inside a closed environment. Results in Engineering, 9, 100196.

[12] Kim, T.-K., Kim, K.-H., & Kwon, H.-B. (2011). Aerodynamic characteristics of a tube train. Journal of Wind Engineering and Industrial Aerodynamics, 99(12), 1187–1196.

[13] Li, T., Zhang, X., Jiang, Y., Zhang, W. (2020). Aerodynamic Design of a Subsonic Evacuated Tube Train System. FDMP-Fluid Dynamics & Materials Processing, 16(1), 121-130.

[14] M. Sivapragasam, S. Ramamurthy, M. D. Deshpande, P. White. "An Air Jet Distortion Generation System", 'Hindawi Limited', 2014.

[15] Mishra, P., & Aharwal, K. R. (2018). A review on selection of turbulence model For CFD analysis of air flow within a cold storage. IOP Conference Series: Materials Science and Engineering, 402, 012145.

[16] Yang, Y., Wang, H., Benedict, M., & Coleman, D. (2017). Aerodynamic Simulation of High-Speed Capsule in the Hyperloop System. 35th AIAA Applied Aerodynamics Conference.

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT).
India Habitat Centre, Lodhi Road, New Delhi, India. Sep 23-25, 2022

Performance Analysis of Non-Ideal DC-DC Positive Output Elementary Superlift Luo Converter

Ritambhra Katoch
Research Scholar
Delhi Technological University
New Delhi, India
ritambhrakatoch_2k20phdee12@dtu.ac.in

Dheeraj Joshi,SMIEEE
Electrical Engineering Department
Delhi Technological University
New Delhi, India
joshidheeraj@dtu.ac.in

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT) | 978-1-6654-9365-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBCONPT57482.2022.9938219

***Abstract*— The increased applications of Power electronic DC to DC converters in various domestic and industrial applications require attention to assess the performance characteristics of converters. The various performance indices of the converters are power transfer efficiency, high voltage gain, low cost, low weight, simple circuitry, and high-power density There are so many power converter topologies which include isolated and non-isolated converters. The isolated converters are compact in size, low weight, economical cost, and minimum losses. In the conversion stages, there are voltage building techniques developed which include voltage lift and super lift techniques. In voltage lift technique, the voltage increases stagewise in arithmetic progression, however in the super lift technique the increase of voltage follows the geometric progression. This results into an increase in the voltage gain of super lift Luo converters largely and so it can be employed in the applications which need the high voltage. In this paper, the Non-ideal POESLL converter is considered, and its performance is analyzed concerning its voltage gain. The performance of POESLLC with and without implementation of PI control has been successfully simulated and verified using MATLAB simulation**

Keywords*— *Super-lift Luo converter, PI, MATLAB, ESR, Nichols -Ziegler

I. INTRODUCTION

DC to DC converters are becoming a very important part of our daily electronic devices which includes mobile phones, laptops, chargers, communication equipment, digital cameras, computer hardware circuits, dental apparatus, regulated power supplies, spacecraft supply system, electric vehicles, and various applications used in industries. There is a requirement of DC-to-DC converter for some devices because the standard available DC sources are either very high or low in voltage Hence, some sort of power conversion, whether it is step-up or is step-down, both are required from available DC source. As an example, generally mobile phones have 3.7 V batteries. Charging of such mobile batteries from a 12 V car battery would need a DC-to-DC converter circuit because a 12 V is high charging voltage for charging of 3.7 V battery. Furthermore, frequently used embedded system circuits in transport vehicles and in industrial equipments need either 3.3 V or 5 V DC power sources. Therefore, step down power converters are used to drive them from a 12V battery. DC-to-DC converters are also used in domestic appliances. There are many DC-to-DC power converters which are categorized into six-generation types: The first generation[1]converters operate in single - Quadrant mode and allow the range of power up to 100W. These converters are categorized further into six categories which include fundamental converters, transformer type converters, developed converters, super-lift converters, and ultra-lift converters. Three fundamental converters are 1: Buck, 2: Boost, and 3: Buck-Boost converters. The power conversion efficiency of these power converters is restricted on account of the effect of ESRs. In Transformer type converters there is isolation between its input and output terminals which results into high isolation voltage properties. They are useful for blocking noise and interferences[1]. So due to these characteristics they can produce a cleaner and desired DC output voltage. Voltage lift converters[2-4] are well known for increasing the voltage transfer gain to a great extent than conventional ones, these converters also minimize the impact of parasitic effects. With these converters the gain increases stage by stage in arithmetic progression.[4]. The Super-lift/Ultra-lift [6] converters are very popular fundamental converters for providing high energy conversion efficiency with very high gain in voltage. With the use of these converters the effect of parasitic elements can be reduced. This results in increase of gain in power series. Two and four quadrant operations are performed by the second-generation converters with an output from 100 to 1000 Watts. Due to the use of passive elements these converters are large. The major application area of these converters is DC motor drives. These include transformer based and developed multi quadrant type Luo power converter. The third-generation of the converters consist of passive elements such as capacitors or inductors and are called switched-component converters. They perform the operations in both two and four quadrant resulting high output power. The fourth Generation Soft Switching converters are also known as ZVS and ZCS converters[6-8]. These converters have the advantage of zero power loss, higher efficiency, and power density due to turning on and off the switch at the moment when voltage and / or currents are zero. The fifth-generation power converters are known as synchronous rectifier DC-to-DC power converters. These are used for applications where power supplies requirement is of low voltage and strong current. As this need cannot be fulfilled by the conventional diode bridge rectifiers but synchronous rectifier is capable of achieving these technical features. The sixth-generation converters include eight topologies of two-element Resonant power converter[9] thirty-eight topologies of three-elements Resonant power converter and ninety-eight topologies of four-element type Resonant power converters. The major application area of these types of converters is military equipment.

There are various modeling techniques used by the researchers for the modeling of Switching converters. Miao Zhu and Fang Lin Luo proposed a method for super-lift converters which is in the form of graphical modeling. This method elaborates the effects caused by equivalent series resistances and forward voltage drop across diodes. The basic

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:11:36 UTC from IEEE Xplore. Restrictions apply.

methods used in the modeling of power converters are averaging and state-space modeling methods. Rodney H.G. Tan and Landon Y. H. Hoo suggested the DC-DC converter model, implementation, and simulation in the state-space space modeling approach and compare it with the circuitry model. Jared A. Baxter and Daniel J. Costinett proposed Converter Analysis using Discrete-Time State-Space Modelling. M. I. Rahman, D. J,ovcic, and K. H. Ahmed presented a model of a dual active bridge DC-DC converter based on the Fourier series of the switching functions of the converter and used the Fundamental harmonic approximation method. Geoffrey N Love and Alan R Wood derived a small-signal model of power electronic converter. . Small signal analysis is used to know the response of a converter device to small signal variations from a given operating value. The small-signal analysis method is very important for designing the robust control strategy which includes sliding mode control and H-infinity control. There are various control techniques applied to converters for efficient and robust operation of converters. The most common control method applied is Proportional-integral-derivative[10] control and about 90% of industrial applications use a PID controller due to its simple implementation, simple and fast calculation as well as higher efficiency and quality of control. Various researchers designed different types of controllers like proportional-Integral-Controllers [10] for different types of DC-DC converters, non-linear controllers like sliding mode controllers[11]H-infinity controllers[12], AI-based controllers named fuzzy logic, and Neural network-based control. All these controllers are employed to improve the stability and transient response of DC-DC converters. A comparative study of these control strategies is carried out further to explore the best control method for a converter.

In this paper, the simulink model of the non-ideal POESLLC is developed and the performance is analysed in both the open-loop as well as the closed-loop configuration using MATLAB. The PI controller is developed and the effect of step-change in input line voltage, as well as load, is observed in a closed-loop system.

II. OPERATION OF POESLLC

The schematic diagram of POESLLC is presented in Figure. 1. The power converter is supposed to work in the mode of continuous conduction. In this mode the duty cycle is *D* and switching period is *T*. The basic circuit elements of POSELLC consist of one inductor *L* ,one MOSFET switch S two capacitors named as C_1 and C_2, two diodes named as *D1* and *D2*, and one *R* as load resistance. In addition to these the equivalent series resistance of inductor and r_{c1}, r_{c2} of capacitors *C1&C2* are also included.

Mode1 operation: During switching on of switch *S* the diode *D1* gets forward biased and *D2* gets reverse biased. The circuit diagram during switching on conditions is shown in Figure. 2. During the switch-on period, the energy is stored in the inductor in the form of a magnetic field and charging of capacitor *C1* is started. The capacitor *C2* is discharged through the load resistor *R*. The switching on mode is shown in Figure. 2.

Mode 2 operation: During switching off of switch S is diode D_2 gets forward biased and D_1 gets reverse biased. The energy stored in the inductor gets dissipates to charge the capacitor C_2 and supply the load. The circuit diagram in switching off mode is shown in Figure. 3.

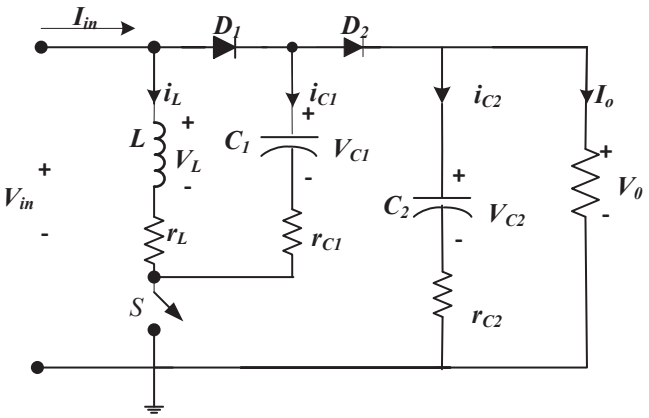


Fig. 1. Circuit diagram of POESLLC

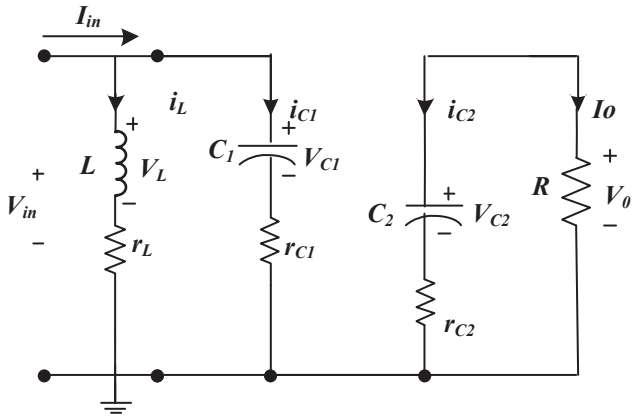


Fig. 2. Equivalent Circuit for switching ON state

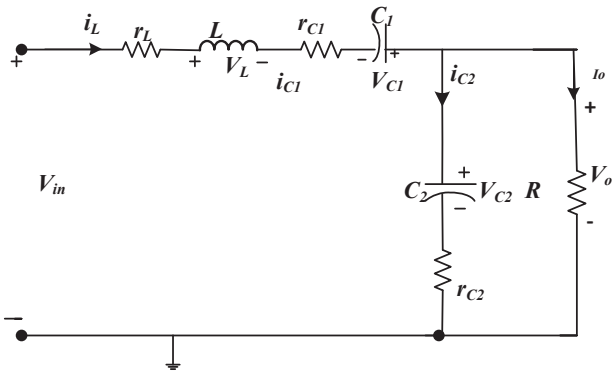


Fig. 3. Equivalent Circuit for switch-OFF state

During switching on period *DT* there is increase of current i_{L1} which is flowing through the inductor with increased input voltage *Vin* and decreases with $(2V_{in}-V_o)$ during off period *(1-D)T*. The ripple factor of the inductor current i_{L1} is given by

$$\triangle i_{L1}=\frac{V_{in}}{L_1}DT=\frac{V_o-2V_{in}}{L_1}(1-D)T \quad (1)$$

The voltage transfer gain is given by

$$G=\frac{V_0}{V_{in}}=\frac{2-D}{1-D} \quad (2)$$

The input current marked as I_{in} equals $(i_{L1}+i_{C1})$ during turn on position and i_{L1}during turn off position. The current i_{C1} in a capacitor is equal to i_{L1} during turn-off. During steady-state

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:11:36 UTC from IEEE Xplore. Restrictions apply.

conditions the total charge inside the capacitors should not change.

$$i_{in-off} = i_{L1-off} = i_{C1-off} \quad (3)$$
$$i_{in-on} = i_{L1-on} + i_{C1-on} \quad (4)$$
$$DTi_{C1-0n} = (1-D)Ti_{C1-off} \quad (5)$$

If L_1 is taken as a very large value then i_{L1} equals to the average value of current I_{L1}. Hence,

$$i_{in-0ff} = i_{C1-off=I_{L1}} \quad (6)$$

$$i_{in-on} = I_{L1} + \frac{1-D}{D} I_{L1} = \frac{I_{L1}}{D} i_{C!-on} = \frac{1-D}{D} I_{L1} \quad (7)$$

The average value of input current is given by

$$I_{in} = Di_{in-on} + (1-D)i_{in-off} = (2-D)I_{L1} \quad (8)$$

The ripple in output voltage V_0 is given by

$$\triangle v_o = \frac{\triangle Q}{C_2} = \frac{D}{f_{C2}} \frac{V_o}{R} \quad (9)$$

Based on equations (1) (2) (3) and (11) &considering 20% and 5% ripple in inductor current and output voltage the value of inductor and capacitors are calculated and the Simulink model of POESLLC is developed with the set of parameters as shown in Table 1.

III. STATE SPACE MODELLING OF POESLLC

During the continuous mode of operation, the dynamics of the converter can be expressed in the form of circuit parameters and state variables.The number of state variables in the POESLLC are equal to the number of inductors and capacitors and the current or voltages of these elements can be identified as state variables. The converter considered in this paper consist of one inductor and two capacitors as a energy storage elements . Therefore three state variables can be defined as $i_{L,1}$,v_{C1} and v_{C2}

In switching mode operation :

The current in an inductor is given by:

$$\frac{di_L}{dt} = \frac{-r_L}{L} i_L + \frac{1}{L} V_{in} \quad (10)$$

The capacitor voltages are given by

$$\frac{dv_{C1}}{dt} = \frac{-1}{r_{C1}C1} v_{C1} + \frac{1}{r_{C1}} V_{in} \quad (11)$$

$$\frac{dv_{C2}}{dt} = \frac{-1}{(R+r_{C2})\,C_2} = v_{C2} \quad (12)$$

The output voltage is given by

In switching-off mode:

$$\frac{di_l}{dt} = \frac{-1}{L}\left(r_L + r_{C1} + \frac{r_{C2}R}{R+r_{C2}}\right) i_L + \frac{1}{L} v_{C1} - \frac{R}{L(R+r_{C2})} v_{C2} + \frac{1}{L} v_{in} \quad (13)$$

Capacitor voltages are given by:

$$\frac{dv_{C1}}{dt} = \frac{-1}{C1} i_L \quad (14)$$

$$\frac{dv_{C2}}{dt} = \frac{R}{(R+r_{C2})C_2} i_L - \frac{1}{(R+r_{C2})C_2} v_{C2} \quad (15)$$

The output voltage is given by

$$v_o = \frac{r_{C2}R}{(R+r_{C2})} i_L + \frac{R}{(R+r_{C2})} v_{C2} \quad (16)$$

Average state-space modeling of converter

For the determination of transfer function model of POESLLC the most common method is the state-space averaging method which can be summarized as follows.

The switch S of converter is on for DT duration and off for(1-D)T duration where D is the duty cycle and T is a time period of switching cycle .The three steps are involved in the average modeling method

Step1: Switching on mode(0<t<DT)

$$\frac{dx(t)}{dt} = A_1 x(t) + B_1 u(t) \quad (17)$$

$$y(t) = C_1 x(t) + D_1 u(t) \quad (18)$$

where,

$$A_1 = \begin{bmatrix} \frac{-r_L}{L} & 0 & 0 \\ 0 & \frac{-1}{r_{C_1C_1}} & 0 \\ 0 & 0 & \frac{R}{(R+r_{C2})C_2} \end{bmatrix}, B_1 = \begin{bmatrix} \frac{1}{L} \\ \frac{1}{r_{C_1}C_1} \\ 0 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 0 & 0 & \frac{R}{(R+r_{C2})} \end{bmatrix}, D_1 = [0]$$

$$A_2 = \begin{bmatrix} \frac{-1}{L}(r_L + r_{C1} + \frac{r_{C2R}}{(R+r_{C2})}) & \frac{1}{L} & \frac{R}{L(R+r_{C2})} \\ \frac{-1}{C_1} & 0 & 0 \\ \frac{R}{(R+r_{C2})C_2} & 0 & \frac{-1}{(R+r_{C2})C_2} \end{bmatrix}, B_2 = \begin{bmatrix} \frac{1}{L} \\ 0 \\ 0 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} \frac{r_{C2}R}{(R+r_{C2})} & 0 & \frac{R}{(R+r_{C2})} \end{bmatrix}, D_2 = 0$$

Step2: Switching-off mode (DT<t<T)

$$\frac{dx(t)}{dt} = A_2 x(t) + B_2 u(t) \quad (19)$$
$$y(t) = C_2 x(t) + D_2 u(t) \quad (20)$$

The average state model equation

$$\frac{dx(t)}{dt} = A' x(t) + B' u(t) \quad (21)$$

$$y = C' x(t) + D' u(t) \quad (22)$$

where,

$$A' = A_1 D + A_2(1-D) \quad (23)$$
$$B' = B_1 D + B_2(1-D) \quad (24)$$
$$C' = C_1 D + C_2(1-D) \quad (25)$$
$$D' = D_1 D + D_2(1-D) \quad (26)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:11:36 UTC from IEEE Xplore. Restrictions apply.

Eq. (23),(24),(25) and (26) represents the non linear average state space model of POSELLC converter.

TABLE I. DESIGN PARAMETERS OF POESLLC

Parameters Name	Symbol	Value
Input voltage	V_{in}	20 V
Output voltage	V_o	60 V
Output power	P_o	72W
Load resistance	R	50 Ω-80 Ω
Inductor	L	100 µH
Capacitor	C1	2200 µF
Capacitor	C2	700 µF
Frequency	f	50KHZ
Duty cycle	D	0.5
ESR inductor	r_L	0.12 Ω
ESR capacitor	r_{C1}	0.2 Ω
ESR capacitor	r_{C2}	0.1 Ω

IV. RESULTS AND DISCUSSION

A. Open Loop Simultion

The Simulink model of POSLC for open-loop configuration is constructed by using a fixed DC source as shown in the Figure below. PWM technique is used for the duty ratio required for the desired voltage. The open-loop configuration is subjected to input line variation and load variation within permissible limits and the scope results are shown in fig below:

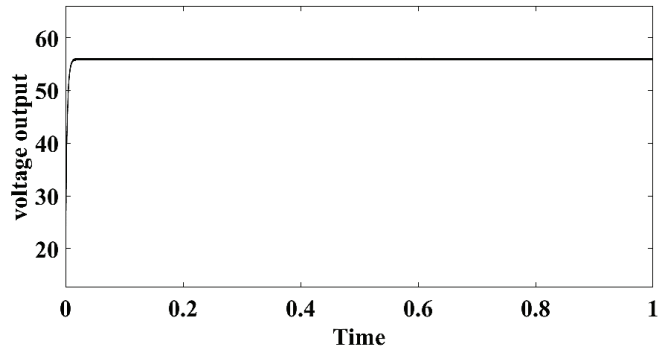


Fig. 4. Open Loop response of POESLLC

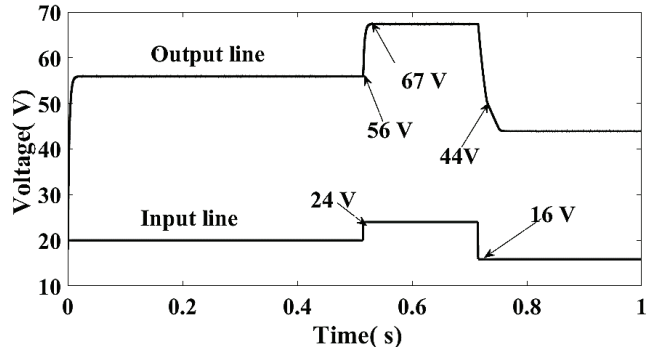


Fig. 5. Open-loop POESLLC voltage output with input voltage variations

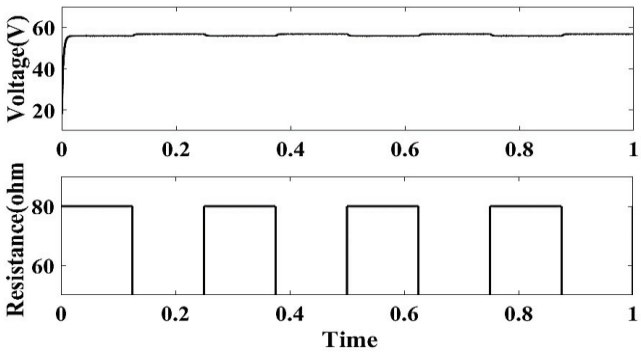


Fig. 6. Open-loop output voltage with load variation

The open loop response of POESLLC is shown in Figure. 4. Though the proposed converter exhibits good voltage gain in open loop, but it is not achieving desired output. Also, when it is subjected to input and load variations it is unstable. In Figure. 5 variation of output voltage with step change in input voltage is shown. It is evident from the figure that in case of open loop system the desired output voltage is not maintained constant and continuously changing. When the input voltage is changed from 20 volt to 24 volt (+20% permissible limit) the output voltage becomes 56 volt to 67 V. On the other hand when input voltage change from 24V to16 V(-20%) the output voltage reaches to 44 V. This leads to highly unstable output voltage whenever there is variations in input line voltage.

Figure.6 shows the response of open loop POESLLC to the step change in load from 80 ohm to 50 ohm. As the load decreases from 80 ohm to 50 ohm the voltage output increases and with increase in load from 50 to 80 there is a drop in voltage. As a result output voltage is not maintained at constant value.

B. Closed-loop Simulation

The MATLAB Simulink model of Closed-loop configuration of POESLLC is constructed. The proportional integral control is employed to form a feedback control system. In the feedback system, the controller output will decide the pulses to the MOSFET switch and the output voltage is maintainedby the PI control action. The PI controller parameters are tuned by Nichols -Ziegler's method as shown in the table to get the desired output. Also, the controller performance is investigated concerning changes in input and load variation. The scope output of Closed-loop POESLLC is shown in Figure. 8.

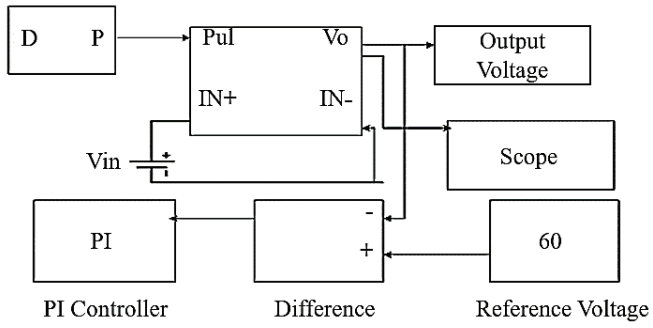


Fig. 7. Closed Loop Block Diagram of converter with PI controller

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:11:36 UTC from IEEE Xplore. Restrictions apply.

TABLE II. [13]: NICHOLS-ZIEGLER'S CHART FOR CONTROLLER GAINS ASSESSMENT

Type of controller	K_p	K_i	K_d
P	$0.5K_{cr}$	0	0
PI	$0.45K_{cr}$	$1/1.2P_cr$	0
PID	$0.6K_{cr}$	$0.5P_{cr}$	$0.125P_{cr}$

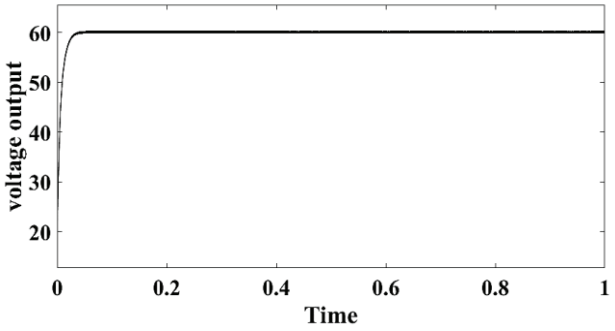


Fig. 8. Closed loop response of POSELLC

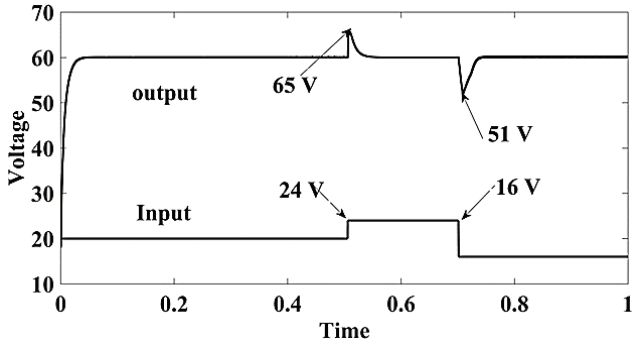


Fig. 9. Closed loop POESLLC output voltage with input variations

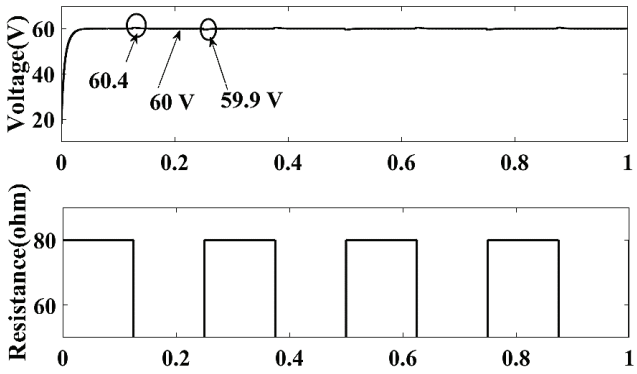


Fig. 10. Closed-loop POESLLC output voltage with a variation in load

The closed loop response of POESLLC is shown in Figure. 8. The output is reaching the steady state output voltage with 0.4% overshoot and 0.02 settling time. Figure.9 shows variation in the output voltage of POESLLC converter when the input voltage changes from 20 V to 24 V (20% of line disturbance). It is found that as there is change of voltage from 20V to 24 V. There is an overshoot of 5.4V which settle quickly to desired output voltage within 0.03 settling time. Similarly, when the input voltage changes from 24 to 16 V the overshoot voltage is equal to 8.3 V with a settling time of 0.03 s.

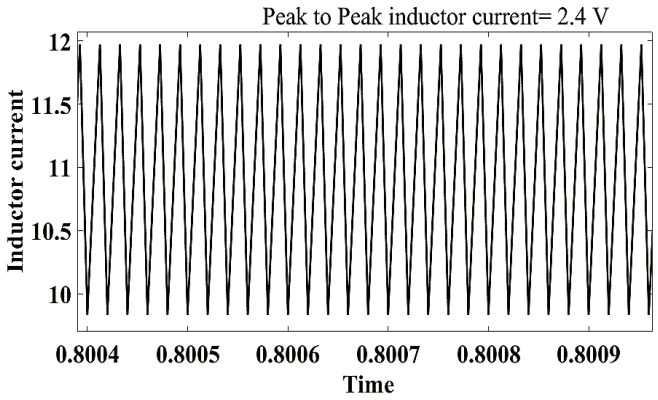


Figure. 11 Steady state Inductor current

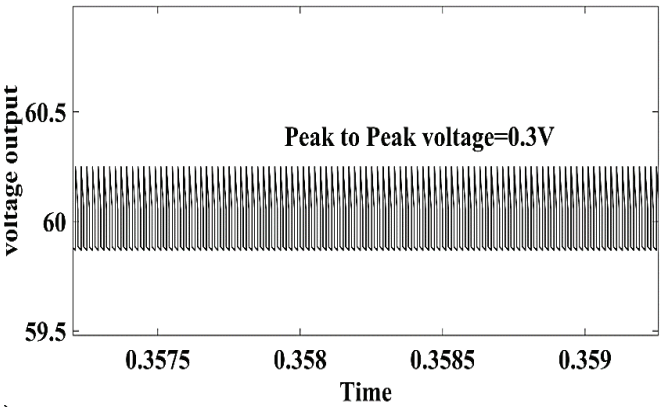


Fig. 11. Figure. 12 Steady- state Closed -loop output voltage

Thus, in the presence of PI controller the converter performance is very much improved. It is providing desired output as well as when it is subjected to input variations the controller quickly exhibits its desired output with a settling time of 0.02s. Also, there is a step-change in load at three instants 0.25,0.5 and 0.75 respectively which leads to a change in output as shown in Figure.10. The overshoot percentage is only 0.4% in case of step change of load from 50ohm-80ohm.The control action of PI pushed the output voltage to the desired level within a settling time of 0.07 s for increase in load and 0.01 s for decrease in load. It is also evident from Figure. 11 and Figure.12 that the ripple in inductor is 2.4 A and voltage ripple is 0.3 V.

V. CONCLUSION

The performance analysis of POESLLC for both open-loop and feedback systems with PI is analyzed. It is being concluded from the results that the proposed converter is exhibiting good voltage gain in open loop configuration but there is enough deviation from the desired value when the open-loop model is subjected to load and input variations and leading to an unstable system which is further improved with PI control where Kp and KI values are provided by the Nichols-Ziegler method. The closed-loop system is subjected to load and input variations and the system quickly maintained the desired output for any change in line input and load with overshoot percentage of 0.4% and settling time of 0.02s.'

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:11:36 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Fang Lin Luo and Hong. Ye: 'Power electronics: Advanced conversion technologies' (CRC Press, LONDON)

[2] Fang Lin Luo, "Positive Output Luo-Converters: Voltage Lift Technique," Proceedings of IEEE Electrical. Power Application., vol. 146, no. 4, pp. 415–432,

[3] Fang Lin Luo and Hong. Ye, "Self-Lift DC-DC Converters," Proceeding. 2nd IEEE International Conference PEDES'98 Australia, Nov. 30–Dec. 3, 1998

[4] Fang Lin Luo, "Re-lift converter: Design, test, simulation and stability analysis," Proc. IEEE Elect. Power Application., vol. 145, no. 4, pp. 315–325

[5] Fang Lin Luo and Hong. Ye, "Positive output super lift converters," IEEE Transaction on power electronics, Vol.18, No. 1, pp. 105-113, January 2003

[6] Aksoy, H. Bodur, and A. Faruk Bakan, "A new ZVT- ZCT-PWM DC–DC converter," IEEE Trans. Power Electron., vol. 25, no. 8, pp. 2093–2105, August 2010

[7] N. Mohan, T. M. Undeland, and W. P. Robbins, Power Electronic Converters, Applications, and Design. Hoboken, NJ: Wiley, 2007

[8] Yie-Tone Chen, Member, IEEE, Shin-Ming Shiu, and Ruey-Hsun Liang, 'Analysis and Design of a Zero-Voltage-Switching and Zero-Current-Switching Interleaved Boost Converter' IEEE Transactions on Power Electronics, Vol. 27, No. 1, January 2012

[9] Y. Yang, D. Huang, F. C. Lee, and Q. Li, "Analysis and reduction of common-mode EMI noise for resonant converters," in Proc. of IEEE 28th Applied Power Electronics Conference and Exposition (APEC), pp. 566–571,2014

[10] M.S. Ali & Kashmira Rathi, "Design and simulation of PID Controller for Power electronics Converter Circuits", International Journal of Innovative and Emerging Research in Engineering, vol. 3, Issue 2, pp 26-31, 2016

[11] Jianwei Zhang, David G. Dorrell, Li and Ahmadreza Argha, Jianwei Zhang, "A Novel Sliding Mode Controller For DC-DC Boost Converters Under Input/Load Variations", IECON2015-Yokohama, page. 001698-001703, November 9-12, year 2015

[12] R. Naim, G. Weiss, B. Y. Shmuel, "H- infinity control applied to boost power converters," IEEE Transactions on Power Electronics, Vol.12, No.4, pp. 677-683, April 1997.

[13] Bambang Pramujati, Hendro Nurhadi, Desmas A Patriawan, "A Study on the Effect of an Attractive and a Repulsive Forces with Feedback Control on a Magnetic Levitation System", International Journal of Scientific & Engineering Research, Volume 6, Issue 6, June-2015.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:11:36 UTC from IEEE Xplore. Restrictions apply.

Wireless Personal Communications
https://doi.org/10.1007/s11277-022-10109-1

PLC and SCADA based Real Time Monitoring and Train Control System for the Metro Railways Infrastructure

Ishu Tomar[1] · **Indu Sreedevi**[1] · **Neeta Pandey**[1]

Accepted: 15 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Industrial Control Systems (ICS) and SCADA systems play a critical role in the management and regulation of critical infrastructure. SCADA systems brings us closer to the real-time application world. All process and equipment control capability is typically provided by a Distributed Control System (DCS) in industries such as power stations, railways, agricultural systems, chemical and water treatment plants. This article proposes to adopt SCADA and PLC technology in the improvement of the performance of real time signaling & train control systems in metro railways infrastructure. The main concern of this work is to minimize the failure in automated metro railways system operator and integrate the information coming from Operational Control Centre (OCC), traction SCADA system, traction power control, and power supply system. This work presents a simulated prototype of an automated metro train system operator that uses PLC and SCADA for real time signaling, monitoring and control of metro railway systems. Here, SCADA is used for the visualization of an automated process operation and then the whole operation is regulated using OMRON (NX1P2-9024DT1) PLC and OMRON's Sysmac studio programming software is used for developing the ladder logic of PLC. The metro railways system has deployed infrastructure based on SCADA from the power supply system, and each station's traction power control is connected to the OCC remotely which commands all the stations and has the highest command priority. An alarm is triggered in the event of an emergency or system congestion. This proposed system overcomes the drawbacks of the current centralized automatic train control (CATC) system. This system provides prominent benefits like augmenting services which may enhance a network's full load capacity and network flexibility, which help in easy modification in the existing program at any time.

Keywords Automation · Train control system · Programmable logic controller · Operations control · Supervisory control and data acquisition

Indu Sreedevi and Neeta Pandey authors have contributed equally to this work.

✉ Ishu Tomar
ishutomar.611@gmail.com

Extended author information available on the last page of the article

Published online: 28 October 2022

1 Introduction

Automation has been developed in the context of industrialization to provide various leverages such as process decentralization, virtualization, interoperability, real-time data acquisition, and an important role in mechanization, which provides the human operator with machinery to assist them with the manual requirements of work [1]. Figure 1 shows that automation reduces power consumption and increases industrial efficiency [2–4]. Energy performance can be improved directly and indirectly by implementing better monitoring, control and optimization strategies by using improved maintenance procedures that reduce the need for additional energy usage due to a process downtime or shutdown processes [5–8]. Automation assures the improved monitoring and control and also enhances the quality of work performed by control systems [9, 10].

Railway systems play a vital role in today's constantly growing transportation sector. Metro railways deliver fast and convenient travels in the metropolitan cities to the daily commuters. The importance of signaling and train control systems in accomplishing these goals cannot be overstated because they offer an automated and connected control to improve train operations safety, efficiency and dependability [11–14]. Earlier, track circuits and transmission-based systems were used for signaling & train control systems in metro railways. But these systems have various drawbacks related to fault of traction power supply system (TPSS), train delays, security etc. Energy conservation in train operation is gaining major attention as issues over energy usage and operational costs in metro

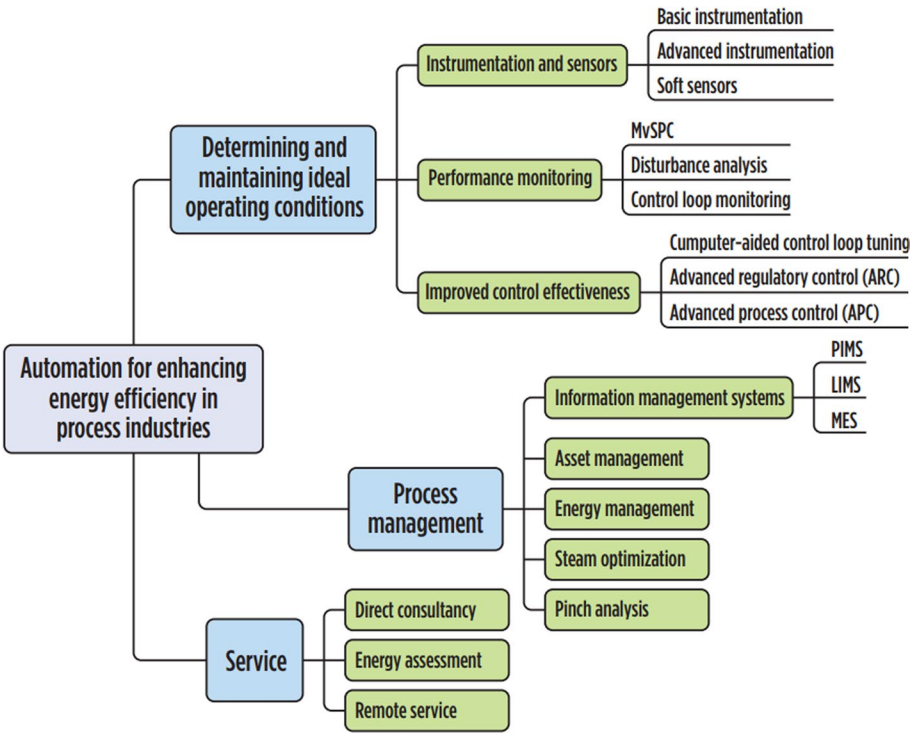


Fig. 1 Automation for energy efficiency [4]

systems grow. Transit operations control plays an important role in the service delivery for determining the impacts on passenger service quality. However, the metro line frequently experiences traction power supply system problems, which result in more than 5-min train delays. Dispatching actions must be made right away during such a disruption to lessen the effects on quality of the passenger service.

With advancements in telecommunications and information technology, the Communication based train control (CBTC) systems are being used for railway signaling and train control systems [15, 16]. But these CBTC systems have various drawbacks [17–19]. Due to unreliable wireless communication and frequency handover, CBTC systems have a significant effect on the functionality, efficiency and usefulness of train control systems. They have a large attack surface. They can be subjected to multiple hacking attacks like network intrusion, data tampering which eventually lead to security hazard. Around the core issue of improving the signaling and train control systems, many researchers have focused on the CBTC techniques. He et al. [20] proposed a mathematical approach for energy conservation in metro trains on the basis of regenerative braking. This approach used the matrix calculus control theory to improve the renewable energy utilization ratio and energy efficiency. Mouna et al. [21] analyzed the vulnerabilities and security risks in monitoring and control of railway systems. Carvajal-Carreño et al. [22] proposed a fuzzy logic-based detection technique for energy-efficient working of CBTC based metro rail. They used the actual attributes of automatic train operations system which uses the coast command to save energy and accounts for the ambiguity around the previous train's speed.

Using information from the 3-axis accelerometer sensors in cellphones, Wang et al. [23] developed a data integration strategy for estimating speed and calibrating position of metro trains in an underground scenario. The lateral, longitudinal and vertical accelerations of a train are measured using multiple cellphones placed in various train cars. The observed accelerations are then converted from the smartphone coordinate systems to those of the metro train.

Yin et al. [24] suggested a technique to reduce operational costs and waiting time on metro lines. This method used a Lagrangian relaxation based heuristic algorithm [25] for reducing the computational complexity. Carrel et al. [26] proposed a framework for studying train operations and control decisions that incorporates service and passenger movement information. On the basis of automatically gathered data, a technique was used to analyze metro line operations to highlight potential weaknesses in particular operations control tactics.

Bai et al. [27] presented a paradigm for cooperative control of multiple trains in real-time using cooperative co-evolutionary algorithm. Using this method, one can find out the current status of metro trains on the extents of locations of train operation mode switches, braking of train acceleration and station stopping. In order to reduce net energy consumption, it seeks to determine the best locations throughout inter-station trials to swap between operation modes of metro train and the degrees of driving and braking for various trains. A cooperative control strategy has been used in [28] to deal with the problem of fault in TPSS in metro rail system. This approach used the hypothetical section mechanism and space-time-speed network technique so that the line capacity can be maximized during disruption.

Most of the algorithms proposed in the literature has mostly concentrated on energy-efficient train scheduling and single-train optimum control. Without energy storage devices, the former cannot guarantee the adequate application of sustainable power on railway lines because it ignores the synchronization of moving and stopping trains. The latter involves planning rail operational processes to synchronize moving and stopping trains for

greater sustainable power consumption. It is required to take dispatching steps in order to maintain line capacity in the area of insufficient power supply during the power supply shortage brought on by the malfunction of the TPSS in the metro line. Therefore, this paper starts with PLC and SCADA based automation systems that can be used in metro railways to minimize the failure. PLCs conduct the majority of the control components required to implement system logic [29]. It is important to understand how real-world stuff works and therefore to assist the user, PLC code is used with SCADA. Input/output, HMI (Human–machine interface), signal hardware controller, networks, communication, database, and software are all part of a SCADA system, which are all connected via a PLC program. PLC is capable of altering or modifying the sequence of operations in accordance with the settings of the purpose to prevent failure. Here, SCADA is used for the visualization of an automated process operation and then the whole operation is regulated using PLC. The PLC used here is OMRON (NX1P2-9024DT1) and OMRON's Sysmac studio programming software is used for developing the ladder logic of PLC. Wonderware Intouch SCADA software is used for the visualization of the operation.

In this article, our motivation is to overcome the drawbacks of the current CBTC and CATC i.e. centralized automatic train control system which comprises of train protection, signaling and operation systems [30–33]. Our motivation is to effectively compensate the effects of fault of traction power supply system (TPSS), train delays and security so that the real time signaling, monitoring and train control can be done effectively in real time. This PLC and SCADA based real time monitoring and train control system for the metro railways helps in detecting exact location, avoiding train collisions and opening emergency exit even after the interruption of power system with the use of programmable logic controller. If the communications link between any of the trains is disrupted then the whole system/network will be affected but the impact is smaller. The following is a summary of this paper's significant contributions.

- We present a simulated prototype of an automated metro train system operator that uses PLC and SCADA for the real time signaling, monitoring and control of metro railways infrastructure.
- Compared to existing systems, this work aims to minimize the failure in automated metro railways system operator and integrate the information coming from Operational Control Centre (OCC), traction SCADA system, traction power control, power supply system and CCTV system.

The framework of this article is designed in following manner: Introduction, Prototype of metro railways system infrastructure, Operations of the metro railways system, Simulation results with PLC ladder logic program, Performance evaluation and Conclusion.

2 Prototype of Metro Railways System Infrastructure

PLC and SCADA are utilized here for monitoring and control of metro railways system. PLC has an internal memory that stores instructions for executing certain operations while scanning the program. The purpose of a PLC is to modify process operations and monitor critical process parameters as needed. PLC receives data/signals from input devices such as sensors, push button switches, contact limit switches, and so on. The information is processed in the input module where logical operations are done on the inputs, and the

required response is supplied to the operator via the output module linked device such as a motor, valve, or alarm (see Fig. 2). The generating and transmitting systems are managed, regulated, optimized, and observed via SCADA. Remote terminal units (RTU) are the core component of these systems, which gather data automatically and link directly to analyzers, sensors, recording devices, or industrial equipment.

SCADA monitor and control the system via a centralized OCC and HMI that controls the traction system equipment and entire power supply through SCADA. Figure 3 shows the working of metro railways system using PLC and SCADA. To execute its tasks, the PLC requires a regulated power source of 24 V DC. IR sensors, stations, and emergencies feed into the PLC, which generates six output signals such as signal/alarm, train running or halting, and door open or close. The PLC's output module will send commands from SCADA to the train. An essential feature is that in the case of an emergency, the alarm will send a status point with the value NORMAL or ALARM.

The prototype of the automated metro railways system operator has four main applications. OCC is the main application of this automated metro train system operator prototype. An operator sits in front of the HMI and controls the metro train system. Each station is linked to the OCC through a remote connection. It has command of all stations and handles all commands such as emergency and congestion with highest priority. CCTV function, metro stations and metro platform view, and electrical control are other sub-applications of the system (traction power control and power supply system).

Figure 4 shows the functioning of automated metro railways system operator along with the main application and sub-applications and Fig. 5 shows OCC runtime window in SCADA Wonderware Intouch software from where the operator may command and monitor. If the metro rail system is not working, then the operator can click to switch control or CCTV or metro station icon or electrical control icon to see the fault. The IR and reflecting sensors on the RTU are used to scan the programs. If there is a problem with the process, the RTU sends a message to the master terminal unit (MTU), which then sends signals or alerts to the operator on the screen.

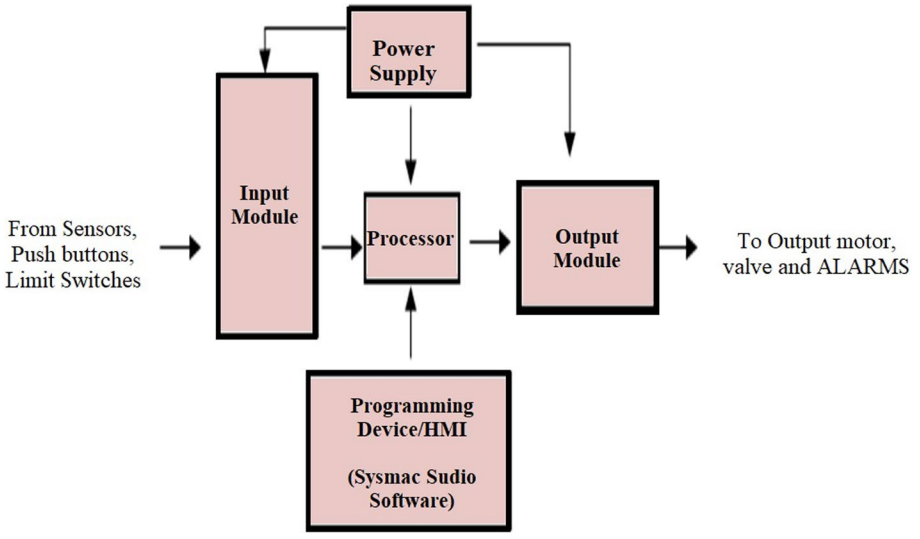


Fig. 2 Working of PLC along with the components

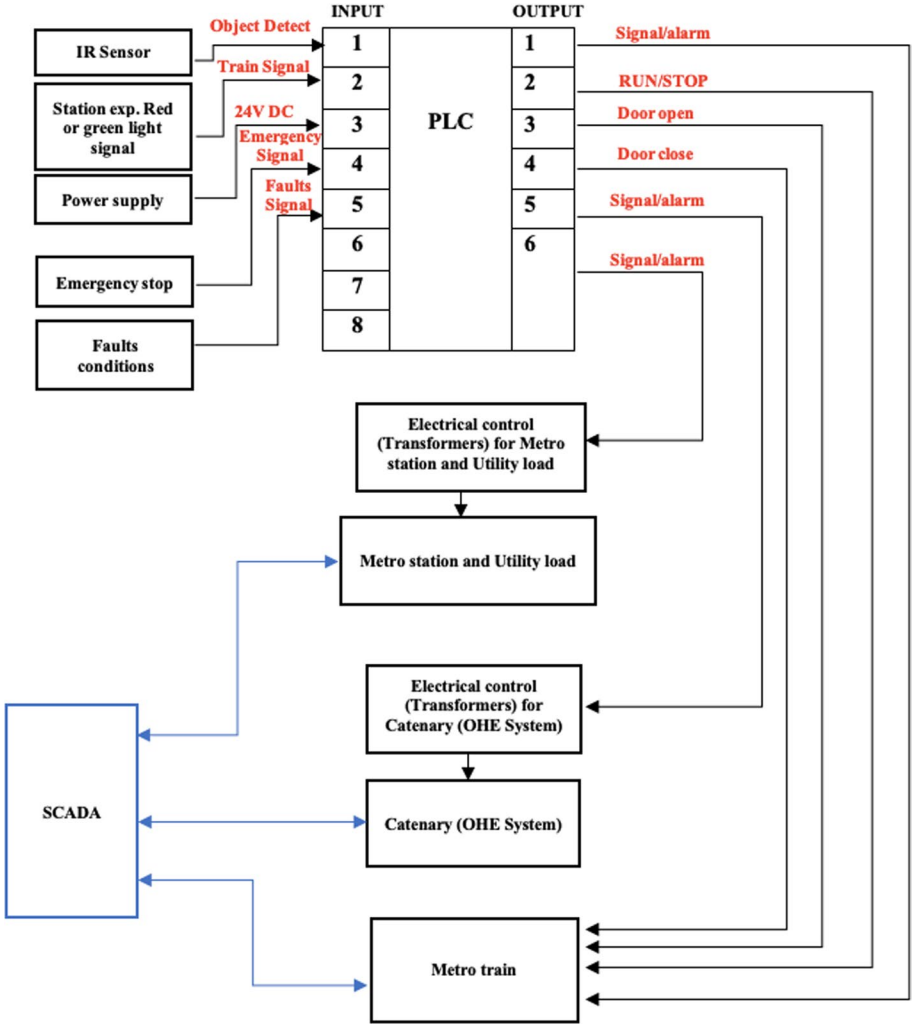


Fig. 3 Working of metro railways system using PLC and SCADA

3 Operations of the Metro Railways System

This section shows the operation of all the four applications of automated metro railways system.

3.1 Metro Platforms View Sub-Application

The driver of the trains controls the metro platforms, and the operator of any platform can manage the traffic of the metro station. Figure 6 and Table 1 show the working of metro platforms view sub-application. Here, t is timing tag of process/tag of slider.

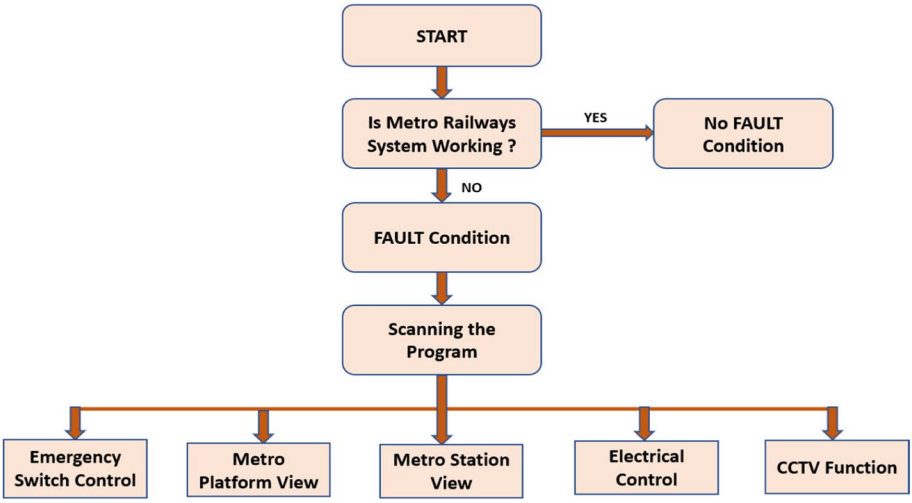


Fig. 4 Functioning of automated metro railways system operator

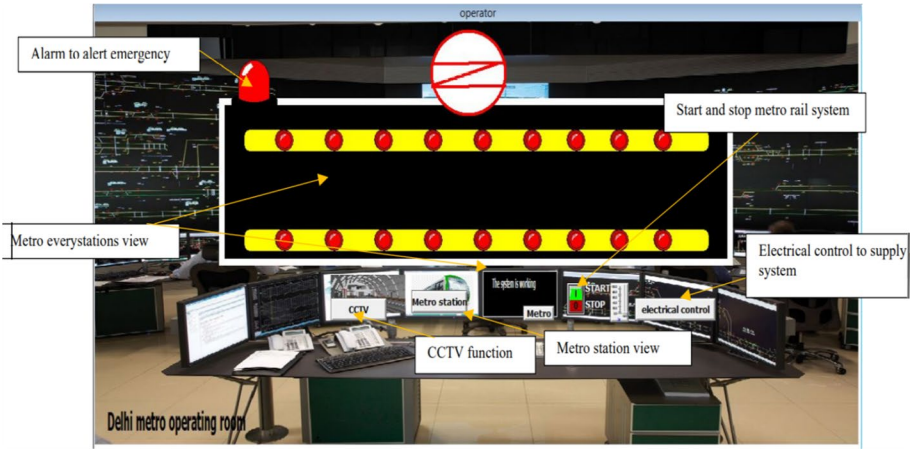


Fig. 5 Operational control center runtime window in Wonderware Intouch SCADA software version 10.0

3.2 Metro Stations (Traction SCADA System) View Sub-Application

Functioning of metro stations (traction SCADA) view sub-application is shown in Fig. 7. Metro train starts moving when the switch is turned ON. The LED in the control panel view indicates the proper location of the metro train. This system can track, control and recognize the location of the train even after a power outage or other breakdown. The emergency light will turn ON in case of any fault station and an alert message is displayed by the OCC in the relevant metro station on the HMI or operation display as shown in Fig. 16.

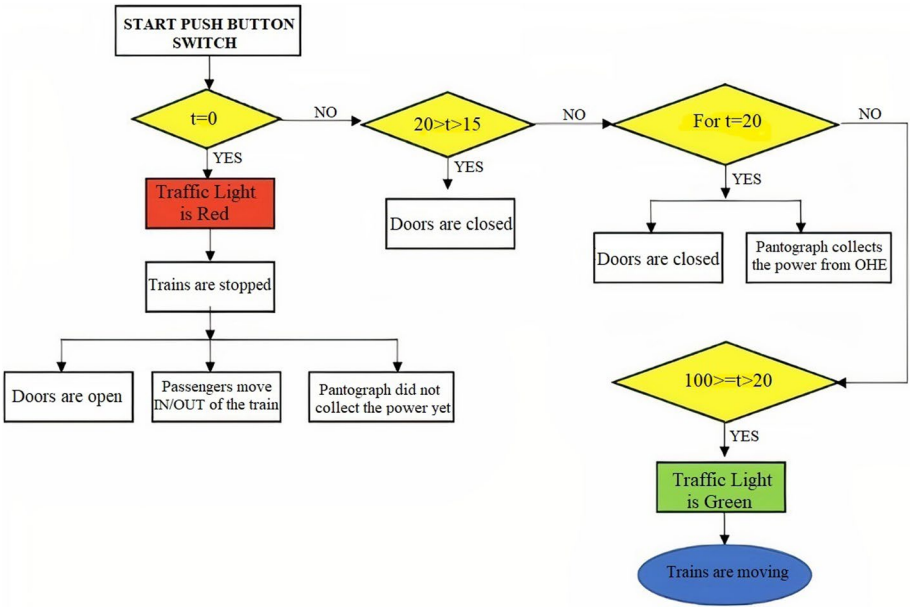


Fig. 6 Working of metro platforms view sub-application

Table 1 Description of metro platform view sub-application

S. No	State	Description
1	t = 0	Traffic light is red. The trains are stopped and passengers move out/in of the train and pantograph has not collected the power yet
2	20 > t > = 15	Metro train doors are closed. Train is ready to move
3	For t = 20	Metro train doors are closed. Pantograph collects the power and electrical signal on overhead system is ON. Train starts running
4	100 > = t > 20	Traffic light is green. The train starts moving forward

3.3 Traction Power Control and Power Supply System (Electrical Control)

Figure 8 shows the functioning of electrical control view sub-application. Metro rail lines must be electrified with 25 kV, single phase, 50 Hz for overhead equipment to deliver electrical energy to locomotives or trains for operating the motor. The substation, feeding post, and other equipment at the control post and switching station make up the power supply system. A catenary i.e. the overhead equipment (OHE) is used to deliver power to a metro train equipped with a pantograph in metro rail traction systems. Overhead equipment will incorporate sectioning posts, sub sectioning posts, and paralleling posts at various positions for maintenance purposes. Transformer1(T1) is a step-down transformer for metro stations and utility load, transformer2(T2) is a stepdown transformer for the catenary, and transformer3(T3) is a step-up transformer for 132 kV incoming line from generating station. This 132 kV supply will be stepped down to

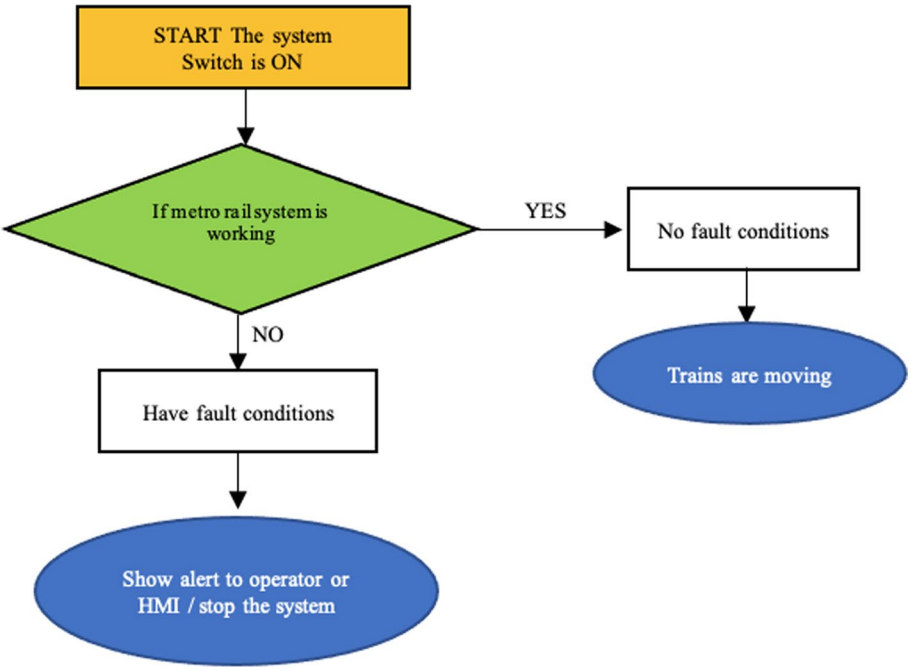


Fig. 7 Working of metro stations (traction SCADA system) view sub-application

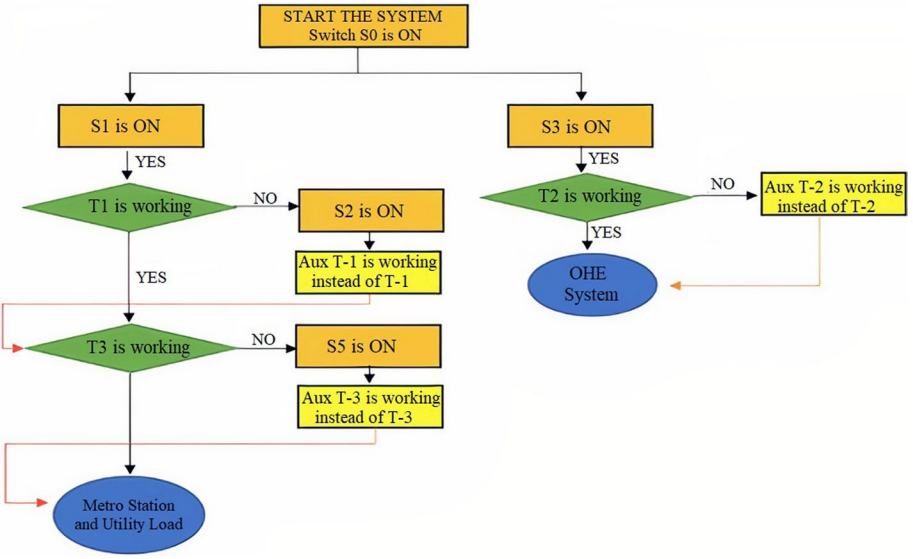


Fig. 8 Working of traction power control system

single phase 25 kV for traction purposes and three phase 33 kV for lighting and general use.

The system is turned on by pressing the button ON S0. The power control and power supply system where t is timing tag of process or the tag of slider is shown here in Table 2.

3.4 Video Surveillance and Control (CCTV) System

A sophisticated intrusion detection system (IDS) will be placed at the platform, with comprehensive CCTV or video surveillance coverage of stations, parking, and public spaces. Internal security monitoring systems such as the SCADA, automatic train supervision system, passenger information system and public address system can be implemented to minimize reaction time from personnel. A smart interface is provided to the control room operator using this CCTV integration with SCADA approach. They can make quick and effective judgements for running the metro system and delivering the required services. The functioning of CCTV view sub-application is shown in Fig. 9. SCADA systems use video data to get more information and a sharper view of the happenings at the remote locations.

CCTV integration with SCADA systems consists of remote video engine and video control center. Video cameras of every remote site are connected to the remote Video Engine (RVE) that record live videos continually. PLC or RTU acts as the local DVR for continuously storing and processing the video data from the cameras at remote locations. Remote PLC/RTU gets the camera-initiated alarms with the help of an integrated Modbus TCP server. Dedicated radio networks are used by remote locations to communicate processed data back to the control centre. Network bandwidth might vary from 9600 baud serial radios to TCP/IP radios of higher bandwidth. Video control centre (VCC) determines whether it has to send still images or video clips to the Wonderware HMI/SCADA system in the control room or not. The technology simultaneously transmits event clips and live video to the VCC at the control centre for the central office. The Wonderware SCADA system and this video surveillance system communicates directly with each other. The operator of the video surveillance system can view and handle video clips, as well as monitor alarms, using a Wonderware HMI. Through an EtherCAT link to the HMI computer, the remote locations can communicate to the main server. Using this CCTV integration with Wonderware HMI SCADA, operators can have the information about the real-time data of the system operations and can view the videos and monitor the status of the system at any time. The remote video engine can be used in a variety of ways, including alarm video clip mode, live video recording, and continuous monitoring. All cameras' video is sent at a user-defined frequency by the system. All cameras are visible to the system operator on a single scrolling system overview screen. In case of an alarm event, the system records a brief motion video clip and sends it to the control room. When the operator clicks on a short clip of video connected with an alarm, it will play and illustrate what was happening right before, during, and after the event that set off the alarm. In the absence of warnings, a system tour is carried out every 25 min, giving the operators a new frame of video and an updated perspective of the distant site. The HMI/SCADA system provides operators with access to a sizable amount of real-time process data. Operators may get live, streaming video from any of the video cameras positioned in useful places with a single mouse click. The live video feeds can show as pop-ups immediately on operators' HMI screens, allowing them to monitor what's happening across the facility.

Table 2 Description of Electrical control view

S. No	State	Description
1	S0 is ON	Power supply is ON
2	T1 and S1 are ON	Power supply(132 kV/33 kV) is fed to the stepdown transformer (T1), then fed to the step-up transformer (T3), and then fed to metro stations and utility load
3	For 50 < = t1 < 100	Transformer (T1) is damaged
4	S2 is ON	When transformer1 (T1) is damaged, the Auxiliary transformer1 (Aux-T1) is operated, and the power supply is fed to the stepdown transformer on standby (132 kV/33 kV). If transformer3 (T3) is also not working, the power supply is running through the Auxiliary transformer3 (Aux-T3 step-up transformer 33 kV/415 kV) and then fed to metro stations and utility load
5	S3 is ON	Stepdown transformer2 (T2, 132 kV/25 kV) is ON and the power supply is fed to catenary overhead electrical lines (OHE) through which the metro train is connected
6	For 50 < = t2 < 100	Transformer2 (T2) is damaged
7	S4 is ON	Auxiliary Transformer2 (Aux-T2) is ON and the power supply (132 kV/25 kV) is fed to the catenary (OHE)

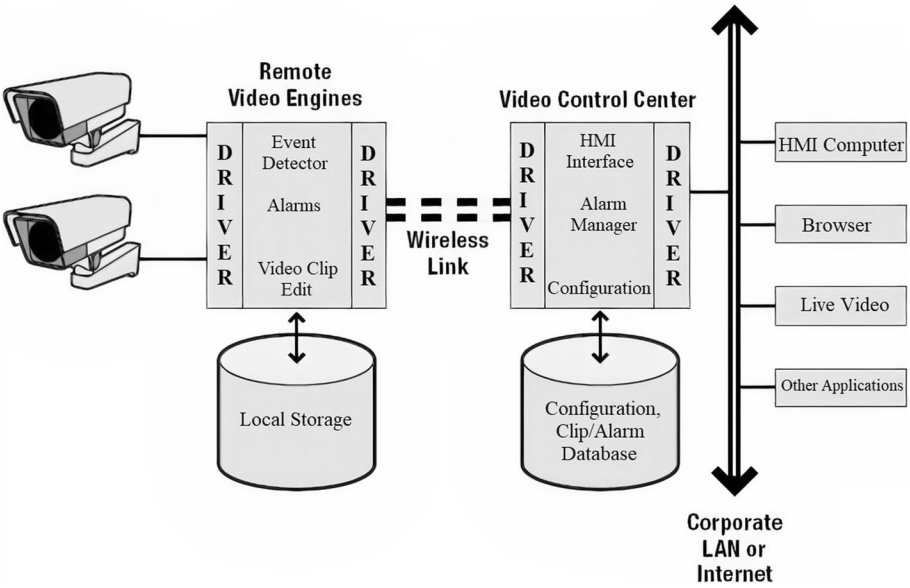


Fig. 9 CCTV integration with SCADA

4 Simulation Results with PLC Ladder Logic Program

A SCADA Wonderware Intouch software window of every operation is shown here along with the ladder logic program of PLC.

4.1 Operational Control Center (OCC)

Figure 10 shows the OCC runtime window in SCADA Wonderware Intouch software from where the operator may command and monitor. If the metro rail system is not working, then the operator can click to switch control or CCTV or metro station icon or electrical control icon. The IR and reflecting sensors on the RTU are used to scan the programs. If

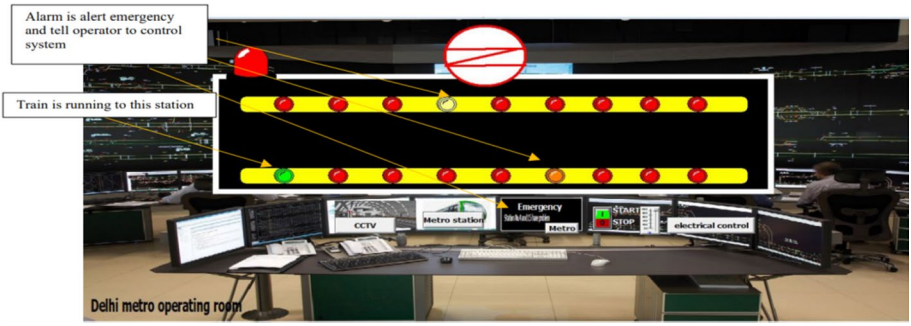


Fig. 10 OCC view of main application with failure in the system

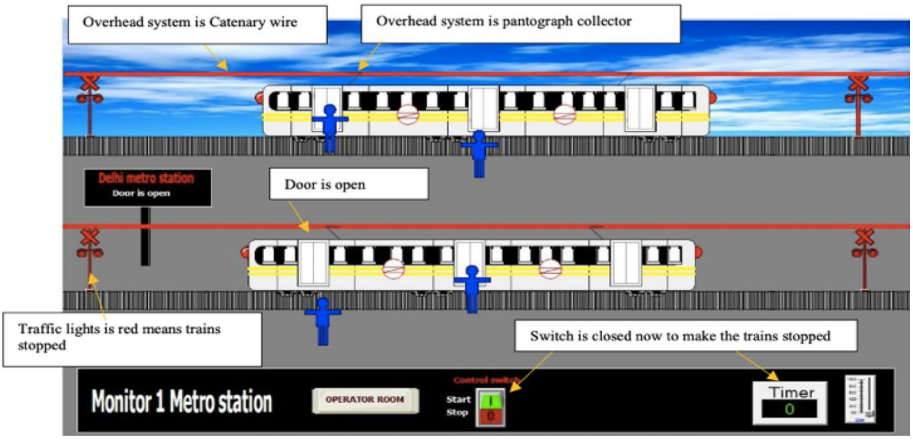


Fig. 11 For t = 0, Trains are stopped and passengers move out/in of the train

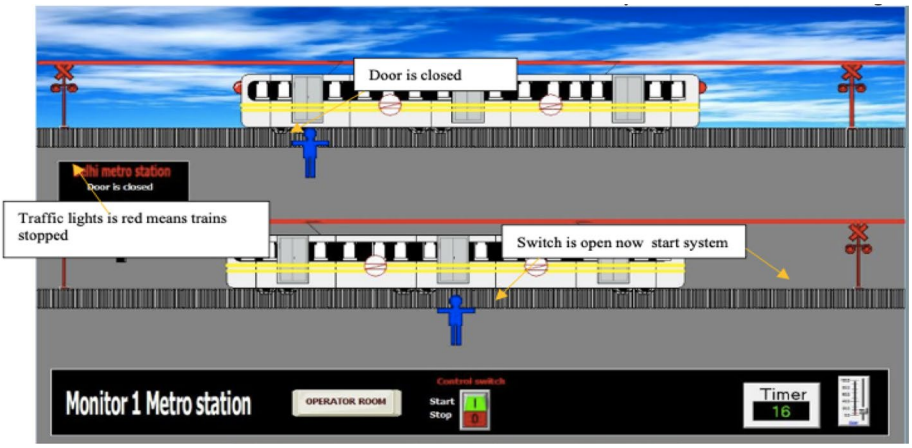


Fig. 12 For 20 > t > = 15, Metro train doors are closed

there is a problem with the process, the RTU sends a message to the master terminal unit, which then sends signals or alerts to the operator on the screen.

4.2 Metro Platforms View Sub-Application

According to the proposed system, the station master has only to give input once, which will go to the PLC, after which the PLC provides its output according to the timer circuit. In this way, the complete task can be done without too much human interference. All the cases of metro platforms view sub-application are shown here with the help of runtime window in SCADA Wonderware Intouch software system and PLC ladder logic. Figure 11, Fig. 12, Fig. 13 were generated using Wonderware Intouch SCADA software version 10.0, https://lab4sys.com/en/download-wonderware-intouch-scada/.

Fig. 13 **a** For t = 20, Metro train doors are closed. Pantograph collects the power and electrical signal on overhead system is ON. **b** For 100 > = t > = 20, Traffic light is green and train starts moving forward

(1) When t = 0 (t is timing tag of process/tag of slider), the traffic light is red. The trains are stopped. Metro train doors are open and passengers move out/in of the train as shown in Fig. 11. Pantograph has not collected the power yet.
(2) When 20 > t > = 15, metro train doors are closed and train is ready to move as shown in Fig. 12.
(3) When 100 > = t > = 20, metro train doors are closed. Pantograph collects the power and electrical signal on overhead system is ON. And train starts running. When the traffic light is green, train starts moving forward as shown in Fig. 13. Figure 14 shows the PLC ladder logic program of metro platforms view sub-application.

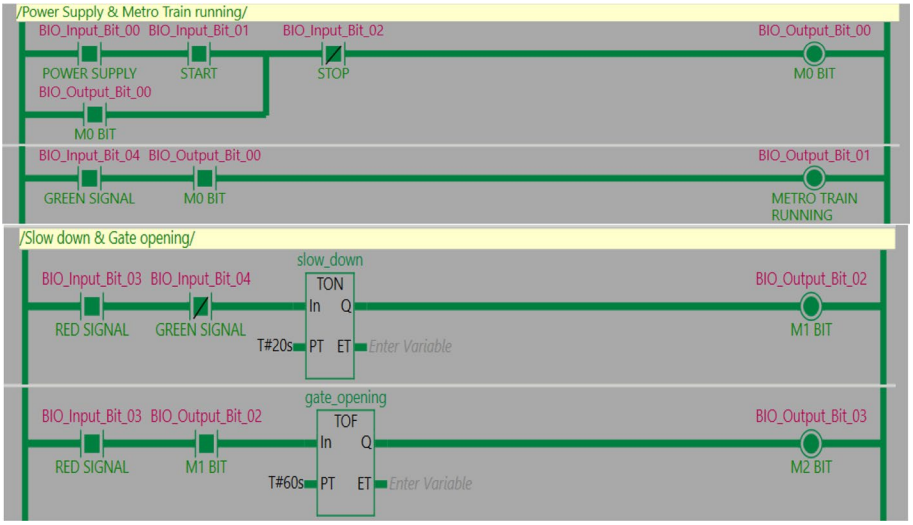


Fig. 14 PLC ladder logic program of metro platforms view sub-application

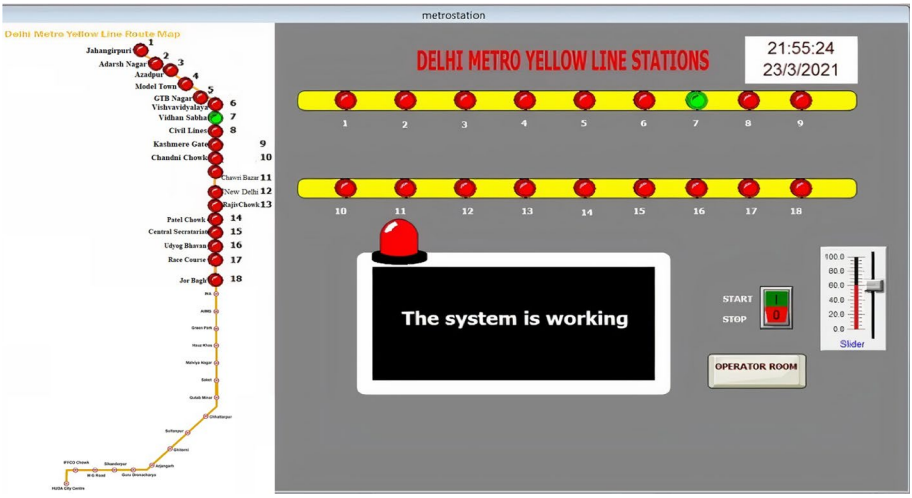


Fig. 15 Metro stations (traction SCADA) of sub-application

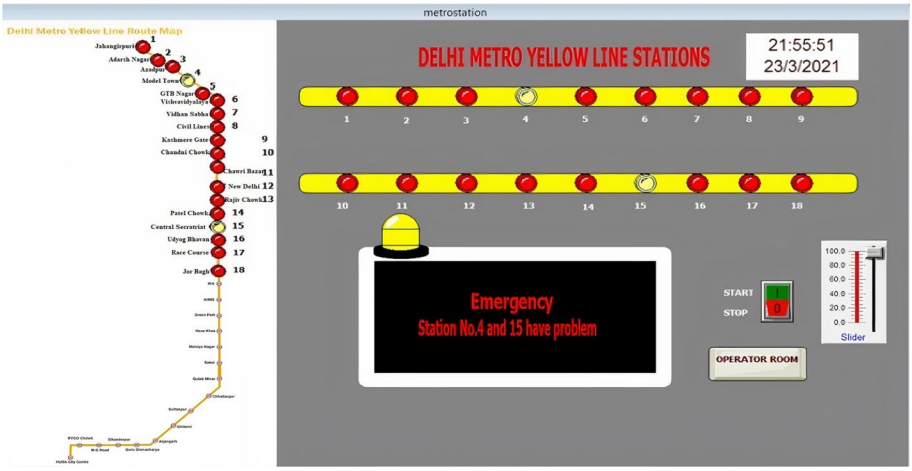


Fig. 16 Metro stations (traction SCADA) of sub-application (failure)

There is a process object in SCADA software. We assign equipment identification, station number, block number, and bit number to link the application in a system with the field. Figure 15 shows the metro stations (traction SCADA) of sub-application. Metro train starts moving when the switch is turned ON. The LED in the control panel view indicates the proper location of metro train. This system can track, control and recognize the location of the train even after a power outage or other breakdown. The emergency light will turn ON in case of any fault station and an alert message is displayed by the OCC in the relevant metro station on the HMI or operation display (see Fig. 16).

4.3 Electrical Control (Traction Power Control and Power Supply System) View Sub-Application

Here, the functioning of Electrical control view sub-application is shown. Metro rail lines must be electrified with 25 kV, single phase, 50 Hz for overhead equipment to deliver electrical energy to locomotives or trains for operating the motor. The substation, feeding post, and other equipment at the control post and switching station make up the power supply system. A catenary i.e. the overhead equipment (OHE) is used to deliver power to a metro train equipped with a pantograph in metro rail traction systems. Transformer1(T1) is a stepdown transformer for metro stations and utility load, transformer2(T2) is a stepdown transformer for the catenary, and transformer3(T3) is a step-up transformer for 132 kV incoming line from generating station. This 132 kV supply will be stepped down to single phase 25 kV for traction purposes and three phase 33 kV for lighting and general use. There is a PLC program implemented in SSP and SP posts. When the OHE 25 kV supply of the mainline fails, paralleling interrupter opens automatically within a predefined time. This program has been loaded in the RTU itself so that the operator can take action faster.

(1) No Fault Condition: For running the system, a power supply of (132 kV/33 kV) is fed to the stepdown transformer (T1), then fed to the step-up transformer (T3), and then fed to metro stations and utility load as shown in Fig. 17. Figure 18 shows the PLC ladder logic program of electrical control view sub-application with no fault condition.

(2) For $50 <= t1 < 100$: Switch ON S0 and S1 means the system are working. 132 kV in coming lines from generating station to transformer1 is stepdown transformer from 132 to 33 kV. Transformer3 is step-up transformer from 33 to 415 kV for metro station and utility load. Alarm tells the operator that transformer1 is damaged. When transformer1 (T1) is damaged, the auxiliary transformer1 (Aux-T1) is operated, and the power supply is fed to the stepdown transformer on standby (132 kV/33 kV). Switch ON S0 and S2 means the system are working. The auxiliary transformer1 provide the power in case of Transformer1 damage as shown in Fig. 19. Figure 20 shows the ladder logic program of PLC for transformer T1 damage.

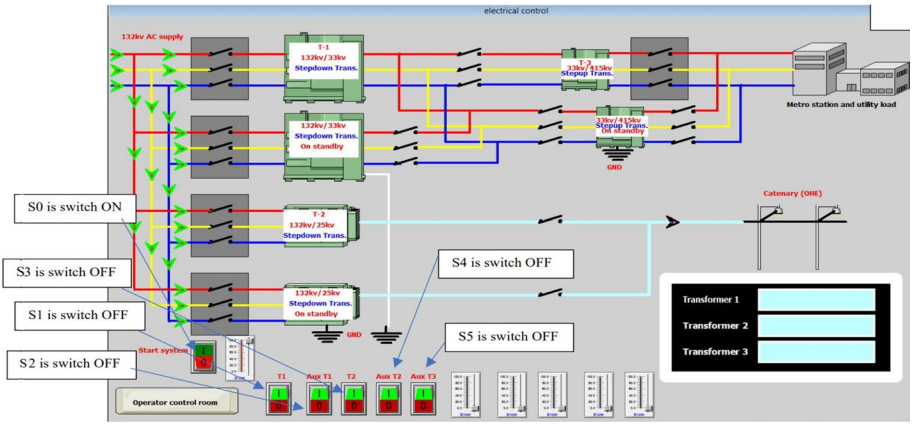


Fig. 17 S0 is ON, Power supply is ON (No fault Condition)

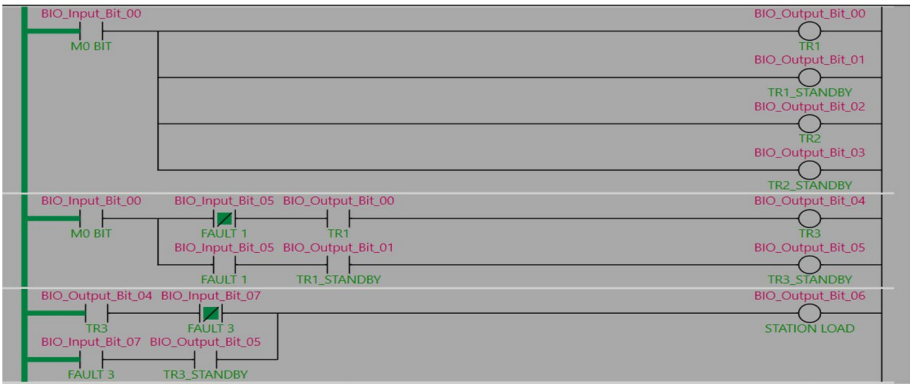


Fig. 18 PLC ladder logic program of electrical control view sub-application (No fault condition)

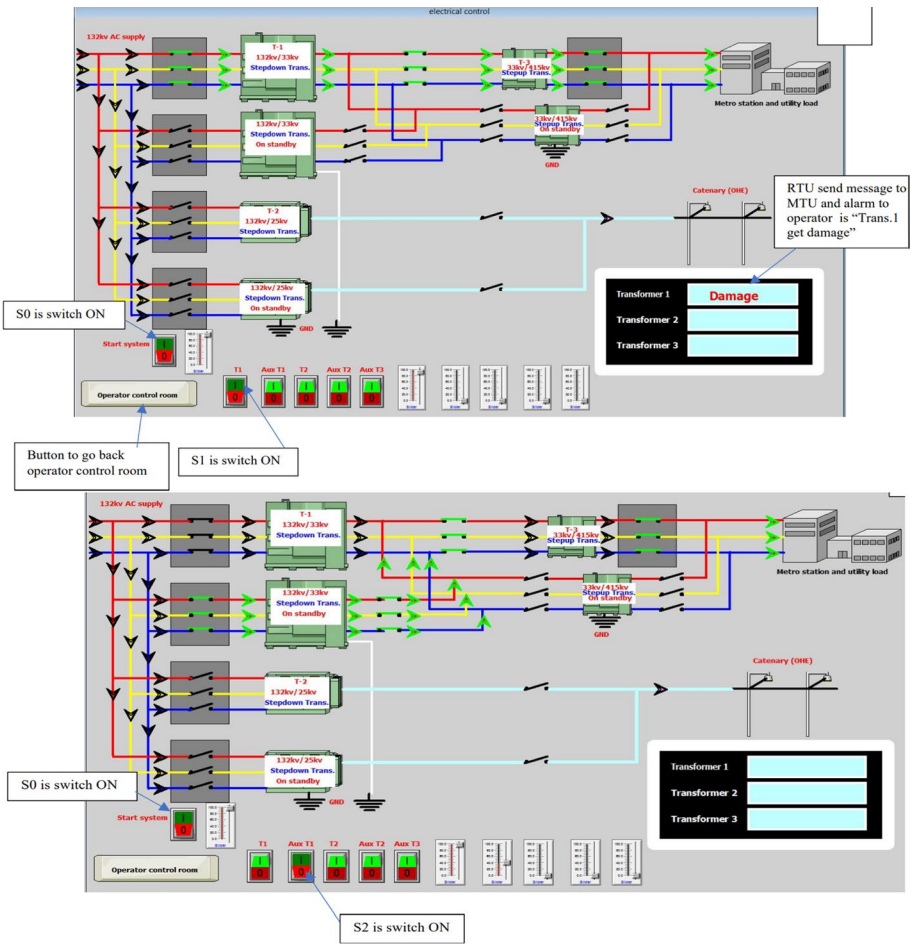


Fig. 19 For 50 < = t1 < 100, When Transformer (T1) is damaged, then Aux-T1 is operated

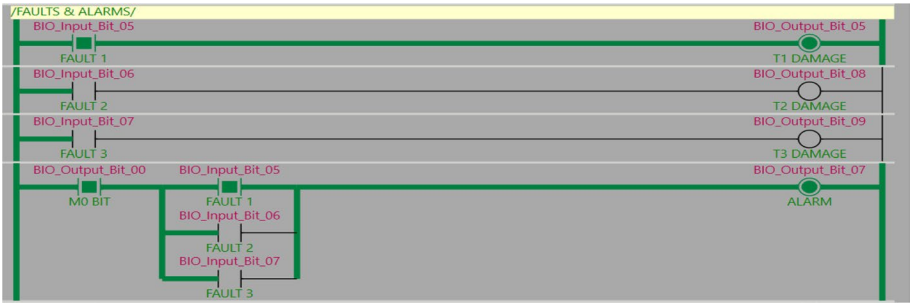


Fig. 20 Ladder logic program of PLC (Transformer T1 damage)

(3) For $50 <= t2 < 100$: Switch ON S0, S2 and S3 means that the system is working. Stepdown transformer2 (T2, 132 kV/25 kV) is ON and the power supply is fed to catenary overhead electrical lines (OHE) through which the metro train is connected. After sometime if transformer2(T2) gets damaged, then Auxiliary transformer2 (Aux-T2) provides the power in case of transformer2(T2) as shown in Fig. 21. Switch ON S0, S2, and S4 means the system is working with Aux-T2. Figure 22 shows the ladder logic program of PLC for transformer T2 damage.

(4) When Stepdown transformer3(T3) gets damaged, then Auxiliary transformer3 (Aux-T3) provide the power in case of T3 as shown in Fig. 23. Switch ON S0, S4, and S5 means the system is working with Aux-T3. Figure 24 shows the ladder logic program of PLC in case of faults (when Aux T1, Aux T2, Aux T3 is standby) and Fig. 25 shows the ladder logic program of PLC in case of T1, T2, T3 damage.

4.4 CCTV

CCTV integration with SCADA systems continuously record the live videos of every remote site. The Wonderware SCADA system and this video surveillance system communicates directly with each other. The Wonderware HMI screen of CCTV sub application is shown is Fig. 26. Through an EtherCAT link to the HMI computer, the remote locations can communicate to the main server. Using this CCTV integration with Wonderware HMI SCADA, operators can have the information about the real-time data of the system operations and can view the videos and monitor the status of the system at any time. All cameras' video is sent at a user-defined frequency by the system.

In the absence of warnings, a system tour is carried out every 25 min, giving the operators a new frame of video and an updated perspective of the distant site. Operators may get live, streaming video from any of the video cameras positioned in useful places with a single mouse click. The live video feeds can show as pop-ups immediately on operators' HMI screens by travelling via the same broadband fiber optic-based EtherCAT network.

Figure 27 shows the input and output mapping of PLC ladder logic for the proposed system in OMRON's Sysmac studio programming software.

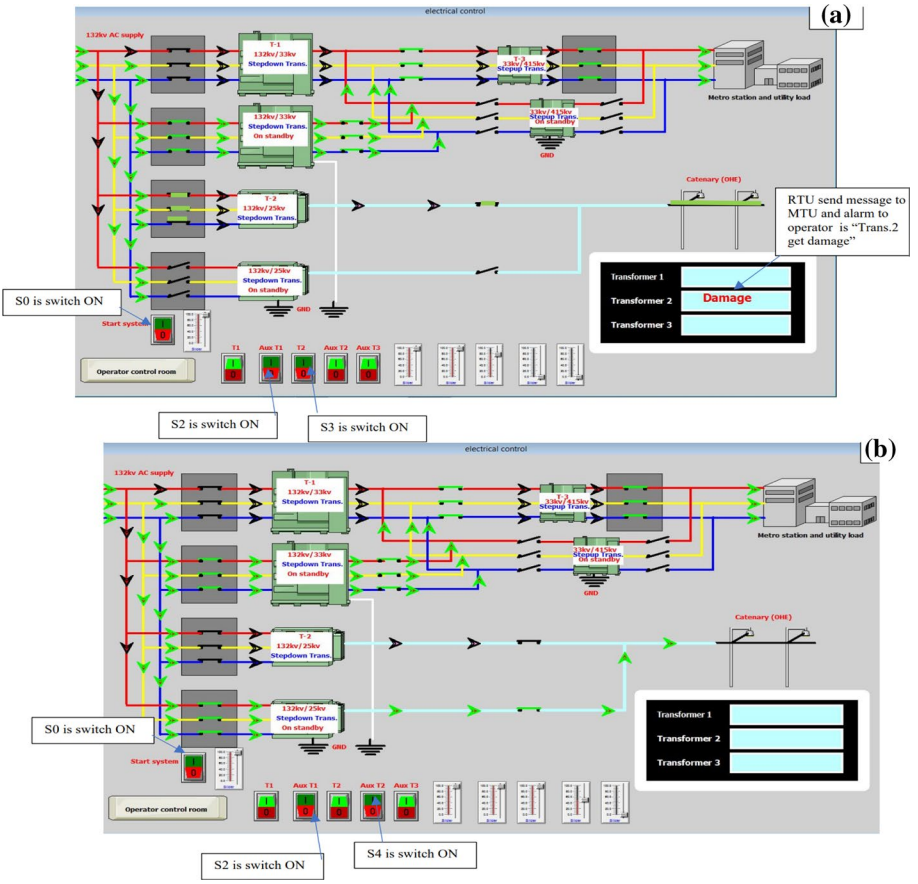


Fig. 21 For 50 < = t2 < 100, When Transformer (T2) is damaged, then Aux-T2 is operated

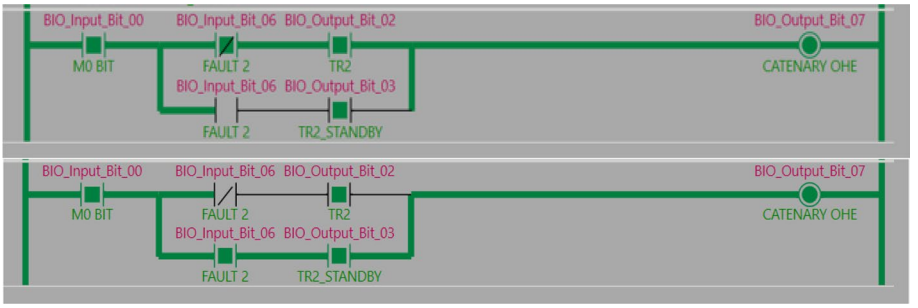


Fig. 22 Ladder logic program of PLC (Transformer T2 damage)

5 Comparative Analysis and Discussion

Industrial control systems and SCADA systems play a critical role in the management and regulation of critical infrastructure. SCADA systems are bringing us closer to the

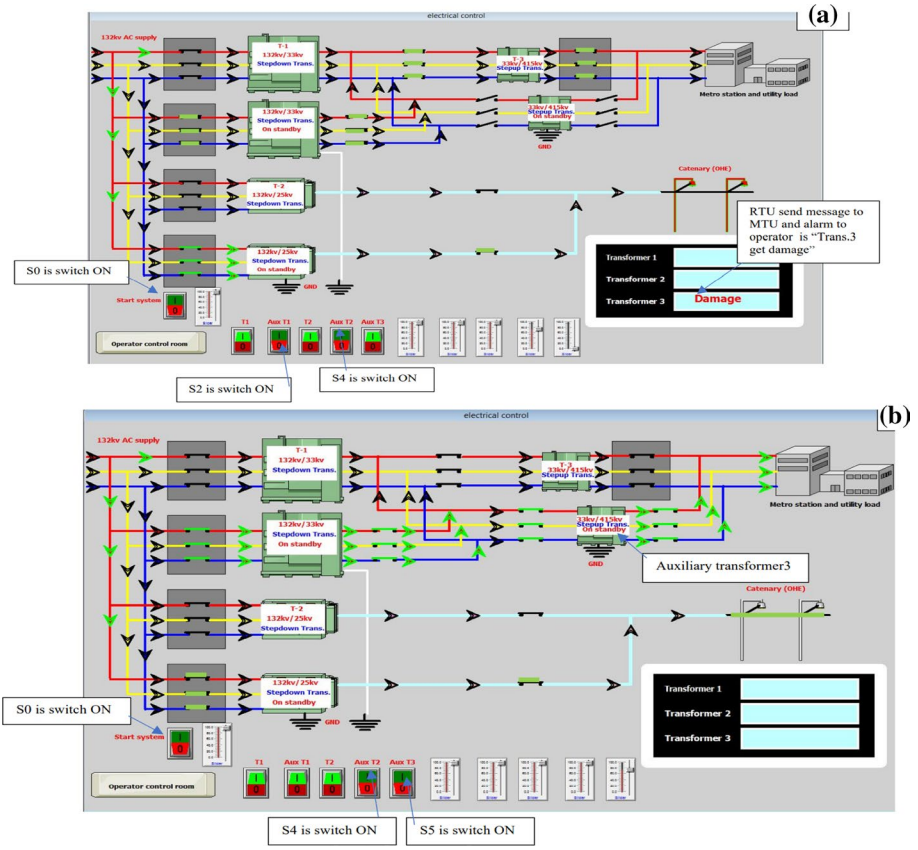


Fig. 23 When T3 gets damaged, then Aux-T3 provides the power in case of T3

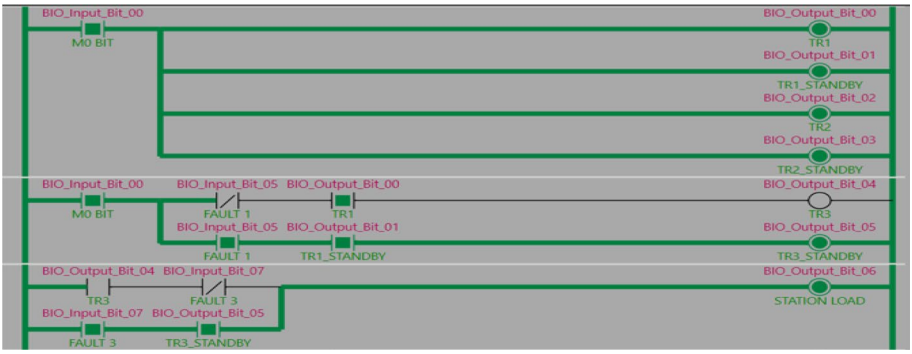


Fig. 24 Ladder logic program of PLC in case of faults (when Aux T-1, Aux T-2, Aux T-3 is standby)

real-time application world. SCADA systems increase the efficiency of a control system operation while also providing improved protection for the equipment it uses. Furthermore, it increases employee productivity. SCADA frameworks use an established monitoring

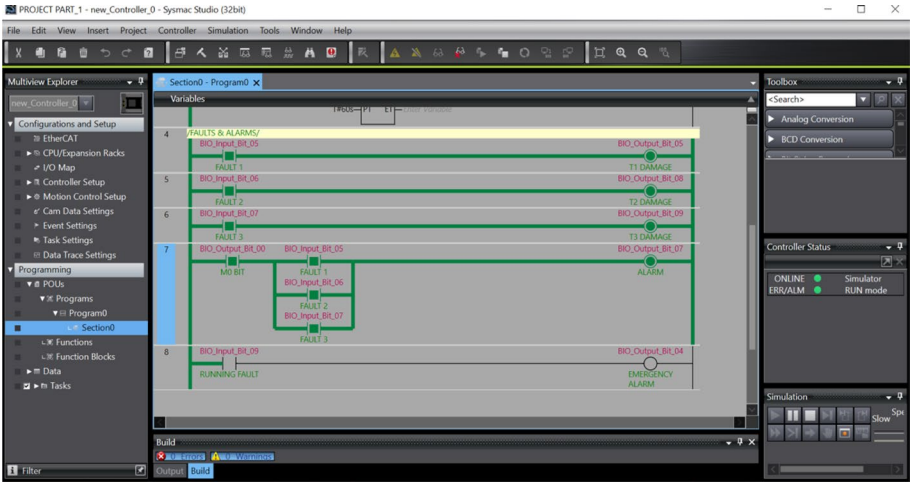


Fig. 25 Ladder logic program of PLC in case of any faults (T1, T2, T3 damage)

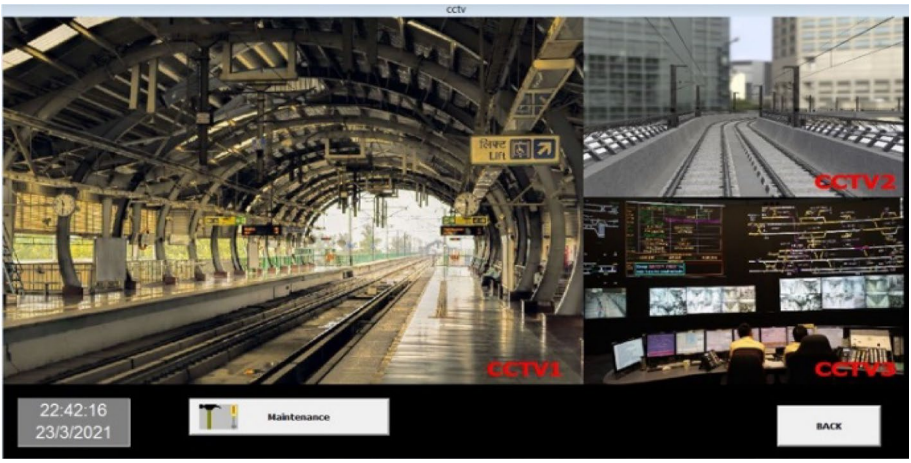
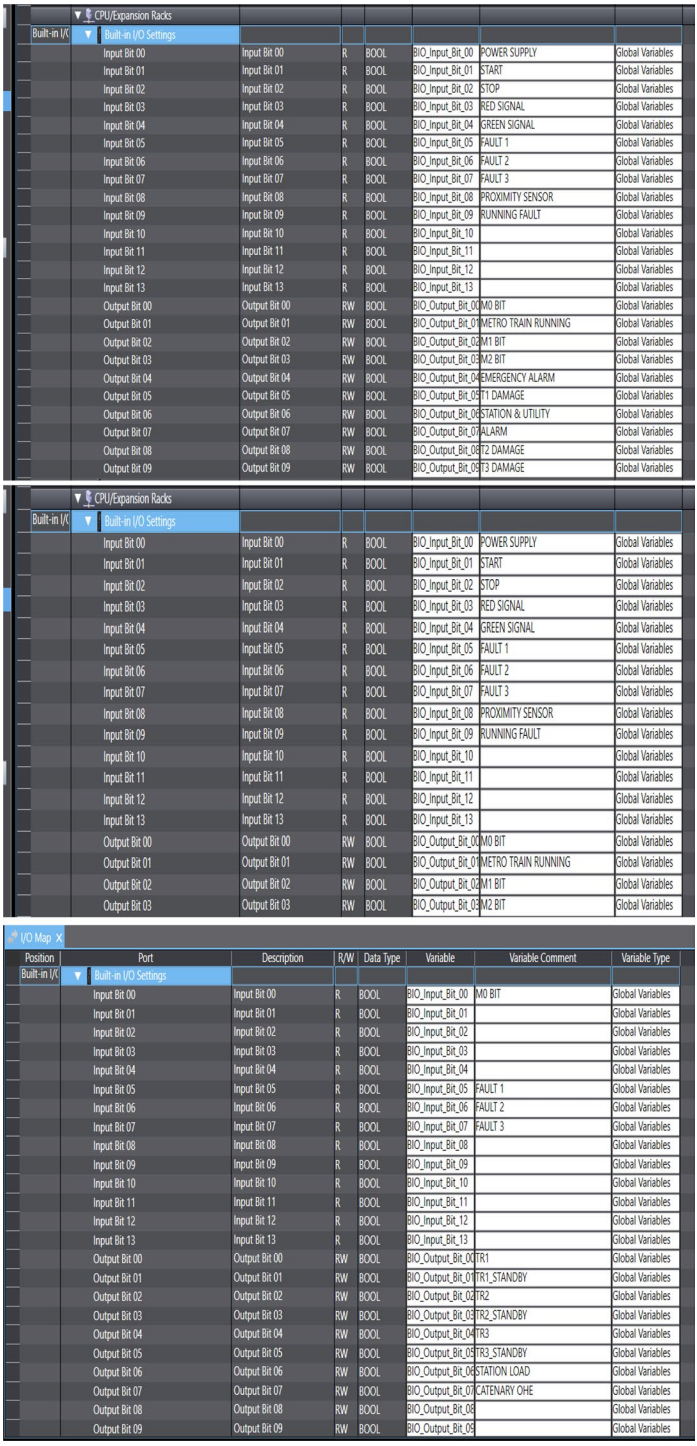


Fig. 26 CCTV view sub-application window in Wonderware Intouch SCADA software version 10.0

platform, advanced communication system, and sensors to provide required information and timely alerts/warnings to observing stations. The designed SCADA system offers continuous access and visualization of all the data in a more efficient and timely manner. It also provides a centralized and thorough display of parameters. The use of "Wonderware Intouch" software facilitates the gathering and visualization of data. The data collecting procedure in the proposed SCADA system is primarily centered on creating adequate communication links between SCADA server and numerous PLCs and sensors that measure and send the readings of a variety of valuable parameters. The performance comparison of the proposed PLC and SCADA based system is done with existing CBTC schemes, cognitive and cooperative CBTC schemes. The information gap does not increase much when the communication latency in network is increased. The suggested technique also reduces the discrepancy between pre-set travel time and actual travel time. In compared to other



Position	Port	Description	R/W	Data Type	Variable	Variable Comment	Variable Type
Built-in I/O	Built-in I/O Settings						
	Input Bit 00	Input Bit 00	R	BOOL	BIO_Input_Bit_00	M0 BIT	Global Variables
	Input Bit 01	Input Bit 01	R	BOOL	BIO_Input_Bit_01		Global Variables
	Input Bit 02	Input Bit 02	R	BOOL	BIO_Input_Bit_02		Global Variables
	Input Bit 03	Input Bit 03	R	BOOL	BIO_Input_Bit_03		Global Variables
	Input Bit 04	Input Bit 04	R	BOOL	BIO_Input_Bit_04		Global Variables
	Input Bit 05	Input Bit 05	R	BOOL	BIO_Input_Bit_05	FAULT 1	Global Variables
	Input Bit 06	Input Bit 06	R	BOOL	BIO_Input_Bit_06	FAULT 2	Global Variables
	Input Bit 07	Input Bit 07	R	BOOL	BIO_Input_Bit_07	FAULT 3	Global Variables
	Input Bit 08	Input Bit 08	R	BOOL	BIO_Input_Bit_08		Global Variables
	Input Bit 09	Input Bit 09	R	BOOL	BIO_Input_Bit_09		Global Variables
	Input Bit 10	Input Bit 10	R	BOOL	BIO_Input_Bit_10		Global Variables
	Input Bit 11	Input Bit 11	R	BOOL	BIO_Input_Bit_11		Global Variables
	Input Bit 12	Input Bit 12	R	BOOL	BIO_Input_Bit_12		Global Variables
	Input Bit 13	Input Bit 13	R	BOOL	BIO_Input_Bit_13		Global Variables
	Output Bit 00	Output Bit 00	RW	BOOL	BIO_Output_Bit_00	TR1	Global Variables
	Output Bit 01	Output Bit 01	RW	BOOL	BIO_Output_Bit_01	TR1_STANDBY	Global Variables
	Output Bit 02	Output Bit 02	RW	BOOL	BIO_Output_Bit_02	TR2	Global Variables
	Output Bit 03	Output Bit 03	RW	BOOL	BIO_Output_Bit_03	TR2_STANDBY	Global Variables
	Output Bit 04	Output Bit 04	RW	BOOL	BIO_Output_Bit_04	TR3	Global Variables
	Output Bit 05	Output Bit 05	RW	BOOL	BIO_Output_Bit_05	TR3_STANDBY	Global Variables
	Output Bit 06	Output Bit 06	RW	BOOL	BIO_Output_Bit_06	STATION LOAD	Global Variables
	Output Bit 07	Output Bit 07	RW	BOOL	BIO_Output_Bit_07	CATENARY OHE	Global Variables
	Output Bit 08	Output Bit 08	RW	BOOL	BIO_Output_Bit_08		Global Variables
	Output Bit 09	Output Bit 09	RW	BOOL	BIO_Output_Bit_09		Global Variables

Fig. 27 Input and output mapping of PLC ladder logic

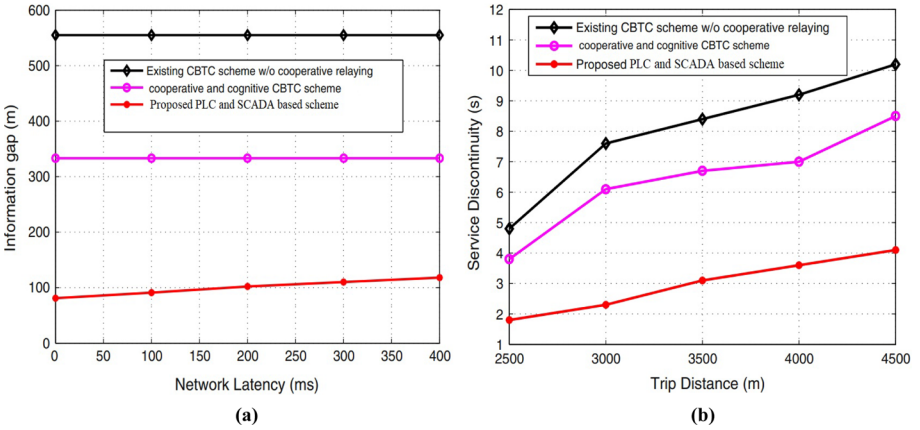


Fig. 28 Comparison of **(a)** Information Gap **(b)** Average service discontinuity

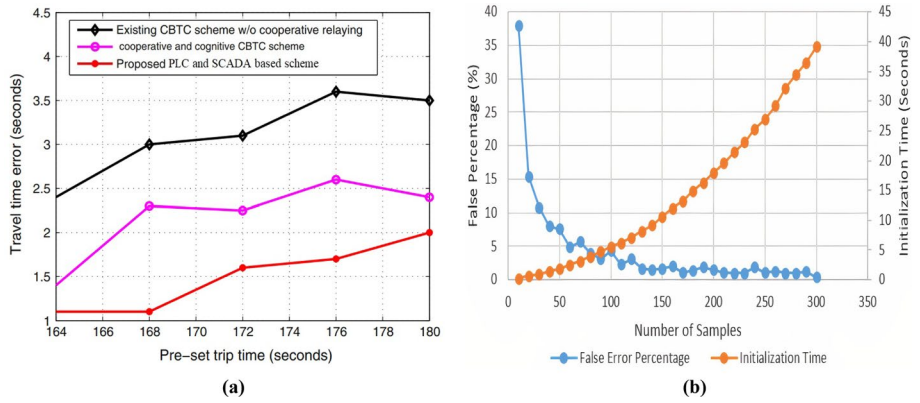


Fig. 29 Comparison of **(a)** Error of Train Travel Time **(b)** Initialization time and false error percentage

approaches, this strategy provides improved average service discontinuity time length and enhances the performance of train control. The proposed system decreases the false error percentage in comparison to the traditional systems (Figs. 28 and 29). Therefore, this system is more stable and reliable for train control in comparison to the traditional systems. Table 3 shows the main characteristics and contribution of the developed platform.

6 Conclusion

In this work, a real time monitoring and train control system has been used in to minimize the failure in automated metro railways system operator and integrate the information coming from operational control centre, traction SCADA system, traction power control, and power supply system. Here, SCADA is used for the visualization of an automated process operation and then the whole operation is regulated using OMRON (NX1P2-9024DT1) PLC. OMRON's Sysmac studio programming software is used for developing the ladder

Table 3 Key features and contribution of the developed platform

S. No	Aspect	Contribution
1	System installation	System installation is easy and any new devices be incorporated easily to traditional power switches with few modifications
2	Cost Saving	Using a single central unit PLC that implements the load sharing algorithm and sends control commands to the basic switching parts Using Wonderware Intouch SCADA software for the visualization of the operation Eliminates the need to purchase extra network units or pay for a service subscription by using a secure remote access platform
3	Interoperability	Safe and uninterrupted monitoring, control and signalling of the metro railways infrastructure
4	Operation sustainability	Good operation sustainability by having a self-restoration capability after a power interruption
5	Convenience	Provides remote access, and displays all parameters of the system in a single SCADA web interface
6	Security	Hierarchical authentication scheme for management and visualization for controlling railways assets in SCADA system
7	Extendibility	Extendible by configuring more devices, switches and screens in the SCADA software
8	Equipment Maintenance	Predictive maintenance practices could be scheduled on the basis of monthly performance evaluations The power quality indices at different nodes of railway system can be used to solve malfunctioning issues

logic of PLC. Wonderware Intouch SCADA software is used for the visualization of the operation. This system helps in detecting exact location, avoiding train collisions and opening emergency exit even after the interruption of power system with the use of programmable logic controller. SCADA frameworks provide required information and timely alerts/warnings to observing stations. The designed SCADA system offers continuous access and visualization of all the data in a more efficient and timely manner. It also provides a centralized and thorough display of parameters. The designed SCADA systems increase the efficiency of a control system operation & overcomes the drawbacks of the current CATC and CBTC system. Based on PLCs linked by industrial EtherCAT networks, operators can monitor and manage the entire operation from any of on-site HMI, or remotely using SCADA software. Surveillance and remote monitoring & control functions using SCADA provides a cost-effective security solution.

Author contributions All authors read and approved the final manuscript.

Funding The authors received no specific funding for this study.

Data availability The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Code availability Code is available.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Valadão, N., Künzel, G., Müller, I., & Pereira, C. E. (2018). Industrial wireless automation: Overview and evolution of WIA-PA. *IFAC-PapersOnLine, 51*(10), 175–180.
2. Adolf, M. (2012) More energy efficiency through process automation. *ZVEI - Zentralverband Elektrotechnikund Elektronikindustrie e.V.,* pp. 1–16, https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2013/januar/More_energy_efficiency_through_process_automation/ZVEI_Energienutzung-englisch.pdf
3. Why industrial automation is necessary to keep pace with rising power consumption (2022). *SAAB RDS Technology Enabler*. https://saabrds.com/why-industrial-automation-is-necessary-to-keep-pace-with-rising-power-consumption/
4. Bonavita, N. (2013). How process automation can increase energy efficiency. *ABB SpA*, Genoa, Italy. https://library.e.abb.com/public/953141f99f2f77f985257b810060bb17/Hydrocarbon%20Processing%20June%202013%20Energy%20Efficiency.pdf
5. Chauhan, R. K., & Chauhan, K. (2019). Building automation system for grid-connected home to optimize energy consumption and electricity bill. *Journal of Building Engineering, 21*, 409–420.
6. Karunanithy, K., & Velusamy, B. (2020). Cluster-tree based energy efficient data gathering protocol for industrial automation using WSNs and IoT. *Journal of Industrial Information Integration, 19*, 1–11.
7. Cao, H., Zhou, J., Jiang, P., Hon, K. K. B., Yi, H., et al. (2020). An integrated processing energy modeling and optimization of automated robotic polishing system. *Robotics and Computer-Integrated Manufacturing, 65*, 101973.
8. Ang, J. H., Goh, C., Saldivar, A. A. F., & Li, Y. (2017). Energy-efficient through-life smart design, manufacturing and operation of ships in an industry 4.0 environment. *Energies, 10*(5), 610.
9. Tomar, B. & Kumar, N. (2020). PLC and SCADA based industrial automated system. In *2020 IEEE International Conference for Innovation in Technology (INOCON)*, pp. 1–5.
10. Nadgauda, N. & Muthukumaraswamy, S. A. (2019). Design and development of industrial automated system using PLC-SCADA. In *2019 IEEE 10th GCC Conference & Exhibition (GCC),* pp. 1–6

11. Longo, M. & Bramani, M. (2015). The automation control systems for the efficiency of metro transit lines. In *2015 AEIT International Annual Conference (AEIT)*, pp. 1–6.
12. Sripriya, S., Divya, V. & Babu, T. (2017). Building management system in metro rail. In 2017 Third International Conference on Sensing, *Signal Processing and Security (ICSSS),* pp. 409–413
13. Shang, F., Zhan, J. & Chen, Y. (2018). Distributed model predictive control for train regulation in urban metro transportation. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC),* pp. 1592–1597
14. Cao, Y., Ma, L., & Zhang, Y. (2018). Application of fuzzy predictive control technology in automatic train operation. *Cluster Computing, 22*(1), 123–132.
15. Standardization of signalling and train control systems for metro railways. (2013). Report of the sub-committee *Ministry of Urban Development, Government of India,* November 2013.
16. Wang, H., Zhao, N., Ning, B., Tang, T., & Chai, M. (2018). Safety monitor for train-centric CBTC system. *IET Intelligent Transport Systems, 12*(8), 931–938.
17. Wang, Y., Chen, L., Wei, J., Kirkwood, D., Xu, Q. et al. (2016). On-line conformance testing of the communication-based train control (CBTC) system. In *IEEE International Conference on Intelligent Rail Transportation (ICIRT)*, pp. 328–333.
18. Q. Wei, R. Niu and T. Tang, "Assessment of metro signaling system resilience," *2018 International Conference on Intelligent Rail Transportation (ICIRT),* pp. 1–5, 2018.
19. Dimitrova, E. (2019). Analysis of automatic control systems for metro trains. In *2019 11th Electrical Engineering Faculty Conference (BulEF),* pp. 1–4.
20. He, D., Yang, Y., Zhou, J., & Chen, Y. (2019). Optimal control of metro energy conservation based on regenerative braking: A complex model study of trajectory and overlap time. *IEEE Access, 7*, 68342–68358.
21. Liu, R., Li, S., & Yang, L. (2018). Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy. *Omega, 90*, 1–33.
22. Rekik, M., Gransart, C. & Berbineau, M. (2018). Analysis of security threats and vulnerabilities for train control and monitoring systems. In Proc. *2018 15th International Multi-Conference on Systems, Signals & Devices (SSD),* pp. 693–698.
23. Carvajal-Carreño, W., Cucala, A. P., & Fernández-Cardador, A. (2016). Fuzzy train tracking algorithm for the energy efficient operation of CBTC equipped metro lines. *Engineering Applications of Artificial Intelligence, 53*, 19–31.
24. Wang, Y., Cong, J., Wang, P., Liu, X., & Tang, H. (2019). A data-fusion approach for speed estimation and location calibration of a metro train based on low-cost sensors in smartphones. *IEEE Sensors Journal, 19*(22), 10744–10752.
25. Yin, J., Yang, L., Tang, T., Gao, Z., & Ran, B. (2017). Dynamic passenger demand-oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. *Transportation Research Part B: Methodological, 97*, 182–213.
26. Carrel, A., Mishalani, R. G., Wilson, N. H. M., & Attanucci, J. P. (2013). A framework for evaluating operations control on a metro line: Integrating multiple perspectives and automatically collected train and passenger movement data. *Public Transport, 5*, 149–176.
27. Bai, Y., Cao, Y., Yu, Z., Ho, T. K., Roberts, C., & Mao, B. (2020). Cooperative control of metro trains to minimize net energy consumption. *IEEE Transactions on Intelligent Transportation Systems, 21*(5), 2063–2077.
28. Wang, X., Su, S. & Tang, T. (2021). Cooperative control among trains in the undersupplied area during the fault of traction power supply system in metro line. In *2021 IEEE Intelligent Transportation Systems Conference (ITSC),* pp. 2411–2416.
29. Alphonsus, E. R., & Abdullah, M. O. (2016). A review on the applications of programmable logic controllers (PLCs). *Renewable and Sustainable Energy Reviews, 60*(8), 1185–1205.
30. Yan, F., Gao, C., & Tang, T. (2017). A safety management and signaling system integration method for communication-based train control system. *Urban Rail Transit, 3*(2), 90–99.
31. Üstündağ, A. & Gençer, Ç. (2021). Designing the clamp system with the emergency braking system in the trains by using PLC and SCADA. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC),* pp. 1436–1441.
32. Wang, X., Li, S., Tang, T. & Yang, L. (2020). Event-triggered predictive control for automatic train regulation and passenger flow in metro rail systems. *IEEE Transactions on Intelligent Transportation Systems,* pp. 1–14.
33. Tabassum, K. (2020). An intelligent metro tracking system for Riyadh smart city. *International Journal of Information Technology, 12*(4), 1103–1109.

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.

Ishu Tomar is a research fellow (Ph.D. research scholar) in the Department of Electronics and Communication Engineering at Delhi Technological University (DTU), Delhi, India. She did her B.Tech. in Electronics and Communication Engineering from Indira Gandhi Delhi Technical University for Women (IGDTUW), Delhi in 2018. Her areas of interest are intelligent transportation systems and real time control systems.

Indu Sreedevi is a Professor in the Department of Electronics and communication Engineering, Delhi Technological University, Delhi, and is currently the Dean of the student welfare department there. She has done Ph.D. in Computer Engineering from University of Delhi and M.Tech., Post B.Tech. from University of Kerala, Kerala. She was awarded the Best Branch Counselor award 2013 IEEE USA. Her research interests are in the field of Computer Vision, Image Processing, and wireless sensor network. She has 3 completed sponsored research projects. She has published around 120 papers in reputed journals, national and international conferences.

Neeta Pandey received her M.E. in Microelectronics from Birla Institute of Technology and Sciences, Pilani and Ph.D. from Guru Gobind Singh Indraprastha University, Delhi. She has served in Central Electronics Engineering Research Institute, Pilani, Indian Institute of Technology, Delhi, Priyadarshini College of Computer Science, Noida and Bharati Vidyapeeth's College of Engineering, Delhi in various capacities. At present, she is a professor in the ECE department, Delhi Technological University. She is a life member of ISTE, and member of IEEE, USA. She has published papers in international, national journals of repute and conferences. Her research interests are in analog and digital VLSI design.

Authors and Affiliations

Ishu Tomar[1] · **Indu Sreedevi**[1] · **Neeta Pandey**[1]

Indu Sreedevi
s.indu@dce.ac.in

Neeta Pandey
neetapadey@dce.ac.in

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT).
India Habitat Centre, Lodhi Road, New Delhi, India. Sep 23-25, 2022

Power Quality Improvement with Seven Level Packed-E Cell for Renewable Applications

[2]Sanjay Upreti, *Member IEEE*, [1]Shivam Kumar Yadav, *Member IEEE*, [1]Bhim Singh, *Fellow IEEE* and [3]Narendra Kumar
[1]Department of Electrical Engineering, Indian Institute of Technology, Delhi, New Delhi-110016, India
[2]Department of Electrical Engineering, BPIT, Rohini, Delhi, India
[3]Department of Electrical Engineering, Delhi Technological University, Delhi-110042, India
(Email: supreti223@gmail.com, shivamyadaviitdelhi@gmail.com, bsingh@ee.iitd.ac.in, narendrakumar@dtu.ac.in)

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT) | 978-1-6654-9365-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBCONPT57482.2022.9938315

***Abstract*—A new seven-level packed E cell (PEC) solar photovoltaic (PV) converter is introduced for grid integration. The three-phase system is modeled with relevant equations for positive and negative half-cycles. The switching states utilize the three PV arrays perfectly to generate seven-level at the converter output. This converter is tested with harmonic standards under different carrier patterns. These carrier patterns provide the different power quality performances. The presented converter is tested with different carrier arrangements like phase disposition(PD), phase opposition disposition (POD), and alternate phase opposition disposition (APOD) with six carriers arranged in level bands. The best performance is investigated for the PEC converter and implemented for the improved power control. A DC link controller for each phase is presented to regulate the voltages. The ratio between the upper U_{1a} and bottom $B_{1a\text{-}2a}$ PV arrays is ternary (1:3) in nature. This system is tested in dynamic solar irradiances to showcase its viability for renewable energy applications.**

Keywords—Packed E Cell Converter, Modulation and Control, Solar Array, Grid-Photovoltaic Integration, Harmonics

I. INTRODUCTION

Multilevel inverters (MLI) are the preferred choices for medium and high power applications of DC-AC power conversion. The MLI are formed from many power semiconductor devices, power sources, auxiliary capacitors. The MLIs offer many advantages over traditional two level inverters; like the rating of the switching devices is low due to peak voltage is attained in multiple steps, the size of harmonic filter is reduced as levels at converter output makes it closure to the sinusoidal wave shape and reduces the size of inverter with increased efficiency.

The MLI development has started with one switch cells fed from decoupled supplies. The evolution of modern multi-level converter begins with the use of clamping diode for creating levels. The neutral clamp diode (NPC), flying capacitor (FC) and cascaded H-Bridge (CHB) are basic MLI topologies. The novel multi-level converter is developed from hybrid combinations of the basic configurations connected in different manner. Packed E-cell (PEC) converter is the result of hybridization of Packed U cell and FC that offers more levels with lower count of switches [1]-[3]. With the easy availability of solar panels at lower cost have encouraged researchers to develop converter with more number of voltage sources so that power capability of converter can be increased.

The harmonic distortion is very important aspect of the inverters. It is lower in MLIs than traditional inverter due to closeness of converter output to the sine wave [4]. The generation of switching logic for staircase voltage output is a very important aspect of a multi-level converter. Various modulation techniques are available for the same. The level-shifted multi-carrier based sine pulse width modulation (MCSPWM) method is one, which is easy to implement and most widely used for MLIs. This modulation is derived from the traditional two-level inverter, and the reference sinusoidal signal is kept within the carrier band. The high-frequency carrier signals keep on crossing the reference signal, and switching signals of the converter are created by comparison of carrier and reference. Different orientations of level-shifted carrier signals create the following categories of MCSPWM,

i) Phase Disposition (PD)

ii) Phase Opposition Disposition (POD)

iii) Alternate Phase Opposition Disposition (APOD)

These methods offer different harmonic spectrum with full utilization of the DC-link voltage. The triangular carrier wave is the normal choice for MCSPWM implementation However, any other shape of carrier wave is used for the modulation. Prominent among them are as follows,

(a) Triangular

(b) Trapezoidal

(c) Parabolic

(d) Modified triangular

Inverter plays a vital role as an interface between the distributed generation and grid or load. In grid integrated system, an inverter is operated in current control mode and current injected to the grid should follow the reference. The injected current from the inverter to the grid is distorted which should be controlled for its harmonic distortion to maintain the power quality standards. The sources of current distortion are mainly power inverter non-linearity controlled by the suitable control strategy. The harmonic filters are used at the converter output to mitigate the high-frequency ripple due to switching frequency [5]. Several current control strategies are employed in grid integrated inverters: slide mode control [6], model predictive control [7], and deadbeat boundary control [8] are nonlinear. The linear current control methods: proportionate-resonant (PR), pseudo-derivative

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:09:32 UTC from IEEE Xplore. Restrictions apply.

feedback, state feedback [9]-[10] are frequently applied for controlling current injection. The nonlinear current control directly switches the inverter and linear control strategies use the pulse-width-modulation (PWM). PWM method ensures constant frequency switching at low switching loss.

This work presents a seven-level solar MLC formed by PEC topology termed as PEC7, that uses seven switches, including one bidirectional and is powered by three solar photovoltaics (PV) array. The PEC is the advance modification of PUC topology where E-cell replaces one U-cell. The voltage level of two bottom PV arrays is one third of the main solar PV array. It keeps ternary ratio among the PV arrays. This system is tied to the grid for supplying active power. The PI controllers employ the DC voltage control while PR controllers control the injected AC current. The PEC7 is analyzed for various combinations of multicarrier signal shapes and their orientation. The presented solar PEC has the following advantages,

- The use of three solar PV arrays enhances the power capacity of new seven level PEC converter.
- It offers smooth and simple control and transfer of power to the grid by AC side PR controller.
- The best modulation method for MLI improves the harmonic performance.

II. SOLAR PEC CONVERTER AND CONTROL

A seven level PEC based topology ideal for three-phase system is depicted in Fig. 1. This is a modification of PUC, where lower U-cell is converted to E-cell by inserting a bi-directional switch G7a-G8a in phase 'a'. There are six another switches S1a-S6a which are similar as in PUC. Three solar PV arrays PVUa, PVB1a and PVB2a feed DC powers to the phase 'a' of the inverter. The PVUa is the main DC source of PEC7 and its voltage magnitude VUa1 is 369V, three times than the VB1a and VB2a voltage magnitudes of 123V of PVB1a and PVB2a . All the solar PV arrays are provided a maximum power point tracker (MPPT) controller to ensure the conversion of maximum solar power to electrical power. The converter output is connected to the grid via R-L filter, which provides the harmonics mitigation by changing the current wave shape to the sinusoidal. The staircase voltage levels of positive and negative cycles in phase 'a' are generated as,

The positive Half cycle is expressed as,

$V_{an} = V_{Ua}$ (G_{1a}, G_{5a} and G_{6a} are ON) (1)

$V_{an} = V_{Ua} - V_{Ba2}$ (G_{1a}, G_{5a}, G_{7a} and G_{8a} are ON) (2)

$V_{an} = V_{Ua} - V_{Ba1} - V_{Ba2}$ (G_{1a}, G_{5a}, and G_{3a} are ON) (3)

$V_{an} = 0$ (G_{1a}, G_{2a}, and G_{3a} are ON) (4)

Similarly negative half cycle equations are given as,

$V_{an} = -V_{Ua1}$ (G_{2a}, G_{3a} and G_{4a} are ON) (5)

$V_{an} = -V_{Ua1} + V_{Ba1}$ (G_{2a}, G_{4a}, G_{7a} and G_{8a} are ON) (6)

$V_{an} = -V_{Ua1} + V_{Ba1} + V_{Ba2}$ (G_{2a}, G_{4a}, and G_{6a} are ON) (7)

$V_{an} = 0$ (G_{4a}, G_{5a}, and G_{6a} are ON) (8)

Detailed switching states of inverter switches are tabulated in Table I. The peak step of converter is 369V and it is achieved by connecting the upper solar array PVUa1 as stated in (1), 2nd peak voltage level is achieved by the connecting PVBa2 in series with PVUa1 so that net converter voltage is 246V (369V-123V) which is expressed in (2) and first voltage level is achieved by connecting both bottom solar arrays PVB1a and PVB2a in series and opposition with PVUa so that net converter voltage becomes 123V (369V-123V-123V). The zero level is achieved by bypassing all the arrays shown in (4). Similarly negative half cycle of stepped output of converter is generated in (4)-(8). Fig. 2 depicts all the modes of the positive half cycle of converter output. Fig 2(a) PVUa is connected through the switches S1a, S5a and S6a and its voltage appears at converter output. Fig 2(b) shows S6a is turned off and S7a S8a are turned ON and it connects PVB1a which opposes the PVUa.

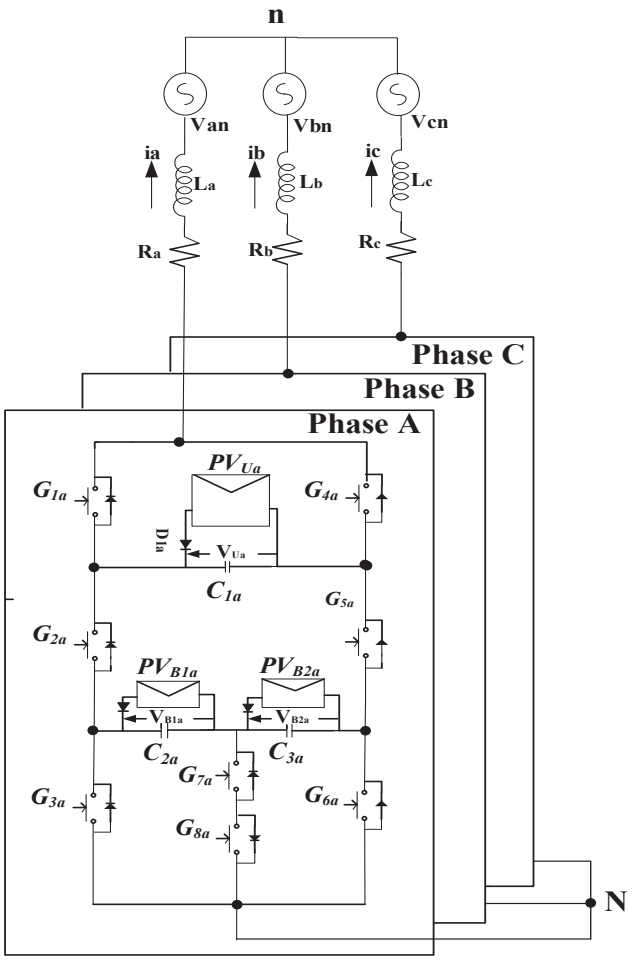


Fig. 1 Seven level PEC grid tied solar system

TABLE I. Switching States For PV-PEC Converter

G_{1a}	G_{2a}	G_{3a}	G_{4a}	G_{5a}	G_{6a}	G_{7a}	G_{8a}	V_{out}(V)
1	0	0	0	1	1	0	0	$+V_{Ua}$
1	0	0	0	1	0	1	1	$+V_{Ua}- V_{Ba2}$
1	0	1	0	1	0	0	0	$+V_{Ua}- V_{Ba1}-V_{Ba2}$
0	0	0	1	1	1	0	0	0
1	1	1	0	0	0	0	0	0
0	1	1	1	0	0	0	0	$-V_{Ua}$
0	1	0	1	0	0	1	1	$-V_{Ua}+ V_{Ba1}$
0	1	0	1	0	1	0	0	$-V_{Ua}+ V_{Ba1}+V_{Ba2}$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:09:32 UTC from IEEE Xplore. Restrictions apply.

So the net converter output voltage 246V. The connection of PVUa achieves the first step of 123 V, PV_{B1a} and PV_{B2a} in series by the switches S_{1a}, S_{5a} and S_{3a}. It connects the bottom solar arrays opposite to the upper solar panel and net voltage is the difference of V_{Ua} and (V_{Ba1} + V_{ba2}), which is 123V. Zero level is achieved by the bypassing of all solar arrays through switches S_{1a}, S_{2a} and S_{3a}. Similarly, the negative half cycle is generated by connecting the solar PV arrays in the opposite direction. All the modes of the voltage level are shown in Table I.

The control strategy's major constituents are the DC-link voltage control, grid current control, and multi-carrier PWM for the switching pulses for the inverter. As shown in the Fig. 2, DC link voltage is compared with the reference voltage from MPPT controller. The PI controllers are employed for the minimization of error signal. The reference DC current is obtained by the summation of all the reference currents from all the solar panels. The orthogonal components of DC reference current i_{dref} and i_{qref} are represented as,

$$i_{dref}(v) = i_{dref}(v-1) + K_{pd}\{e_{dc}(v) - e_{dc}(v-1)\} - + K_{id}e_{dc}(v) \quad (9)$$

$$i_{qref}(v) = i_{qref\,(v-1)} + K_{pq}\{e_{dc}(v) - e_{dc}(v-1)\} + K_{iq}e_{dc}(v) \quad (10)$$

Where $K_{pd\text{-}q}$ and $K_{id\text{-}q}$ are control parameters of PI controllers and e_{dc} is error signal.

The direct-quadrature (D-Q) control of AC side is represented in Fig. 2(b). The grid voltage v_{abc} and current i_{abc} are transformed into direct and quadrature components v_d, v_q, i_d and i_q. The i_d and i_q are compared to i_{dref} and zero respectively which gives the modulating reference signals m_{abc} and these are compared with low frequency multi-carriers to generate the switching sequence of the PEC7 inverter. The mathematical model of the current control is represented as,

$$u_d = \left(k_p + \frac{k_i}{s}\right)\left(i_{dref} - i_d\right) - \omega L i_q + v_d \quad (11)$$

$$u_q = \left(k_p + \frac{k_i}{s}\right)\left(i_{qref} - i_q\right) + \omega L i_d + v_q \quad (12)$$

Here u_d and uq are the direct and quadrature axis components of the converter output voltage. The PWM switching of inverter is generated by the comparison of reference signal with the high frequency level shifted multi-carrier signals. Six triangular carrier signals of same amplitude and frequency are used to create the seven levels. There are three various type of level shifted modulation techniques: phase disposition (PD), phase opposition disposition (POD) and alternate phase opposition disposition (APOD), these modulations are shown in Fig. 3.

In the PD Multicarrier modulation as shown in Fig. 2 all six multi-carrier signals are vertically shifted top most carrier signal varies between 0.67 to 1.0, second signal between 0.33-0.67 and third signal oscillates in between 0.0- 0.33 for positive half cycle. Frequency of all the signals is kept at 1 kHz. Similarly, other three carriers are vertically shifted equally between 0 to -1. In the APOD modulation, alternate carrier signals are 180^0 phase apart which depicted in the middle part of Fig. 3. The number of carriers and their frequency remains same as that of PD modulation. The POD modulation as shown in the bottom part of Fig. 3, the three carrier signals of positive side are in phase and equally shifted vertically between 0-1 however the three carrier signals in the negative side are in phase with each other but 180^0 phase shifted with the signals of positive half.

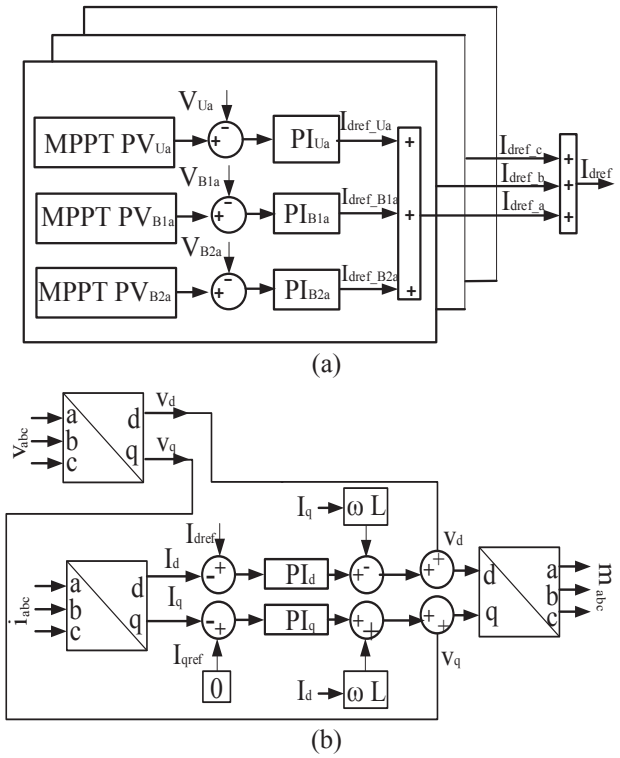


Fig. 2 (a) DC link voltage control (b) AC side current control

The comparators of all multicarrier with sinusoidal reference give response when the value of carrier is less than that of sinusoidal reference.

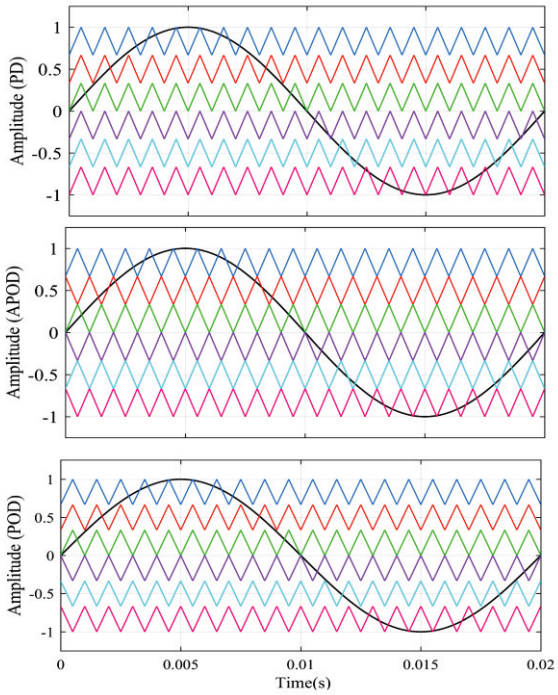


Fig. 3 PD, APOD and POD Multi-carrier PWM

III. Results and Discussion

Figs. 4-8 depict the grid integrated simulated performance solar system. The solar system is operated at solar irradiation of 1000 W/m^2 with the objective active

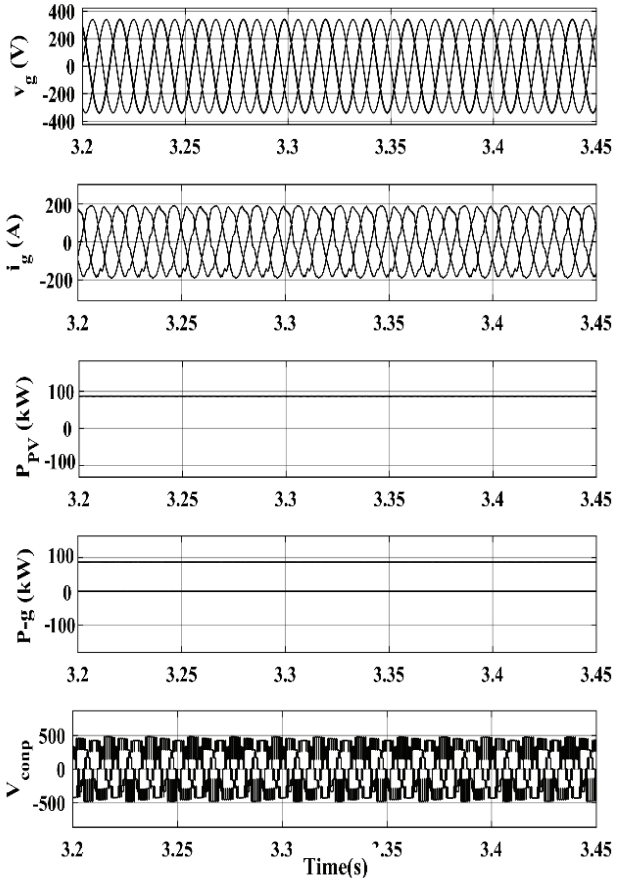
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:09:32 UTC from IEEE Xplore. Restrictions apply.

power injection to the grid and response at the steady state is shown in Fig 4. The total extracted solar power P_{PV} is 82.9 kW and grid power P_g is 81.3 kW at unity power factor and efficiency of 98%. The grid voltage V_g and current i_g are shown in the diagram.

The dynamic response of the PEC7 when solar irradiation falls to 50% from 1000 W/m^2 the solar power P_{PV} falls to half 41.5 kW and grid power also follows the solar power and reduces to half 40.6 kW at unity power factor. The converter voltage V_{conp} is maintained and grid current reduces to 50%. The system dynamics during the irradiation fall are shown in Fig. 5.

The rise of solar irradiation to normal at 1000 W/m^2 and the system's response is shown in the Fig. 6. The solar power P_{PV} resumes to 82.9 kW and grid power follows to settle at 81.3 kW, reactive power is zero. The converter voltage maintains its shape while the grid current i_g rises by 100%. Fig. 7 shows performance in temperature variation to 50 ^{o}C from 25^{o}C. It reduces power injection from solar PV source to grid.

The harmonic analysis of the PEC7 converter phase and line voltage is done for three modulations methods: PD, POD and APOD and results are tabulated in Table II-III. The PD modulation technique is selected for the analysis as it has given the best harmonic response. Fig. 8 displays the harmonic response of the PEC7 with PD multi-carrier PWM, voltage harmonic response is shown in Fig. 8(a). The fundamental component of the phase voltage is 369 V and THD is 16.85%. Fig. 8(b) shows the current harmonics, which stand at the 2.47% that meets the grid IEEE standard.



Fig, 4 Steady-state response of grid integrated PEC7

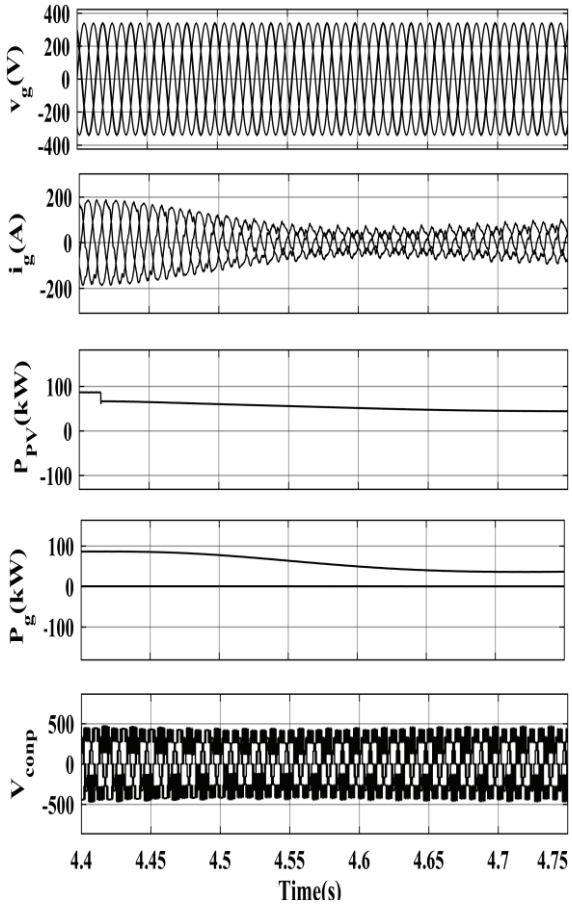


Fig. 5 Dynamic response of grid integrated PEC7 during the fall of irradiation

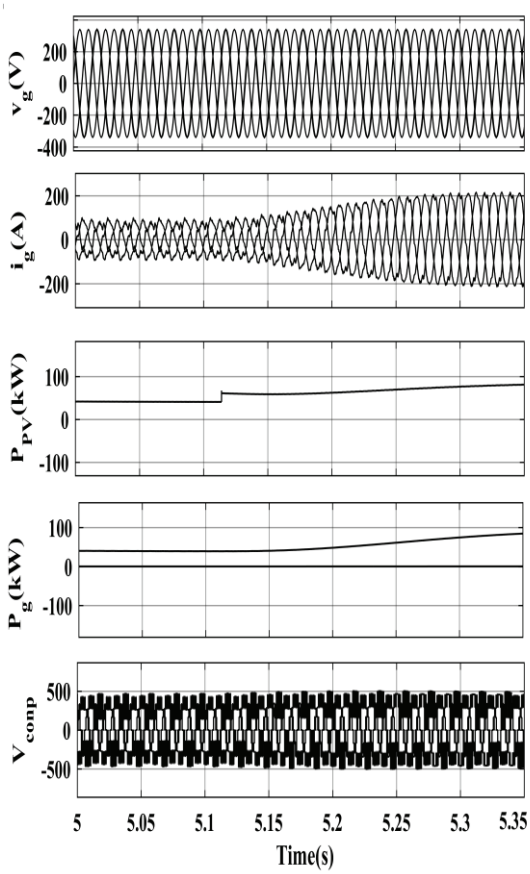


Fig.6 PEC7 during rise of irradiation

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:09:32 UTC from IEEE Xplore. Restrictions apply.

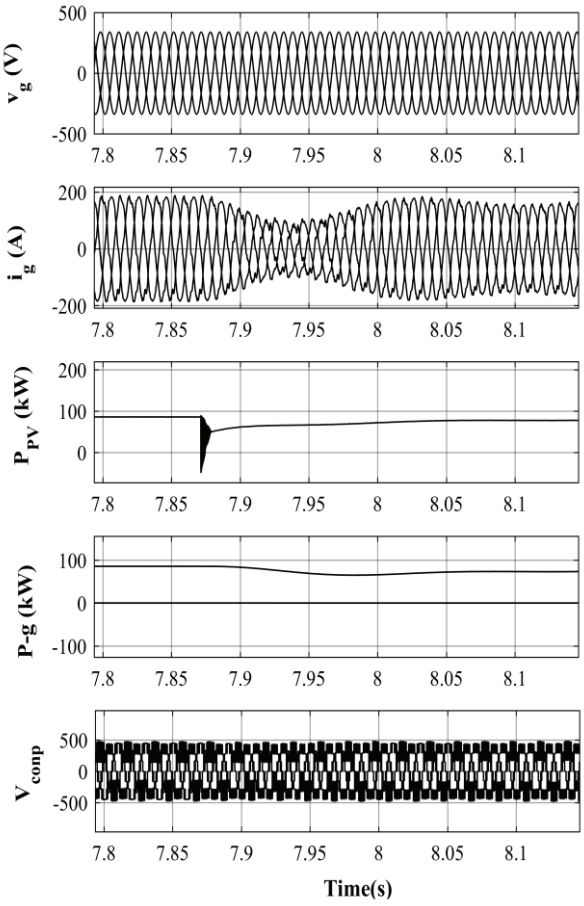


Fig.7 PEC7 during rise of temperature

TABLE II. PEC Converter Phase Voltage Harmonics

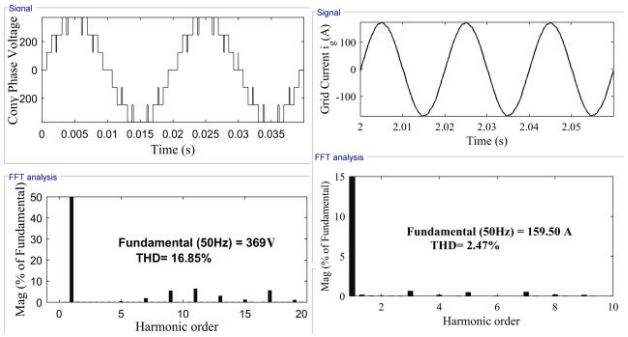
Irradiance (W/m^2)	PD (%)	POD (%)	APOD (%)
800	19.22	20.50	20.00
850	18.67	19.20	18.9
900	17.52	18.25	17.50
950	16.31	17.01	16.89
1000	15.80	16.95	16.50

TABLE III. PEC Converter Line Voltage Harmonics

Irradiance (W/m^2)	PD (%)	POD (%)	APOD (%)
800	12.32	12.80	12.65
850	12.20	12.90	12.5
900	11.91	12.20	12.01
950	10.54	10.90	10.78
1000	9.51	9.89	9.65

IV. CONCLUSIONS

This PEC7 is simulated in MATLAB/SIMULINK in grid integration mode for medium power application. It has used the ternary configuration to generate seven levels. It is evaluated for the PD, POD and APOD multi-carrier techniques. It is found that the PD modulation is the best for harmonic reduction and used to operate the PEC system. This system's performance is excellent in constant and varying solar irradiance/temperature. This system is evaluated for temperature variation and its performance found proper. This work is done for seven level with the future scope of increasing levels by addition of more power-cell and solar PV arrays.



(a) (b)

Fig, 8 Harmonic response of PEC7 (a) voltage THD (b) Current THD

ACKNOWLEDGMENTS

The authors acknowledge Department of Science and Technology (DST), to fund the research in the EED, at IIT, Delhi with grant: RP31095 (FIST scheme), by OPAL-RT(under: RP0253) and by SERB National Science Chair Fellowship.

APPENDIX

V_{U1a}=369 V, $V_{B1a\text{-}B2a}$ = 123V, P_{pv}=82.9 kW, R=0.02 ohm L= 2 mH, f_{sw}= 1kHz, Grid Voltage, v_s = 415 V. kpd=1, kid=0.02, kpq=1, kid=0.02, k_{pDC}=0.9, k_{iDC}=0.005

REFERENCES

[1] L. K. Haw, N. A. Jefry and W. K. Ing, "The New Hybrid Multilevel Inverter with Reduced Number of Switches," *2021 IEEE 11th International Conference on System Engineering and Technology (ICSET)*, 2021, pp. 337-341

[2] T. Qanbari and B. Tousi, "Single-Source Three-Phase Multilevel Inverter Assembled by Three-Phase Two-Level Inverter and Two Single-Phase Cascaded H-Bridge Inverters," *IEEE Trans. on Power Electron.*, vol. 36, no. 5, pp. 5204-5212, May 2021.

[3] M. S. Sadabadi, M. Sharifzadeh, M. Mehrasa, H. Karimi and K. Al-Haddad, "Decoupled dq Current Control of Grid-tied Packed E-Cell Inverters in Vehicle-to-Grid Technologies," *IEEE Trans. on Ind. Electron.(Early Access).*

[4] S. K. Yadav, N. Mishra and B. Singh, "Multilevel Converter with Nearest Level Control for Integrating Solar Photovoltaic System," *IEEE Trans. on Ind. App., 2022. (Early Access).*

[5] S. Upreti, S. K. Yadav, B. Singh and N. Kumar, "New Multicarrier Modulation Scheme for Harmonics Mitigation of T-Type Solar Multilevel Inverter", *J. Inst. Eng. India Ser. B*, vol. 103, pp. 903–911 2022.

[6] Y. He, H. S. -H. Chung, C. N. -M. Ho and W. Wu, "Use of Boundary Control With Second-Order Switching Surface to Reduce the System Order for Deadbeat Controller in Grid-Connected Inverter," in *IEEE Transactions on Power Electronics*, vol. 31, no. 3, pp. 2638-2653, March 2016.

[7] A. Bemporad, "Model Predictive Control Design: New Trends and Tools," *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006.

[8] Ramaiah, L. N and M. K. Mishra, "Loss Modulated Deadbeat Control for Grid Connected Inverter System," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2022.

[9] R. M. Essefi, M. Souissi and H. H. Abdallah, "Linear current control technique for grid connected photovoltaic system," *14th International Conference on Sciences and Techniques of Automatic Control & Computer Engineering - STA'2013*, 2013, pp. 432-437.

[10] M. Sabarimuthu, N. Senthilnathan, N. Priyadharshini, M. A. Kumar, N. Telagam and S. K. Sree, "Comparison of Current Control Methods for a Three Phase Shunt Active Filter," *2021 7th International Conference on Electrical Energy Systems (ICEES)*, 2021, pp. 32-37.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:09:32 UTC from IEEE Xplore. Restrictions apply.

International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-022-01697-2

ORIGINAL ARTICLE

Probabilistic multiplicative unbalanced linguistic term set and its application in matrix games

Tanya Malhotra[1] · Anjana Gupta[1]

Received: 21 September 2021 / Accepted: 13 October 2022
© Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Probabilistic linguistic term sets (PLTSs) are suitable for enunciating evaluators' complex linguistic perceptions more accurately within the intricate qualitative setting. Usually, the PLTS is based on a balanced concept and does not serve as a good information representation for an unbalanced qualitative concept. Therefore, to reflect experts' distinct preferences and uncertainties, this paper proposes a new PLTS called the probabilistic multiplicative unbalanced linguistic term set (PM-ULTS), where both probabilities and non-uniform spacing of the linguistic labels are considered simultaneously. Afterwards, we put forward specific operational laws for the newly constructed PLTS to preserve the resultant linguistic labels and corresponding probability information. Some elementary aggregation operators beneficial in aggregating probabilistic linguistic information in decision-making problems are also constructed, and their excellent properties are addressed. Furthermore, based on the proposed concept, this study initiates the design of a unified two-person linguistic matrix game model with the new PLTS as a parameter. It addresses the imprecise information by the information measure function. Such a two-player probabilistic unbalanced linguistic matrix game is considered a convenient technique for multiple decision scenarios. Additionally, the proposed game model involves a re-translation process to convert the output back into the original probabilistic unbalanced linguistic domain without information loss, thereby escalating the interpretability of the game model compared with other existing uncertain matrix game methodologies. Finally, we discuss the significance of the proposed methodology and concept to question its validity and usefulness by presenting suitable examples.

Keywords Probabilistic multiplicative unbalanced linguistic term set · Aggregation operators · Operational laws · Information measure function · Two-person matrix game

1 Introduction

The formulation of traditional game theory was done 80 years ago, which was ideally introduced by Von Neumann and Morgenstern [57]. Since then, the advancement in game theory has gone beyond limitations and has been conclusively applied in diverse fields like economics, finance, sociology, and so forth. It is prefigured that within the traditional setup of games, the payoff values are well-known with certainty. Nevertheless, with the rapid development of society, the assumption of certainty becomes increasingly complex and challenging. Therefore, multifold extensions of conventional game theory have been conducted by scholars to capture the uncertainty in the context of payoffs and goals. To quantify the inherent uncertainty of strategic expression within the game problems, the payoffs are represented in the form of interval numbers [36, 37, 46], fuzzy numbers [4, 5, 7, 10, 47], and stochastic payoffs [16, 50, 51].

Although the foundation of fuzzy games can overcome the difficulties envisioned in conventional game-theoretic models, it is still incapable of addressing qualitative information persisting within real-world problems. However, in most imprecise games, the best expression in strategic communication naturally takes a linguistic form because of the underlying problem's complexity and the imperfectness of human preferences. Linguistic labels [71]

✉ Anjana Gupta
anjanagupta@dce.ac.in

[1] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India

Published online: 24 November 2022

are practical tools because incorporating linguistic information strengthens the adaptability and reliability of traditional decision models [45]. Recent years have seen extensive study and utility of linguistic variables, which take the values of words or phrases from native or artificial languages. To annex the dimension of linguistics in the field of matrix games, Arfi [1, 2] anchored game theory in terms of fuzzy linguistic logic, which is a more relaxed and straightforward approach to expressing information and payoff opinions reasonably without restraining precision. After that, many investigations have been directed toward building a more robust endowment of linguistic game-theoretic methodologies. Papers in this direction of research are [44, 52, 53].

However, to cope with the uncertainty and complexity of decision information, it is difficult to completely express the opinions of game players via conventional single and unique linguistic terms. To understand, let us consider a basic illustration. Assume that two life insurance companies are competing strategically to sell their policies to the customers for two consecutive years, i.e., Year 1 and Year 2. Here, a game player may think that "the number of policies booked and issued for the companies is "average" for one year and "absolutely high" for the second year." The hesitant fuzzy linguistic term set (HFLTS) representing players' hesitant qualitative preference information in this particular case can be identified as $LT = \{``Average,'' ``High''\}$. In this instance, all possible linguistic labels have equal importance, i.e., the weights of "average" and "high" are 50% and 50%, respectively, thereby losing preference information. Nevertheless, the number of policies booked and issued for the companies in year two is "absolutely high", which means it should be considered more important than "average". To express such kind of preference information realistically, Pang et al. [48] proposed the notion of a probabilistic linguistic term set (PLTS). With the PLTS, the players or decision-makers (DMs) assess using several linguistic terms instead of a single term more appropriately by allocating a weight through probability for each linguistic term. For the above instance, a PLTS can be fixed as $LT = \{``Average''(0.2), ``High''(0.8)\}$, which conveys that the precedence of "average" is 0.2 while the precedence of "high" is 0.8. It testifies that the "high" is more significant than the "average". In this manner, the affluence and flexibility of linguistic expression information have been accelerated to a huge extent compared to other linguistic models. From this realistic viewpoint, it is meaningful and consequential to develop linguistic game models with PLTS as parameter to express the collected payoff information accurately in the complex decision environments.

Since PLTS provides players with a more extensible and comprehensive way to express complex cognitive-linguistic information, thus Mi et al. [42] put forward a two-person zero-sum matrix game model taking PLTS as input and imparting imprecise information by using the triangular membership function. However, it is envisioned that the existing PLTS is primarily based on the predefined symmetrically distributed linguistic label set. It is well known that in a real-world decision-making environment, experts may prefer to represent more linguistic information on the right side rather than on the left side of the mid-term so that the linguistic variables have an irregular and non-symmetrical distribution. So, to tackle this issue, it is crucial to develop an imprecise game model that can efficiently represent the payoff information involving an unbalanced concept that has not been investigated so far.

Although in literature, Han et al. [19] introduced the notion of probabilistic unbalanced linguistic term sets (PULTSs), some limitations persist with the model. In that paper, the authors developed a novel computational model based on Archimedean copula and co-copula to manage unbalanced linguistic information. However, the representation of the information is still provided by an additive scale set as a result the literal meaning of unbalanced linguistic term set (ULTS) is missing. Another core problem associated with the PULTSs is the high complexity in different application scenarios. Based on this analysis, we constructed a new PLTS called probabilistic multiplicative unbalanced linguistic term set (PM-ULTS), which reflects not just the probabilistic information of linguistic variables but also emphasizes the irregular and non-symmetric dispersion of each linguistic label. This is the pioneer research gap filled by the research work conducted to model probabilistic unbalanced information. Further, we introduced some new operational laws and elementary aggregation operators to aggregate probabilistic multiplicative unbalanced linguistic terms. The operational laws presented in the paper can easily sustain the operation results without exceeding the boundaries of multiplicative ULTSs and can keep probabilistic information complete after the operations.

The significant contributions offered by this paper can be outlined as follows:

1. To fill the research gap, the PM-ULTS is proposed, which can take into account the hesitancy of DMs' preferences and reflect their opinions more accurately by effectively distinguishing the spacing difference for each linguistic label.
2. On comparing PM-ULTS with PULTS proposed by Han et al.[19], it is viewed that the PM-ULTS can able to handle both the type of unbalanced linguistic

information, i.e., Type I ULTS [22] and Type II ULTS [61].

3. The novel operational laws developed for PM-ULTS are closed, i.e., the outcome obtained after applying the operation does not exceed the bounds of LTSs, thereby overcoming the unreasonableness in logic proficiency. In addition, these laws can efficiently deal with the persisting decision information by expressing it using the PM-ULTSs and can avoid losing the given priority information validly.
4. To address practical issues, the above-mentioned theoretical concept is incorporated into an imprecise game model to address probabilistic multiplicative unbalanced linguistic circumstances, enhancing the feasibility and flexibility of the game theory scope. This is the second aspect of this study. In our work, we formulate a probabilistic multiplicative unbalanced linguistic linear programming (PMULLP) model to solve a constant sum unbalanced linguistic matrix game problem. Further, the PMULLP model is transformed into a classical linear programming problem (LPP) by defining the information measure of the given probabilistic multiplicative unbalanced linguistic terms. The defined information measure can easily avoid loss of information since it is based on the bijection operator.
5. The computational unbalanced linguistic game model proposed in the paper is applied to solve the multi-attribute decision-making (MADM) problem modelled as a constant sum unbalanced linguistic game problem. Furthermore, comparative analyses of the proposed method with the existing method are made to highlight the advantages.

Finally, the study is structured as follows: Section 2 briefly outlines works related to the existing probabilistic linguistic term set and ULTS. Section 3 presents some preliminary definitions and concepts about the new 2-tuple unbalanced linguistic computational model, probabilistic linguistic term sets, and conventional two-player zero-sum matrix games. Section 4 proposes a novel concept of probabilistic multiplicative unbalanced linguistic term sets. The new ranking method, operational laws with their properties, and aggregation operators are also given. Next, based on the proposed concept, the methodology to solve uncertain linguistic matrix games with probabilistic multiplicative unbalanced linguistic information is presented in Sect. 5. Furthermore, we illustrate an application example about companies' stock selection problem to present the method's effectiveness in a physical realm, followed by the comparison analyses with the existing methods in Sect. 6. Finally, some concluding remarks are given in Sect. 7.

2 Related works

This section will give a brief literature overview of existing PLTS and ULTS.

To date, considerable research has been devoted to reflecting the uncertainty and preference degree of linguistic terms in the form of PLTS in different facets like medical level assessment, supplier selection, venture capital, and so on. In the follow-up research, many scholars have been dedicated to examining mathematical operations and measures of PLTSs [27, 65]. Pang et al. [48] introduced the study of mathematical operations and aggregation operators for PLTSs. However, the final results obtained by the authors exceed the bounds and are not PLTS. To overcome the drawback so that final results are not exceeding the bounds, Gou and Xu [18] proposed new probabilistic linguistic operators. Afterwards, Zhang [68] presented the new probabilistic linguistic weighted arithmetic average (WAA) operator for PLTS. Liao et al. [26] proposed a new ELECTRE III methodology to address MADM problems primarily based on novel operations of PLTSs, including the addition, subtraction, and division of PLTSs considering the disparity degrees of linguistic terms. Beyond that, Bai et al. [6] proposed the Interval-valued probabilistic linguistic term set (IVPLTS) by extending the corresponding probability of linguistic terms to interval values. Zhai et al. [70] extended the PLTS to introduce the probabilistic linguistic vector-term sets. Feng et al. [17] developed a probabilistic linguistic qualitative multiple criteria method along with a possibility degree comparison technique for handling any group decision-making problems, such that HFLTSs present the alternative evaluation information, and weights information are partially known. Liu et al. [29] suggested entropy measures to address uncertainties within PLTSs. Liao et al. [25] introduce a linear programming method to solve the MADM problems where the experts' preferences over alternatives concerning the given criteria are presented in the form of PLTSs. Mao et al. [41] developed a novel method for MADM with PLTSs as input parameters. In that paper, the authors initially defined the possibility degree and range value of PLTSs. Afterwards, a possibility degree algorithm to rank PLTSs is circumscribed. Also, a Euclidean distance measure between PLTSs is represented and outstretched to probabilistic linguistic matrices. Finally, based on Archimedean t-norm and s-norm, some newly defined operational laws for PLTSs are given, and their properties are discussed.

For the moment, the application prospect of PLTS is comprehensive and valuable as it plays a significant role in addressing real emergency cases. Therefore, in recent years, the development in the study of PLTS has been rapid

and has attracted several researchers around the globe. However, the literature on PLTS is so vast and extensive that it is difficult to cite every contribution, thus for a more comprehensive insight into the idea of PLTS, one can also refer to the papers [8, 29–33, 38, 54, 56, 58–60, 64, 69].

The literature review above shows that the existing PLTS and its extension are based on the balanced concept. Henceforth it cannot support an unbalanced and irregular difference in spacing for each linguistic term. In literature, to represent information distributed irregularly, Herrera et al. [22] proposed a methodology to address unbalanced linguistic information by using non-uniformly distributed linguistic labels. In addition, many functional studies have been established ideally concerning both theoretical and practical aspects of unbalanced linguistic representation models. For instance, Numerical scale model [11, 12], Proportional linguistic 2-tuple model [61], linguistic 3-tuple model [49], Personalized individual semantic model [34, 35] and so forth. In sequence, one can also refer to the papers [9, 13–15, 21, 62] for a deeper understanding of the literature based on the context of ULTSs. The methodology proposed in these papers is capable of addressing unbalanced linguistic information. Furthermore, to aggregate unbalanced linguistic labels in the process of computing with words, some mathematical operators are introduced suchlike, the linguistic proportional 2-tuple power average operator [24], ordered weighted average (OWA) operator [23], type-I OWA operator [39], weighted unbalanced linguistic aggregation operator [40] and Heronian mean operator [20]. It is noted that managing unbalanced linguistic labels usually involves three steps. At first, it is required to transform unbalanced linguistic labels into the form of classic linguistic 2-tuples, followed by an aggregation operation to aggregate the altered linguistic 2-tuples. Finally, the computational results are transformed into unbalanced linguistic labels to avoid information loss. The first step is anticipated to play a prominent role in the processes mentioned above to comprehend a suitable linguistic value presenting the unbalanced linguistic information within the computational process.

Meanwhile, it is viewed that the representation of the unbalanced linguistic information in the existing unbalanced linguistic models is given by using an additive symmetric linguistic term set. As a result, the literal meaning of the ULTS is missing. Also, the computational cost of performing the three steps is usually very high and time-consuming in large-scale complicated decision problems. Therefore, to overcome this issue and contribute to the ongoing research Malhotra and Gupta [43] introduced a novel 2-tuple unbalanced linguistic approach where the linguistic information is assessed in a multiplicative unbalanced linguistic scale set such that the terms are unevenly distributed. The proposed method in the paper is less complicated and computationally inexpensive. Hence it is practically helpful in real-world problems.

The ULTS can reflect the non-uniformity of distinct linguistic labels, while the PLTS can consist of different linguistic terms associated with different weights by utilizing the probabilities. The PLTS promotes the affluence and flexibility of representing convoluted cognitive-linguistic information. In addition, to the best of our knowledge, the synergy of PLTS with multiplicative ULTS cannot be found in the current literature. Therefore, to bridge the research gap, we introduce the notion of PM-ULTS that focuses on explicating the non-uniform and asymmetrical distribution of linguistic assessment values along with the probability associated with each unbalanced linguistic label. Conclusively, it ameliorates the practicality of linguistic assessments. Lastly, considering the influence of different circumstances on linguistic variables' meaning and expression, we summarise the differences and similarities of the linguistic expressions in Table 1.

3 Preliminaries

The present section briefly reviews some elementary definitions and concepts related to the new 2-tuple unbalanced linguistic approach, probabilistic linguistic term set and conventional zero-sum matrix games necessary to understand the proposal presented in the paper.

3.1 2-tuple unbalanced linguistic computational model

Let $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, m\}$ where, $a > 1$, $n, m \in \mathbb{Z}^+$ be a predefined multiplicative ULTS with cardinality $n + m + 1$ such that n and m may or may not be equal. We call the term t_{a^i} as a plausible value for an unbalanced linguistic variable such that the following properties are satisfied:

(i) The set LS is ordered: $t_{a^i} \geq t_{a^j}$ for $i \geq j$; (ii) A maximization operator: $\max(t_{a^i}, t_{a^j}) = t_{a^i}$ for $i \geq j$; (iii) A minimization operator: $\min(t_{a^i}, t_{a^j}) = t_{a^i}$ for $i \leq j$. The new unbalanced 2-tuple linguistic model, proposed by Malhotra and Gupta [43], manifests unbalanced linguistic information utilizing an unbalanced linguistic 2-tuple (t_λ, γ), where $t_\lambda \in LS$ and γ represents the value of "*symbolic translation*". The translation function that supports converting a numeric value β into an equivalent 2-tuple unbalanced linguistic value is given in the following definition.

Definition 1 Let $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, m\}$, be a multiplicative ULTS having $n + m + 1$ as cardinality. Let $\beta \in [a^{-n}, a^m]$ be a generic numeric value. To obtain the 2-

Table 1 A summary on distinct linguistic expression term sets and their characteristic

Representation model	Paper	Characteristic	Probabilistic information
HFLTS	[63]	A collection of several consecutive linguistic terms	N/A
Proportional linguistic term	[61]	A proportion of two linguistic terms	Known completely
PLTS	[48]	A probabilistic distribution of several symmetrically distributed linguistic terms	Known partially
PULTS	[19]	A probabilistic distribution of several irregularly distributed linguistic terms	Available partially and unbalanced linguistic expression representation is based on the symmetrically distributed linguistic label set. Hence, the literal meaning of ULTS is missing
PM-ULTS	Present paper	A probabilistic distribution of several irregularly distributed linguistic terms	Known incompletely and unbalanced linguistic expression representation is based on the multiplicative ULTS. Hence, the literal meaning of ULTS is preserved

tuple unbalanced linguistic value equivalent to a numeric value is shown below:

$$\Omega : [a^{-n}, a^{m}] \longrightarrow LS \times [-0.5, 0.5]$$
$$\Omega(\beta) = (t_{\lambda}, \gamma)$$

in which

$$\begin{cases} \lambda = a^{i_L} \text{ or } \lambda = a^{i_R} & \text{for } i_L,\ i_R \in \{-n, \ldots, 0, \ldots, m\}; \\ \gamma = \dfrac{\beta - \lambda}{d_L + |d_R|} & \text{for } \gamma \in [-0.5, 0.5]. \end{cases}$$

The decision rule for the computation of values λ and γ defined above in the definition can be found in the paper [43].

Definition 2 Let $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, m\}$ be an ULTS and let (t_{λ}, γ) be a 2-tuple corresponding to numerical value $\beta \in [a^{-n}, a^{m}] \subset \mathbb{R}$. An inverse translation function Ω^{-1} returning equivalent numerical information from a corresponding 2-tuple linguistic information is defined as:

$$\Omega^{-1} : LS \times [-0.5, 0.5] \longrightarrow [a^{-n}, a^{m}]$$
$$\Omega^{-1}(t_{\lambda}, \gamma) = \gamma \times (d_L + |d_R|) + \lambda = \beta.$$

Definition 3 [43] Let $LS = \{t_{a^{(-n)}}, \ldots, t_{a^0}, t_{a^1}, \ldots, t_{a^m}\}$ be an ULTS with cardinality $n + m + 1$. The negation operator for a 2-tuple is given as:

$$\text{neg}(t_{a^i}, \gamma) = (\text{neg}(t_{a^i}), -\gamma)$$

The details about the existence of a $\text{neg}(t_{a^i})$ for an unbalanced linguistic term in all the plausible cases are provided in the paper [43].

Remark 1 Based on the property of linearity and monotonicity of the Ω operator, it is easily observed that for the case $n = m$, the following holds:

(i) $\text{neg}(\text{neg}(\tilde{v}_1)) = \tilde{v}_1$;
(ii) $\min\{\text{neg}(\tilde{v}_1), \text{neg}(\tilde{v}_2)\} = \text{neg}(\max\{\tilde{v}_1, \tilde{v}_2\})$;
(iii) $\max\{\text{neg}(\tilde{v}_1), \text{neg}(\tilde{v}_2)\} = \text{neg}(\min\{\tilde{v}_1, \tilde{v}_2\})$.

where $\tilde{v}_1 = (t_{\lambda_1}, \gamma_1)$, $\tilde{v}_2 = (t_{\lambda_2}, \gamma_2)$ are two 2-tuple unbalanced linguistic variables.

3.2 Probabilistic linguistic term sets

Pang et al. [48] extended the HFLTS by introducing the notion of PLTS, where probability is assigned to each linguistic term. Conclusively, compared to other extensions of the linguistic label set, the added probability amends the flexibility and capability of complex linguistic expressions. Henceforth, PLTS has gained the attention of researchers [25, 66, 67]. The notion of PLTS is defined as follows:

Definition 4 [48] Let $LT = \{s_{(-\tau)}, \ldots, s_0, \ldots, s_{\tau}\}$ be a finitely ordered LTS. The PLTS is defined as follows:

$$LT(p) = \left\{ s_i^{(k)}(p^{(k)}) \,|k = 1 \text{ to } \#LT(p) \right\}$$

where $s_i^{(k)} \in LT\ (i = -\tau, \ldots, 0, \ldots, \tau)$, the term $s_i^{(k)}(p^{(k)})$ expresses the linguistic term $s_i^{(k)}$ associated with the probability $p^{(k)}$, $0 \leq \sum_{k=1}^{\#LT(p)} p^{(k)} \leq 1$, and $\#LT(p)$ represents the number of all given distinct linguistic terms in $LT(p)$.

It is noted that if $\sum_{k=1}^{\#LT(p)} p^{(k)} = 1$ then complete information of probabilistic distribution of each plausible linguistic terms is obtained; if $\sum_{k=1}^{\#LT(p)} p^{(k)} < 1$ then the existence of partial ignorance is presented which persists in practical decision-making problems and if $\sum_{k=1}^{\#LT(p)} p^{(k)} = 0$ means complete ignorance. Handling the ignorance of $LT(p)$ is crucial work for the utility of PLTS.

Next, the details of the normalization method for PLTS $LT(p)$ are given in the definition below:

Definition 5 [48] Given a PLTS $LT(p)$ with $\sum_{k=1}^{\#LT(p)} p^{(k)} < 1$, the normalized PLTS $\overline{LT}(p)$ is defined as: $\overline{LT}(p) = \left\{ s_i^{(k)}(\bar{p}^{(k)}) \,|\, k = 1 \text{ to } \#LT(p) \right\}$ where $\bar{p}^{(k)} = \frac{p^{(k)}}{\sum_{k=1}^{\#LT(p)} p^{(k)}}$ for all $k = 1$ to $\#LT(p)$.

Next, Mao et al. [41] gave a possibility degree algorithm to rank PLTSs. The details of a possibility degree algorithm are stated below:

Definition 6 Given a symmetric LTS $LT = \{s_\alpha \,|\, \alpha = -\tau, \ldots, 0, \ldots, \tau\}$. Assume

$$LT_1(p) = \left\{ s_i^{(k_1)}(\bar{p}_1^{(k_1)}) \,|\, k_1 = 1 \text{ to } \#LT_1(p) \right\}$$

and

$$LT_2(p) = \left\{ s_j^{(k_2)}(\bar{p}_2^{(k_2)}) \,|\, k_2 = 1 \text{ to } \#LT_1(p) \right\}$$

for $(i, j = -n, \ldots, 0, \ldots, m)$ are two normalized PLTSs with $\#LT_1(p) = \#LT_2(p)$. A binary relation $R(s_i^{(k_1)}, s_j^{(k_2)})$ is defined as:

$$R(s_i^{(k_1)}, s_j^{(k_2)}) = \begin{cases} \bar{p}_1^{k_1} \bar{p}_2^{k_2}, & \text{if } i^{k_1} > j^{k_2} \\ \frac{1}{2} \bar{p}_1^{k_1} \bar{p}_2^{k_2}, & \text{if } i^{k_1} = j^{k_2} \\ 0, & \text{if } i^{k_1} < j^{k_2} \end{cases} \tag{1}$$

where the range of $k_1\,(k_2)$ is expressed as $(k_1(k_2) = 1, 2, \ldots, \#LT_1(p))$.

From Eq. (1) we can infer that the possibility degree of $s_i^{(k_1)}$ in $LT_1(p)$ is not less than $s_j^{(k_2)}$ in $LT_2(p)$. Henceforth, the possibility degree is defined as:

$$P(LT_1(p) \geq LT_2(p)) = \sum_{k_1=1}^{\#LT_1(p)} \sum_{k_2=1}^{\#LT_2(p)} R\left(s_i^{(k_1)}, s_j^{(k_2)} \right) \tag{2}$$

Definition 7 [41] Let $LT = \{s_i | i = -\tau, \ldots, 0, \ldots, \tau\}$ be a finite LTS and $LT(p) = \left\{ s_i^{(k)}(\bar{p}^{(k)}) | k = 1 \text{ to } \#LT(p) \right\}$ be a normalized PLTS. Let $r^{(k)}$ $(i = -n, \ldots, 0, 1, \ldots, m)$ be the subscript of a linguistic term $s^{(k)}$. Let $r^{(-)} = \min_k \{r^k\}$ and $r^{(+)} = \max_k \{r^k\}$ be the lower and upper boundaries of the term $s_i^{(k)}$, $\bar{p}^-$ and $\bar{p}^+$ be the respective probabilities. Then, the range value is given as:

$$R(LT(p)) = |r^{(+)}\bar{p}^+ - r^{(-)}\bar{p}^-|. \tag{3}$$

After combining the possibility degree and the range value, Mao et al. [41] defined the ordered relation for two classic PLTSs.

Definition 8 Let $LT_1(p)$ and $LT_2(p)$ be any two normalized PLTS. The comparison of PLTS is given as follows:

(i) If $P(LT_1(p) \geq LT_2(p)) > 0.5$, then $LT_1(p) > LT_2(p)$;
(ii) If $P(LT_1(p) \geq LT_2(p)) = 0.5$, then
  (a) if $R(LT_1(p)) < R(LT_2(p))$, then $LT_1(p) > LT_2(p)$;
  (b) if $R(LT_1(p)) = R(LT_2(p))$, then $LT_1(p) = LT_2(p)$.

Based on the above-mentioned analysis, Mao et al. [41] states a possibility algorithm to rank a given list of classic PLTSs.

Step 1: Compute the possibility degrees $P_{ij} = P(LT_i(p) \geq LT_j(p))$ $(i, j = 1, 2, \ldots, n)$ and the range value $R(LT_i(p))$ $(i = 1, 2, \ldots, n)$.

Step 2: The superior index P_i attained for $LT_1(p)$ is given by $P_i = \sum_{j=1}^{n} P_{ij}$.

Step 3: Based on the descending values of P_i the desired ranking order of $LT_i(p)$ $(i = 1, 2, \ldots, n)$ is generated. However, if for some $LT_i(p)$ the value of the superior index is equal, then the reordering of $LT_i(p)$ is obtained by the ascending value of $R(LT_i(p))$.

3.3 A zeros-sum matrix game

We review some elementary definitions from conventional matrix game theory taken from famous text Barron [3].

Definition 9 [3] A two-person zero-sum matrix game G is described as a triplet (S_1, S_2, A), where S_1 and S_2 termed as a finite strategy set for player I (PI) and player II (PII), respectively, and A is a real-valued payoff matrix of PI against PII while $-A$ is taken as the real-valued payoff matrix for PII.

Definition 10 [3] Given matrix $A = [a_{ij}]_{n \times m}$, let v^- and v^+ are the values corresponding to PI and PII, respectively such that: $v^- = \max_{i=1,\ldots,n} \min_{j=1,\ldots,m} a_{ij}$; $v^+ = \min_{j=1,\ldots,m} \max_{i=1,\ldots,n} a_{ij}$.

Definition 11 [3] Suppose $(x, y) \in S_1 \times S_2$ is considered to be an ordered pair of mixed strategy, such that

$$S_1 = \left\{ (x_1, \ldots, x_n) \, : \, x_i \geq 0, \, i = 1, \ldots, n, \, \sum_{i=1}^{n} x_i = 1 \right\};$$

$$S_2 = \left\{ (y_1, \ldots, y_m) \, : \, y_j \geq 0, \, j = 1, \ldots, m, \, \sum_{j=1}^{m} y_j = 1 \right\}.$$

Here, x_i is taken as the probability of PI selecting strategy i and y_j is the probability of PII choosing strategy j.

For more comprehensive overview about crisp zero-sum matrix game, one can refer to the text in [3].

4 Probabilistic multiplicative unbalanced linguistic term set

The present section is devoted to the novel concept of PM-ULTSs, followed by proposing a novel comparison method to rank PM-ULTS, basic operational laws, and the aggregation operators to aggregate PM-ULTS. Some properties of probabilistic unbalanced linguistic operational laws and aggregation operators are also discussed there.

4.1 The notion of the probabilistic multiplicative unbalanced linguistic term set

In literature, the idea of reflecting the probabilistic information for the evenly distributed linguistic terms is proposed by Pang et al. [48]. Analogously, we define the concept of PM-ULTSs by considering the case where the linguistic terms are unevenly distributed in the following definition:

Definition 12 Let $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, m\}$ be an ULTS where n and m may or may not be equal and $a > 1$, a probabilistic multiplicative ULTS is defined as: $LS(p) = \left\{ t_{a^i}^{(k)}(p^{(k)}) | \, k = 1, 2, \ldots, \#LS(p) \right\}$ where $t_{a^i}^{(k)} \in LS$, $p^{(k)} \geq 0$, $\sum_{k=1}^{\#LS(p)} p^{(k)}$ may or may not be equal to 1 and $\#LS(p)$ is the cardinality of all the given distinct multiplicative unbalanced linguistic terms in $LS(p)$.

Note 1 The term $t_{a^i}^{(k)}(p^{(k)})$ is interpreted as unbalanced linguistic term $t_{a^i}^{(k)}$ associated with the probability $p^{(k)}$.

It is noted that if $\sum_{k=1}^{\#LS(p)} p^{(k)} < 1$ or $\sum_{k=1}^{\#LS(p)} p^{(k)} > 1$ then the probability information within the set is incomplete whereas if $\sum_{k=1}^{\#LS(p)} p^{(k)} = 1$ then the information is complete.

Remark 2 We perform the normalization process for the incomplete PM-ULTS to transform the set into complete PM-ULTS. The normalization of the PM-ULTS can be acquired by using the following steps:

(i) Normalize the probability of PM-ULTS. If $\sum_{k=1}^{\#LS(p)} p^{(k)} \neq 1$, then the given PM-ULTS is defined as $\widetilde{LS}(p) = \left\{ t_{a^i}^{k}(\bar{p}^{k}) | \sum_{k=1}^{\#LS(p)} \bar{p}_k = 1 \right\}$ where $\bar{p} = \frac{p^k}{\sum_{k=1}^{\#LS(p)} p^{(k)}}$, $t_{a^i} \in LS$, $p^{(k)} \geq 0$, and range of k is expressed as $k = 1$ to $\#\widetilde{LS}(p)$;

(ii) Normalize the granularity of PM-ULTS. Let $LS_1(p)$ and $LS_2(p)$ are two PM-ULTS such that $\#LS_1(p)$, $\#LS_2(p)$ represent the cardinalities of the set respectively. If $\#LS_1(p) < \#LS_2(p)$, we add $\#LS_2(p) - \#LS_1(p)$ multiplicative unbalanced linguistic terms to the set $LS_1(p)$, so that both PM-ULTSs have equal cardinalities. It is noted that the added multiplicative unbalanced linguistic terms are the least one having the probability as zero in $LS_1(p)$.

Henceforth, the resultant PM-ULTSs are termed as the normalized PM-ULTSs. Hereafter, for the sake of convenience, we refer to $LS_1(p)$ and $LS_2(p)$ as normalized PM-ULTS with $\#LS_1(p) = \#LS_2(p)$.

The demonstration of the normalization process is presented in Example 1.

Example 1 Let $LS = \{t_{1/2} : \text{`` Very low''}, t_{1/2} : \text{`` low''}, t_1 : \text{`` avg ''}, t_2 : \text{`` high''}, t_4 : \text{`` Very high''}\}$ be a multiplicative ULTS. Let $LS_1(p) = \left\{ t_{1/2}^1(0.6), t_2^2(0.4) \right\}$ and $LS_2(p) = \left\{ t_{1/4}^1(0.3), t_{1/2}^2(0.2), t_4^3(0.1) \right\}$ are PM-ULTS. According to the normalization process, since $\#LS_1(p) < \#LS_2(p)$, then we add the linguistic terms $t_{1/4}$ to $LS_1(p)$ with probability 0 so that the numbers of linguistic terms in both the PM-ULTS are identical, and thus, we have $\widetilde{LS}_1(p) = \left\{ t_{1/4}^1(0), t_{1/2}^2(0.6), t_2^3(0.4) \right\}$. Therefore, after normalization, $\widetilde{LS}_1(p) = \left\{ t_{1/4}^1(0), t_{1/2}^1(0.6), t_2^2(0.4) \right\}$ and $\widetilde{LS}_2(p) = \left\{ t_{1/4}^1(0.5), t_{1/2}^2(0.333), t_4^3(0.16) \right\}$.

Unless otherwise mentioned, the PM-ULTS is always presupposed to be normalized PM-ULTS. Furthermore, it is viewed that, generally, PM-ULTS are not ordered sets; therefore, to fix the position of probabilistic multiplicative unbalanced linguistic terms, we define ordered PM-ULTSs in the following fashion.

Definition 13 Given a PM-ULTS $LS(p) = \left\{ t_{a^i}^{(k)}(p^{(k)}) \mid k = 1, 2, \ldots, \#LS(p) \right\}$, then $LS(p)$ is termed as an ordered PM-ULTS if the unbalanced linguistic terms $t_{a^i}^{(k)}(p^{(k)})(k = 1, \ldots, \#LS(p))$ are arranged in a descending order of the values $a^i \cdot p$ for each k.

Next, in literature, Mao et al. [41] designed a possibility degree algorithm to rank a series of classic PLTSs, which is a robust and beneficial approach. Similarly, the simplified ranking method to rank PM-ULTSs can also be stated. We begin with the following definition.

Definition 14 Given ULTS LS. Let $LS_1(p) = \left\{ t_{a^i}^{(k_1)}(\bar{p}_1^{(k_1)}) \mid k_1 = 1 \text{ to } \#LS_1(p) \right\}$ and $LS_2(p) = \{ t_{a^j}^{(k_2)}(\bar{p}_2^{(k_2)}) \mid k_2 = 1 \text{ to } \#LS_1(p) \}$ for $(i, j = -n, \ldots, 0, \ldots, m)$ are two normalized PM-ULTSs with $\#LS_1(\bar{p}_1) = \#LS_2(\bar{p}_2)$. An unbalanced binary linguistic relation $\hat{R}(t_{a^i}^{(k_1)}, t_{a^j}^{(k_2)})$ is defined as:

$$\hat{R}(t_{a^i}^{(k_1)}, t_{a^j}^{(k_2)}) = \begin{cases} \bar{p}_1^{k_1} \bar{p}_2^{k_2}, & \text{if } a^i > a^j \\ \frac{1}{2} \bar{p}_1^{k_1} \bar{p}_2^{k_2}, & \text{if } a^i = a^j \\ 0, & \text{if } a^i < a^j \end{cases} \tag{4}$$

where the range of $k_1\,(k_2)$ is expressed as $(k_1(k_2) = 1, 2, \ldots, \#LS_1(p))$.

From Eq. (4) we can infer that the possibility degree of $t_{a^i}^{(k_1)}$ in $LS_1(p)$ is greater than $t_{a^j}^{(k_2)}$ in $LS_2(p)$. Hence, the possibility degree is given as:

$$P(LS_1(p) \geq LS_2(p)) = \sum_{k_1=1}^{\#LS_1(p)} \sum_{k_2=1}^{\#LS_2(p)} \hat{R}(t_{a^i}^{(k_1)}, t_{a^j}^{(k_2)}) \tag{5}$$

Mao et al. [41] have defined some desirable properties of the possibility degree for the classic normalized PLTSs. However, it is anticipated that these properties hold even for the case of normalized PM-ULTSs. Meanwhile, for the case if $P(LS_1(p) \geq LS_2(p)) = P(LS_2(p) \geq LS_1(p)) = 0.5$, then it can be concluded that the classic PLTSs cannot be compared and hence, range value of PLTSs is calculated representing the degree of dispersion. On similar grounds, the range value for normalized PM-ULTSs $LS(p)$ is stated below:

Definition 15 Let $LS = \{ t_{a^i} | i = -n, \ldots, 0, \ldots, m \}$ be an ULTS and $LS(p) = \left\{ t_{a^i}^{(k)}(\bar{p}^{(k)}) | k_1 = 1 \text{ to } \#LS(p) \right\}$ be a normalized PM-ULTS. Let $a_i^{(k)}$ $(i = -n, \ldots, 0, 1, \ldots, m)$ be the subscript of unbalanced linguistic term $t_{a^i}^{(k)}$. Let $a^{i(-)} = \min_k \{a^i\}$ and $a^{i(+)} = \max_k \{a^i\}$ be the lower and upper boundaries of the term $t_{a^i}^{(k)}$, $\bar{p}^-$ and $\bar{p}^+$ be the respective probabilities. The range value is given as:

$$R(LS(p)) = |a^{i(+)} \bar{p}^+ - a^{i(-)} \bar{p}^-|. \tag{6}$$

After combining the possibility degree and the range value, Mao et al. [41] defined the ordered relation for two classic PLTSs. Similarly, a simplified proposal of the ordered relation for two normalized PM-ULTS $LS_1(p)$ and $LS_2(p)$ can be stated below:

Definition 16 Let $LS_1(p)$ and $LS_2(p)$ be any two normalized PM-ULTS. The comparison of which is given as follows:

(i) If $P(LS_1(p) \geq LS_2(p)) > 0.5$, then $LS_1(p) > LS_2(p)$;
(ii) If $P(LS_1(p) \geq LS_2(p)) = 0.5$, then
  (a) if $R(LS_1(p)) < R(LS_2(p))$, then $LS_1(p) > LS_2(p)$;
  (b) if $R(LS_1(p)) = R(LS_2(p))$, then $LS_1(p) = LS_2(p)$.

Next, based on the abovementioned analysis, Mao et al. [41] states a possiblity algorithm to rank a given list of classic PLTSs. Analogously, we propose a possibility algorithm to rank list of normalized K PM-ULTSs $LS_i(p)$ $(i = 1, 2, \ldots, K)$:

Step 1: Calculate the possibility degrees $P_{ij} = P(LS_i(p) \geq LS_j(p))$ $(i, j = 1, 2, \ldots, K)$ using Eq. (5) and the range value $R(LS_i(p))$ $(i = 1, 2, \ldots, K)$ using Eq. (6), respectively.
Step 2: Calculate the superior index $P_i = \sum_{j=1}^{n} P_{ij}$.
Step 3: The descending values of P_i generates the desired ranking order of $LS_i(p)$ $(i = 1, 2, \ldots, K)$. However, if for some $LS_i(p)$ the value of superior index is equal, then the reordering of $LS_i(p)$ is given by the ascending value of $R(LS_i(p))$.

For clarity of the ranking algorithm mentioned above, we give the following example:

Example 2 Consider the list of four PM-ULTS i.e., $LS_1(p) = \left\{ t_{1/2^3}(0.1), t_{1/2^2}(0.4), t_{2^0}(0.3), t_2(0.2) \right\}$, $LS_2(p) = \left\{ t_{1/2}(0.5), t_{2^0}(0.5) \right\}$, $LS_3(p) = \{ t_{2^2}(0.2), t_{2^3}(0.8) \}$, $LS_4(p) = \{ t_{2^2}(0.4), t_{2^3}(0.6) \}$ based on ULTS $LS = \left\{ t_{1/2^3}, t_{1/2^2}, t_{1/2}, t_{2^0}, t_{2^1}, t_{2^2}, t_{2^3} \right\}$. We will apply the proposed ranking algorithm to rank the given list of PM-ULTSs.

Step 1: Calculate the possibility degrees P_{ij} $(i, j = 1, 2, 3, 4)$ using Eq. (5) and the range value $R(LS_i(p))$ $(i = 1, 2, 3, 4)$ using Eq. (6) as demonstrated: $P_{11} = 0.5$, $P_{12} = 0.425$, $P_{13} = 0$, $P_{14} = 0$, $P_{21} = 0.575$, $P_{22} = 0.5$, $P_{23} = 0$, $P_{24} = 0$, $P_{31} = 1$, $P_{32} = 1$, $P_{33} = 0.5$, $P_{34} = 0.68$, $P_{41} = 1$, $P_{42} = 1$, $P_{43} = 0.32$, $P_{44} = 0.5$. $R(LS_1(p)) = 0.3875$, $R(LS_2(p)) = 0.25$, $R(LS_3(p)) = 5.6$, $R(LS_4(p)) = 3.2$.

Step 2: The corresponding superior index $P_i = \sum_{j=1}^{4} P_{ij} (i = 1, 2, 3, 4)$ is obtained as: $P_1 = 0.925$, $P_2 = 1.075$, $P_3 = 3.18$, $P_4 = 2.82$.

Step 3: From the values of $P_i (i = 1, 2, 3, 4)$ and $R(LS_i(p))$ $(i = 1, 2, 3, 4)$ obtained in the previous steps, the ranking order is given as: $LS_3(p) \succ LS_4(p) \succ LS_2(p) \succ LS_1(p)$.

Next, we define the negation of PM-ULTS in the following manner.

Definition 17 Let $LS(p) = \left\{ t_{a^i}^{(k)}(\bar{p}^{(k)}) | k = 1 \text{ to } \#LS(p) \right\}$ be a normalized PM-ULTS. The negation of PM-ULTS is defined as: $\text{neg}(LS(p)) = \left\{ \text{neg}(t_{a^i})^{(k)}((\bar{p})^{(k)}) \,|\, k = 1, 2, \ldots, \#LS(p) \right\}$.

4.2 Operational laws for probabilistic multiplicative unbalanced linguistic terms

In this subsection, we will present some basic operational laws primarily based on the given PM-ULTSs. For this, we will assume that all PM-ULTSs are normalized and have equal cardinality, i.e., for each PM-ULTSs $LS_1(p_1)$ and $LS_2(p_2)$, we have $\sum_{k=1}^{\#LS_1(p_1)} \bar{p}_1^k = \sum_{k=1}^{\#LS_2(p_2)} \bar{p}_2^k = 1$ and $\#LS_1(p_1) = \#LS_2(p_2)$.

Definition 18 Let $LS_1(p_1)$ and $LS_2(p_2)$ be any two arbitrary ordered PM-ULTSs, $LS_1(p_1) = \left\{ t_{a^i}^{(k)}(\bar{p}_2^k) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$ and $LS_2(p_2) = \left\{ t_{a^j}^{(k)}(\bar{p}_2^k) \,|\, k = 1, 2, \ldots, \#LS_2(p_2) \right\}$ for $(i, j = -n, \ldots, 0, \ldots, m)$. The operational laws based on minimum distance measures are given as follows:

(1) Addition operation

$$LS_1(p_1) \oplus LS_2(p_2) = \left\{ t_{a^i \oplus a^j}^{(k)}(\bar{p}_1^k \oplus \bar{p}_2^k) | \, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$$

(2) Multiplication operation

$$LS_1(p_1) \otimes LS_2(p_2) = \left\{ t_{a^i \otimes a^j}^{(k)}(\bar{p}_1^k \otimes \bar{p}_2^k) | \, k = 1, 2, \ldots, \#LS_1(p_1) \right\}.$$

(3) Scalar multiplication operation

For $\lambda \in (0, 1)$ being any arbitrary scalar, $\lambda \odot LS(p) = \left\{ t_{\lambda \odot a^i}^{(k)}(\bar{p}^k) | \, k = 1, 2, \ldots, \#LS(p) \right\}$.

(4) Power operation

$$(LS(p))^{\lambda} = \left\{ t_{a^{\lambda i}}^{(k)}(\bar{p}^{(k)}) \,|\, k = 1, 2, \ldots, \#LS(p) \right\}$$

Based on the two arbitrary PM-ULTSs $LS_1(p_1)$ and $LS_2(p_2)$, the rule-based method to compute values for the aforementioned operational laws can be outlined in the Algorithms 1, 2, 3, and 4 given in Appendix 1.

Remark 3 The outcomes obtained after using the operational laws should preserve the probability information.

Based on the Definition 18, the following properties of the operational laws can be defined.

Theorem 1 (*Closure properties*). *Let Γ be the set of all PM-ULTSs, i.e., $LS_1(p_1) \in \Gamma$ and $LS_2(p_2) \in \Gamma$. Then,*

(1) $LS_1(p_1) \oplus LS_2(p_2) \in \Gamma$;
(2) $LS_1(p_1) \otimes LS_2(p_2) \in \Gamma$;
(3) $\lambda \odot LS_1(p_1) \in \Gamma$;
(4) $(LS_1(p_1))^{\lambda} \in \Gamma$;
(5) $\text{neg}(LS_1(p_1) \oplus LS_2(p_2)) \in \Gamma$;
(6) $\text{neg}(LS_1(p_1) \otimes LS_2(p_2)) \in \Gamma$;
(7) $\text{neg}(\lambda \odot LS_1(p_1)) \in \Gamma$;
(8) $\text{neg}((LS_1(p_1))^{\lambda}) \in \Gamma$.

Proof The proof of Theorem 1 follows directly from Definition 18 based on the rule-based method given in Algorithms 1–4 (see Appendix 1). □

Theorem 2 (*Commutative law*). *For any $LS_1(p_1)$ and $LS_2(p_2)$ we have*:

(1) $LS_1(p_1) \oplus LS_2(p_2) = LS_2(p_2) \oplus LS_1(p_1)$;
(2) $LS_1(p_1) \otimes LS_2(p_2) = LS_2(p_2) \otimes LS_1(p_1)$.

Proof Let $LS_1(p_1) = \left\{ t_{a^i}^{(k)}(\bar{p}_1^k) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$ and $LS_2(p_2) = \left\{ t_{a^j}^{(k)}(\bar{p}_2^k) \,|\, k = 1, 2, \ldots, \#LS_2(p_2) \right\}$, $(i, j = -n, \ldots, 0, 1, \ldots, m)$ where $a > 1$ and n may or may not be equal be the two normalized PM-ULTS with $\#LS_1(\bar{p}_1) = \#LS_2(\bar{p}_2)$. According to the Definition 18, we have $LS_1(p_1) \oplus LS_2(p_2) = \left\{ t_{a^i \oplus a^j}^{(k)}(\bar{p}_1^k \oplus \bar{p}_2^k) | \, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$.

Next, based on the rule-based method given in Algorithm 1 we consider the following cases:

Case 1: If $t^{(k)}_{a^i \oplus a^j} \in \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^{(m)}}\}$ for any $k = 1, 2, \ldots, \#LS_1(p_1)$. Then we have, $t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^i + a^j}$ and therefore,

$$\begin{aligned} &LS_1(p_1) \oplus LS_2(p_2) \\ &= \left\{ t^{(k)}_{a^i + a^j}(\bar{p}^k_1 \oplus \bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \\ &= \left\{ t^{(k)}_{a^j + a^i}(\bar{p}^k_1 \oplus \bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \\ &= \left\{ t^{(k)}_{a^j \oplus a^i}(\bar{p}^k_2 \oplus \bar{p}^k_1) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \\ &= LS_2(p_2) \oplus LS_1(p_1). \end{aligned}$$

Case 2: If $t^{(k)}_{a^i \oplus a^j} \notin \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^{(m)}}\}$ for any $k = 1, 2, \ldots, \#LS_1(p_1)$. Then we have, $t^{(k)}_{a^i \oplus a^j} = t^{(k)}_d$ where $d = \begin{cases} a^i + a^j \text{ if } d \in [a^{-n}, a^m]; \\ a^i + a^j - a^m \text{ otherwise.} \end{cases}$. Now, for $d \in [a^{-n}, a^m]$ or $d \notin [a^{(-n)}, a^m]$. Then, we have $t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^i{}_L} = t^{(k)}_{a^j \oplus a^i}$ or $t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^i{}_R} = t^{(k)}_{a^j \oplus a^i}$.

Hence, $LS_1(p_1) \oplus LS_2(p_2) = LS_2(p_2) \oplus LS_1(p_1)$.

Similarly, we can prove the second part. This proves the theorem. □

Theorem 3 (*Associative law*). *Let* $LS_1(p_1)$, $LS_2(p_2)$, $LS_3(p_3)$ *be any three normalized PM-ULTSs, then*

(i) $(LS_1(p_1) \oplus LS_2(p_2)) \oplus LS_3(p_3) = LS_1(p_1) \oplus (LS_2(p_2) \oplus LS_3(p_3))$;

(ii) $(LS_1(p_1) \otimes LS_2(p_2)) \otimes LS_3(p_3) = LS_1(p_1) \otimes (LS_2(p_2) \otimes LS_3(p_3))$.

Proof Let $LS_1(p_1) = \left\{ t^{(k)}_{a^i}(\bar{p}^k_1) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$, $LS_2(p_2) = \left\{ t^{(k)}_{a^j}(\bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$, and $LS_3(p_3) = \left\{ t^{(k)}_{a^l}(\bar{p}^k_3) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$ $(i, j, l = -n, \ldots, 0, 1, \ldots, m)$ where $a > 1$ be three normalized PM-ULTS with $\#LS_1(p_1) = \#LS_2(p_2) = \#LS_3(p_3)$.

Consider $(LS_1(p_1) \oplus LS_2(p_2)) \oplus LS_3(p_3) = \left\{ t^{(k)}_{a^i \oplus a^j}(\bar{p}^k_1 \oplus \bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \oplus \left\{ t^{(k)}_{a^l}(\bar{p}_3{}^k) \,|\, k = 1, 2, \ldots, \#LS_2(p_1) \right\}$.

Based on the rule based method discussed in Algorithm 1, for any k $(k = 1, 2, \ldots, \#LS_1(p_1))$, we have the following cases:

Case 1: If $t^{(k)}_{a^i \oplus a^j} \in \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^i + a^j}$ and $p = \bar{p}_1 \oplus \bar{p}_2 = \begin{cases} \bar{p}_1 + \bar{p}_2, \text{ if } \bar{p}_1 + \bar{p}_2 \in [0, 1]; \\ \bar{p}_1 + \bar{p}_2 - 1, \text{ otherwise} \end{cases}$.

Therefore, $(LS_1(p_1) \oplus LS_2(p_2)) \oplus LS_3(p_3)$

$$\begin{aligned} &= \left\{ t^{(k)}_{a^i \oplus a^j}(\bar{p}^k_1 \oplus \bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \\ &\oplus \left\{ t^{(k)}_{a^l}(\bar{p}^k_3) \,|\, k = 1, 2, \ldots, \#LS_2(p_1) \right\} \\ &= \left\{ t^{(k)}_{a^i + a^j}(\bar{p}^k_1 \oplus \bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \\ &\oplus \left\{ t^{(k)}_{a^l}(\bar{p}^k_3) \,|\, k = 1, 2, \ldots, \#LS_2(p_1) \right\}. \end{aligned}$$

Next, again using rule-based method we consider the following subcases:

Subcase i: If $t^{(k)}_{a^i + a^j \oplus a^l} \in \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i + a^j \oplus a^l} = t^{(k)}_{a^i + a^j + a^l}$ and $\bar{p} \oplus \bar{p}_3 = \begin{cases} \bar{p} + \bar{p}_3, \text{ if } \bar{p}_1 + \bar{p}_2 \in [0, 1]; \\ \bar{p} + \bar{p}_3 - 1, \text{ otherwise} \end{cases}$.

Therefore, $(LS_1(p_1) \oplus LS_2(p_2)) \oplus LS_3(p_3) = \left\{ t^{(k)}_{a^i + a^j + a^l}(p^k \oplus \bar{p}^k_3) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$.

Subcase ii: If $t^{(k)}_{a^i + a^j \oplus a^l} \notin \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i + a^j \oplus a^l} = t^{(k)}_{a^i{}_L}$ or $t^{(k)}_{a^i + a^j \oplus a^l} = t^{(k)}_{a^i{}_R}$ and $\bar{p} \oplus \bar{p}_3 = \begin{cases} \bar{p} + \bar{p}_3, \text{ if } \bar{p}_1 + \bar{p}_2 \in [0, 1]; \\ \bar{p} + \bar{p}_3 - 1, \text{ otherwise} \end{cases}$.

Therefore, $(LS_1(p_1) \oplus LS_2(p_2)) \oplus LS_3(p_3) = \left\{ t^{(k)}_{a^i{}_L}(p^k \oplus \bar{p}^k_3) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$.

Case 2: If $t^{(k)}_{a^i \oplus a^j} \notin \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^i{}_L}$ or $t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^i{}_R}$ and $p = \bar{p}_1 \oplus \bar{p}_2 = \begin{cases} \bar{p}_1 + \bar{p}_2, \text{ if } \bar{p}_1 + \bar{p}_2 \in [0, 1]; \\ \bar{p}_1 + \bar{p}_2 - 1, \text{ otherwise} \end{cases}$.

Therefore, $(LS_1(p_1) \oplus LS_2(p_2)) \oplus LS_3(p_3)$

$$\begin{aligned} &= \left\{ t^{(k)}_{a^i \oplus a^j}(\bar{p}^k_1 \oplus \bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \\ &\oplus \left\{ t^{(k)}_{a^l}(\bar{p}^k_3) \,|\, k = 1, 2, \ldots, \#LS_2(p_1) \right\} \\ &= \left\{ t^{(k)}_{a^i{}_L}(\bar{p}^k_1 \oplus \bar{p}^k_2) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\} \\ &\oplus \left\{ t^{(k)}_{a^l}(\bar{p}^k_3) \,|\, k = 1, 2, \ldots, \#LS_2(p_1) \right\}. \end{aligned}$$

Next, again using rule-based method we consider the following subcases:

Subcase i: If $t^{(k)}_{a^{i_L}\oplus a^l} \in \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^{i_L}\oplus a^l} = t^{(k)}_{a^{i_L}+a^l}$ and $\bar{p}\oplus\bar{p}_3 = \begin{cases} \bar{p}+\bar{p}_3, \text{ if } \bar{p}_1+\bar{p}_2 \in [0,1]; \\ \bar{p}+\bar{p}_3-1, \text{ otherwise} \end{cases}$.
Therefore, $(LS_1(p_1)\oplus LS_2(p_2))\oplus LS_3(p_3) = \left\{t^{(k)}_{a^{i_L}+a^l}(p^k\oplus\bar{p}^k_3)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Subcase ii: If $t^{(k)}_{a^{i_L}\oplus a^l} \notin \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^{i_L}\oplus a^l} = t^{(k)}_{a^{i_L}}$ or $t^{(k)}_{a^{i_L}\oplus a^l} = (t^{(k)}_{a^{i_R}})$ and $\bar{p}\oplus\bar{p}_3 = \begin{cases} \bar{p}+\bar{p}_3, \text{ if } p+\bar{p}_3 \in [0,1]; \\ \bar{p}+\bar{p}_3-1, \text{ otherwise} \end{cases}$.
Therefore, $(LS_1(p_1)\oplus LS_2(p_2))\oplus LS_3(p_3) = \left\{t^{(k)}_{a^{i_L}}(p^k\oplus\bar{p}^k_3)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Next, consider $LS_1(p_1)\oplus(LS_2(p_2)\oplus LS_3(p_3)) = \left\{t^{(k)}_{a^i}(\bar{p}^k_1)\,|\,k=1,2,\ldots,\#LS_2(p_1)\right\}\oplus\left\{t^{(k)}_{a^j\oplus a^l}(\bar{p}^k_2\oplus\bar{p}^k_3)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Based on the rule based method discussed in Algorithm 1, for any $k\,(k=1,2,\ldots,\#LS_1(p_1))$ we have the following cases:

Case 1: If $t^{(k)}_{a^j\oplus a^l} \in \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^j\oplus a^l} = t^{(k)}_{a^j+a^l}$ and $p = \bar{p}_2\oplus\bar{p}_3 = \begin{cases} \bar{p}_2+\bar{p}_3, \text{ if } \bar{p}_2+\bar{p}_3 \in [0,1]; \\ \bar{p}_2+\bar{p}_3-1, \text{ otherwise} \end{cases}$.
Therefore, $LS_1(p_1)\oplus(LS_2(p_2)\oplus LS_3(p_3))$
$= \left\{t^{(k)}_{a^i}(\bar{p}^k_1)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$
$\oplus\left\{t^{(k)}_{a^j\oplus a^l}(\bar{p}^k_2\oplus\bar{p}^k_3)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$
$= \left\{t^{(k)}_{a^i}(\bar{p}^k_1)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$
$\oplus\left\{t^{(k)}_{a^j+a^l}(p^k)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Next, again using rule-based method we consider the following subcases:

Subcase i: If $t^{(k)}_{a^i\oplus(a^j+a^l)} \in \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i\oplus(a^j+a^l)} = t^{(k)}_{a^i+a^j+a^l}$ and $\bar{p}_1\oplus p = \begin{cases} \bar{p}_1+p, \text{ if } \bar{p}_1+p \in [0,1]; \\ \bar{p}_1+p-1, \text{ otherwise} \end{cases}$.
Therefore, $LS_1(p_1)\oplus(LS_2(p_2)\oplus LS_3(p_3)) = \left\{t^{(k)}_{a^i+a^j+a^l}(\bar{p}^k_1\oplus p^k)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Subcase ii: If $t^{(k)}_{a^i\oplus(a^j+a^l)} \notin \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i\oplus(a^j+a^l)} = t^{(k)}_{a^{i_L}}$ or $t^{(k)}_{a^i\oplus(a^j+a^l)} = t^{(k)}_{a^{i_R}}$ and $\bar{p}_1\oplus p = \begin{cases} \bar{p}_1+p, \text{ if } \bar{p}_1+p \in [0,1]; \\ \bar{p}_1+p-1, \text{ otherwise} \end{cases}$.
Therefore, $(LS_1(p_1)\oplus(LS_2(p_2)\oplus LS_3(p_3)) = \left\{t^{(k)}_{a^{i_L}}(\bar{p}^k_1\oplus p^k)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Case 2: If $t^{(k)}_{a^j\oplus a^l} \notin \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^j\oplus a^l} = t^{(k)}_{a^{j_L}}$ or $t^{(k)}_{a^j\oplus a^l} = t^{(k)}_{a^{j_R}}$ and $p = \bar{p}_2\oplus\bar{p}_3 = \begin{cases} \bar{p}_2+\bar{p}_3, \text{ if } \bar{p}_2+\bar{p}_3 \in [0,1]; \\ \bar{p}_2+\bar{p}_3-1, \text{ otherwise} \end{cases}$.
Therefore, $LS_1(p_1)\oplus(LS_2(p_2)\oplus LS_3(p_3))$
$= \left\{t^{(k)}_{a^i}(\bar{p}^k_1)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$
$\oplus\left\{t^{(k)}_{a^j\oplus a^l}(\bar{p}^k_2\oplus\bar{p}^k_3)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$
$= \left\{t^{(k)}_{a^i}(\bar{p}^k_1)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$
$\oplus\left\{t^{(k)}_{a^{j_L}}(p^k)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Next, again using rule-based method we consider the following subcases:

Subcase i: If $t^{(k)}_{a^i\oplus a^{j_L}} \in \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i\oplus a^{j_L}} = t^{(k)}_{a^i+a^{j_L}}$ and $\bar{p}_1\oplus\bar{p} = \begin{cases} \bar{p}_1+\bar{p}, \text{ if } \bar{p}_1+p \in [0,1]; \\ \bar{p}_1+\bar{p}-1, \text{ otherwise} \end{cases}$.
Therefore, $LS_1(p_1)\oplus(LS_2(p_2)\oplus LS_3(p_3)) = \left\{t^{(k)}_{a^i+a^{j_L}}(\bar{p}^k_1\oplus p^k)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Subcase ii: If $t^{(k)}_{a^i\oplus a^{j_L}} \notin \{t_{a^{(-n)}}, \ldots, t_{a^0}, \ldots, t_{a^m}\}$. Then, $t^{(k)}_{a^i\oplus a^{j_L}} = t^{(k)}_{a^{j_L}}$ and $\bar{p}_1\oplus p = \begin{cases} \bar{p}_1+p, \text{ if } \bar{p}_1+p \in [0,1]; \\ \bar{p}_1+p-1, \text{ otherwise} \end{cases}$.
Therefore, $LS_1(p_1)\oplus(LS_2(p_2))\oplus LS_3(p_3)) = \left\{t^{(k)}_{a^{j_L}}(\bar{p}^k_1\oplus p^k)\,|\,k=1,2,\ldots,\#LS_1(p_1)\right\}$.

Clearly, it is seen that $(LS_1(p_1)\oplus LS_2(p_2))\oplus LS_3(p_3) = LS_1(p_1)\oplus(LS_2(p_2))\oplus LS_3(p_3))$ and hence associative law holds.

Similarly, we can prove the second part. This proves the theorem. □

Remark 4 It is noteworthy to derive the subsequent conclusions based on the aforementioned operational laws of PM-ULTSs.

(1) Linking the similar unbalanced linguistic terms having different probabilities associated together. For instance, consider the terms $t_{2^2}(0.2)$ and $t_{2^2}(0.3)$,

then we can connect them into $t_{2^2}(0.5)$ by summing up their corresponding probabilities. Nevertheless, if the sum total of two probabilities is not within the bounds then it can be re-translated by subtracting 1 from the resultant, i.e., if $\sum_{i=1}^{n} p_i > 1$ where n is the number of similar unbalanced linguistic terms, then $\sum_{i=1}^{n} p_i - 1$ would be the resultant value.

(2) We rank the resultant probabilistic multiplicative unbalanced linguistic terms obtained after employing the operational laws in increasing order of its subscript.

(3) The operational laws are defined in such a way that the resultant probabilistic information is preserved.

Example 3 can illustrate the whole mechanism of the above-mentioned operational laws and algebra operations.

Example 3 Let $LS = \{t_{a^i} \mid i = -4, \ldots, 0, \ldots, 3\}$ be a ULTS. For convenience, we take $a = 2$. Let $LS_1(p_1) = \{t_{1/8}(0.1276),\ t_1(0.8147),\ \ t_4(0.9572)\}$ and $LS_2(p_2) = \{t_{1/16}(0.9134),\ t_{1/2}(0.9572),\ t_8(0.4854)\}$. Since the given sets are not normalized, therefore we will firstly normalize the given PM-ULTSs such that $\widetilde{LS}_1(p_1) = \{t_{1/8}(0.0669),\ t_1(0.4290),\ \ t_4(0.5041)\}$ and $\widetilde{LS}_2(p_2) = \{t_{1/16}(0.3877),\ t_{1/2}(0.4063),\ t_8(0.2060)\}$, then

(1)

$$\begin{aligned}&\widetilde{LS}_1(p_1) \oplus \widetilde{LS}_2(p_2)\\ &= \{t_{1/8 \oplus 1/16}(0.4546), t_{1 \oplus 1/2}(0.8353), t_{4 \oplus 8}(0.7101)\}\\ &= \{t_{0.1875}(0.4546), t_{1.5}(0.8353), t_{12}(0.7101)\}\\ &= \{t_{0.1250}(0.2273), t_1(0.41765), t_4(0.3551)\}.\end{aligned}$$

(2)

$$\begin{aligned}&\widetilde{LS}_1(p_1) \otimes \widetilde{LS}_2(p_2)\\ &= \{t_{1/2^3 \otimes 1/2^4}(0.0259), t_{1 \oplus 1/2}(0.1743), t_{2^2 \oplus 2^3}(0.1038)\}\\ &= \{t_{1/8}(0.0259), t_{1/2}(0.1743), t_{2^2}(0.1038)\}\\ &= \{t_{1/8}(0.0852), t_{1/2}(0.5734), t_4(0.3414)\}.\end{aligned}$$

(3) Take $\lambda = 0.8$.

$$\begin{aligned}&0.8 \odot \widetilde{LS}_1(p_1)\\ &= 0.8 \odot \{t_{1/8}(0.0669),\ t_1(0.4290),\ t_4(0.5041)\}\\ &= \{t_{0.1250}(0.0669),\ t_1(0.4290),\ t_4(0.5041)\}.\end{aligned}$$

(4) Take $\lambda = 0.8$.

$$\begin{aligned}&(\widetilde{LS}_1(p_1))^{0.8}\\ &= \left(\{t_{1/2^3}(0.0669),\ t_{2^0}(0.4290),\ t_{2^2}(0.5041)\}\right)^{0.8}\\ &= \{t_{1/2^2}(0.0669),\ t_1(0.4290),\ t_{2^2}(0.5041)\}.\end{aligned}$$

(5)

$$\begin{aligned}&\mathrm{neg}(\widetilde{LS}_1(p_1) \oplus \widetilde{LS}_2(p_2))\\ &= \{\mathrm{neg}(t_{1/2^3})(0.7727), \mathrm{neg}(t_1)(0.58235), \mathrm{neg}(t_4)(0.6449)\}\\ &= \{t_{2^3}(0.7727),\ t_1(0.58235),\ t_{1/2^2}(0.6449)\}\\ &= \{t_{1/2^2}(0.3225),\ t_1(0.2912),\ t_{2^3}(0.3864)\}.\end{aligned}$$

4.3 Aggregation operators for PM-ULTSs

Considering the significance of aggregation operators in practical decision-making or matrix game problems useful for the emulsion of information, we will propose some elementary aggregation operators to aggregate PM-ULTSs.

Before we define the aggregation operators, the PM-ULTSs utilized during the operation are always presumed to be normalized PM-ULTSs having equal cardinalities.

Definition 19 Let $\{LS_i(p_i) | i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. The probabilistic multiplicative unbalanced linguistic averaging (PMULA) operator is defined as a mapping $LS_1(p_1) \times LS_2(p_2) \times \ldots \times LS_K(p_K) \to LS(p)$ such that

$$\begin{aligned}&PMULA(LS_1(p_1),\ LS_2(p_2), \ldots,\ LS_K(p_K))\\ &= \frac{1}{K}(LS_1(p_1) \oplus LS_2(p_2) \oplus \ldots \oplus LS_K(p_K)).\end{aligned}$$

The desired properties of the PMULA operator can be listed as follows.

Theorem 4 *Let $\{LS_i(p_i) | i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. Then, for the PMULA operator we have*

(1) Idempotency
If for all K $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ are equal i.e., $LS_i(p_i) = LS(p)$, then, $PMULA(LS_1(p_1), \ldots, LS_K(p_K)) = LS(p)$.

(2) Monotonicity
Let $\{LS_i'(p_i')|i = 1, 2, \ldots, K\}$ be a collection of K PM-ULTS. If for any $i(i = 1, \ldots, K)$ we have $LS_i'(p_i') \geq LS_i(p_i)$ then, $PMULA(LS_1'(p_1), \ldots, LS_K'(p_K)) \geq PMULA(LS_1(p_1), \ldots, LS_K(p_K))$.

(3) Boundary
$\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq PMULA(LS_1(p_1), \ldots, LS_K(p_K)) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$.

Proof

(1) Since $LS_i(p_i) = LS(p)$ for all $i = 1, \ldots, K$ then, we have $PMULA(LS_1(p_1), LS_2(p_2), \ldots, LS_K(p_K)) = \frac{1}{K}(LS_1(p_1) \oplus LS_2(p_2) \oplus \ldots \oplus LS_K(p_K)) = \frac{1}{K}(LS(p) \oplus LS(p) \oplus \ldots \oplus LS(p)) = LS(p)$.

(2) Since $LS_i'(p_i') \geq LS_i(p_i)$ for each $i(i = 1, \ldots, K)$, i.e.,

$$LS_1'(p_1') \geq LS_1(p_1)$$
$$LS_2'(p_2') \geq LS_2(p_2)$$
$$\vdots$$
$$LS_K'(p_K') \geq LS_K(p_K).$$

After adding all the above K inequalities and multiplying both sides by the factor $\frac{1}{K}$, we get

$$\frac{1}{K}(LS_1'(p_1') \oplus LS_2'(p_2') \oplus \ldots \oplus LS_K'(p_K')) \geq \frac{1}{K}(LS_1(p_1) \oplus LS_2(p_2) \oplus \ldots \oplus LS_K(p_K)) \Rightarrow PMULA(LS_1'(p_1), \ldots, LS_K'(p_K)) \geq PMULA(LS_1(p_1), \ldots, LS_K(p_K)).$$

(2) Since $\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ and $\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq LS_i(p_i) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ for each $i(i = 1, \ldots, K)$, i.e.,

$$\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq LS_1(p_1) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$$
$$\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq LS_2(p_2) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$$
$$\vdots$$
$$\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq LS_K(p_K) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}.$$

After adding all the above K inequalities, we get

$$\begin{aligned} &K \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\leq LS_1(p_1) \oplus \ldots \oplus LS_K(p_K) \\ &\leq K \max\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\Rightarrow \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\leq \frac{1}{K}(LS_1(p_1) \oplus \ldots \oplus LS_K(p_K)) \\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\Rightarrow \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\leq PMULA(LS_1(p_1), \ldots, LS_K(p_K)) \\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}. \end{aligned}$$

This proves the theorem. □

Definition 20 Let $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. Assign a relevant weight vector $W = (w_1, w_2, \ldots, w_K)$ such that $w_i \in [0, 1]$ and $\sum_{i=1}^{K} w_i = 1$. The probabilistic multiplicative unbalanced linguistic weighted averaging (PMULWA) operator is defined as:

$$\begin{aligned} &PMULWA(LS_1(p_1), LS_2(p_2), \ldots, LS_K(p_K)) \\ &= (w_1 \odot LS_1(p_1)) \oplus \ldots \oplus (w_K \odot LS_K(p_K)) \\ &= \left(w_1 \odot \left\{t_{a^{i_1}}^{(k)}(\bar{p}_1^{(k)})|k = 1, \ldots, \#LS_1(p_1)\right\}\right) \oplus \ldots \\ &\quad \oplus \left(w_K \odot \left\{t_{a^{i_K}}^{(k)}(\bar{p}_K^{(k)})|k = 1, \ldots, \#LS_1(p_1)\right\}\right). \end{aligned}$$

The desired properties of the PMULWA operator can be listed as follows.

Theorem 5 *Let $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. Let $W = (w_1, w_2, \ldots, w_K)$ be the associated weight vector such that $w_i \in [0, 1]$ and $\sum_{i=1}^{K} w_i = 1$. Then, for the PMULA operator, we have*

(1) Idempotency
If for all K $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ are equal i.e., $LS_i(p_i) = LS(p)$, then, $PMULWA(LS_1(p_1), \ldots, LS_K(p_K)) = LS(p)$.

(2) Monotonicity
Let $\{LS_i'(p_i')|i = 1, 2, \ldots, K\}$ be a collection of K PM-ULTS. If for any $i(i = 1, \ldots, K)$ we have $LS_i'(p_i') \geq LS_i(p_i)$ then, $PMULWA(LS_1'(p_1), \ldots, LS_K'(p_K)) \geq PMULWA(LS_1(p_1), \ldots, LS_K(p_K))$.

(3) Boundary

$$\begin{aligned}&\min\{LS_i(p_i)|i = 1, 2, \ldots, K\}\\ &\leq PMULWA(LS_1(p_1), \ldots, LS_K(p_K))\\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}.\end{aligned}$$

Proof

(1) Since $LS_i(p_i) = LS(p)$ for all $i = 1, \ldots, K$ then, we have $PMULWA(LS_1(p_1), LS_2(p_2), \ldots, LS_K(p_K)) = (w_1 \odot LS_1(p_1) \oplus w_2 \odot LS_2(p_2) \oplus \ldots w_K \odot \oplus LS_K(p_K)) = (w_1 \odot LS(p) \oplus w_2 \odot LS(p) \oplus w_K \ldots \oplus LS(p))$. Since $\sum_{i=1}^{K} w_i = 1$ thus, $PMULWA(LS_1(p_1), \ldots, LS_K(p_K)) = LS(p)$.

(2) Since $LS_i'(p_i') \geq LS_i(p_i) \Rightarrow w_i \odot LS_i'(p_i') \geq w_i \odot LS_i(p_i)$ for each $i(i = 1, \ldots, K)$, i.e.,

$$\begin{aligned}&w_1 \odot LS_1'(p_1') \geq w_1 \odot LS_1(p_1)\\ &w_2 \odot LS_2'(p_2') \geq w_2 \odot LS_2(p_2)\\ &\vdots\\ &w_K \odot LS_K'(p_K') \geq w_K \odot LS_K(p_K).\end{aligned}$$

Clearly, the inequalities mentioned are well defined and after adding all of K inequalities, we get

$$\begin{aligned}&(w_1 \odot LS_1'(p_1') \oplus w_2 \odot LS_2'(p_2') \oplus \cdots \oplus w_K \odot LS_K'(p_K'))\\ &\geq (w_1 \odot LS_1(p_1) \oplus w_2 \odot LS_2(p_2) \oplus \cdots \oplus w_K \odot LS_K(p_K))\\ &\Rightarrow PMULWA(LS_1'(p_1), \ldots, LS_K'(p_K))\\ &\geq PMULWA(LS_1(p_1), \ldots, LS_K(p_K)).\end{aligned}$$

(3) Since $\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ and $\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq LS_i(p_i) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ for each i. Multiplying by weight vector $w_i(i = 1, \ldots, K)$ throughout the inequality expression we get $w_i \odot \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq w_i \odot LS_i(p_i) \leq w_i \odot \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ for each $i(i = 1, \ldots, K)$, i.e.,

$$\begin{aligned}&w_1 \odot \min\{LS_i(p_i)|i = 1,2,\ldots,K\}\\ &\leq w_1 \odot LS_1(p_1) \leq w_1 \odot \max\{LS_i(p_i)|i = 1,2,\ldots,K\}\\ &w_2 \odot \min\{LS_i(p_i)|i = 1,2,\ldots,K\}\\ &\leq w_2 \odot LS_2(p_2) \leq w_2 \odot \max\{LS_i(p_i)|i = 1,2,\ldots,K\}\\ &\vdots\\ &w_K \odot \min\{LS_i(p_i)|i = 1,2,\ldots,K\}\\ &\leq w_K \odot LS_K(p_K) \leq w_K \odot \max\{LS_i(p_i)|i = 1,2,\ldots,K\}.\end{aligned}$$

Since $\sum_{i=1}^{K} w_i = 1$, therefore adding all the above K inequalities, we get

$$\begin{aligned}&\sum_{i=1}^{K} w_i \odot \min\{LS_i(p_i)|i = 1, 2, \ldots, K\}\\ &\leq w_1 \odot LS_1(p_1) \oplus \ldots \oplus w_K \odot LS_K(p_K)\\ &\leq \sum_{i=1}^{K} w_i \odot \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}\\ &\Rightarrow \min\{LS_i(p_i)|i = 1, 2, \ldots, K\}\\ &\leq PMULWA(LS_1(p_1), \ldots, LS_K(p_K))\\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}.\end{aligned}$$

This proves the theorem. □

Definition 21 Let $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. The probabilistic multiplicative unbalanced linguistic geometric (PMULG) operator is defined as:

$$\begin{aligned}&PMULG(LS_1(p_1), LS_2(p_2), \ldots, LS_K(p_K))\\ &= (LS_1(p_1) \otimes \ldots LS_K(p_K))^{1/K}\\ &= (LS_1(p_1))^{1/K} \otimes \ldots (LS_K(p_K))^{1/K}.\end{aligned}$$

The desired properties of the PMULG operator can be listed as follows.

Theorem 6 *Let $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. Then, for the PMULG operator, we have*

(1) Idempotency
If for all K $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ are equal i.e., $LS_i(p_i) = LS(p)$, then, $PMULG(LS_1(p_1), \ldots, LS_K(p_K)) = LS(p)$.

(2) Monotonicity
Let $\{LS_i'(p_i')|i = 1, 2, \ldots, K\}$ be a collection of K PM-ULTS. If for any $i(i = 1, \ldots, K)$, $LS'(p_i') \geq LS_i(p_i)$ then,

$$PMULG(LS'_1(p_1), \ldots, LS'_K(p_K)) \geq PMULG(LS_1(p_1), \ldots, LS_K(p_K)).$$

(3) Boundary

$$\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq PMULG(LS_1(p_1), \ldots, LS_K(p_K)) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}.$$

Proof

(1) Since $LS_i(p_i) = LS(p)$ for all $i = 1, \ldots, K$ then, we have

$$\begin{aligned} &PMULG(LS_1(p_1), LS_2(p_2), \ldots, LS_K(p_K)) \\ &= (LS_1(p_1))^{1/K} \otimes \ldots (LS_K(p_K))^{1/K} \\ &= (LS(p))^{1/K} \otimes \ldots (LS(p))^{1/K} \\ &= LS(p)^{\underbrace{1/K + \ldots + 1/K}_{(K \text{ times})}} = LS(p). \end{aligned}$$

(2) Since $LS'_i(p'_i) \geq LS_i(p_i) \Rightarrow (LS'_i(p'_i))^{1/K} \geq (LS_i(p_i))^{1/K}$ for each $i(i = 1, \ldots, K)$, i.e.,

$$\begin{aligned} &(LS'_1(p'_1))^{1/K} \geq (LS_1(p_1))^{1/K} \\ &(LS'_2(p'_2))^{1/K} \geq (LS_2(p_2))^{1/K} \\ &\vdots \\ &(LS'_K(p'_K))^{1/K} \geq (LS_K(p_K))^{1/K}. \end{aligned}$$

After rearranging all the above K inequalities, we get

$$\begin{aligned} &(LS'_1(p'_1))^{1/K} \otimes (LS'_2(p'_2))^{1/K} \otimes \ldots \otimes (LS'_K(p'_K))^{1/K} \\ &\geq (LS_1(p_1))^{1/K} \otimes (LS_2(p_2))^{1/K} \otimes \ldots \otimes (LS_K(p_K))^{1/K} \\ &\Rightarrow PMULG(LS'_1(p_1), \ldots, LS'_K(p_K)) \\ &\geq PMULG(LS_1(p_1), \ldots, LS_K(p_K)). \end{aligned}$$

(3) Since $\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ and for each $i(i = 1, \ldots, K)$, we have

$$\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq LS_i(p_i) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\} \Rightarrow (\min\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{1/K} \leq (LS_i(p_i))^{1/K} \leq (\max\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{1/K}$$ i.e.,

$$\begin{aligned} &(\min\{LS_i(p_i)|i=1,2,\ldots,K\})^{1/K} \\ &\leq (LS_1(p_1))^{1/K} \leq (\max\{LS_i(p_i)|i=1,2,\ldots,K\})^{1/K} \\ &(\min\{LS_i(p_i)|i=1,2,\ldots,K\})^{1/K} \\ &\leq (LS_2(p_2))^{1/K} \leq (\max\{LS_i(p_i)|i=1,2,\ldots,K\})^{1/K} \\ &\vdots \\ &(\min\{LS_i(p_i)|i=1,2,\ldots,K\})^{1/K} \\ &\leq (LS_K(p_K))^{1/K} \leq (\max\{LS_i(p_i)|i=1,2,\ldots,K\})^{1/K}. \end{aligned}$$

After rearranging all the terms in the above K inequalities, we get

$$\begin{aligned} &(\min\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{\underbrace{1/K + \ldots + 1/K}} \\ &\leq (LS_1(p_1))^{1/K} \otimes \ldots \otimes (LS_K(p_K))^{1/K} \\ &\leq (\max\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{\underbrace{1/K + \ldots + 1/K}} \\ &\Rightarrow \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\leq (LS_1(p_1))^{1/K} \otimes \ldots \otimes (LS_K(p_K))^{1/K} \\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\Rightarrow \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\leq PMULG(LS_1(p_1), \ldots, LS_K(p_K)) \\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}. \end{aligned}$$

This proves the theorem. □

Definition 22 Let $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. Assign a relevant weight vector $W = (w_1, w_2, \ldots, w_K)$ such that $w_i \in [0, 1]$ and $\sum_{i=1}^{K} w_i = 1$. The probabilistic multiplicative unbalanced linguistic weighted geometric (PMULWG) operator is defined as:

$$\begin{aligned} &PMULWG(LS_1(p_1), LS_2(p_2), \ldots, LS_K(p_K)) \\ &= (LS_1(p_1))^{w_1} \otimes \ldots (LS_K(p_K))^{w_K}. \end{aligned}$$

In particular, on setting $w = (1/K, \ldots, 1/K)^T$, then the proposed PMULWG operator reduces to PMULG operator. Next, we will outline the desired properties of the PMULWG operator in the following fashion.

Theorem 7 *Let $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ be the cluster of K normalized PM-ULTSs. Let $W = (w_1, w_2, \ldots, w_K)$ be the associated weight vector such that $w_i \in [0, 1]$ and $\sum_{i=1}^{K} w_i = 1$. Then, for the PMULWG operator, we have*

(1) Idempotency

If for all K $\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ are equal

i.e., $LS_i(p_i) = LS(p)$, then, $PMULWG(LS_1(p_1), \ldots, LS_K(p_K)) = LS(p)$.

(2) Monotonicity

Let $\{LS_i'(p_i')|i = 1, 2, \ldots, K\}$ be a collection of K PM-ULTS. If for any $i(i = 1, \ldots, K)$, $LS_i'(p_i') \geq LS_i(p_i)$ then,

$$PMULWG(LS_1'(p_1), \ldots, LS_K'(p_K)) \geq PMULWG(LS_1(p_1), \ldots, LS_K(p_K)).$$

(3) Boundary

$$\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq PMULWG(LS_1(p_1), \ldots, LS_K(p_K)) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}.$$

Proof

(1) Since $LS_i(p_i) = LS(p)$ for any $i = 1, \ldots, K$ then, we have

$$\begin{aligned} &PMULWG(LS_1(p_1), LS_2(p_2), \ldots, LS_K(p_K)) \\ &= (LS_1(p_1))^{w_1} \otimes \ldots (LS_K(p_K))^{w_K} \\ &= (LS(p))^{w_1} \otimes \ldots (LS(p))^{w_K} \\ &= LS(p)^{\sum_{i=1}^{K} w_i} = LS(p). \end{aligned}$$

(2) Since $LS_i'(p_i') \geq LS_i(p_i) \Rightarrow (LS_i'(p_i'))^{w_i} \geq (LS_i(p_i))^{w_i}$ for each $i(i = 1, \ldots, K)$, i.e.,

$$\begin{aligned} &(LS_1'(p_1'))^{w_1} \geq (LS_1(p_1))^{w_1} \\ &(LS_2'(p_2'))^{w_2} \geq (LS_2(p_2))^{w_2} \\ &\vdots \\ &(LS_K'(p_K'))^{w_K} \geq (LS_K(p_K))^{w_K}. \end{aligned}$$

After rearranging all the above K inequalities, we get

$$\begin{aligned} &(LS_1'(p_1'))^{w_1} \otimes (LS_2'(p_2'))^{w_2} \otimes \ldots \otimes (LS_K'(p_K'))^{w_K} \\ &\geq (LS_1(p_1))^{w_1} \otimes (LS_2(p_2))^{w_2} \otimes \ldots \otimes (LS_K(p_K))^{w_K} \\ &\Rightarrow PMULWG(LS_1'(p_1), \ldots, LS_K'(p_K)) \\ &\geq PMULWG(LS_1(p_1), \ldots, LS_K(p_K)). \end{aligned}$$

(3) Since $\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}$ and for each $i(i = 1, \ldots, K)$, we have $\min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \leq LS_i(p_i) \leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\} \Rightarrow (\min\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_i} \leq (LS_i(p_i))^{w_i} \leq (\max\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_i}$ i.e.,

$$\begin{aligned} &(\min\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_1} \leq (LS_1(p_1))^{w_1} \leq (\max\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_1} \\ &(\min\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_2} \leq (LS_2(p_2))^{w_2} \leq (\max\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_2} \\ &\vdots \\ &(\min\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_K} \leq (LS_K(p_K))^{w_K} \leq (\max\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{w_K}. \end{aligned}$$

After rearranging all the terms in the above K inequalities, we get

$$\begin{aligned} &(\min\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{\sum_{i=1}^{K} w_i} \\ &\leq (LS_1(p_1))^{w_1} \otimes \ldots \otimes (LS_K(p_K))^{w_K} \\ &\leq (\max\{LS_i(p_i)|i = 1, 2, \ldots, K\})^{\sum_{i=1}^{K} w_i} \\ &\Rightarrow \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\leq (LS_1(p_1))^{w_1} \otimes \ldots \otimes (LS_K(p_K))^{w_K} \\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\Rightarrow \min\{LS_i(p_i)|i = 1, 2, \ldots, K\} \\ &\leq PMULWG(LS_1(p_1), \ldots, LS_K(p_K)) \\ &\leq \max\{LS_i(p_i)|i = 1, 2, \ldots, K\}. \end{aligned}$$

This proves the theorem. □

Remark 5 It is observed that on comparing the average arithmetical operator and the geometric mean operator, the value of the geometric mean is always lower than the arithmetic mean because of the compounding effect. However, the geometric mean is more accurate than the arithmetic mean for the positively skewed data. It is influenced by extreme values and is often reported as a consistent function to aggregate the rate of increase/decrease. Therefore, if the input probabilistic unbalanced data represents some growth rate, then the proposed PMULWG and PMULG operators are used. Otherwise, the PMULA and PMULWA operators may be the better choice, especially for computing the average value of the input information.

Considering the flexibility and interpretability of PLTS to handle complex uncertainties, numerous studies in the literature have been devoted to its application and achievements. Recently, Mi et al. [42] introduced a two-person zero-sum matrix game model taking probabilistic linguistic information as input. However, real-life problems are not always subjected to uniformly distributed LTSs. Henceforth, the subsequent section is explicitly dedicated to introducing a constant-sum matrix game problem with probabilistic unbalanced linguistic information. It is envisioned that the probabilistic multiplicative unbalanced linguistic information has not been established

in the current matrix game models to our optimum knowledge.

5 Matrix games with probabilistic multiplicative unbalanced linguistic information

This section proposes the notion of a constant-sum matrix game problem with probabilistic unbalanced linguistic information. Section 5.1 is dedicated to the terminology related to the constant-sum probabilistic unbalanced linguistic matrix game. Then, the probabilistic unbalanced linguistic linear programming model is developed to acquire the optimal strategy set of corresponding players and the probabilistic unbalanced linguistic game value in Section 5.2.

5.1 Terminology of a constant-sum probabilistic multiplicative unbalanced linguistic information

Due to the perplexity and uncertainty present within the practical decision-making problems where DMs prefer to present linguistic labels distributed non-uniformly, the probabilistic unbalanced linguistic information is introduced to model approximate payoff values. We start with the following definition.

Definition 23 A two-person constant-sum unbalanced linguistic game $\widetilde{G}$ is described as a quadruple $(S_1, S_2, LS, \widetilde{R})$, where S_1 and S_2 represents the strategy set of the player I (PI) and player II (PII), $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, m\}$ having $n + m + 1$ cardinality is a presupposed ULTS for both the PI and PII, $\widetilde{R} = [LS_{cd}(p)]_{r\times q}$ is the payoff matrix for PI with probabilistic unbalanced linguistic information denoted as:

$$\widetilde{R} = \begin{pmatrix} LS_{11}(p) & LS_{12}(p) & \ldots & LS_{1q}(p) \\ LS_{21}(p) & LS_{22}(p) & \ldots & LS_{2q}(p) \\ \vdots & \vdots & \ddots & \vdots \\ LS_{r1}(p) & LS_{r2}(p) & \ldots & LS_{rq}(p) \end{pmatrix}$$

where $LS_{cd}(p) = \left\{(t_{a^i})_{cd}(p^{(k)}) | k = 1, \ldots, \#LS(p)\right\}$, $\#LS_{11}(p) = \#LS_{12}(p) = \ldots = \#LS_{rq}(p) = \#LS(p)$. On the contrary $\text{neg}(\widetilde{R})$ is the payoff matrix for PII.

It is worth observing that the probabilistic unbalanced linguistic payoff matrix $\widetilde{R} = [LS_{cd}(p)]_{r\times q}$ for PI is defined with the intention that each $LS_{cd}(p)$ represents response to strategy c of PI, $c = 1, \ldots, r$ when PII selects strategy d, $d = 1, \ldots, q$, to play the game. Consequently, the corresponding payoff of PI is $LS_{cd}(p)$ while the payoff of PII is $\text{neg}(LS_{cd}(p))$.

A probabilistic unbalanced linguistic payoff implies the importance of a particular strategy played by the game player that can be easily expressed on a multiplicative unbalanced linguistic scale set, i.e., LS.

Since the ranking of PM-ULTSs is possible and is defined in the previous section 4.1. Henceforth, the idea of the classical value of the game can be extended to the probabilistic multiplicative unbalanced linguistic matrix game $\widetilde{G}$. Furthermore, PI and PII are examined from the prospects of the gain floor and the lost ceiling, respectively. Finally, the goal of players is esteemed as follows:

Definition 24 Let $\widehat{v}$ and $\widehat{\omega}$ are the values corresponding to PI and PII, respectively such that:

Table 2 The comprehensive evaluation payoff matrix $\widetilde{R}$ in form of PM-ULTSs

	C_1	C_2	C_3
R_1	$\{t_{2^0}(1), t_{2^0}(0), t_{2^0}(0)\}$	$\{t_2(0.6), t_{2^2}(0.4), t_{2^1}(0)\}$	$\{t_{2^2}(0.3), t_{2^3}(0.7), t_{2^2}(0)\}$
R_2	$\{t_{1/8}(0.8), t_{1/4}(0.2), t_{1/4}(0)\}$	$\{t_{1/4}(0.6), t_{1/2}(0.4), t_{1/4}(0)\}$	$\{t_{2^0}(1), t_{2^0}(0), t_{2^0}(0)\}$
R_3	$\{t_{1/8}(0.8), t_{1/4}(0.2), t_{1/4}(0)\}$	$\{t_{2^2}(0.3), t_{2^3}(0.7), t_{2^2}(0)\}$	$\{t_{2^2}(0.3), t_{2^3}(0.7), t_{2^2}(0)\}$

Table 3 The comprehensive evaluation payoff matrix $\text{neg}(\widetilde{R})$ in form of PM-ULTSs

	C_1	C_2	C_3
R_1	$\{t_{2^0}(1), t_{2^0}(0), t_{2^0}(0)\}$	$\{t_{1/2}(0.6), t_{1/4}(0.4), t_{1/2}(0)\}$	$\{t_{1/4}(0.3), t_{1/8}(0.7), t_{1/4}(0)\}$
R_2	$\{t_8(0.8), t_4(0.2), t_4(0)\}$	$\{t_4(0.6), t_2(0.4), t_4(0)\}$	$\{t_{2^0}(1), t_{2^0}(0), t_{2^0}(0)\}$
R_3	$\{t_8(0.8), t_4(0.2), t_4(0)\}$	$\{t_{1/4}(0.3), t_{1/8}(0.7), t_{1/4}(0)\}$	$\{t_{1/4}(0.3), t_{1/8}(0.7), t_{1/4}(0)\}$

$$\widehat{v}^{1-}(p) = \max_{c=1,\ldots,r} \min_{d=1,\ldots,q} LS_{cd}^{(k)}(p^{(k)});$$
$$\widehat{\omega}^{2-}(p) = \max_{d=1,\ldots,q} \min_{c=1,\ldots,r} \mathrm{neg}(LS_{cd}^{(k)}(p^{(k)})).$$

Note 2 From Definition 24, it is noted that $\widehat{v}(p) = \widehat{v}^{1-}(p)$ and $\widehat{\omega}(p) = \mathrm{neg}(\widehat{\omega}^{2-}(p))$ refers to the lower and upper value of the game $\widetilde{G}$, respectively. The value of the game is defined if $\widehat{v}(p) = \widehat{\omega}(p)$. Meanwhile, the strategies c^* and d^* offer the payoff value $LS_{c^*d^*}(p)$ such that the equality condition holds are optimal pure strategies for PI and PII. Also, the ordered strategy pair (c^*, d^*) is ascertained as the saddle point of the game $\widetilde{G}$.

Note 3 Obviously from Remark 1, it is clear that

$$\begin{aligned}\widehat{\omega}(p) &= \mathrm{neg}(\widehat{\omega}^{2-}(p))\\ &= \mathrm{neg}(\max_{d=1,\ldots,q} \min_{c=1,\ldots,r} \mathrm{neg}(LS_{cd}^{(k)}(p^{(k)})))\\ &= \min_{d=1,\ldots,q} \max_{c=1,\ldots,r} LS_{cd}^{(k)}(p^{(k)}).\end{aligned}$$

Since the above analysis holds only for the case $n = m$, we consider the following example having predefined ULTS for $n = m$ to understand the above concept of the matrix game.

Example 4 Let $LS = \{t_{1/8}, t_{1/4}, t_{1/2}, t_{2^0}, t_2, t_{2^2}, t_{2^3}\}$ be the presumed ULTS. Suppose for PI the probabilistic unbalanced linguistic payoff matrix $\widetilde{R}$ having all the PM-ULTSs as payoff entries is stated in the Table 2 and additionally, the payoff matrix for PII, i.e., $\mathrm{neg}(\widetilde{R})$ is given in Table 3.

From the given payoff matrices, it is anticipated that each payoff entry provided in the form of PM-ULTSs is normalized with equal cardinalities. Hence, we can proceed further. Now, based on the ranking algorithm and Definition 16 given in the previous Sect. 4.1, one can rank the PM-ULTSs $LS_{cd}(p)$ as well as $\mathrm{neg}(LS_{cd}(p))$.

Therefore, the lower and upper values of the game are given as follows:

For gain floor: PI

$$\begin{aligned}\widehat{v}^{1-}(p) &= \widehat{v}(p)\\ &= \max_{c=1,2,3} \min_{d=1,2,3,4} LS_{cd}(p_{cd})\\ &= \{t_{2^0}(1), t_{2^0}(0), t_{2^0}(0)\}\end{aligned}$$

For loss ceiling: PII

$$\begin{aligned}\widehat{\omega}^{2-}(p) &= \max_{d=1,2,3,4} \min_{c=1,2,3} \mathrm{neg}(LS_{cd}(p_{cd}))\\ &= \{t_{2^0}(1), t_{2^0}(0), t_{2^0}(0)\}\end{aligned}$$

Hence, in this instance $(1, 1)$ is the saddle point and probabilistic unbalanced linguistic value of the game is $\widehat{v}(p) = \widehat{\omega}(p) = \{t_{2^0}(1), t_{2^0}(0), t_{2^0}(0)\}$.

However, it is not always plausible for a probabilistic unbalanced linguistic matrix game to have equality, i.e., $\widehat{v}(p) = \widehat{\omega}(p)$. Generally, the following inequality exists for any probabilistic unbalanced linguistic matrix game.

Theorem 8 *Suppose $\widehat{v}(p)$ and $\widehat{\omega}(p)$ be the probabilistic unbalanced linguistic lower and upper values of probabilistic unbalanced linguistic matrix game $\widetilde{G}$ such that both values exist. Then, $\widehat{v}(p) \le \widehat{\omega}(p)$.*

Proof Without loss of generality, let $\widehat{v}(p)$ and $\widehat{\omega}(p)$ both exist, so that for some column d and fixed row c, we have,

$$\min_{d=1,\ldots,q} LS_{cd}(p) \le LS_{cd}(p),$$

Now, take max over $c = 1, \ldots, r$ on both sides, such that,

$$\begin{aligned}&\widehat{v}(p) \equiv \max_{c=1,\ldots,r} \min_{d=1,\ldots,q} LS_{cd}(p) \le \max_{c=1,\ldots,r} LS_{cd}(p)\\ &\Rightarrow \widehat{v}(p) \le \max_{c=1,\ldots,r} LS_{cd}(p).\end{aligned}$$

Seeing that the above inequality claims for any d. Hence, the subsequent inequality is obtained.

$$\widehat{v}(p) \le \min_{d=1,\ldots,q} \max_{c=1,\ldots,r} LS_{cd}(p)$$

Hence, $\widehat{v}(p) \le \widehat{\omega}(p)$. This proves the theorem. □

Considering the crisp setup of a 2-person zero-sum matrix game, the expected payoff is described as the statistical expectation of information present within the payoff matrix. However, the probabilistic unbalanced linguistic matrix game has ambiguity and uncertainty within the assessment of the payoff entries such that the veracity of the expected value is incomprehensible. Thus, it is necessary to give the following definition explaining the probabilistic unbalanced linguistic expected payoffs of corresponding players.

Definition 25 Consider $x^{(k)} = (x_1^{(k)}, \ldots, x_r^{(k)}) \in S_1$ and $y^{(k)} = (y_1^{(k)}, \ldots, y_q^{(k)}) \in S_2$ as a mixed strategy pair associated with the degree of importance for corresponding PI

Table 4 Unbalanced linguistic payoff matrix for the FY2016-2017

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	Avg	G	PR	Avg	G	VG
A_2	W	VG	Avg	Avg	W	VW
A_3	W	W	G	G	VG	Avg
A_4	PR	G	PR	W	W	W

Table 5 Unbalanced linguistic payoff matrix for the FY2017-2018

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	Avg	Avg	VG	Avg	G	G
A_2	G	G	W	G	W	VW
A_3	VW	G	Avg	Avg	VG	G
A_4	VG	VG	PR	Avg	W	Avg

and PII, respectively. Then, PI's probabilistic unbalanced linguistic expectation payoff is defined by

$$\widetilde{E}_{\widetilde{R}}(x^{(k)}, y^{(k)}) = \bigoplus_{c=1}^{r}\left(x_c^{(k)}\left(\bigoplus_{d=1}^{q} LS_{cd}(p) \odot y_d^{(k)}\right)\right).$$

On the other hand PII's probabilistic unbalanced linguistic expectation payoff is $\mathrm{neg}(\widetilde{E}_{\widetilde{R}}(x^{(k)}, y^{(k)}))$.

5.2 Methodology for probabilistic multiplicative unbalanced linguistic matrix game

This subsection discusses the probabilistic multiplicative unbalanced linguistic linear programming (PMULLP) method for matrix game problems having probabilistic multiplicative unbalanced linguistic information.

Based on the presumed multiplicative ULTS, $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, n\}$, the probabilistic unbalanced linguistic payoff matrix is viewed as:

$$\widetilde{R} = \left(LS_{cd}(p)\right)_{r\times q}.$$

Let $\widehat{v}_{(p)}$ refers to the minimal probabilistic unbalanced linguistic payoff value of PI and let $x^{(k)} = (x_1^{(k)}, \ldots, x_r^{(k)}) \in S_1$ denote its mixed strategy set associated with certain degree of importance. Now, based on the goal that PI is having, the unbalanced linguistic expectation of PI for the dth strategy of PII is presented as the weighted average of the probabilistic multiplicative unbalanced linguistic variables $LS_{1d}(p), LS_{2d}, \ldots, LS_{rd}$, i.e., $LS_{1d}(p)x_1^{(k)} \oplus LS_{2d}(p)x_2^{(k)} \oplus \cdots \oplus LS_{rd}(p)x_r^{(k)}$. Similarly, $\widehat{\omega}_{(p)}$ can be viewed as the maximal probabilistic unbalanced linguistic payoff value of the PII, $y^{(k)} = (y_1^{(k)}, \ldots, y_q^{(k)}) \in S_2$ be its mixed strategy set, and the unbalanced linguistic expectation of PII against PI's cth strategy is presented as $LS_{c1}(p)y_1^{(k)} \oplus LS_{c2}(p)y_2^{(k)} \oplus \cdots \oplus LS_{cq}(p)y_q^{(k)}$.

Hence, we have the required PMULLP problem for PI:

$$\text{subject to}\begin{cases} \max \quad \widehat{v}(p) \\ \bigoplus_{c=1}^{r} LS_{cd}(p)x_c \geq \widehat{v}(p),\ d = 1, \ldots, q \\ \sum_{c=1}^{r} x_c = 1;\ x_c \geq 0,\ c = 1, \ldots, r. \end{cases} \quad \text{(PMULLP1)}$$

where the probabilistic unbalanced linguistic objective value can be outlined as $\widehat{v}(p) = \left\{\widetilde{v}^{(k)}(p^{(k)}) | k = 1, \ldots, \#LS(p)\right\}$, $\widetilde{v} \in LS$, and, $\sum_{k=1}^{\#LS(p)} p^{(k)} = 1$.

Next, for PII, we have the following system:

$$\text{subject to}\begin{cases} \min \quad \widehat{\omega}(p) \\ \bigoplus_{d=1}^{q} LS_{cd}(p)y_d \leq \widehat{\omega}(p),\ c = 1, \ldots, r \\ \sum_{d=1}^{q} y_d = 1;\ y_d \geq 0,\ d = 1, \ldots, q. \end{cases} \quad \text{(PMULLP2)}$$

where the probabilistic unbalanced linguistic objective value can be outlined as $\widehat{\omega}(p) = \left\{\widetilde{\omega}^{(k)}(p^{(k)}) | k = 1, \ldots, \#LS(p)\right\}$, $\widetilde{\omega} \in LS$, and, $\sum_{k=1}^{\#LS(p)} p^{(k)} = 1$.

After setting up the mathematical models of a probabilistic multiplicative unbalanced linguistic matrix game, namely PMULLP1 and PMULLP2, it is necessary to describe a unified mechanism to solve these models. Before we define the algorithm's main steps, we will give the following definition required in the procedure.

Definition 26 Let $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, n\}$ be the presumed ULTSs and $LS(p) = \left\{t_{a^i}^{(k)}(p^{(k)}) | k = 1, \ldots, \#LS(p)\right\}$ be the PM-ULTSs. Then, the information measure of the PM-ULTSs is defined as a mapping:

$IM : LS(p) \rightarrow \mathbb{R}$

such that $IM(t_{a^i}(p)) = p \cdot a^i$.

Table 6 Unbalanced linguistic payoff matrix for the FY2018-2019

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	G	VG	VG	G	VG	PR
A_2	Avg	Avg	G	Avg	VW	W
A_3	W	W	G	Avg	G	-
A_4	G	VG	VG	G	Avg	G

Table 7 Unbalanced linguistic payoff matrix for the FY2019-2020

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	VG	G	G	VG	PR	PR
A_2	W	-	Avg	VG	Avg	Avg
A_3	VG	Avg	VG	VG	PR	VG
A_4	G	PR	G	Avg	G	Avg

For the sake of convenience, while solving the probabilistic unbalanced linguistic game problem, we rewrite $p \cdot a^i$ as $p \cdot \Omega^{-1}(t_{a^i})$. The definition of Ω^{-1} is the same as given in the paper [43].

We take into consideration certain assumptions before we conduct the necessary steps of the newly designed method, i.e.,

- Firstly, we need to normalize the payoff matrix so that each payoff entry has equal cardinality.
- Secondly, the probability of the unbalanced linguistic terms within the set $LS(p)$ should be nonzero.
- Finally, the cardinality of the probabilistic unbalanced linguistic game value of both players should be the same as that of each payoff entry in the payoff matrix.

Now we provide the main steps involved in the working process of solving the probabilistic unbalanced linguistic matrix game problem:

Step 1: Analyze the original two-player constant sum probabilistic unbalanced linguistic matrix game problem and determine the mixed strategy set $x = (x_1^{(k)}, x_2^{(k)}, \ldots, x_r^{(k)}) \in S_1$ and $y = (y_1^{(k)}, y_2^{(k)}, \ldots, y_q^{(k)})$ of the players I and II, respectively. Based on the ULTS $LS = \{t_{a^i} | i = -n, \ldots, 0, \ldots, n\}$, PI and PII express their multiplicative unbalanced linguistic evaluation estimates over the strategies $x_c^{(k)} (c = 1, \ldots, r)$ with respect to the strategies $y_d^{(k)} (d = 1, \ldots, q)$. All the different multiplicative unbalanced linguistic evaluation estimates $LS_{cd}^{(k)} (k = 1, \ldots, \#LS(p))$ having the corresponding probability $p^{(k)} (k = 1, \ldots, \#LS(p))$ are represented as PM-ULTS $LS_{cd}(p) = \left\{ (t_{a^i})_{cd}^{(k)} (p^{(k)}) | k = 1, \ldots, \#LS_{cd}(p) \right\}$, where $p^{(k)} > 0$, $k = 1, \ldots, \#LS(p)$, $\#LS_{11}(p) = \#LS_{12}(p) = \ldots = \#LS_{rq}(p) = \#LS(p)$, and $\sum_{k=1}^{\#LS(p)} p^{(k)} = 1$. All the PM-ULTSs $LS_{cd}(p) (c = 1, \ldots, r;\ d = 1, \ldots, q)$ are recorded within the probabilistic multiplicative unbalanced linguistic payoff matrix $\widetilde{R} = [LS_{cd}(p)]_{r \times q}$.**Step 2:**Formulate PMULLP1 and PMULLP2 for PI and PII, respectively.**Step 3:**For PI, due to the fact that there exist ordering relation of probabilistic multiplicative unbalanced linguistic term, therefore in PMULLP1, $\max\{\widehat{v}(p)\}$ is easily written as $\max\left\{ \widetilde{v}^{(k)}(p^{(k)}) | k = 1, \ldots, \#LS(p) \right\}$, which is considered as a multi-objective linguistic linear optimization problem (MOLLOP). Similarly, for PII we have $\min\{\widetilde{\omega}(p)\}$ which is also considered as a MOLLOP $\min\left\{ \widetilde{\omega}^{(k)}(p^{(k)}) | k = 1, \ldots, \#LS(p) \right\}$ where $\widetilde{v}^{(k)}, \widetilde{\omega}^{(k)} \in LS$.**Step 4:**Now decompose the problem PMULLP1 as follows:

For each $k = 1, \ldots, \#LS(p)$, we have

$$\max \quad \widetilde{v}^{(k)}(p^{(k)})$$
$$\text{subject to} \begin{cases} \bigoplus_{c=1}^{r} (t_{a^i})_{cd}^{(k)} (p^{(k)}) x_c^{(k)} \geq \widetilde{v}^{(k)}(p^{(k)}), d = 1, \ldots, q \\ \sum_{c=1}^{r} x_c^{(k)} = 1; x_c \geq 0, c = 1, \ldots, r. \end{cases} \quad \text{(PMULLP1.1)}$$

Similarly, for PII, we have

For each $k = 1, \ldots, \#LS(p)$

$$\min \quad \widetilde{\omega}^{(k)}(p^{(k)})$$
$$\text{subject to} \begin{cases} \bigoplus_{d=1}^{q} (t_{a^i})_{cd}^{(k)} (p^{(k)}) y_d^{(k)} \leq \widetilde{\omega}^{(k)}(p)^{(k)}, c = 1, \ldots, r \\ \sum_{d=1}^{q} y_d^{(k)} = 1; y_d \geq 0, d = 1, \ldots, q. \end{cases} \quad \text{(PMULLP2.1)}$$

Step 5: Since Ω^{-1} is determined as a monotonically increasing transformation function, henceforth, on the application of Ω^{-1} on both the sides of the constraint of PMULLP1.1 we obtain the following system of linear optimization problem:

$$\begin{cases} \sum_{c=1}^{r} \Omega^{-1}((t_{a^i})_{cd}^{(k)}) \cdot p^{(k)} x_c^{(k)} \geq \Omega^{-1}(\widetilde{v}^{(k)}) \cdot p^{(k)}, d = 1, \ldots, q \\ \sum_{c=1}^{r} x_c^{(k)} = 1; x_c \geq 0, c = 1, \ldots, r. \end{cases}$$

Now, following from the Definition 26, we can rewrite the term $\Omega^{-1}((t_{a^i})_{cd}^{(k)}) \cdot p^{(k)}$ as $(a^i)_{cd}^{(k)} p^{(k)} (k = 1, \ldots, \#LS(p))$. Next, by Assuming that $\Omega^{-1}(\tilde{v}^{(k)}) \cdot p^{(k)} = v^{(k)} p^{(k)} > 0$, we make the following substitution as $X_c^{(k)} = \frac{x_c^{(k)}}{v^{(k)} p^{(k)}} (k = 1, \ldots, \#LS(p)), (c = 1, \ldots, r)$ and $V^{(k)} = \frac{1}{v^{(k)} p^{(k)}}$. Therefore, the transformed system of the linear optimization problem is given as follows:

$$\min \quad V^{(k)} = X_1^{(k)} + \ldots + X_r^{(k)}$$
$$\text{subject to} \begin{cases} \sum_{c=1}^{r} (a^i)_{cd}^{(k)} p^{(k)} X_c^{(k)} \geq 1, \ d = 1, \ldots, q \\ X_c^{(k)} \geq 0, \ c = 1, \ldots, r. \end{cases} \quad \text{(PMULLP1.2)}$$

Step 6: We repeat the above step for PII and the problem can be constructed in an analogous fashion

$$\begin{cases} \sum_{d=1}^{q} \Omega^{-1}((t_{a^i})_{cd}^{(k)}) \cdot (p^{(k)}) y_d^{(k)} \le \Omega(\tilde{\omega}^{(k)}) \cdot (p)^{(k)}, \ c = 1, \ldots, r \\ \sum_{d=1}^{q} y_d^{(k)} = 1; \ y_d \ge 0, \ d = 1, \ldots, q. \end{cases}$$

Recall that, $\Omega^{-1}((t_{a^i})_{cd}^{(k)}) \cdot p^{(k)}$ as $(a^i)_{cd}^{(k)} p^{(k)} (k = 1, \ldots, \#LS(p))$. Also, by setting $Y_d^{(k)} = \frac{y_d^{(k)}}{\omega^{(k)} p^{(k)}} (k = 1, \ldots, \#LS(p)), \quad (d = 1, \ldots, q)$ and $V^{(k)} = \frac{1}{\omega^{(k)} p^{(k)}}$ since $\Omega^{-1}(\tilde{\omega}^{(k)}) \cdot p^{(k)} = \omega^{(k)} p^{(k)} > 0$.

Therefore, the transformed system of linear optimization problem is given as follows:

$$\max \quad V^{(k)} = Y_1^{(k)} + \ldots + Y_q^{(k)}$$
$$\text{subject to} \begin{cases} \sum_{d=1}^{q} (a^i)_{cd}^{(k)} p^{(k)} Y_d^{(k)} \le 1, \ c = 1, \ldots, r \\ Y_d^{(k)} \ge 0, \ d = 1, \ldots, q. \end{cases} \quad \text{(PMULLP2.2)}$$

Step 7: After solving the mathematical models PMULLP1.2 and PMULLP2.2, we can obtain the optimal strategies $(x_1^{(k)*}, \ldots, x_r^{(k)*})$ and $(y_1^{(k)*}, \ldots, y_q^{(k)*})$ and the value of the game $p^{(k)} v^{(k)} = p^{(k)} \omega^{(k)} = \frac{1}{V^{(k)*}}$.

Step 8: After representing probabilistic unbalanced linguistic information by information measure function and obtaining the corresponding efficient game value mentioned in the previous step, we can then apply the retranslation process to retransform the optimal values into probabilistic unbalanced linguistic information. The retranslation process can be performed in the following two steps:

(1) For the value $p^{(k)} v^{(k)*} = p^{(k)} \omega^{(k)*} = \frac{1}{V^{(k)*}}$, if $\frac{1}{V^{(k)*}} \in \{a^i | i = -n, \ldots, 0, \ldots, n\}$ then, $\tilde{v}^{(k)} = \tilde{\omega}^{(k)} = \Omega(\frac{1}{V^{(k)*}})$ $(k = 1, \ldots, \#LS(p))$. If $\frac{1}{V^{(k)*}} \notin \{a^i | i = -n, \ldots, 0, \ldots, n\}$ then either $\frac{1}{V^{(k)*}} \in [a^{-n}, a^n]$ or otherwise $\frac{1}{V^{(k)*}} = \frac{1}{V^{(k)*}} + a^{-n}$. Finally, $\tilde{v}^{(k)} p^{(k)} = \tilde{\omega}^{(k)} p^{(k)} = \Omega(\frac{1}{V^{(k)*}})$.

It is noted that if $\gamma = 0 \Rightarrow \Omega(\frac{1}{V^{(k)*}}) = t_\lambda^{(k)}$ and for $\gamma \neq 0$ then Ω has its usual meaning and (t_λ, γ).

(2) Finally, the optimal mixed strategy for PI and PII is represented as $x^* = (x_1^*, \ldots, x_r^*)$ where $x_c^* = \left\{ x_c^{(k)} | k = 1, \ldots, \#LS(p) \right\}, (c = 1, \ldots, r)$, $y^* = (y_1^*, \ldots, x_q^*)$ where $y_d^* = \left\{ y_d^{(k)} | k = 1, \ldots, \#LS(p) \right\}, (d = 1, \ldots, q)$, respectively. The probabilistic unbalanced linguistic game value can be represented as $\widehat{v}(p) = \left\{ \widetilde{v}^{(k)}(\theta) | k = 1, \ldots, \#LS(p) \right\}$ and $\widehat{\omega}(p) = \left\{ \widetilde{\omega}^{(k)}(\theta) | k = 1, \ldots, \#LS(p) \right\}$ where

$$\theta^{(k)} = \begin{cases} 1, & \text{if } \gamma = 0; \\ \gamma + 0.5, & \text{otherwise.} \end{cases}$$

It is noteworthy that in the proposed linear programming models (PMULLP1.2) and (PMULLP2.2), the word linear stands for representing the relationship between distinct variables of degree one. In contrast, the word programming represents the technique of choosing the best strategies from the given mixed strategy set in an optimum manner.

Remark 6 The chief characteristics of the models (PMULLP1.2) and (PMULLP2.2) can be briefly stated. Here, the objective functions are well-defined and given quantitatively. All constraints and objective functions follow a linear relationship. The values of the computed variables are non-negative, and a finite number of input and output variables must exist to have an optimal solution for the probabilistic unbalanced linguistic game problem. Moreover, it is observed that both the linear programming models are primal-dual linear optimization models in the crisp setup.

6 A numerical illustration and comparison analyses

This section illustrates the practicality of PM-ULTS and highlight the usefulness and applicability of the two-player constant-sum game under probabilistic multiplicative unbalanced linguistic circumstance with no saddle point by considering a practical companies selection problem to invest and further manifest the strength of the proposed method through a comparative analysis.

6.1 Decision problem description

Investing is a fundamental mode of increasing wealth over time. Nevertheless, investing withholds the risk of dropping money, especially while selecting the companies in which one may wish to invest. To become a successful investor and avoid costly mistakes, it is significant to perform thorough research to understand the fundamentals of the companies before investing. Therefore, in this study, we consider a decision-making problem involving two

competitive players: the first is an investor who aims to invest his money by choosing stocks after carefully analyzing a company's fundamentals, and the second is Nature. The idea of presenting a game model with Nature originated from the portfolio choice problem given in the paper [55]. However, in this paper, it is presumed that the DM is reactionary, expecting that Nature is contrary to compete against him to minimize his payoffs.

The underground working conditions underlying companies' selection process involve several perils and uncertainties. Consequently, experts often have a hesitant perception of evaluation information. This situation is considered appropriate for employing qualitative tools instead numeric values to direct assessments. In this context, the proposed probabilistic multiplicative unbalanced game model described in the previous section can be utilized to address the real-life problem of companies' selection to invest in.

Based on preliminary screening, an investor has shortlisted four companies as an alternative set, i.e., $A_c\,(c = 1, 2, 3, 4)$ where he wants to invest. Since investing decision is tough and is not always easy for a person to invest his hard-earned money without conducting research. In order to compete among these shortlisted companies to choose the most stable of them and rank them from the viewpoint of their significance degree, it is crucial for an investor to perform the evaluation and selection operation primarily based on various essential criteria. Nevertheless, in this case study, we consider the following six criteria based on the detailed investigation: the performance of company (C_1); the market value of company (C_2); the efficiency level of the company (C_3); the business model of the company (C_4); Employee satisfaction level (C_5); shareholders funds (C_6). Since an investor is careful about strategic choice of Nature, so he will select his mixed strategy set as $x = (x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, x_4^{(k)}),\ x_c \geq 0,\ c = 1, 2, 3, 4,\ \sum_{c=1}^{4} x_c^{(k)} = 1$ for each $(k = 1, \ldots, \#LS)$ over the alternative set $\{A_c | c = 1, 2, 3, 4\}$. On the contrary, the investor view nature as a non-cooperative player, thus to counter the choice of his mixed strategy, nature will choose the mixed strategy set as $y = (y_1^{(k)}, y_2^{(k)}, y_3^{(k)}, y_4^{(k)}, y_5^{(k)}, y_6^{(k)})\, y_d \geq 0,\ d = 1, 2, 3, 4, 5, 6,\ \sum_{d=1}^{6} y_d^{(k)} = 1,\ (k = 1, \ldots, \#LS)$ over the criteria set $\{C_d | d = 1, 2, 3, 4, 5, 6\}$.

To deal with this problem by employing PM-ULTSs, an investor utilize the following ULTS to evaluate the company's alternative set A_c:

$$LS = \{\text{P, VW, W, Avg, G, VG, PR}\}$$

Note: Poor (P), Very weak (VW), Weak (W), Average (Avg), Good (G), Very Good (VG), Perfect (PR). Based on the representation algorithm stated in paper [43], we have

$$LS = \{t_{1/8} : (\text{P}),\ t_{1/4} : (\text{VW}),\ t_{1/2} : (\text{W}),\ t_1 : (\text{Avg}),\ t_2 : (\text{G}),\ t_4 : (\text{VG}),\ t_8 : (\text{PR})\}$$

Next, as we proceed further it is perceived that in the present case study the investor provide his preferences in respect of several multiplicative unbalanced linguistic terms on the response alternatives $A_c\,(c = 1, 2, 3, 4)$ subject to the six criteria $C_d\,(d = 1, 2, 3, 4, 5, 6)$. Since investors studied the overall companies' performance based on the mentioned criteria for the past four financial years (FY), i.e., $2016 - 2020$. Thus, the original multiplicative unbalanced linguistic payoff matrices for the four consecutive years are given in Tables 4-7. It is noted that the blanks mentioned in Tables 6 and 7 represent that for a given FY, the company's evaluation for that particular criteria is very weak. As a result, it can be neglected. By summing up, these four tables directly, the assessment information in Table 8 can be exhibited in terms of PM-ULTSs, which is regarded as a payoff matrix for this decision analysis problem modelled as a probabilistic unbalanced linguistic matrix game. To understand the representation of the linguistic evaluation of A_2 concerning C_2, i.e., $\{VG(0.5), G(0.25), Avg(0.25)\}$ is derived based on the analysis made by the investor for four consecutive years that for two out of four years companies overall assessment is very good (VG), for one out of four years is good (G), and also for one year it is average (Avg).

6.2 Probabilistic linguistic matrix game solution process

The primary solution procedure for solving the probabilistic linguistic game problem is described in the subsequent steps.

Step 1: Normalize the payoff matrix.
We normalize the payoff matrix and obtain the final normalized probabilistic unbalanced linguistic matrix, which is given in Table 9.

Step 2: Construct the PMULLP problem.
After attaining the normalized probabilistic unbalanced linguistic payoff matrix, we can construct the PMULLP problem (see Eq. (M1)) for PI by using model (PMULLP1).

Table 8 Comprehensive probabilistic multiplicative unbalanced linguistic payoff matrix

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	$\{Avg(0.5), G(0.25), VG(0.25)\}$	$\{G(0.5), Avg(0.25), VG(0.25)\}$	$\{PR(0.25), VG(0.5), G(0.25)\}$	$\{Avg(0.5), G(0.25), VG(0.25)\}$	$\{G(0.5), VG(0.25), PR(0.25)\}$	$\{VG(0.25), G(0.25), PR(0.5)\}$
A_2	$\{W(0.5), G(0.25), Avg(0.25)\}$	$\{VG(0.25), G(0.25), Avg(0.25)\}$	$\{Avg(0.5), W(0.25), G(0.25)\}$	$\{Avg(0.5), G(0.25), VG(0.25)\}$	$\{W(0.5), VW(0.25), Avg(0.25)\}$	$\{VW(0.5), W(0.25), Avg(0.25)\}$
A_3	$\{W(0.5), VW(0.25), VG(0.25)\}$	$\{W(0.5), G(0.25), Avg(0.25)\}$	$\{G(0.5), Avg(0.25), VG(0.25)\}$	$\{G(0.25), Avg(0.5), VG(0.25)\}$	$\{VG(0.5), PR(0.25), G(0.25)\}$	$\{Avg(0.25), G(0.25), VG(0.25)\}$
A_4	$\{PR(0.25), VG(0.25), G(0.5)\}$	$\{G(0.25), VG(0.5), PR(0.25)\}$	$\{PR(0.5), VG(0.25), G(0.25)\}$	$\{W(0.25), Avg(0.5), G(0.25)\}$	$\{W(0.5), Avg(0.25), G(0.25)\}$	$\{W(0.25), Avg(0.5), G(0.25)\}$

Table 9 Comprehensive normalized probabilistic multiplicative unbalanced linguistic payoff matrix

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	$\{Avg(0.5), G(0.25), VG(0.25)\}$	$\{G(0.5), Avg(0.25), VG(0.25)\}$	$\{PR(0.25), VG(0.5), G(0.25)\}$	$\{Avg(0.5), G(0.25), VG(0.25)\}$	$\{G(0.5), VG(0.25), PR(0.25)\}$	$\{VG(0.25), G(0.25), PR(0.5)\}$
A_2	$\{W(0.5), G(0.25), Avg(0.25)\}$	$\{VG(0.33), G(0.33), Avg(0.33)\}$	$\{Avg(0.5), W(0.25), G(0.25)\}$	$\{Avg(0.5), G(0.25), VG(0.25)\}$	$\{W(0.5), VW(0.25), Avg(0.25)\}$	$\{VW(0.5), W(0.25), Avg(0.25)\}$
A_3	$\{W(0.5), VW(0.25), VG(0.25)\}$	$\{W(0.5), G(0.25), Avg(0.25)\}$	$\{G(0.5), Avg(0.25), VG(0.25)\}$	$\{G(0.25), Avg(0.5), VG(0.25)\}$	$\{VG(0.5), PR(0.25), G(0.25)\}$	$\{Avg(0.33), G(0.33), VG(0.33)\}$
A_4	$\{PR(0.25), VG(0.25), G(0.5)\}$	$\{G(0.25), VG(0.5), PR(0.25)\}$	$\{PR(0.5), VG(0.25), G(0.25)\}$	$\{W(0.25), Avg(0.5), G(0.25)\}$	$\{W(0.5), Avg(0.25), G(0.25)\}$	$\{W(0.25), Avg(0.5), G(0.25)\}$

$$\max \quad \widehat{v}(p) = \left\{\widetilde{v}^{(k)}(p^{(k)}) | k = 1, 2, 3\right\}$$
$$\text{subject to} \begin{cases} \{Avg(0.5), G(0.25), VG(0.25)\}x_1^{(k)} \oplus \{W(0.5), G(0.25), Avg(0.25)\}x_2^{(k)} \oplus \{W(0.5), VW(0.25), VG(0.25)\}x_3^{(k)} \oplus \{PR(0.25), VG(0.25), G(0.5)\}x_4^{(k)} \geq \widehat{v}(p) \\ \{G(0.5), Avg(0.25), VG(0.25)\}x_1^{(k)} \oplus \{VG(0.33), G(0.33), Avg(0.33)\}x_2^{(k)} \oplus \{W(0.5), G(0.25), Avg(0.25)\}x_3^{(k)} \oplus \{G(0.25), VG(0.5), PR(0.25)\}x_4^{(k)} \geq \widehat{v}(p) \\ \{PR(0.25), VG(0.5), G(0.25)\}x_1^{(k)} \oplus \{Avg(0.5), W(0.25), G(0.25)\}x_2^{(k)} \oplus \{G(0.5), Avg(0.25), VG(0.25)\}x_3^{(k)} \oplus \{PR(0.5), VG(0.25), G(0.25)\}x_4^{(k)} \geq \widehat{v}(p) \\ \{Avg(0.5), G(0.25), VG(0.25)\}x_1^{(k)} \oplus \{Avg(0.5), G(0.25), VG(0.25)\}x_2^{(k)} \oplus \{G(0.25), Avg(0.5), VG(0.25)\}x_3^{(k)} \oplus \{W(0.25), Avg(0.5), G(0.25)\}x_4^{(k)} \geq \widehat{v}(p) \\ \{G(0.5), VG(0.25), PR(0.25)\}x_1^{(k)} \oplus \{W(0.5), VW(0.25), Avg(0.25)\}x_2^{(k)} \oplus \{VG(0.5), PR(0.25), G(0.25)\}x_3^{(k)} \oplus \{W(0.5), Avg(0.25), G(0.25)\}x_4^{(k)} \geq \widehat{v}(p) \\ \{VG(0.25), G(0.25), PR(0.5)\}x_1^{(k)} \oplus \{VW(0.5), W(0.25), Avg(0.25)\}x_2^{(k)} \oplus \{Avg(0.25), G(0.5), VG(0.25)\}x_3^{(k)} \oplus \{W(0.33), Avg(0.33), G(0.33)\}x_4^{(k)} \geq \widehat{v}(p) \\ \sum_{c=1}^{4} x_c^{(k)} = 1;\ x_c^{(k)} \geq 0,\ c = 1, 2, 3, 4. \end{cases} \tag{M1}$$

Step 3: Formulation of game problem for PI.
Based on the methodology presented in the previous section, the MOLLOP is decomposed into three single objective PMULLP problems and further, based on the transformation function, the system is converted into three crisp linear optimization problems (see Eqs. (M1.1), (M1.2), (M1.3)).

$$\min V^1 = X_1^1 + X_2^1 + X_3^1 + X_4^1$$
$$\text{subjectto} \begin{cases} 0.5X_1^1 + 0.0625X_2^1 + 0.25X_3^1 + 2X_4^1 \geq 1 \\ X_1^1 + 0.165X_2^1 + 0.25X_3^1 + 0.5X_4^1 \geq 1 \\ 2X_1^1 + 2X_2^1 + X_3^1 + 4X_4^1 \geq 1 \\ 0.5X_1^1 + 0.5X_2^1 + 0.5X_3^1 + 0.125X_4^1 \geq 1 \\ X_1^1 + 0.25X_2^1 + 0.2X_3^1 + 0.25X_4^1 \geq 1 \\ X_1^1 + 0.125X_2^1 + 0.33X_3^1 + 0.125X_4^1 \geq 1 \\ X_1^1, X_2^1, X_3^1, X_4^1 \geq 0. \end{cases} \tag{M1.1}$$

$$\min \quad V^2 = X_1^2 + X_2^2 + X_3^2 + X_4^2$$
$$\text{subject to} \begin{cases} 0.5X_1^2 + X_2^2 + 0.0625X_3^2 + X_4^2 \geq 1 \\ 0.25X_1^2 + 0.66X_2^2 + 0.5X_3^2 + 2X_4^2 \geq 1 \\ 2X_1^2 + 0.5X_2^2 + 0.25X_3^2 + X_4^2 \geq 1 \\ 0.5X_1^2 + 0.125X_2^2 + 0.5X_3^2 + 0.5X_4^2 \geq 1 \\ X_1^2 + 0.0625X_2^2 + 2X_3^2 + 0.25X_4^2 \geq 1 \\ 0.5X_1^2 + 0.125X_2^2 + 0.66X_3^2 + 0.5X_4^2 \geq 1 \\ X_1^2, X_2^2, X_3^2, X_4^2 \geq 0. \end{cases} \tag{M1.2}$$

$$\min \quad V^3 = X_1^3 + X_2^3 + X_3^3 + X_4^3$$
$$\text{subject to} \begin{cases} X_1^3 + 0.25X_2^3 + X_3^3 + X_4^3 \geq 1 \\ X_1^3 + 0.33X_2^3 + 0.25X_3^3 + 2X_4^3 \geq 1 \\ 0.5X_1^3 + 0.25X_2^3 + X_3^3 + 0.5X_4^3 \geq 1 \\ X_1^3 + 0.5X_2^3 + X_3^3 + 0.5X_4^3 \geq 1 \\ 2X_1^3 + 0.25X_2^3 + 0.5X_3^3 + 0.5X_4^3 \geq 1 \\ 4X_1^3 + 0.25X_2^3 + 0.132X_3^3 + 0.5X_4^3 \geq 1 \\ X_1^3, X_2^3, X_3^3, X_4^3 \geq 0. \end{cases} \tag{M1.3}$$

Step 4: Calculate the probabilistic linguistic value of the game.
The optimal solution obtained for the mentioned problem of PI are $X^{(1)*} = (2, 0, 0, 0)$, $X^{(2)*} = (2, 0, 0.2857, 1.7143)$, $X^{(3)*} = (0.2449, 0, 0.7347, 0.2857)$ and $V^* = (2, 2, 1.2653)$. Hence, for PI, the optimal mixed strategy is obtained as $x_1^* = \{1, 0, 0.1935\}$, $x_2^* = 0$, $x_3^* = \{0, 0.14285, 0.5806\}$, $x_4^* = \{0, 0.85715, 0.2258\}$. After the re-translation process the optimal probabilistic unbalanced linguistic value is $\widehat{v}(p) = \left\{t_{1/2}(0.9612), t_{2^0}(0.0387), t_{1/2}(0)\right\}$.

Step 5: Formulation of game problem for PII.
Repeat the similar steps to formulate the PMULLP (see Eq. (M2)) problem for PII by using model (PMULLP2). After the transformation, we obtain the corresponding three linear optimization problem (see Eq. (M2.1), (M2.2), (M2.3)).

$$\min \quad \widehat{\omega}(p) = \left\{\widetilde{\omega}^{(k)}(p^{(k)}) | k = 1, 2, 3\right\}$$

$$\text{subject to} \begin{cases} \{Avg(0.5), G(0.25), VG(0.25)\}y_1^{(k)} \oplus \{G(0.5), Avg(0.25), VG(0.25)\}y_2^{(k)} \\ \oplus \{PR(0.25), VG(0.5), G(0.25)\}y_3^{(k)} \oplus \{Avg(0.5), G(0.25), VG(0.25)\}y_4^{(k)} \\ \oplus \{G(0.5), VG(0.25), PR(0.25)\}y_5^{(k)} \oplus \{VG(0.25), G(0.25), PR(0.5)\}y_6^{(k)} \le \widehat{\omega}(p) \\ \{W(0.5), G(0.25), Avg(0.25)\}y_1^{(k)} \oplus \{VG(0.33), G(0.33), Avg(0.33)\}y_2^{(k)} \\ \oplus \{Avg(0.5), W(0.25), G(0.25)\}y_3^{(k)} \oplus \{Avg(0.5), G(0.25), VG(0.25)\}y_4^{(k)} \\ \oplus \{W(0.5), VW(0.25), Avg(0.25)\}y_5^{(k)} \oplus \{VW(0.5), W(0.25), Avg(0.25)\}y_6^{(k)} \le \widehat{\omega}(p) \\ \{W(0.5), VW(0.25), VG(0.25)\}y_1^{(k)} \oplus \{W(0.5), G(0.25), Avg(0.25)\}y_2^{(k)} \\ \oplus \{G(0.5), Avg(0.25), VG(0.25)\}y_3^{(k)} \oplus \{G(0.25), Avg(0.5), VG(0.25)\}y_4^{(k)} \\ \oplus \{VG(0.5), PR(0.25), G(0.25)\}y_5^{(k)} \oplus \{Avg(0.25), G(0.5), VG(0.25)\}y_6^{(k)} \le \widehat{\omega}(p) \\ \{PR(0.25), VG(0.25), G(0.5)\}y_1^{(k)} \oplus \{G(0.25), VG(0.5), PR(0.25)\}y_2^{(k)} \\ \oplus \{PR(0.5), VG(0.25), G(0.25)\}y_3^{(k)} \oplus \{W(0.25), Avg(0.5), G(0.25)\}y_4^{(k)} \\ \oplus \{W(0.5), Avg(0.25), G(0.25)\}y_5^{(k)} \oplus \{W(0.33), Avg(0.33), G(0.33)\}y_6^{(k)} \le \widehat{\omega}(p) \\ \sum_{d=1}^{6} y_d^{(k)} = 1;\ y_d^{(k)} \ge 0,\ d = 1, 2, 3, 4, 5, 6. \end{cases} \tag{M2}$$

$$\max \quad V^1 = Y_1^1 + Y_2^1 + Y_3^1 + Y_4^1 + Y_5^1 + Y_6^1$$

$$\text{subject to} \begin{cases} 0.5Y_1^1 + Y_2^1 + 2Y_3^1 + 0.5Y_4^1 + Y_5^1 + Y_6^1 \le 1 \\ 0.0625Y_1^1 + 0.165Y_2^1 + 2Y_3^1 + 0.5Y_4^1 + 0.25Y_5^1 + 0.125Y_6^1 \le 1 \\ 0.25Y_1^1 + 0.25Y_2^1 + Y_3^1 + 0.5Y_4^1 + 0.2Y_5^1 + 0.33Y_6^1 \le 1 \\ 2Y_1^1 + 0.5Y_2^1 + 4Y_3^1 + 0.125Y_4^1 + 0.25Y_5^1 + 0.125Y_6^1 \le 1 \\ Y_1^1,\ Y_2^1,\ Y_3^1,\ Y_4^1,\ Y_5^1,\ Y_6^1 \ge 0. \end{cases} \tag{M2.1}$$

$$\max \quad V^2 = Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 + Y_5^2 + Y_6^2$$

$$\text{subject to} \begin{cases} 0.5Y_1^2 + 0.25Y_2^2 + 2Y_3^2 + 0.5Y_4^2 + Y_5^2 + 0.5Y_6^2 \le 1 \\ Y_1^2 + 0.66Y_2^2 + 0.5Y_3^2 + 0.125Y_4^2 + 0.0625Y_5^2 + 0.125Y_6^2 \le 1 \\ 0.0625Y_1^2 + 0.5Y_2^2 + 0.25Y_3^2 + 0.5Y_4^2 + 2Y_5^2 + 0.66Y_6^2 \le 1 \\ Y_1^2 + 2Y_2^2 + Y_3^2 + 0.5Y_4^2 + 0.25Y_5^2 + 0.5Y_6^2 \le 1 \\ Y_1^2,\ Y_2^2,\ Y_3^2,\ Y_4^2,\ Y_5^2,\ Y_6^2 \ge 0. \end{cases} \tag{M2.2}$$

$$\max \quad V^3 = Y_1^3 + Y_2^3 + Y_3^3 + Y_4^3 + Y_5^3 + Y_6^3$$

$$\text{subject to} \begin{cases} Y_1^3 + Y_2^3 + 0.5Y_3^3 + Y_4^3 + 2Y_5^3 + 4Y_6^3 \le 1 \\ 0.25Y_1^3 + 0.33Y_2^3 + 0.25Y_3^3 + 0.5Y_4^3 + 0.25Y_5^3 + 0.25Y_6^3 \le 1 \\ Y_1^3 + 0.25Y_2^3 + Y_3^3 + Y_4^3 + 0.5Y_5^3 + 0.132Y_6^3 \le 1 \\ Y_1^3 + 2Y_2^3 + 0.5Y_3^3 + 0.5Y_4^3 + 0.5Y_5^3 + 0.5Y_6^3 \le 1 \\ Y_1^3,\ Y_2^3,\ Y_3^3,\ Y_4^3,\ Y_5^3,\ Y_6^3 \ge 0. \end{cases} \tag{M2.3}$$

Step 6: Compute the probabilistic linguistic matrix game value.

After solving, the optimal solution obtained for the mentioned problem for PII are $Y^{(1)*} = (0.4, 0, 0, 1.6, 0, 0)$, $Y^{(2)*} = (0, 0, 0, 2, 0, 0)$, $Y^{(3)*} = (0, 0.2449, 0.8571, 0, 0.1633, 0)$ and $V^* = (2, 2, 1.2653)$. Hence, for PII, the optimal mixed strategy is obtained as $y_1^* = \{0.2, 0, 0\}$, $y_2^* = \{0, 0, 0.1935\}$, $y_3^* = \{0, 0, 0.6774\}$, $y_4^* = \{0.8, 1, 0\}$, $y_5^* = \{0, 0, 0.1291\}$, $y_6^* = 0$. After the re-translation process the optimal probabilistic unbalanced linguistic value is $\widehat{\omega}(p) = \{t_{1/2}(0.9612), t_{20}(0.0387), t_{1/2}(0)\}$.

Here, for the investor, the criteria for which $y_d^* = 0$ are of least significance. Analogously, the alternatives with $x_c^* = 0$ are irrelevant to the investor.

Finally, to rank the given alternatives we calculate the total score of the alternatives as follows: $A_1 = 1.1935$, $A_2 = 0$, $A_3 = 0.72345$, $A_4 = 1.08295$. Hence, the four alternatives are ordered as $A_1 > A_4 > A_3 > A_2$, and alternative A_1 is the recommended company for the investor to invest his money.

6.3 Comparative analyses

The sequel gives comparative analyses to validate the feasibility of the PM-ULTSs and the proposed probabilistic linguistic game-theoretic model.

Compared with the studies of Han et al. [19], the work presented in our paper has various advantages. Our study develops a novel concept, namely PM-ULTSs, to overcome the limitation of PULTS [19]. Both PM-ULTS and PULTS can represent the unbalanced linguistic information and the weight information of each unbalanced linguistic term. However, the irregular information in PULTS is based on the unbalanced linguistic computational model proposed by Herrera et al. [22]. On the other hand, the information in PM-ULTS is based on multiplicative ULTS proposed in the paper [43]. It is envisioned that the unbalanced linguistic computational model given in the paper [22] is computationally expensive and is capable of dealing with Type I unbalanced linguistic information. By contrast, the linguistic computational model for multiplicative ULTS is computationally less expansive and can address both Type I and II unbalanced linguistic information [22, 61].

Moreover, the PULTS is based on the symmetrically distributed LTS in terms of information representation. As a result, the literal meaning of ULTS is obscured. The PM-ULTS, on the other hand, is based on the multiplicative ULTS to indicate the variation in spacing for each linguistic term while maintaining the literal meaning. Furthermore, the proposed strategy eliminates data loss and can produce more accurate and reliable decision outcomes. The aggregation operators proposed in the paper [19] to aggregate PULTS is based on Archimedean copula and co-copula, which increases complexity when applied to aggregate information in large-scale group decision-making problem. Conversely, the aggregation operator to aggregate PM-ULTS does not rely on Archimedean copula and co-copula, effectively reducing computing complexity. Additionally, the novel operational laws constructed for PM-ULTS are closed and reduce the computational time when applied to decision problems.

Next, it is perceived that the imprecise game model using PM-ULTS as an input parameter, a flexible tool for computing with words, has not been explored before within the game theory framework. Moreover, in comparison to the methodologies given in literature [18, 25, 28], the proposed game-theoretic approach to solving multi-decision analysis problems is more advantageous in the sense that the latter requires prior knowledge of the criteria's weights, whereas our proposed approach evaluates the weights as an intermediate step. Additionally, in contrast to game-theoretic models with single linguistic terms [1, 2], and linguistic 2-tuple information [52, 53], the probabilistic unbalanced linguistic information allows players to provide evaluations on multiple linguistic terms irregularly distributed along with their probability information.

Table 10 The strengths and weaknesses of the proposed PM-ULTSs and probabilistic unbalanced linguistic game model

Strengths	Weaknesses
PM-ULTSs generalized the existing unbalanced computing with word models to represent non-uniformity of the labels and corresponding probabilities	It's necessary to capture probabilistic information of the linguistic lables
Incompleteness within the probabilistic information is allowed	The elicitation of the probabilistic information is not known
The weights of alternatives are calculated in an intermedial step while solving the game problem	
The proposed PM-ULTSs preserves the literal meaning of the ULTS alongwith the probabilistic information	

Furthermore, the game models having linguistic 2-tuple information do not handle an incomplete evaluation. As a result, players' burden and difficulty level increases. On the other hand, probabilistic linguistic games efficiently handle challenges with inadequate information. Mi et al. [42] recently suggested a methodology for solving matrix game problems in terms of probabilistic linguistic information based on symmetric LTSs. In that paper, the authors have approximated the probabilistic linguistic information into a classical triangular membership function. Consequently, it leads to information loss. In our suggested strategy, we are not translating the probabilistic linguistic information into some approximation function. Instead, we define an information measure of the probabilistic linguistic term using the bijective function Ω and Ω^{-1}. As a result, there will be no information loss in the final re-translation process.

Lastly, the developed method for solving matrix game problems is primarily based on PMULTS, consisting of probabilistic unbalanced linguistic elements such that each element is independent. Hence, the probabilistic linguistic matrix game problem can be easily transformed into MOLLOP, which is decomposed into a single objective PMULLP problem. The objective function of these transformed linear problems runs in $O(n)$ time, where n is the cardinality of the PMULTS. Therefore, the proposed model has linear time complexity making it more reasonable and convincible in handling probabilistic aunbalanced linguistic game problems.

In general, the probabilistic unbalanced linguistic game methodology described in the paper has its advantages and disadvantages. The DMs should carefully analyze their needs and decision-making scenarios and select a suitable evaluation method.

7 Conclusion

This paper introduces the concept of PM-ULTSs, which considers the probability of linguistic variables and the non-uniformity of linguistic labels. After then, some new operational laws for managing probabilistic unbalanced linguistic information are also developed, which could produce valid results and keep the property of operational laws closed. Furthermore, some elementary aggregation operators to aggregate PM-ULTSs have been constructed for further consideration, and their properties are also discussed. These operators need to be beneficial in situations involving information that cannot be expressed with an actual number but rather the probabilistic multiplicative unbalanced linguistic information. Finally, we have applied our proposed theoretical concept to examine the two-player matrix game problem having probabilistic unbalanced linguistic information as input. The information measure for the probabilistic multiplicative unbalanced linguistic term is calculated to transform the probabilistic linguistic information. The information measure is primarily based on the bijective function Ω and Ω^{-1} such that it avoids the loss of information in the re-translation process. Thus, the linear optimization model for both the players is formulated to obtain optimal game value and corresponding mixed strategies. The calculated game value is further re-translated into the original PM-ULTS to maintain the interpretability of the probabilistic unbalanced linguistic game. Furthermore, the validity and advantages of the probabilistic unbalanced linguistic matrix game were verified and justified by applying it to solve real-life multi-criteria decision analysis probalems about the companies selected by a general investor to invest his sum of money from the perspective of the game with Nature.

It is perceived that the proposed matrix game methodology perfectly handles complicated decision-analysis problems under the probabilistic unbalanced linguistic environment. In addition, the method is consistent and more acquirable to solve a decision-making problem as it can generate the optimal weights for the given alternatives in an intermedial step, unlike the other existing methods where the weights are presupposed to be well-known a priori. The optimal game value obtained is viewed as a PM-ULTSs that represents the uncertainty more reasonably using several unbalanced linguistic terms instead of a single linguistic term such that it closely aligns human thoughts. We believe that the wide-ranging application of our proposed concept makes it promising. It is readily apparent that the present study brings a new viewpoint to the uncertain linguistic matrix game solution and strategic analysis. We summarize the strengths and weaknesses of the proposed technique in Table 10.

Although the practical example of companies' selection problem by a general investor to invest his money is provided to validate the expediency and efficacy of the suggested method, it can be extended to address numerous case studies and large-scale group decision-making problems. In addition, the practicality of the proposed PM-ULTS concept and probabilistic unbalanced linguistic game model can be outstretched to employ them in other fields through a combination of advanced theories. In the future, our work will attempt to widen the scope of the present study to define unbalanced linguistic bimatrix and cooperative games in a probabilistic environment. Our research will also plan to extend the proposed concept of PM-ULTS to develop an asymmetric cloud model, interval-valued PM-ULTS, and other distinct application fields.

1. Appendix

Algorithm 1 Computation of $t^{(k)}_{a^i \oplus a^j}$ and $\bar{p}_1^{(k)} \oplus \bar{p}_2^{(k)}$

Input: $LS_1(p_1) = \left\{ t^{(k)}_{a^i}(\bar{p_1}^k) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$ and $LS_2(p_2) = \left\{ t^{(k)}_{a^j}(\bar{p_2}^k) \,|\, k = 1, 2, \ldots, \#LS_2(p_1) \right\}$, $(i, j = -n, \ldots, 0, 1, \ldots, m)$ where $a > 1$ and n may or may not be equal.

Output: $LS_1(p_1) \oplus LS_2(p_2) = \left\{ t^{(k)}_{a^i \oplus a^j}(\bar{p}_1^k \oplus \bar{p}_2^k) | k = 1, 2, \ldots, \#LS_1(p_1) \right\}$,

Begin:

for $k \leftarrow 1$ to $\#LS_1(p_1)$ **do**

if $t^{(k)}_{a^i \oplus a^j} \in \left\{ t_{a^{-n}}, \ldots, t_{a^0}, \ldots, t_{a^m} \right\}$ **then**

$t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^i + a^j}$;

end if

if $t^{(k)}_{a^i \oplus a^j} \notin \left\{ t_{a^{-n}}, \ldots, t_{a^0}, \ldots, t_{a^m} \right\}$ **then**

$t^{(k)}_{a^i \oplus a^j} \leftarrow t^{(k)}_d$ where

$$d \leftarrow \begin{cases} a^i + a^j \text{ if } d \in [a^{-n}, a^m]; \\ a^i + a^j - a^m \text{ otherwise.} \end{cases}.$$

if $d \in [a^{-n}, a^m]$ or $d \notin [a^{-n}, a^m]$ **then**

$i_L \leftarrow \arg\min_i \left\{ d - a^i \,|\, d - a^i \geq 0 \right\}$;

$i_R \leftarrow \arg\max_i \left\{ d - a^i \,|\, d - a^i \leq 0 \right\}$, for $(i = -n, \ldots, 0, \ldots, m)$.

Compute:

$D_{left} \leftarrow d - a^{i_L}$ and $D_{right} \leftarrow d - a^{i_R}$ where $i_L, i_R \in (-n, \ldots, 0, \ldots, m)$

Define: $D = \min(D_{left}, |D_{right}|)$

if $D \leftarrow D_{left}$ **then**

$t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^{i_L}}$

else

$t^{(k)}_{a^i \oplus a^j} = t^{(k)}_{a^{i_R}}$

end if

end if

end if

Next compute for $\bar{p}_1^{(k)} \oplus \bar{p}_2^{(k)}$**:**

$\bar{p}_1^{(k)} \oplus \bar{p}_2^{(k)} \leftarrow p_3^{(k)}$, where $p_3^{(k)} \leftarrow \begin{cases} \bar{p}_1^k + \bar{p}_2^k, & \text{if } \bar{p}_1 + \bar{p}_2 \leq 1; \\ \bar{p}_1^k + \bar{p}_2^k - 1, & \text{if } \bar{p}_1 + \bar{p}_2 > 1. \end{cases}$

if $\sum_{k=1}^{\#LS_{(p_1)}} p_3^{(k)} \leftarrow 1$ **then**

done

else

$p_3^{(k)} \leftarrow \frac{p_3^{(k)}}{\sum_{k=1}^{\#LS_1(p_1)} p_3^{(k)}}$

end if

end for

Algorithm 2 Computation of $t^{(k)}_{a^i \otimes a^j}$ and $\bar{p}_1^{(k)} \otimes \bar{p}_2^{(k)}$

Input: $LS_1(p_1) = \left\{ t^{(k)}_{a^i}(\bar{p_1}^k) \,|\, k = 1, 2, \ldots, \#LS_1(p_1) \right\}$ and $LS_2(p_2) = \left\{ t^{(k)}_{a^j}(\bar{p_2}^k) \,|\, k = 1, 2, \ldots, \#LS_2(p_1) \right\}$, $(i, j = -n, \ldots, 0, 1, \ldots, m)$ where $a > 1$ and n may or may not be equal.

Output: $LS_1(p_1) \otimes LS_2(p_2) = \left\{ t^{(k)}_{a^i \otimes a^j}(\bar{p}_1^k \otimes \bar{p}_2^k) | k = 1, 2, \ldots, \#LS_1(p_1) \right\}$,

Begin:

$k \leftarrow 1$

while $k \leq \#LS_1(p_1)$ **do**

$$t^{(k)}_{a^i \otimes a^j} \leftarrow \begin{cases} t^k_{a^{i+j+n}} \text{ if } i + j < -n; \\ t^k_{a^{i+j}} \text{ if } i + j \in \left\{ -n, -(n-1), \ldots, 0, \ldots, m \right\} \\ t^k_{a^{i+j-m}} \text{ if } i + j > m. \end{cases} \quad \textbf{and}$$

$\bar{p}_1^k \otimes \bar{p}_2^k \leftarrow \bar{p}_3^{(k)}$, $\bar{p}_3^{(k)} = \frac{p_3^{(k)}}{\sum_{k=1}^{\#LS_1(p_1)} p_3^{(k)}}$

end while

Algorithm 3 Computation of $t^{(k)}_{\lambda\odot a^i}$

Input: $LS(p)=\left\{t^{(k)}_{a^i}(\bar{p}^k)\,|\,k=1,2,\ldots,\#LS(p)\right\}$, where $a>1$ and n may or may not be equal and $\lambda\in(0,1)$.

Output: $\lambda\odot LS(p)=\left\{t^{(k)}_{\lambda\odot a^i}(\bar{p}^k)|\,k=1,2,\ldots,\#LS(p)\right\}$

Begin:

for $k\leftarrow 1$ to $\#LS_1(p_1)$ **do**
  if $t^{(k)}_{\lambda\odot a^i}\in\left\{t_{a^{-n}},\ldots,t_{a^0},\ldots,t_{a^m}\right\}$ **then**
    $t^{(k)}_{\lambda\odot a^i}=t^{(k)}_{\lambda\cdot a^i}$;
  else
    if $t^{(k)}_{\lambda\odot a^i}\notin\left\{t_{a^{-n}},\ldots,t_{a^0},\ldots,t_{a^m}\right\}$ **then**
      $t^{(k)}_{\lambda\odot a^i}\leftarrow t^{(k)}_{d}$ where

$$d\leftarrow\begin{cases}t^{k}_{\lambda\cdot a^i+a^{-n}}\text{ if }d\notin[a^{-n},a^m];\\ t^{k}_{\lambda\cdot a^i}\text{ if }d\in[a^{-n},a^m]\end{cases}$$

      if $d\in[a^{-n},a^m]$ or $d\notin[a^{-n},a^m]$ **then**
        $i_L\leftarrow\arg\min_i\left\{d-a^i\,|\,d-a^i\geq 0\right\}$;
        $i_R\leftarrow\arg\max_i\left\{d-a^i\,|\,d-a^i\leq 0\right\}$, for $(i=-n,\ldots,0,\ldots,m)$.
        Compute:
        $D_{left}\leftarrow d-a^{i_L}$ and $D_{right}\leftarrow d-a^{i_R}$ where $i_L,i_R\in(-n,\ldots,0,\ldots,m)$
        Define: $D=\min(D_{left},|D_{right}|)$
        if $D\leftarrow D_{left}$ **then**
          $t^{(k)}_{\lambda\odot a^i}=t^{(k)}_{a^{i_L}}$
        else
          $t^{(k)}_{\lambda\odot a^i}=t^{(k)}_{a^{i_R}}$
        end if
      end if
    end if
  end if
end for

Algorithm 4 Computation of $t^{(k)}_{a^{\lambda\cdot i}}$

Input: $LS(p)=\left\{t^{(k)}_{a^i}(\bar{p}^k)\,|\,k=1,2,\ldots,\#LS(p)\right\}$, where $a>1$ and n may or may not be equal and $\lambda\in(0,1)$.

Output: $LS(p)^{\lambda}=\left\{t^{(k)}_{a^{\lambda\cdot i}}(\bar{p}^k)|\,k=1,2,\ldots,\#LS(p)\right\}$,

Begin:

for $k\leftarrow 1$ to $\#LS_1(p_1)$ **do**
  if $\lambda\cdot i\in\left\{-n,-(n-1),\ldots,0,\ldots,m\right\}$ **then**
    done
  else
    if $\lambda\cdot i\notin\left\{-n,-(n-1),\ldots,0,\ldots,m\right\}$ **then**
      $i_L\leftarrow\arg\min_i\left\{(\lambda\cdot i)-a^i\,|\,(\lambda\cdot i)-a^i\geq 0\right\}$;
      $i_R\leftarrow\arg\max_i\left\{(\lambda\cdot i)-a^i\,|\,(\lambda\cdot i)-a^i\leq 0\right\}$, for $(i=-n,\ldots,0,\ldots,m)$.
      Define: $I=\min(i_L,|i_R|)$
      if $I\leftarrow i_L$ **then**
        $t^{(k)}_{a^{\lambda\cdot i}}=t^{(k)}_{a^{i_L}}$
      else
        $t^{(k)}_{a^{\lambda\cdot i}}=t^{(k)}_{a^{i_R}}$
      end if
    end if
  end if
end for

Funding Not applicable.

Data availability statement Not applicable.

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.

Code availability Not applicable.

References

1. Arfi B (2006) Linguistic fuzzy-logic game theory. J Confl Resolut 50(1):28–57
2. Arfi B (2006) Linguistic fuzzy-logic social game of cooperation. Ration Soc 18:471–537
3. Barron EN (2009) Game theory: an introduction. Wiley, New Jersy
4. Bector CR, Chandra S (2005) Fuzzy mathematical programming and fuzzy matrix games. Springer, Heidelberg
5. Bector CR, Chandra S, Vidyottama V (2004) Duality in linear programming with fuzzy parameters and matrix games with fuzzy payoffs. Fuzzy Sets Syst 146:253–269
6. Bai C, Ren Z, Shuang S et al (2018) Interval-valued probabilistic linguistic term sets in multi-criteria group decision making. Int J Intell Syst 33(6):1301–1321
7. Chen YW, Larbani M (2006) Two-person zero-sum game approach for fuzzy multiple attribute decision making problems. Fuzzy Sets Syst 157(1):34–51
8. Chen J, Hai W, Xu Z (2019) Uncertain probabilistic linguistic term sets in group decision making. Int J Fuzzy Syst 21(3):1241–1258
9. Cabrerizo FJ, Pérez IJ, Herrera-Viedma E (2010) Managing the consensus in group decision making in an unbalanced fuzzy linguistic context with incomplete information. Knowl Based Syst 23:169–181
10. Dubois D, Prade H (1980) Fuzzy sets and systems. Theory and applications. Mathematics in sciences and engineering series, vol 144. Academic Press, New York
11. Dong Y, Xu Y, Yu S (2009) Computing the numerical scale of the linguistic term set for the 2-tuple fuzzy linguistic representation model. IEEE Trans Fuzzy Syst 17(6):1366–1378
12. Dong Y, Li CC, Herrera F (2016) Connecting the linguistic hierarchy and the numerical scale for the 2-tuple linguistic model and its use to deal with hesitant unbalanced linguistic information. Inf Sci 367:259–278
13. Dong YC, Li CC, Herrera F (2015) An optimization-based approach to adjusting unbalanced linguistic preference relations to obtain a required consistency level. Inf Sci 292:27–38
14. Dong YC, Li CC, Xu Y, Gu X (2015) Consensus-based group decision making under multi-granular unbalanced 2-tuple linguistic preference relations. Group Decis Negot 24:217–242
15. Dong YC, Wu YZ, Zhang HJ, Zhang GG (2015) Multi-granular unbalanced linguistic distribution assessments with interval symbolic proportions. Knowl Based Syst 82:139–151
16. Fernández FR, Puerto J, Zafra MJ (2002) Cores of stochastic cooperative games. Int Game Theory Rev 4(3):265–280
17. Feng X, Liu Q, Wei C (2019) Probabilistic linguistic QUALIFLEX approach with possibility degree comparison. J Intell Fuzzy Syst 36(1):710–730
18. Gou XJ, Xu ZS (2016) Novel basic operational laws for linguistic terms, hesitant fuzzy linguistic term sets and probabilistic linguistic term sets. Inf Sci 372:407–427
19. Han B, Tao Z, Chen H, Zhou L, Liu J (2020) A new computational model based on Archimedean copula for probabilistic unbalanced linguistic term set and its application to multiple attribute group decision making. Comput Ind Eng 140:106–264
20. Han B, Chen HY, Zhu JM, Liu JP (2018) An approach to linguistic multiple attribute decision-making based on unbalanced linguistic generalized Heronian mean aggregation operator. Comput Intell Neurosci. https://doi.org/10.1155/2018/1404067
21. Han B, Tao ZF, Chen HY, Zhou LG (2018) Intuitionistic unbalanced linguistic generalized multiple attribute group decision making and its application to green products selection. Math Probl Eng 1–24
22. Herrera F, Herrera-Viedma E, Martínez L (2008) A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. IEEE Trans Fuzzy Syst 16(2):354–370
23. Isern D, Marin L, Valls A, Moreno A (2010) The unbalanced linguistic ordered weighted averaging operator. In: IEEE International conference on fuzzy systems (FUZZ), p 1–8
24. Jiang L, Liu HB, Cai JF (2015) The power average operator for unbalanced linguistic term sets. Inf Fusion 22:85–94
25. Liao HC, Jiang LS, Xu ZS, Xu JP, Herrera F (2017) A probabilistic linguistic linear programming method in hesitant qualitative multiple criteria decision making. Inf Sci 415–416:341–355
26. Liao HC, Jiang L, Lev B et al (2019) Novel operations of PLTSs based on the disparity degrees of linguistic terms and their use in designing the probabilistic linguistic ELECTRE III method. Appl Soft Comput 80:450–464
27. Liu P, Li Y (2019) Multi-attribute decision making method based on generalized Maclaurin symmetric mean aggregation operators for probabilistic linguistic information. Comput Ind Eng 131:282–294
28. Liu PD, Teng F (2019) Probabilistic linguistic TODIM method for selecting products through online product reviews. Inf Sci 485:441–455
29. Liu HB, Jiang L, Xu ZS (2018) Entropy measures of probabilistic linguistic term sets. Int J Comput Intell Syst 11:45–57
30. Lin M, Chen Z, Liao HC et al (2019) ELECTRE II method to deal with probabilistic linguistic term sets and its application to edge computing. Nonlinear Dyn 96(3):2125–2143
31. Lin MW, Wang HB, Xu ZS, Yao ZQ, Huang JL (2018) Clustering algorithms based on correlation coefficients for probabilistic linguistic term sets. Int J Intell Syst 33:2402–2424
32. Lin M, Chen Z, Xu Z, Gou X, Herrera F (2021) Score function based on concentration degree for probabilistic linguistic term sets: an application to TOPSIS and VIKOR. Inf Sci 551:270–290
33. Lin M, Xu Z et al (2018) Multi-attribute group decision-making under probabilistic uncertain linguistic environment. J Oper Res Soc 69(2):157–170
34. Li CC, Dong Y, Herrera F, Herrera-Viedma E, Martínez L (2017) Personalized individual semantics in computing with words for supporting linguistic group decision making. An application on consensus reaching. Inf Fusion 33:29–40
35. Li CC, Rodríguez RM, Herrera-Viedma E, Martínez L, Dong YC, Herrera F (2018) Personalized individual semantics based on consistency in hesitant linguistic group decision making with comparative linguistic expressions. Knowl Based Syst 145:156–165
36. Li DF, Nan JX, Zhang MJ (2012) Interval programming models for matrix games with interval payoffs. Optim Methods Softw 27(1):1–16
37. Li DF (2011) Linear programming approach to solve interval-valued matrix games. Omega 39(6):655–666
38. Ma Z, Zhu J, Chen Y (2019) A probabilistic linguistic group decision-making method from a reliability perspective based on evidential reasoning. IEEE Trans Syst Man Cybern Syst. https://doi.org/10.1109/TSMC.2018.2815716
39. Mata F, Pérez LG, Chiclana F, Herrera-Viedma E (2015) Aggregation of unbalanced fuzzy linguistic information in decision problems based on Type-1 OWA operator. In: IEEE international conference on in fuzzy systems (FUZZ-IEEE), p 1–6
40. Meng D, Pei Z (2013) On weighted unbalanced linguistic aggregation operators in group decision making. Inf Sci 223:31–41

41. Mao XB, Wu M, Dong JY, Wan SP, Jin Z (2019) A new method for probabilistic linguistic multi-attribute group decision making: application to the selection of financial technologies. Appl Soft Comput 77:155–175
42. Mi X, Liao H, Zeng XJ, Xu Z (2021) The two-person and zero-sum matrix game with probabilistic linguistic information. Inf Sci 570:487–499
43. Malhotra T, Gupta A (2020) A new 2-tuple linguistic approach for unbalanced linguistic term sets. IEEE Trans Fuzzy Syst 29:2158–2168
44. Malhotra T, Gupta A, Singh A (2019) Methodology for interval-valued matrix games with 2-tuple fuzzy linguistic information. In: International conference on numerical computations: theory and algorithms, Springer, Cham, p 154–168
45. Martinez L, Ruan D, Herrera F, Herrera-Viedma E, Wang PP (2009) Linguistic decision making: tools and applications. Inf Sci 179(14):2297–2298
46. Nan JX, Li DF (2014) Linear programming technique for solving interval-valued constraint matrix games. J Ind Manag Optim 10(4):1059–1070
47. Nishizaki I, Sakawa M (2001) Fuzzy and multiobjective games for conflict resolution. Physica-Verlag, Heidelberg
48. Pang Q, Wang H, Xu Z (2016) Probabilistic linguistic term sets in multi-attribute group decision making. Inf Sci 369:128–143
49. Pei Z, Zheng L (2017) New unbalanced linguistic scale sets: the linguistic information representations and applications. Comput Ind Eng 105:377–390
50. Suijs J, Borm P, De Waegenaere A (1998) Stochastic cooperative games in insurance. Insurance 22:209–228
51. Suijs J, Borm PP, De Waegenaere A, Tijs S (1999) Cooperative games with stochastic payoffs. Eur J Oper Res 113:193–205
52. Singh A, Gupta A, Mehra A (2020) Matrix games with 2-tuple linguistic information. Ann Oper Res 287(2):895–910
53. Singh A, Gupta A (2018) Matrix games with interval-valued 2-tuple linguistic information. Games 9(3):62
54. Song Y, Li G (2019) A large-scale group decision-making with incomplete multi-granular probabilistic linguistic term sets and its application in sustainable supplier selection. J Oper Res Soc 70(5):827–841
55. Sengupta JK (1989) A portfolio decisions as games. Int J Syst Sci 20(8):1323–1334
56. Tang M, Long Y, Liao H, Xu Z (2019) Inclusion measures of probabilistic linguistic term sets and their application in classifying cities in the Economic Zone of Chengdu Plain. Appl Soft Comput 82:105–572
57. Von Neumann J, Morgenstern O (1953) Theory of games and economic behavior, 3rd edn. Princeton University Press, Princeton
58. Wu XL, Liao HC, Xu ZS et al (2018) Probabilistic linguistic multimoora: a multicriteria decision making method based on the probabilistic linguistic expectation function and the improved borda rule. IEEE Trans Fuzzy Syst 26(6):3688–3702
59. Wu XL, Liao HC (2018) An approach to quality function deployment based on probabilistic linguistic term sets and ORESTE method for multi-expert multi-criteria decision making. Inf Fusion 43:13–26
60. Wu XL, Liao HC (2019) A consensus-based probabilistic linguistic gained and lost dominance score method. Eur J Oper Res 272(3):1017–1027
61. Wang JH, Hao J (2006) A new version of 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans Fuzzy Syst 14(3):435–445
62. Wang B, Liang J, Qian Y, Dang C (2015) A normalized numerical scaling method for the unbalanced multi-granular linguistic sets. Int J Uncertain Fuzziness Knowl Based Syst 23(2):221–243
63. Wang H, Xu Z, Zeng XJ (2018) Hesitant fuzzy linguistic term sets for linguistic decision making: current developments, issues and challenges. Inf Fusion 43:1–12
64. Xu ZS, He Y, Wang X (2019) An overview of probabilistic-based expressions for qualitative decision-making: techniques, comparisons and developments. Int J Mach Learn Cybern 10(6):1513–1528
65. Yu W, Zhang H, Li B (2019) Operators and comparisons of probabilistic linguistic term sets. Int J Intell Syst 34(7):1476–1504
66. Zhang YX, Xu ZS, Liao HC (2017) A consensus process for group decision making with probabilistic linguistic preference relations. Inf Sci 414:260–275
67. Zhang Y, Xu Z, Wang H, Liao H (2016) Consistency-based risk assessment with probabilistic linguistic preference relation. Appl Soft Comput 49:817–833
68. Zhang XL (2018) A novel probabilistic linguistic approach for large-scale group decision making with incomplete weight information. Int J Fuzzy Syst 20(7):2245–2256
69. Zhang XL, Xing XM (2017) Probabilistic linguistic VIKOR method to evaluate green supply chain initiatives. Sustainability 9(7):1231
70. Zhai Y, Xu Z, Liao H (2016) Probabilistic linguistic vector-term set and its application in group decision making with multi-granular linguistic information. Appl Soft Comput 49:801–816
71. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning-I. Inf Sci 8(3):199–249

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.

Circular Economy and Sustainability
https://doi.org/10.1007/s43615-022-00231-4

ORIGINAL PAPER

Product Packaging by E-commerce Platforms: Impact of COVID-19 and Proposal for Circular Model to Reduce the Demand of Virgin Packaging

Tanya Arora[1] · **Sarvani Reddy Chirla**[1] · **Nimisha Singla**[1] · **Lovleen Gupta**[1]

Received: 24 June 2022 / Accepted: 26 October 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract
E-commerce packaging waste is a matter of concern, especially with the increasing popularity of online shopping due to the COVID-19 pandemic. This leads to the generation of a massive amount of e-commerce packaging waste as well as resource utilisation and CO_2 emissions that go into the production of packaging. The aim of this study is to analyse the impact of COVID-19 on consumer trends in the e-commerce industry, quantitatively analyse the carbon emissions of packaging used, and present a circular model to reduce the demand for virgin packaging. Using a convenience sampling method, an online questionnaire was administered to 285 respondents to gather data on changes in shopping practices due to the COVID-19 pandemic, consumer awareness levels and observations on packaging materials, and practices employed by e-commerce companies. It was found that the number of orders placed per month increased after the onset of the pandemic and that most households dispose of packaging with household wastes as opposed to reusing or recycling. Primary data (study participants packaging waste production) in combination with secondary data (emission factors) was further used to calculate emissions due to mixed packaging waste, which came out to be 2,705.94 kg CO_2e per metric tonne of waste produced. In addition, the paper presents a practical solution to reducing virgin packaging material production, as well as modifications in packaging used to ensure efficient working of the packaging reuse model when implemented by the e-commerce companies.

Keywords COVID-19 · E-commerce packaging questionnaire · Carbon footprint assessment · Packaging reuse model · Sustainable E-commerce · Packaging waste minimisation

✉ Lovleen Gupta
lgupta@dce.ac.in

[1] Department of Civil and Environmental Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, 110042 Delhi, India

Published online: 22 November 2022

Introduction

In India, e-commerce platforms were first introduced in the year 2004 and ever since, the industry has grown exponentially [1]. The COVID-19 pandemic played a crucial role in the shift from brick-and-mortar shopping [2]–3. This was primarily due to three factors: Internet penetration and digital transformation, health and safety concerns, and an increase in comfort, variety, and affordability provided by the e-commerce industry [4]–5.

Research into the environmental impact of e-commerce and traditional in-store shopping has revealed that in-store shopping can reduce CO_2 emissions by up to 70% or even 84% in some cases [1]. It has emerged that the present model of the e-commerce industry is hence flawed and has not entirely been optimised [6]. Several studies have succeeded in establishing a direct relationship between the increase in the usage of e-commerce platforms, CO_2 emissions and energy use [1, 7]. Infrastructure, product packaging and design, and transportation are the three major factors that influence the impact of the e-commerce industry on the environment [4, 8]. Since these topics are widely dispersed in nature, this paper focuses on one of the factors, i.e. product packaging and design in the e-commerce industry.

The utilisation of packaging has seen an evident increase due to its three-fold function: aesthetic and appealing nature; health, hygiene, and safety; and preventing damage to the product during the multiple levels of manual handling [2, 9]–10. A single parcel sometimes can use up to 7 different types of packaging materials such as envelopes, cardboard, plastic bags, woven bags, tape, buffer material, and paper waybills [6]. According to the Indian Institute of Packaging, packaging consumption seems to have doubled in the past decade from 4.3 kg per person per annum to 8.6 kg per person per annum [11].

Globally, packaging waste management is an essential issue in urban areas—one of its main causes being, the lack of an effective and well-planned packaging waste management strategy. The packaging problem is one of the most visible problems related to land pollution and has begun to pose one of the biggest challenges for the e-commerce industry on its route to sustainability, especially due to the global pandemic [6, 12, 13]–14. According to the World Bank report 'What a Waste 2.0', India at present generates over a tenth of the global waste produced (highest globally) and in South Asia, packaging materials such as paper and cardboard account for 10%, followed by plastic which accounts for 8% of the net waste [15]. Despite the presence of solid waste management laws and regulations, in India, most of the waste is discarded without segregation at the source with 75% being openly dumped, 4% landfilled, 16% composted, and 5% recycled [15]–16. Packaging waste is also responsible for microplastic marine and land pollution, overconsumption of resources, and has a significant impact on the energy use for transport [1, 4, 11, 12]. Certain types of packaging are also known to contain chemical residue from their manufacturing that is hazardous to human health. On subsequent use, disposal, and recycling, these chemicals may transfer to the products being used by an individual, and thereby increase the exposure of humans and environments to toxins [4, 5, 12, 17, 18].

As such unsustainable packaging practices continue to thrive and threaten the concept of sustainable development, it has become essential to develop the concept of a circular economy for the e-commerce packaging industry. Exploring packaging systems through the circular economy perspective would also help shape innovative restorative and regenerative models that do not require as much technological and financial input as other means such as redesigning packaging and developing new packaging materials [14]. Such systems would also further allow maximum utilisation of resources and win the support of

consumers who are known to prefer circular and incentive-based systems over government tax levying solutions, in addition to reducing the overall environmental impact of the packaging over the long term [19].

This study set out to answer the following questions: (1) whether the COVID-19 pandemic led to an increase in the usage of e-commerce platforms for purchases; and (2) subsequent impact on the environment, if any, due to the increased use of virgin packaging materials as well as possible solutions.

In view of all of the above, the objectives of this study are to (1) analyse the impact of COVID-19 on consumer trends in the e-commerce industry; (2) quantitatively assess the carbon footprint of packaging material used; and (3) present a circular model to reduce the demand for virgin packaging.

This paper is organised as follows: the "Methodology" section describes the methodology used for data collection and analysis. As part of the study, a questionnaire was floated, the results of which have been discussed in the "Results and Discussion" section. In addition to this, it also contains a computation of the carbon footprint of e-commerce packaging used, based on the online shopping practices of the survey respondents. The "Proposal and Recommendations" section presents a possible solution to the problem raised in the form of a packaging reuse model along with modifications that can be made to existing packaging. Finally, in the "Theoretical and Practical Implications" section, several conclusions are drawn and suggestions for future research are made.

Methodology

Study Settings

The study was conducted in India. Being a developing country, it is currently seeing a rise in industrial growth rate and urban population [20, 21]. Both of these factors have together played a crucial role in the growth of the e-commerce industry. However, the above two factors may also result in an increased rate of packaging waste generation in India. Consequently, the country may be seen as an ideal location to conduct the survey. There is a need for a study to provide ways to manage packaging waste and mitigate environmental problems. The study involves a comparative analysis of the amount of packaging used by e-commerce companies pre and during COVID-19 as a result of the difference in the number of orders in both periods. Pre-COVID-19 is taken as the period prior to March 2020 and during COVID-19 as post-March 2020. This division was done due to the start of the lockdown period in India in March 2020 which is the period being monitored for an increase in the usage of e-commerce platforms [22].

Data Collection

Descriptive and remedial research approaches are both used in this study. The sources of data for the study were both primary and secondary. Questionnaire surveys were used to collect the primary data sources. The secondary data sources for this study have been collected by performing a literature review of relevant books, established news media, and journals published over the past two decades.

The survey was collected through a structured questionnaire comprising 17 questions, arranged and segmented into four sections as follows: (1) demographic attributes of study

participants; (2) comparison of shopping practices prior to March 2020 and post-March 2020; (3) awareness levels and observations of the consumer on packaging materials and practices employed by e-commerce companies; and (4) views of the consumer on online shopping. The questionnaire used for the study is attached in the Online Resource along with the survey results. A pilot survey was conducted amongst research colleagues and acquaintances prior to sharing them with the participants; this measure was performed to test the accuracy of the survey questions.

The survey was performed using Google forms between September 9, 2021, and September 27, 2021. This platform was preferred because of the easy user interface. Our target population involves customers of e-commerce websites and includes the old and young population residing all over India in rural, suburban, and metropolitan cities. The gender of the respondents was considered irrelevant to our study since a collective response was collected from each household. Convenience sampling was used to float the questionnaire and a total of 285 responses were collected, thereby satisfying the rule of thumb that 200 responses provide a fairly good survey accuracy [23]. There was no reward system in place for taking part in the survey.

All the questions were made compulsory to answer. The survey included an attention check question as well. Respondents were given sufficient time to study the questionnaire, ask questions, and obtain clarification, if necessary, on issues associated with the research and questionnaire before completion.

Data Analysis

For data analysis, the primary method of collecting data was through an online questionnaire, and the data was stored in MS Excel. This helped with the data processing and representation in graphical and tabular forms. Out of the 285 respondents, 213 answered the attention check question correctly and only these responses were taken into consideration while conducting the analysis.

The survey was carried out by the means of convenience sampling, due to which there is an unequal representation of respondents in terms of region and locality. Nevertheless, it has provided us with some reliable insights that contributed considerably to the paper.

Results and Discussion

Demographic Features and Age Groups of Study Participants

Table 1 shows the demographic features and age groups of study participants. The study included individuals from a wide range of age groups from diverse localities spread across India.

The largest group in terms of age (66.7%, $n=142$) was between 18 and 24 years, whereas those under 17 (1.90%, $n=4$) and above 55 (2.30%, $n=5$) were under-represented. 74.60% of the surveyors were situated in Northern India, whereas few were from Northeast India (0.92%, $n=2$) and twice that amount from East India (1.88%, $n=4$). Many people from the sample group (76.10%, $n=162$) resided in urban/metropolitan cities, whereas few study participants were from rural areas (4.2%, $n=9$). Unequal distribution of the sample space may be observed due to the utilisation of convenience sampling.

Table 1 Demographic features and age groups of study participants

Criteria		Frequency	Percent
Age	Under 17	4	1.90%
	18–24	142	66.70%
	25–35	33	15.50%
	35–55	29	13.60%
	Above 55	5	2.30%
Part of India	North India	159	74.60%
	South India	23	10.80%
	Central India	7	3.30%
	East India	4	1.88%
	Northeast India	2	0.92%
	West India	18	8.50%
Locality	Urban/metropolitan	162	76.10%
	Suburban	42	19.70%
	Rural/village	9	4.20%

Comparison of Shopping Practices Prior to and Post-March 2020

Figure 1a displays the number of orders placed by consumers via e-commerce companies per month prior to and post-March 2020. This data is further used to compute the difference in the average number of orders purchased by the study participants on a monthly basis for both cases. It may be noted that for the case of 'over five times', both the maximum and minimum number of orders are taken as 6.

The minimum and maximum number of orders placed by the survey respondents have been computed as follows:

Minimum number of orders:
Σ (Lower limit of frequency of orders per month × number of respondents who chose each option).
Maximum number of orders:
Σ (Upper limit of frequency of orders per month × number of respondents who chose each option).

Table 2 shows the computed average number of orders prior to and post-March 2020. The average has been computed by calculating the minimum and maximum number of orders per month based on the survey responses and calculating the mean. The average number of monthly orders was observed to be more for the 'prior to March 2020' period than for the 'post-March 2020' period by a value of 199.

Figure 1b shows the different kinds of products ordered prior to March 2020 and post-March 2020. A significant increase was observed in the frequency of products purchased from the groceries sector (by 122%). Following this, there was an increase in the purchase of household products (by 48%) and footwear, cosmetics, and personal care (by 28%). This trend of increasing frequency of purchase was displayed by the rest of the product types as well. The COVID-19 pandemic has thus led to an increase in society's dependence on the e-commerce market for a variety of products ranging from groceries to fashion, and even furniture, electronics, and household products.

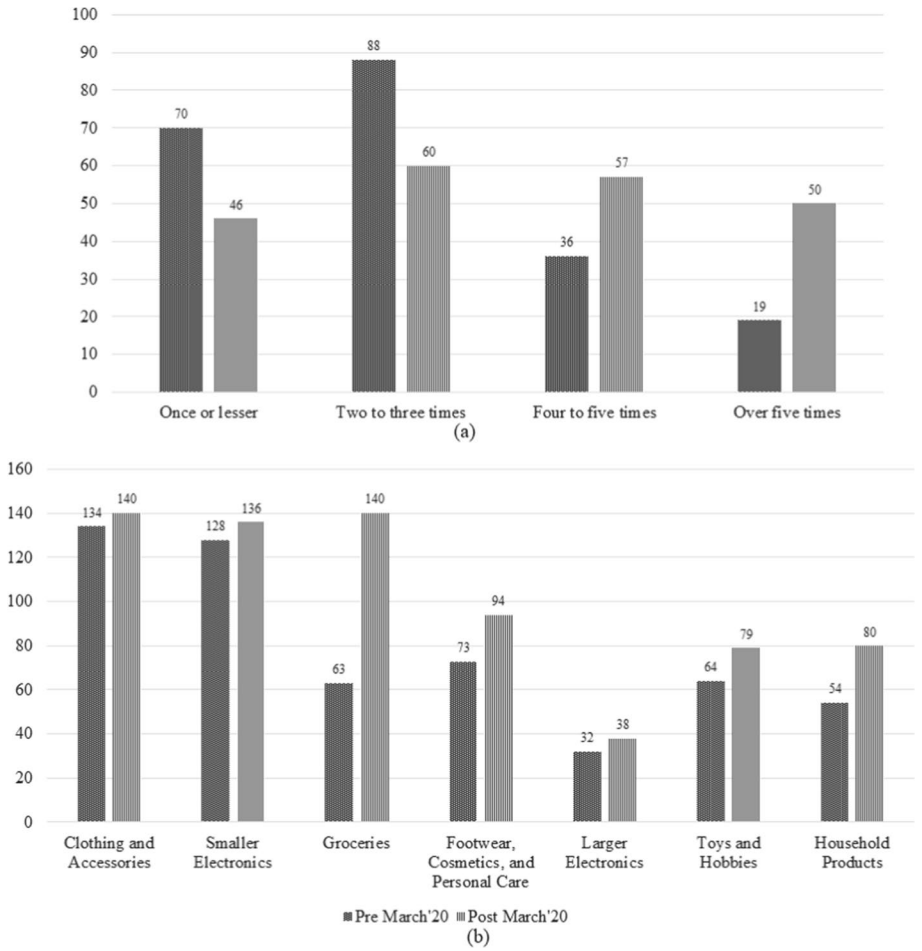


Fig. 1 **a** Graph displaying the number of orders placed by consumers via e-commerce companies per month prior to and post-March 2020. **b** Graph displaying the frequency of different products purchased prior to and post-March 2020

Table 2 Calculating the average number of orders per month prior to and post-March 2020

Net number of orders	Prior to March 2020	Post-March 2020
Minimum Number of Orders	434	648
Maximum Number of Orders	628	811
Average	531	729.5

Consumer Observations on Packaging Materials and Practices Employed by E-commerce Companies

Table 3 displays the consumer observations on packaging materials and practices employed by e-commerce companies. Study participants were allowed to choose multiple options for

Table 3 Frequency of most common kinds of packaging materials

Type of packaging material	Frequency
Poly mailers	138
Bubble wrap	132
Corrugated boxes	124
Tapes	114
Paper bags	93
Plastic packaging sheets	92
Bubble mailers	75
Sustainable packaging materials	31
Cloth bags	27

this question. It was observed that most of the orders of the study participants came packaged in poly mailers (138), bubble wrap (132), and finally corrugated boxes (124).

E-commerce companies enwrap their products in multiple layers of (excessive) packaging materials to ensure health and safety conditions as well as increase the aesthetic nature of their product [2, 9–10, 14]. As displayed in Fig. 2a, majority (72.3%, $n=154$) responded that they do get an average amount of excessive packaging materials with their products, such as bubble wrap, styrofoam, packing peanuts, and air pockets.

Consumer Awareness, Practices, and Opinion on Handling E-commerce Packaging Waste and the Packaging Waste Problem

Figure 2b shows the different methods that were selected by the study participants for the handling and disposing of packaging materials that come along with their order. In the results, 38.5% ($n=82$) stored the items that could be reused and disposed of the rest with household wastes, 25.4% ($n=54$) disposed of the waste without carrying out segregation at the source, 22.5% ($n=48$) stored items that could be reused and segregated the rest in order to dispose of it carefully by means of recycling and composting, 7.5% ($n=16$) dropped off recyclable items for recycling and disposed of the rest with other household wastes, and 0.5% ($n=1$) returned the packaging to the e-commerce company or any other third party for reusing. Disposal of packaging material is observed to be the chosen method of dealing with packaging wastes for a majority of the study participants.

Furthermore, according to the survey result analysis, most study participants (70.4%, $n=150$) were aware of the impacts of packaging, and of this majority, 77.9% ($n=117$) considered it to be a major problem in the present world. From this, it can be understood that consumers are in fact aware of the problem and would possibly be well receptive to solutions. It should be noted however that this would be dependent on multiple factors such as the efforts that would have to be put in from the consumer's side, additional expenses, convenience, etc. [24].

As displayed in Fig. 2c, the respondents identified the bodies they hold responsible for the ongoing excessive packaging problem. Apart from the options given below, there were 4 individuals who had mentioned through the 'others' option, who they thought was responsible for the excessive packaging problem. These included the following: inefficient distribution, transport and delivery companies, the logistics team, and the lack of feasible alternative options in the market.

Fig. 2 **a** Pie chart displaying the average amount of packaging material received per order. **b** Pie chart displaying census of different modes of disposing of e-commerce packaging. **c** Pie chart displaying consumers' perspective on which body should be held accountable for the packaging problem

It can hence be observed that the majority of the respondents believe that the e-commerce company and sellers have played a major role in the packaging problem and would want the solutions to be mainly e-commerce company and seller centric.

Calculating Carbon Footprint

From the survey results, the most common packaging materials were poly mailers and corrugated boxes. Poly mailers are made out of plastic which is known to have a surprisingly low carbon footprint [25]. But on investigating the environmental impact of the non-biodegradable product made out of fossil fuels, it is evident that it has a significant impact on the waste problem in the country affecting both the environment and biodiversity [6, 12, 13]–26. Corrugated boxes may be thought of as an eco-friendly option since they're biodegradable, unlike plastic. However, according to a life cycle analysis of the two, corrugated boxes exceed poly mailers in emission release, fossil fuel, and water consumption by a margin of 595%, 135%, and 404%, respectively [25].

In India, only 48% of the waste is segregated and only 37.2% of waste is processed [26]. The rest of the waste is disposed of in dumping zones and landfill sites. The survey results presented above Fig. 2b also indicate this result, by revealing that almost 69.4% of the individuals dump most of their wastes in the general waste bin indicating unsegregated waste in their households.

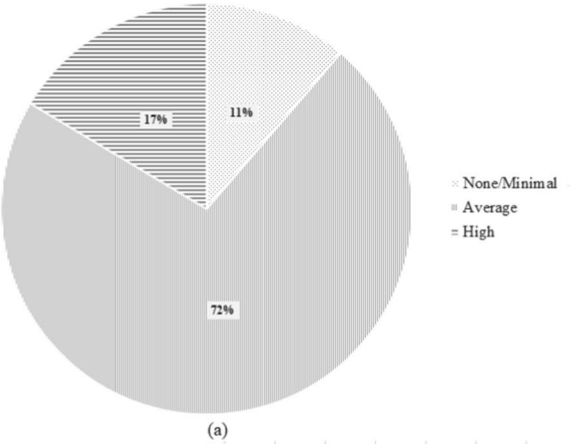
Carbon footprint is a measure of the amount of equivalent carbon dioxide released due to an action and hence acts as an environmental indicator [27, 28]. The purpose of carrying out a carbon footprint analysis of the discarded packaging which is usually disposed of by landfilling in India is to display that packaging often leads to excessive resource consumption and greater greenhouse gas emissions thereby contributing to climate change.

The carbon footprint calculation was done in accordance with the Greenhouse Gas Protocol [29, 30]. Through the method, one can estimate the emissions by computing the product of the activity data and the emission factor.

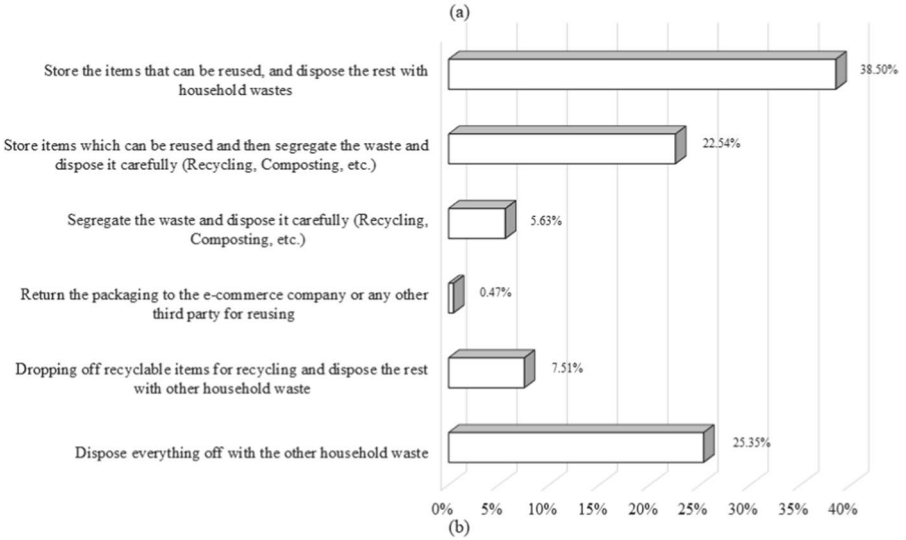
The activity data, as observed in various studies, usually indicates the activity, in this scenario, the quantity of waste. To estimate the quantity of corrugated boxes and poly mailers respectively, a unique methodology was employed. Firstly, the survey results were used to calculate the ratio of the number of orders received in corrugated boxes and plastic polybags (two of the most used packaging materials), respectively; the ratio was then used to calculate the number of corrugated boxes and poly mailers and their proportions (by weight) in one collective ton of waste.

Secondly, the emission factors for landfilled poly mailers and corrugated boxes were derived from secondary data sources.

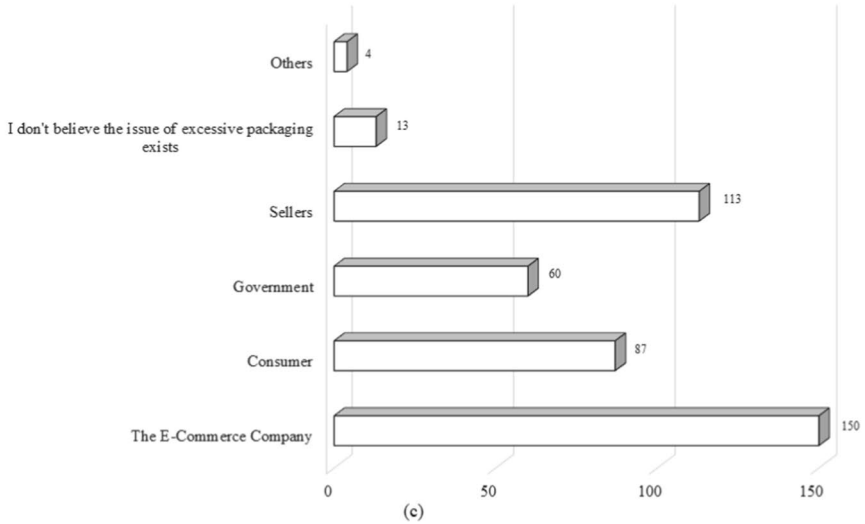
As mentioned in a comparative study, there exist four different methodologies to compute the carbon footprint of a product; these differ based on greenhouse gas choice, system settings, method of quantification, and other aspects. However, quantisation based on GHG activity data multiplied by GHG emission factors is the most widely used and recommended [31].


11%
17%
72%
None/Minimal
Average
High

(a)


Store the items that can be reused, and dispose the rest with household wastes
38.50%
Store items which can be reused and then segregate the waste and dispose it carefully (Recycling, Composting, etc.)
22.54%
Segregate the waste and dispose it carefully (Recycling, Composting, etc.)
5.63%
Return the packaging to the e-commerce company or any other third party for reusing
0.47%
Dropping off recyclable items for recycling and dispose the rest with other household waste
7.51%
Dispose everything off with the other household waste
25.35%
0%
5%
10%
15%
20%
25%
30%
35%
40%

(b)


Others
4
I don't believe the issue of excessive packaging exists
13
Sellers
113
Government
60
Consumer
87
The E-Commerce Company
150
0
50
100
150

(c)

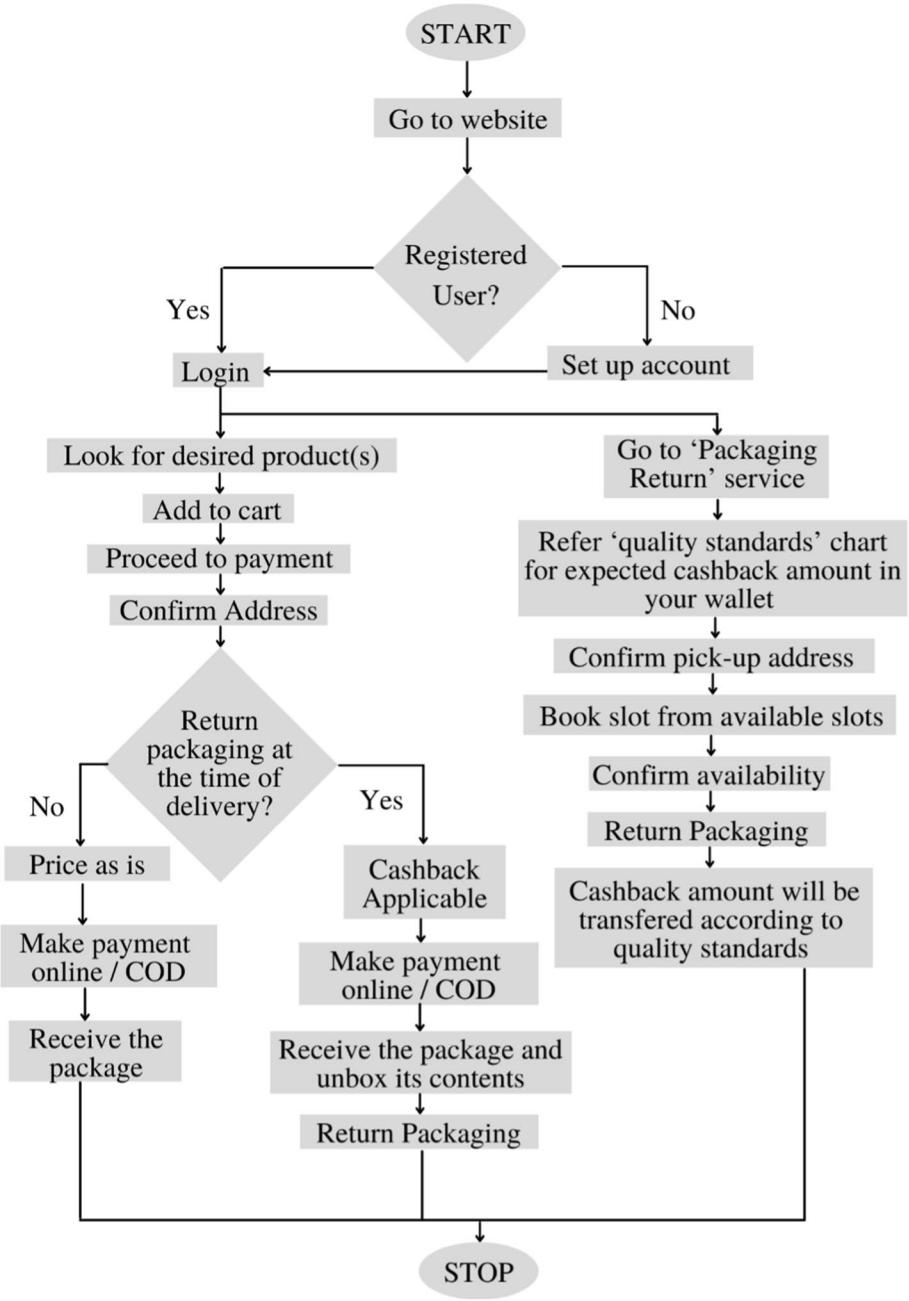


Fig. 3 User interface flow chart for the packaging reuse model

Assumptions

(A) Each order is delivered in one single poly mailer or one single corrugated box. (B) The poly mailer and corrugated box are both assumed to be medium-sized, and their weights have been estimated with the help of secondary data we received on an e-commerce website that sells them. (C) Assume that at the end of their life they are discarded by the user (without recycling) and are disposed of by landfilling (majority result). (D) Carbon footprint is calculated for 1 tonne of waste generated due to disposal of poly mailers and corrugated boxes by landfilling, using the ratio we have derived from our survey.

Table 4 displays the material characteristics taken into consideration for the calculation of the carbon footprint of e-commerce packaging. A majority of the respondents (73%, $n = 156$ for corrugated boxes; 56%, $n = 114$ for poly mailers) chose medium as the most common size of the packaging they received. The corresponding material weight for medium-size packaging was decided upon after consulting the product packaging catalogues of various e-commerce platforms.

Steps Involved in Evaluating Carbon Footprint of 1 Tonne of Waste Comprising of Poly Mailers and Corrugated Boxes in the Ratio We Have Obtained from Our Survey

Step 1 Find the ratio of the number of orders received in poly mailers to the number of orders received in corrugated boxes:

$$Ratio = \frac{Number\ of\ Corrugated\ Boxes}{Number\ of\ Polymailers} = \frac{124}{138} = \frac{62}{69}$$

Step 2 Extrapolate the ratio using the respective weight of the packaging material to obtain 1 tonne of mixed waste (poly mailers and corrugated boxes). Assume x to be the common number to multiply the ratio with. In this case, the net number of poly mailers and corrugated boxes would be $69\times$ and $62x$, and their net weight is $69\times x\times 15$ and $62\times x\times 150$, respectively.

Finding x:

$$Net\ weight\ of\ Polymailers + Net\ weight\ of\ Corrugated\ Boxes = 1\ tonne = 10^6 grams$$

$$(69 \times x \times 15) + (62 \times x \times 150) = 1,000,000$$

$$1035x + 9300x = 1,000,000$$

$$10335x = 1,000,000$$

$$x = \frac{1,000,000}{10335} = 96.76$$

Table 4 Preliminary data presentation for carbon footprint computation

Material type	Material weight	Number of people who chose it (81 overlap)	Emission factor
Plastic (HDPE), landfilled	Medium size = 15 g	138	1200 kg CO_2e per metric tonne (production) + 40 kg CO_2e per metric tonne (landfilling) = 1240 kg CO_2e per metric tonne (Jusuf ZEKIRI & Vjollca Visoka HASANI, 2015)
Corrugated boxes, landfilled	Medium size = 150 g	124	2869 kg CO_2e per metric tonne (Flexpack, 2017)

Step 3 Find the net number of poly mailers and corrugated boxes by substituting the value of x in the table above

$$Net\ number\ of\ Polymailers = 69x = 69 \times 96.76 = 6676$$

$$Net\ number\ of\ Corrugated\ Boxes = 62x = 62 \times 96.76 = 5999$$

Step 4 Net weight of poly mailer and corrugated boxes respectively in tonnes

$$Net\ weight\ of\ Polymailers\ in\ Tonnes = \frac{Net\ number\ of\ Polymailers \times Weight\ of\ 1\ Polymailers}{10^6}$$
$$= \frac{6676 \times 15}{10^6} = 0.1001 tonne$$

$$Net\ weight\ of\ Corrugated\ Boxes\ in\ Tonnes = \frac{Net\ number \times Weight\ of\ 1 Corrugated\ box}{10^6}$$
$$= \frac{5999 \times 15}{10^6} = 0.8999 tonne$$

Step 5 Net emissions due to 1 tonne of landfilled poly mailers and corrugated boxes

$$\begin{aligned} Net\ emissions &= Emissions\ due\ to 0.1001\ tonne\ of\ Polymailers \\ &+ Emissions\ due\ to\ 0.8999\ tonne\ of\ Corrugated\ boxes \\ &= (0.1001 \times 1240) + (0.8999 \times 2869) = (124.124 + 2581.813) \\ &= 2,705.937 kg\ CO2e\ per\ Metric\ Tonne \end{aligned}$$

According to the above results, it is visible that one tonne of waste consists of poly mailers and cardboard boxes, when discarded in the landfills result in the emission of 2705.937 kg CO2 equivalent. Table 5 summarises the carbon footprint computation.

Considering the majority of our sample has begun to realise that packaging poses a major problem and that they mostly place responsibility on the e-commerce companies, it is evident that people would support action to be taken up by the higher authorities (e-commerce companies).

Hence, keeping in line with the Governments' revised Solid Waste Management Rules (mandating Extended Producer Responsibility—for the management of product and packaging waste) and our survey, we have put forth a model through which companies can not

only reduce virgin packaging and hence reduce their carbon footprint but also become more economical and comply with the government rules [1, 5, 16, 28, 32, 33].

Proposal and Recommendations

Reusing packaging materials as many times as possible before disposal can cut down on the emissions associated with producing virgin packaging material [5]. So far, there is no reusable packaging system available for e-commerce that is less expensive than single-use options [34]. In order to tackle this problem, we designed a packaging reuse model for e-commerce companies through which they can reuse packaging as long as it is in good condition.

Overview

E-commerce companies can take back the packaging if it is in usable condition (refer to quality standards section) and reuse the packaging for future orders. The customer will receive a cashback (amount to be determined by the company) according to the condition of the packaging returned.

Packaging Pickup

The user interface will be as in Fig. 3. The customer is provided with the following options:

Case 1: Returning packaging at the time of delivery.

A. Return packaging from the previous order when the new order is delivered.
B. Return packaging from previous as well as current orders when the new order is delivered.

Case 2: Scheduling an appointment for package pickup. In case the customer is not able to return the packaging at the time of delivery (e.g. due to not being at home or the package being a gift), the customer can schedule a time for the packaging to be picked up. Pickup timings displayed to the customer will be when packages would be delivered to nearby areas (radius can be determined by the company) in order to reduce extra trips to pick up packaging (to avoid higher fuel consumption and emissions).

Table 5 Summary of carbon footprint computation

Quantity	Poly mailer	Corrugated boxes
Number of orders received in each packaging	138	124
Weight of one mailer/box	15 g	150 g
Net number of the packaging to make 1 tonne in the same ratio as derived in survey	6676	5999
Net weight in tonnes of each type of packaging	0.1001 tonne	0.8999 tonne
Net emission due to each packaging type for respective net weight	1240	2869

Quality Standards Chart

The quality standards correlate the condition of the packaging to the expected cashback. It serves the purpose of conveying the preferred condition of the packaging to the customers and maintaining a standard in terms of the packaging returned.

Three cases may be considered: (A) for packaging as good as new or in reusable condition, the cashback would be as determined by the company; (B) for packaging in recyclable condition, the cashback may be 25–50% lower than cashback in case A; and (C) for packaging in unusable condition, no refund may be offered by the company.

Cashback

The cashback is an incentive to ensure that customers handle packaging responsibly and encourages customers to return packaging.

Advantages of the Packaging Reuse Model

A. Reduces costs for the e-commerce company: The e-commerce company would have to invest substantially less in new packaging once the model stabilises. There will not be significant fuel costs to pick up packaging as delivery men would only be picking up packaging when delivering orders to nearby areas.
B. Reduces energy consumption and emissions: With recycling rates of packaging being low due to contamination and/or lack of resources, this model prioritises poly mailers over recycling thereby reducing energy used at recycling facilities and consequently emissions produced.
C. Ensures responsible handling of packaging: The customers would handle the packaging carefully, keeping the incentive in mind. This further increases the rate of reuse.
D. Increased customer loyalty: With the option of returning packaging at the time of the next delivery, customers are more likely to choose that particular e-commerce platform for subsequent orders.
E. Attract environmentally conscious customers: Customers who care about the environment are more likely to order from e-commerce companies adopting such models [5, 9, 14, 34–36].

Modifications to Packaging

There are a few changes to packaging that can be made in order to make the circular packaging system more efficient [5, 14, 25, 35, 37]. They are listed below.

Resealable Plastic Poly Mailers

Poly mailers are cut or torn open due to the adhesive being too strong. These cut poly mailers can only be reused a couple of times if they have to be cut every time to open, as the size of the poly mailer would keep decreasing. Along with this, the reused packaging would not look as neat as virgin packaging and customers may not opt for reused packaging. Resealable plastic poly mailers would enable the packaging to be reused multiple times. In order to ensure that the packaging has not been opened by someone else and the

contents emptied, there can be a seal such as a sticker across the opening which serves as an indication.

Translucent Pocket for Order Invoice

The order invoice is usually a sticker that is stuck on the packaging. Removal of the sticker might cause the plastic poly mailer to tear or the top skin of corrugated boxes peeling off. This disables the plastic poly mailers from being reused and reduces the aesthetic appeal of the cardboard boxes and consequently reduces its structural integrity when the sticker gets peeled off multiple times as a result of reusing. New stickers cannot be placed on old stickers as this is a privacy concern in addition to the limitation of the number of times new stickers can be repeatedly placed on top of old ones. In order to avoid this, translucent resealable pockets can be provided on the packaging into which the order invoice can be slid.

Reusable Packaging

Cardboard boxes and poly mailers are designed to be single-use packaging materials. They are hence subject to wear and tear easily and are limited in terms of the number of times they can be reused. There are several private foreign companies that have launched reusable packaging products (that last between 20 and 1000 cycles) specifically for e-commerce. However, research still needs to be done to identify whether emissions released by these products are in fact lesser than those by the conventional ones. And whether they are more economical.

Theoretical and Practical Implications

This study has contributions to the research on e-commerce virgin and overpackaging solutions. First, this study connected the COVID-19 pandemic to an increase in the usage and disposal of virgin e-commerce packaging. Previous literature has discussed the impact of COVID-19 on the e-commerce industry, most papers focusing on the increased popularity of e-commerce platforms [3, 38, 39], but very few linking it with an increase in e-commerce packaging [1]. Some papers have discussed COVID-19 and packaging but explored the aspect of food-related e-commerce only [2, 13, 40].

Second, this study also proposes a novel solution to the e-commerce packaging problem. Solutions proposed by previous studies focus on reusable materials and alternative materials [1, 24, 35] and changes in packaging design [1]. Although the adoption of reusable packaging is a promising solution, there are some inconveniences such as e-commerce companies having to restructure their business model, reluctance of customers to pay more for reusable alternatives, etc. [24]. This implies that there could be a possible delay with regard to the implementation of these solutions. Not only does this study propose a solution that involves extending the lifespan of the single-use packaging materials already in use, it also offers an incentive to encourage customers in the form of cashbacks. E-commerce companies would not have to make drastic changes in the packaging process thereby enabling easy and quick implementation and they would also gain customer support.

Third, this study uses a survey questionnaire to gain consumer insight and feedback on multiple aspects of e-commerce packaging such as changes in e-commerce practices due to the pandemic, packaging received by customers, how consumers deal with the packaging waste produced, and finally, consumer awareness and feedback. Although survey centric studies related to packaging have been carried out previously, they were to establish only qualitative results [6, 19]. To our knowledge, this is the first study related to e-commerce packaging to conduct both quantitative and quantitative analyses on the survey results. The survey results were used to conduct carbon footprint calculations of the estimated packaging waste produced by the respondents. The method although logical is not something that has been applied previously. It not only took the survey responses into account but also accounted for the weight of the packaging material. The method would hence be applicable for quantifying waste breakdowns in future studies.

To the best of our knowledge, this is the first study to analyse the consumption of packaging by the e-commerce industry with the help of a household survey and propose a circular model for packaging reuse in India. From the viewpoint of developing countries, this research would contribute to increased awareness and the subsequent adoption of sustainable practices by e-commerce companies on the handling of delivery packaging. It also sheds light on public awareness and attitude towards solid waste management practices in the country. The current paper's findings will benefit the following groups: manufacturers, who on reusing their packaging would eventually cut down on their raw material expenditure; and informal agents, who would be engaged in job creation to allow ease in the collection of packaging material from the consumers. This will help in limiting the environmental impact of virgin packaging, while also reducing the load on solid waste treatment plants.

Conclusion

The packaging problem is associated with environmental degradation, land pollution, and a negative impact on the health of humans and biodiversity alike. It has been further exacerbated due to the COVID-19 pandemic and is one of the biggest drawbacks of the e-commerce industry. On analysing the valid responses (by 213 respondents), the following conclusions were laid down on e-commerce packaging trends and consumers' purchasing behaviour: (A) the average monthly orders and sales of e-commerce companies have increased considerably when compared to the pre-Covid era; (B) with respect to packaging waste management, the majority of the individuals stored the items that could be reused, and disposed of the rest with household wastes; (C) the most common packaging materials included poly mailers, bubble wrap, and corrugated boxes; and (D) majority of the sample were aware of the impacts of packaging on the environment, and believed that the E-commerce company played a crucial role in the packaging problem. For our survey respondents, it was computed that 2705.937 kg CO_2e is released for 1 tonne of waste consisting of poly mailers and corrugated boxes when discarded in the landfills.

The incentivised packaging reuse model can prove to be an effective method in order to cut down on the emissions associated with producing virgin packaging material. To optimise the model, certain modifications can be made to current packaging materials like resealable plastic poly mailers and translucent pockets for the order invoice. Furthermore, reusable packaging materials can be introduced into the packaging reuse model to extend the longevity of the packaging materials. Large enterprises that are not able to manage

their own packaging waste may hire PROs (Producer Responsibility Organisations), to manage the packaging waste for them [32].

There are a few limitations of the study: (1) the utilisation of a convenience sample as the primary data set and the limited number of responses (213 valid responses) has resulted in an unequal representation of individuals and hence prevent the study from being generalised to the population of India; and (2) non-inclusion of materials like tape, bubble wrap, and styrofoam while calculating the carbon footprint, due to lack of concrete data. But nevertheless, the study gives direction to future researchers regarding the e-commerce packaging problem and possible solutions. Future studies aimed at calculating and reducing the environmental impact of packaging for an increasingly hi-tech world, should be oriented towards analysing the feasibility of launching a packaging reuse model, understanding the consumers' propensity to abide by such packaging reuse models, and tackling the over-packaging problem through innovations in packaging design and material.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43615-022-00231-4.

Acknowledgements We would like to thank the professors at different colleges, our peers, and survey respondents who helped circulate and responded to the survey, which provided as the primary data base for our research paper.

Author Contribution TA and SRC drafted the article, designed and circulated the survey, and analysed and interpreted the survey responses. NS designed the packaging reuse model and aided in the designing the survey and editing the manuscript. LG offered guidance, performed critical revisions of the article at several stages, and aided in designing and circulating the survey. All authors read and approved the final manuscript.

Data Availability The data that support the findings of the study are the survey responses. They have been included in the supplementary material. The spreadsheet containing the responses is also linked here.

Declarations

Ethics Approval and Consent to Participate The study involved the use of human participants in the form of an anonymous survey. All participants were fully informed that their anonymity is ensured, why the research is being conducted, as well as how the data would be utilised. This information was included in the email body which was sent with the survey link as well as the introductory section of the questionnaire. Personal data collected involved age, which part of India they reside in as well as the locality. This was purely for demographic analysis and to observe any trends. It was made clear to the survey respondents that participation in the survey was voluntary and as per the participant's consent.

Consent for publication Not applicable.

Competing Interests The authors declare no competing interests.

References

1. Escursell S, Llorach-Massana P, Roncero MB (2021) Sustainability in e-commerce packaging: a review. J Clean Prod 280:124314. https://doi.org/10.1016/J.JCLEPRO.2020.124314
2. Filho WL, Voronova V, Kloga M, Paço A, Minhas A, Salvia AL, Ferreira CD, Sivapalan S (2021) COVID-19 and waste production in households: a trend analysis. Sci Total Environ 777:145997. https://doi.org/10.1016/J.SCITOTENV.2021.145997
3. Elrhim MA, Elsayed A (2020) The effect of COVID-19 spread on the e-commerce market: the case of the 5 largest e-commerce companies in the world. SSRN Electron J. https://doi.org/10.2139/SSRN.3621166

4. Tiwari S, Singh P, "Environmental impacts of e-commerce," *IPCBEE*, 2011. [Online]. Available: https://www.researchgate.net/profile/Sunita-Tiwari-2/publication/268343962_Environmental_Impacts_of_E-Commerce/links/553b515c0cf245bdd7646ad0/Environmental-Impacts-of-E-Commerce.pdf. Accessed 14 Apr 2022.
5. Xie G, Huang L, Apostolidis C, Huang Z, Cai W, Li G (2021) Assessing consumer preference for overpackaging solutions in e-commerce. Int J Environ Res Public Health 18(15). https://doi.org/10.3390/IJERPH18157951
6. Chueamuangphan K, Kashyap P, Visvanathan C (2020) Packaging waste from e-commerce: consumers' awareness and concern. Sustainable waste management: Policies and Case Studies 27–41. https://doi.org/10.1007/978-981-13-7071-7_3
7. Pålsson H, Pettersson F, Winslott Hiselius L (2017) Energy consumption in e-commerce versus conventional trade channels - insights into packaging, the last mile, unsold products and product returns. J Clean Prod 164, 765–778. https://doi.org/10.1016/J.JCLEPRO.2017.06.242
8. Sui DZ, Rejeski DW (2002) Environmental impacts of the emerging digital economy: the e-for-environment e-commerce? Environ Manag 29:2, 29(2):155–163. https://doi.org/10.1007/S00267-001-0027-X
9. Gurumoorthy KS, Sanyal S, Chaoji V (2021) Think out of the package: recommending package types for e-commerce shipments. Lect Notes Comput Sci (Incl Subseries Lect Notes Artif Intell Lect Notes Bioinforma) LNAI 12461:290–305. https://doi.org/10.1007/978-3-030-67670-4_18
10. Anderson A, Chandralingam R, PraveenKumar TR (2021) Impact of COVID-19 pandemic on plastic surge and environmental effects. Energy Sources Part A: Recover Utilization Environ Effects. https://doi.org/10.1080/15567036.2021.1900456
11. Mordor Intelligence Industry Report (2022) Packaging industry in India size, growth, analysis (2022 - 27)," 2021. Accessed: Apr. 14. [Online]. Available: https://www.mordorintelligence.com/industry-reports/packaging-industry-in-india
12. Susan E. M. Selke (2022) Packaging and the environment : alternatives, trends, and solutions. Lancaster, Pa. : Technomic Pub. Co., 1994. https://archive.org/details/packagingenviron0000selk_a8d7 . Accessed 14 Apr 2022.
13. van Fan Y, Jiang P, Hemzal M, Klemeš JJ (2021) An update of COVID-19 influence on waste management. Sci Total Environ 754 https://doi.org/10.1016/J.SCITOTENV.2020.142014
14. Meherishi L, Narayana SA, Ranjani KS (2019) Sustainable packaging for supply chain management in the circular economy: a review. J Clean Prod 237:117582. https://doi.org/10.1016/J.JCLEPRO.2019.07.057
15. Kaza S, Yao L, Bhada-Tata P, van Woerden F (2018) What a waste 2.0 a global snapshot of solid waste management to 2050
16. Pani SK, Pathak AA (2021) Managing plastic packaging waste in emerging economies: the case of EPR in India. J Environ Manag 288:112405. https://doi.org/10.1016/J.JENVMAN.2021.112405
17. Ncube LK, Ude AU, Ogunmuyiwa EN, Zulkifli R, Beas IN (2021) An overview of plastic waste generation and management in food packaging industries. Recycling 6(1):12. https://doi.org/10.3390/RECYCLING6010012
18. Groh KJ, Backhaus T, Carney-Almroth B, Geueke B, Inostroza PA, Lennquist A, Leslie HA, Maffini M, Slunge D, Trasande L, Warhurst AM, Muncke J (2019) Overview of known plastic packaging-associated chemicals and their hazards. Sci Total Environ 651:3253–3268. https://doi.org/10.1016/J.SCITOTENV.2018.10.015
19. Lu S, Yang L, Liu W, Jia L (2020) User preference for electronic commerce overpackaging solutions: implications for cleaner production. J Clean Prod 258:120936. https://doi.org/10.1016/J.JCLEPRO.2020.120936
20. Datt G, Ravallion M, Murgai R (2016) Growth urbanization and poverty reduction in India. https://doi.org/10.3386/W21983
21. Vij D (2012) Urbanization and solid waste management in India: present practices and future challenges. Procedia Soc Behav Sci 37:437–447. https://doi.org/10.1016/J.SBSPRO.2012.03.309
22. Singh RP, Chauhan A (2020) Impact of lockdown on air quality in India during COVID-19 pandemic. https://doi.org/10.1007/s11869-020-00863-1/Published
23. 2022 Statistical Sampling and Survey Design Services Global Market Size Growth Report with COVID-19 Impact. (n.d.). Retrieved November 17, 2022, from https://www.researchandmarkets.com/reports/5481490/2022-statistical-sampling-and-survey-design
24. Coelho PM, Corona B, ten Klooster R, Worrell E (2020) Sustainability of reusable packaging–current situation and trends. Resour Conserv Recycl X 6:100037. https://doi.org/10.1016/J.RCRX.2020.100037

25. Flexpack (2022) Packaging industry in India," 2017. Accessed: [Online]. Available: https://www.walkwatertalent.com/wp-content/uploads/2020/01/Industry-Paper-Packaging-Industry-in-India.pdf
26. Shukla P, Sharma PK, Pandey S, Chintala V (2021) Unsegregated municipal solid waste in India - current scenario, challenges and way forward. Nature Environ Pollut Technol 20(2):851–863. https://doi.org/10.46488/NEPT.2021.V20I02.048
27. Lo-Iacono-Ferreira VG, Viñoles-Cebolla R, Bastante-Ceca MJ, Capuz-Rizo SF (2020) Transport of Spanish fruit and vegetables in cardboard boxes: a carbon footprint analysis. J Clean Prod 244 https://doi.org/10.1016/J.JCLEPRO.2019.118784
28. Su Y, Duan H, Wang Z, Song G, Kang P, Chen D (2020) Characterizing the environmental impact of packaging materials for express delivery via life cycle assessment. J Clean Prod 274:122961. https://doi.org/10.1016/J.JCLEPRO.2020.122961
29. Scope 3 Calculation Guidance | Greenhouse Gas Protocol. (n.d.). Retrieved November 17, 2022, from https://ghgprotocol.org/scope-3-technical-calculation-guidance
30. World Resources Institute and World Business Council for Sustainable Development (2022) The greenhouse gas protocol - revised edition." Accessed: [Online]. Available: https://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf
31. Gao T, Liu Q, Wang J (2014) A comparative study of carbon footprint and assessment standards. Int J Low Carbon Technol 9(3):237–243. https://doi.org/10.1093/IJLCT/CTT041
32. Duan H, Song G, Qu S, Dong X, Xu M (2019) Post-consumer packaging waste from express delivery in China. Resour Conserv Recycl 144:137–143. https://doi.org/10.1016/J.RESCONREC.2019.01.037
33. Ragazzi M, Katsoyiannis I, Magaril E, Viotti P, Al-Kayiem HH, Schiavon M, Ionescu G, Sliusar N, Hossain R, Islam T, Shanker R, Khan D, Elizabeth K, Locock S, Ghose A, Schandl H, Dhodapkar R, Sahajwalla V (2022) Plastic waste management in India: challenges, opportunities, and roadmap for circular economy. https://doi.org/10.3390/su14084425
34. Zimmermann Till, Bliklen Rebecca (2020) Single-use vs. reusable packaging in e-commerce: comparing carbon footprints and identifying break-even points. GAIA - Ecological Perspectives for Science and Society 29(3). https://doi.org/10.14512/gaia.29.3.8
35. Baskoro ML (2020) Exploring the emerging trends of reusable shipping packaging for e-commerce. JISA(Jurnal Informatika Dan Sains) 3(2). https://doi.org/10.31326/JISA.V3I2.832
36. Ross S, Evans D (2003) The environmental effect of reusing and recycling a plastic-based packaging system. J Clean Prod 11(5):561–571. https://doi.org/10.1016/S0959-6526(02)00089-6
37. Jusuf ZEKIRI, Vjollca Visoka HASANI (2015) The role and impact of the packaging effect on consumer buying behaviour. Ecoforum 4(1). https://www.ecoforumjournal.ro/index.php/eco/article/viewFile/189/167
38. Bhatti A, et al. (2020) E-commerce trends during COVID-19 pandemic. Int J Future Gener Commun Networking 13(2)
39. Galhotra B, Dewan A (2020) Impact of COVID-19 on digital platforms and change in E-commerce shopping trends. Proceedings of the 4th international conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 861–866. https://doi.org/10.1109/I-SMAC49090.2020.9243379
40. Oliveira WQ, de, Azeredo HMC, de, Neri-Numa IA, Pastore GM (2021) Food packaging wastes amid the COVID-19 pandemic: Trends and challenges.Trends Food Sci Technol 116:1195–1199. https://doi.org/10.1016/J.TIFS.2021.05.027
41. IBEF, "E-commerce," 2019. . [Online]. Available: https://www.ibef.org. Accessed 14 Apr 2022
42. (PDF) TERI Analysis on Challenges and opportunities - Plastic Waste Management in India. (n.d.). Retrieved November 17, 2022 from https://www.researchgate.net/publication/325780999_TERI_Analysis_on_Challenges_and_opportunities_-_Plastic_Waste_Management_in_India
43. Joshi R, Ahmed S (2016) Status and challenges of municipal solid waste management in India: a review. Cogent Environ Sci 28:1139434. https://doi.org/10.1080/23311843.2016.1139434

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.

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT).
India Habitat Centre, Lodhi Road, New Delhi, India. Sep 23-25, 2022

Pulse splitting cum width control for transient handling of LLC Resonant Converter

Prajeet Shukla
Electrical Engineering Department,
Delhi Technological University,Delhi
prajeetshukla95@gmail.com

Vishal Verma,
Electrical Engineering Department,
Delhi Technological University,Delhi
vishalverma@dce.ac.in

2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT) | 978-1-6654-9365-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBCONPT57482.2022.9938336

***Abstract*— The Proposed Converter discusses the control algorithm to suppress the inrush starting current of the resonant converter with fast dynamics. The Proposed converter is able to handle the inrush current by switching the converter at the start with twice the resonant frequency and reducing the switching frequency to near resonant frequency while at the steady state condition the variable switching frequency is achieved using look up table method. Initial transients are eliminated using proposed modified hysteresis controller. The 3.3KW, 380V/96V LLC resonant converter outcomes and performance are confirmed using MATLAB/SIMULINK.**

Keywords—LLC, Inrush Current, Transient Control, State Plane, Gain Plot

I. INTRODUCTION

Requirement of highly efficient, compact size SMPS with low EMI emissions have been the focal point for researchers round the globe. Majority applications include Isolated DC-DC Power Supplies for servers, high end audio, LED lighting and variety of home appliances, etc.

LLC Resonant converter has become the usual choice with the advent of high frequency converter technology as it operates with low switching losses, utilizing sinusoidal flux coupling for efficient magnetic circuit operation and exhibits operation with wider input voltage range and load perturbations. However, control of LLC resonant converter is still complex to implement as output is controlled by resonant tank circuitry having fast transient response. Out of many control techniques reported, frequency control is the simpler and easy to implement for LLC converter as the resonant tank impedance can be directly controlled by switching frequency [1,4]. Also, the relationship between the control signal and frequency shows linear characteristics, which can alleviate the need of complex non-linear controllers.

With variable frequency control, resonant tank continuously exchanges pulsating energy with the source resulting into circulating losses which becomes worst at light load conditions where system efficiency is most adversely affected. Burst mode control is often employed to improve efficiency at light loads . Beside this, at staring the converter witness a large inrush current and voltage stresses on primary switches which may lead to burning of the switches [6,7]. Optimal trajectory control is presented in the literature to demonstrate effective operation, but due to involvement of complex mathematical expressions, duly aided by multiple sensors become tedious to implement on fixed point microcontrollers [5,13]. Average geometric control is discussed in literature [6,7] which is based on average large signal model to predict large signal behavior and consequently controls the transients at the cost of large computations.

The proposed scheme focuses on simplifying the control scheme for implementation on inexpensive fixed-point microcontroller by avoiding the complex mathematical expressions and geometric modeling-based algorithm by using preventive measures to mitigate the inrush current during transients. The transient conditions are handled through new method by splitting the gate pulses into two halves. The propose scheme, thus operating the converter at switching frequency twice the resonant frequency along with variable pulse width, utilizing the increased impedance due to double frequency to limit current and pulse width modulation for voltage magnitude control. Variable Frequency control of the LLC converter is exercised through 1-D look up table along with simple PI controller to customize the operation amidst load perturbation.

II. CIRCUIT DESCRIPTION

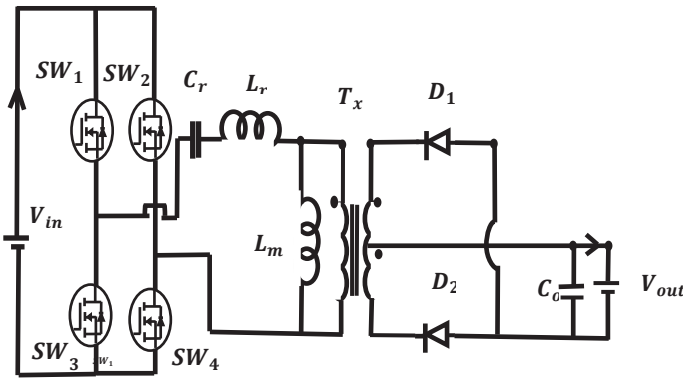


Fig.1 Full Bridge LLC Resonant Converter

The LLC resonant converter circuit is shown in Fig 1, which consists of Class D inverter and center taped Class D rectifier. The resonant tank C_r-L_r-L_m is trapped between the legs of the H bridge inverter. The rectifier section is center tapped and load is connected in parallel to the output capacitive filter C_0. Switches SW_1, SW_4 and SW_2, SW_3 are alternately switched with a duty cycle of almost 50% under steady state operation, thereby handing over the command to

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:10:17 UTC from IEEE Xplore. Restrictions apply.

variable frequency control for tweaking the impedance of resonant tank. DC voltage transfer function of LLC resonant converter is almost insensitive to the load variations. In addition, the circuit has high efficiency over a wide range of load perturbations.

III. DESIGN OF LLC CONVERTER

The equivalent circuit of LLC resonant converter is as shown in the figure. Here, 'V_{ab}' is the generated square wave voltage across resonant tank, 'V_{cd}' is the reflected output voltage, 'R_L' is the reflected load across resonant tank, 'I_m' is the magnetizing current, 'I_r' is the resonant tank current and 'I_o' is the reflected output current.

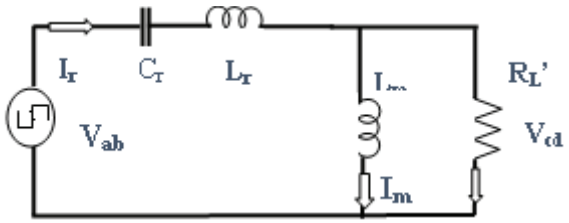


Fig. 2 Equivalent circuit of LLC Converter

A. Gain Equation:

The input to output voltage gain function is defined as [13],

$$M_g^{dc} = \frac{n*V_{cd}}{V_{ab}/2} \quad (1)$$

Based on the Fundamental harmonic approximation (FHA) method, the gain function is approximated by their fundamental voltages,

$$M_{g,1}^{ac} = \frac{V_{cd,1}}{V_{ab,1}} \quad (2)$$

FHA-based gain function can be expressed with circuit parameters as,

$$M_g^{dc} = M_{g,1}^{ac} = \left| \frac{(j\omega L_m)\, R_e}{(j\omega L_m)\, R_e + (j\omega L_r) + 1/(j\omega C_r)} \right| \quad (3)$$

Equation (3) can be expressed in a normalized form, accordingly the normalized gain function is obtained as:

$$M_g^{dc} = M_{g,1}^{ac} = \left| \frac{L_n * {f_n}^2}{[(L_n+1)*{f_n}^2-1]+j({f_n}^2-1)*L_n*f_n*Q_e} \right| \quad (4)$$

where, f_n is normalized switching frequency: $f_n = \frac{f_{sw}}{f_o} = \frac{2\pi f_{sw}}{2\pi f_o} = \frac{\omega_{sw}}{\omega_o} = \omega_n$, L_n is Normalized inductance ratio: $L_n = \frac{L_m}{L_r}$ and Q_e is the quality factor of the series resonant tank: $Q_e = \frac{\sqrt{L_r/C_r}}{R_e}$.

B. Parameter Design:

Transformer's turns ratio:

$$n = M_g * \frac{V_{in}}{V_{out}} = \frac{V_{in}}{V_{out}}\Big|_{M_g=1} = 4.17 \approx 4 \quad (5)$$

$$M_{gmin} = n * \frac{(V_{o_min}+V_F)}{V_{in_max}} = 4 * \frac{(96*(1-5\%)+0.7)}{415} = 0.885 \quad (6)$$

where, V_F is forward voltage drop across switch

$$M_{gmax} = n * \frac{(V_{o_max}+V_F)}{V_{in_min}} = 4 * \frac{(96*(1+5\%)+0.7)}{380} = 1.08 \quad (7)$$

The values of the Quality factor and inductance ratio are selected such that the resulting gain curve plot drawn should cut the maximum gain value and minimum gain value designed for frequency range above the resonant frequency for ensuring ZVS operation. The gain plot is presented in Fig. 2. Accordingly Quality factor, $Q_e = 1.1$; Inductance ratio, $L_n = 3.2$; equivalent resistance, $R_e = \frac{8*n^2}{\pi^2} * \frac{V_o}{I_o} = 36.22\Omega$ and at 110% Loading $R_e = 32.92\Omega$; the chosen resonant frequency $f_r = 150\text{kHz}$; Hence the resonant capacitor $C_r = \frac{1}{2\pi*Q_e*f_r*R_e} = 26.6\text{ nF}$; resonant inductor $L_r = \frac{1}{(2\pi*f_r)^2*C_r} = 42.3\ \mu\text{H}$; and magnetizing inductor $L_m = L_n * L_r = 135.36\mu\text{H}$.

The switch is selected as per the feasible dead time which is given by:

$$t_d \geq 16 * C_{eq} * f_{sw} * L_m \geq 104nsec$$

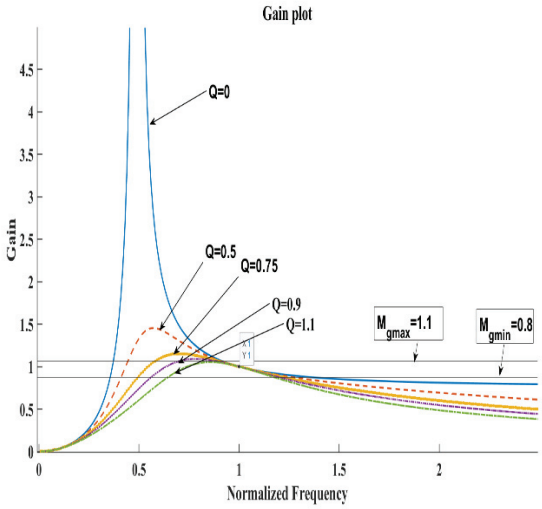


Fig. 3 Gain Plot for the designed LLC converter

For the design of the tank parameters, the switching frequency is kept greater than resonant frequency, thus magnetizing inductance gets clamped by the output voltage and the circuit behaves as a series resonant tank and aids ZVS. Thus, with the change in line voltages and/or load condition the output voltage changes. To keep the output voltage fixed at the required level the feedforward control becomes requisite to modulate the switching frequency at different line and load conditions.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:10:17 UTC from IEEE Xplore. Restrictions apply.

The switching frequency for different load conditions can be precalculated from the gain plot and thus operation of the converter at this precalculated switching frequency is stored in a 2-Darray for lending support to closed loop control without losing time for calculations at different load condition offer proper load regulation. Similarly, line regulation can also be controlled by precalculated switching frequency for different line voltages acquiring input from 2-D array and closed loop proposed control.

TABLE I. CONVERTER DESIGN PARAMETER

Parameters	Value
V_{in}	400 V
V_O	96 V
L_n	3.2
Q_e	1.1
f_r	150 kHz
n	5:1:1
L_r	42.3 uH
C_r	26.6 nF
L_m	136 uH

C. Control Scheme:

The proposed controller employs closed loop frequency control since the relationship between frequency and control signal is linear, thereby giving liberty to use linear PI controller effectively, and avoiding complexity implemented on costly DSP controllers. The block diagram depicting the control logic is as shown in the Fig.4. The steady state control is implemented as voltage controlled derived frequency control realized through the PI controller. f_{error} computed by PI controller in addition to feedforward switching frequency $f_{feedforward}$ component derived from the 2-D look up table holding the data of switching frequency corresponding to different load current values and different input voltage loading forms the reference frequency value designated for resonant tank excitation. Feedforward control offer faster dynamics of the converter. Finally using the VCO, the gate pulses are generated at near 50% duty ratio including the dead time value requisite for discharge from output capacitance of the switch provoking the antiparallel diode to invoke ZVS of primary side switches.

Under steady state the converter is operated at the resonant frequency as efficiency of LLC resonant converter is maximum at this frequency. During transients, if converter is continued to operate near resonant frequency it is bound to experience large inrush current and voltage levels which if not managed may result into the failure of switches. To manage such situation, the reported literature suggests the converter is operated at 2-3 times the resonant frequency, thereby bringing in substantial impedance in the resonant tank. To operate the converter at such high frequency, gate pulses are generated under closed loop control, thus during transient period only reduced or clipped level of inrush currents are allowed in the resonant tank. The proposed control during transient conditions is invoked by chopping the generated gate pulses into two from the middle as shown in the Fig. 5.

The transient state control is made to work in current control by averaging the sensed current and its level monitored thereof. Any transient condition at near resonant frequency raises the current level which is readily detected, and the same is differentiated from steady state condition. The VCO, which earlier was entrusted to generate 50% duty at particular frequency now generates a tapering duty output also starting from near 40% to 0%. Such tapered duty were ANDed and subtracted from 50% duty waveform producing the splitting of pulses at double frequency as shown in the Fig. 5. Such pulse width modulated double frequency pulse not only act on limiting the inrush current through increased resonant tank impedance, but also control the voltage due to modulated pulses.

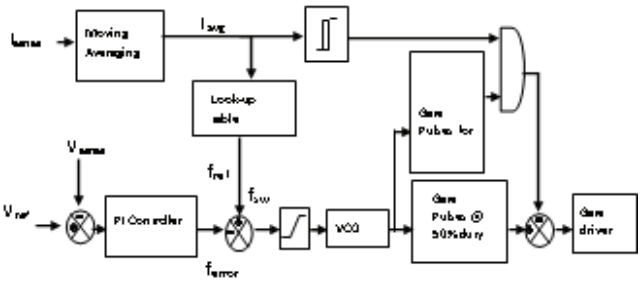


Fig. 4 Block Diagram of control scheme incorporating proposed transient handling control

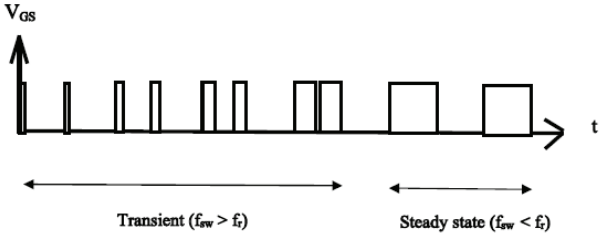


Fig. 5 Gate Pulse chopping for handling transient conditions

IV. PERFORMANCE EVALUATION

During on-load start of high-power converter system, the issue of handling inrush current dominates. The proposed control logic is validated through MATLAB simulation result for transient conditions. The converter is shown supplying the load of 3.3kW at 96V from 380V DC Bus. The parameters of resonant tank circuits are L_r = 42.3μH, L_m = 136 μH and C_r =26.6nF which is operating closely above the resonant frequency of 150kHz. Two types of transient responses are investigated in our study; the starting response where the magnetic circuit is routed along with sudden loading of source, transition from no load to full load and full load to no load. Fig,7 depicts converter tank response waveform while looping the operation of tank circuit at around resonant frequency. It may be observed that transient peak voltage may go up to 3Kv, and the peak current may mount up to 75A, while diode current may witness two peak curves of 300A. Usually frequency is changed from 2-3 times the resonant frequency and a complex tracking control is employed to bring the frequency back near resonant frequency to limit the inrush current by increasing impedance. The transient response of LLC resonant converter using the proposed control logic is shown in the Fig. 6(a). It may be observed that by splitting the pulses eventually frequency goes double, besides reduction in pulse width. It is clear from the Fig. 6(a) that inrush current is very well contained below 50A and voltage across output rises to its steady state without any ripple. As proposed in the control logic it can be observed in the Fig. 6(a) at no load start of LLC resonant converter, the gate pulse V_{gs} is split into two

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:10:17 UTC from IEEE Xplore. Restrictions apply.

half indicating operation at the twice the resonant frequency with limiting the inrush tank current with fast transient settling time, t_s = 400 µsec and rise time of t_r = 350 µsec. Similarly transient current limitation is depicted in Fig. 6(b) where load transient of LLC resonant converter is made from no load to full load using proposed control logic. It may be observed that the control logic enables the transient settlement in 1msec and keeping the peak current well within 20A. While observing similar response shown in Fig. 6(c), where load transition is made from full load to no load and its transient settling is made from full load to no load and its transient settling in time, t_s = 50 µsec. For depiction of the outlined control the state plane trajectory is shown in Fig. 6(d) with decreasing radii depicting decrease in transient energy, while settling down to normal operating conditions.

A. RESULTS:

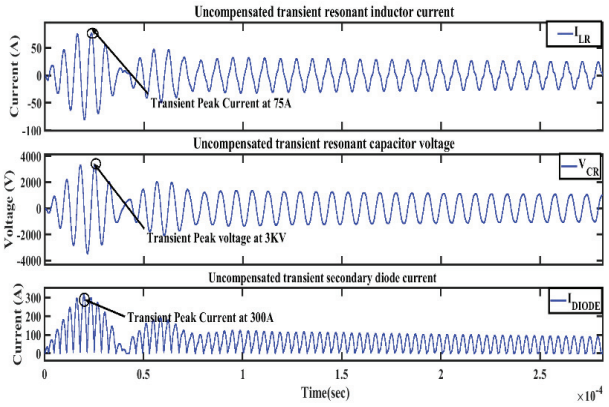
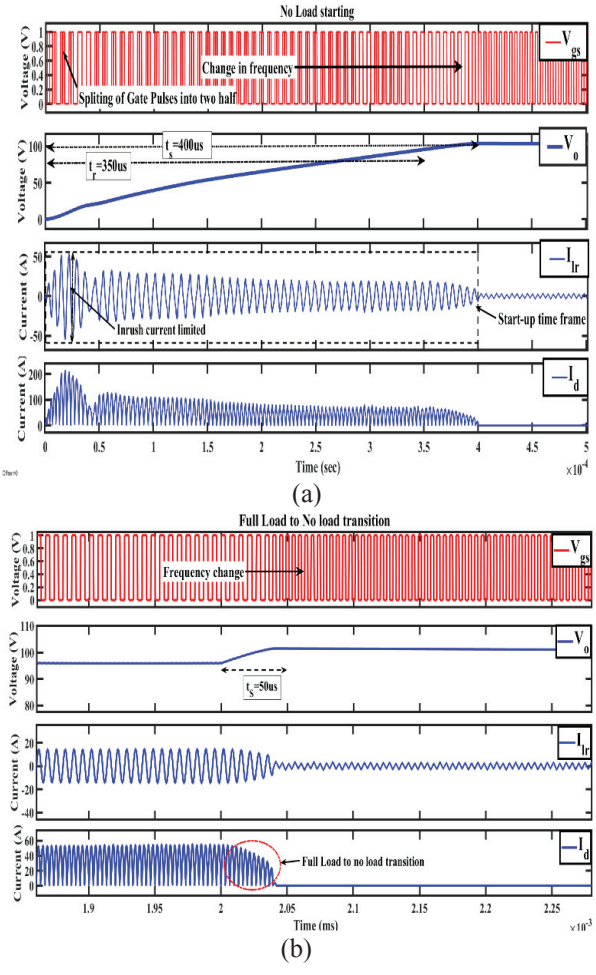
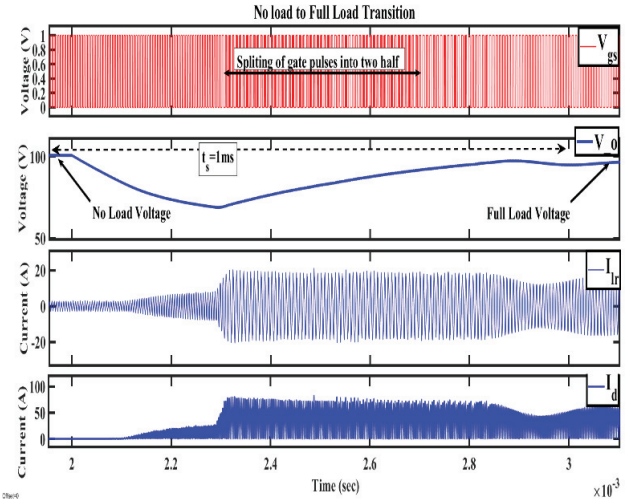


Fig.6 Converter tank response waveform without transient controller





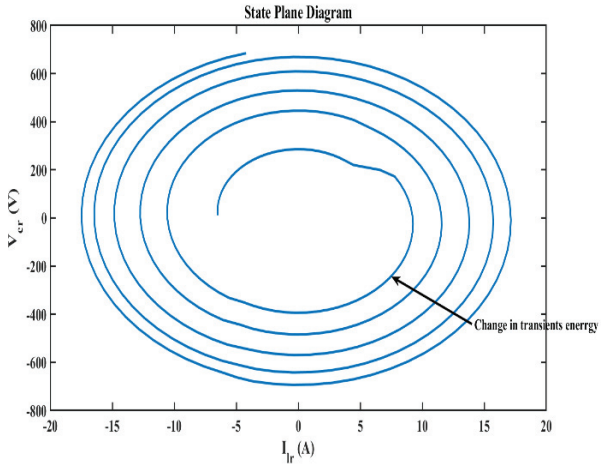


Fig. 7 Transient response of LLC resonant converter under the proposed control logic: (a) at No load start-up; (b) full load to no load transition; (c) no load to full load transition; and (d) State Plane Trajectory.

V. CONCLUSION:

For handling transients in an LLC resonant converter, a novel pulse splitting control method with hybrid control has been studied. The observed performance of the pulse splitting cum PWM has shown that transients can be handled smoothly even under conditions of maximum load. The results that were demonstrated support the controller ability to contain inrush current in the tank circuit within acceptable bounds without the use of complicated controls or expensive dsp controllers. A simulation study has been conducted and has validated the effectiveness of the proposed scheme. It has also been proposed that the proposed technique responds quickly to load perturbations and starting transients and is able to limit the flow of unwarranted current in the tank circuit by swiftly splitting the pulses and controlling the width of the pulses for the safety of the switches. To achieve optimal efficiency during steady state operation, the switching frequency smoothly transacts close to resonant frequency. The 3.3KW LLC resonant converter simulation results confirm the theoretical analysis. The system's flexibility in adjusting the tank circuit's impedance while regulating pulse width to accommodate transient power is an advantage. The scheme in general may be applicable to other topologies of resonant converters.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:10:17 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] J. Zeng, G. Zhang, S. S. Yu, B. Zhang and Y. Zhang, "LLC resonant converter topologies and industrial applications — A review," in Chinese Journal of Electrical Engineering, vol. 6, no. 3, pp. 73-84, Sept. 2020.

[2] F. Musavi, M. Craciun, D. S. Gautam and W. Eberle, "Control Strategies for Wide Output Voltage Range LLC Resonant DC–DC Converters in Battery Chargers," in IEEE Transactions on Vehicular Technology, vol. 63, no. 3.

[3] R. Oruganti and F. C. Lee, "Resonant Power Processors, Part II-Methods of Control," in IEEE Transactions on Industry Applications, vol. IA-21, no. 6, pp. 1461-1471, Nov. 1985.

[4] M. Mohammadi and M. Ordonez, "Fast transient response of series resonant converter using an average large signal model," 2015 IEEE Energy Conversion Congress and Exposition (ECCE), 2015, pp. 187-192.

[5] W. Feng, F. C. Lee and P. Mattavelli, "Simplified Optimal Trajectory Control (SOTC) for LLC Resonant Converters," in IEEE Transactions on Power Electronics, vol. 28, no. 5, pp. 2415-2426, May 2013.

[6] M. Mohammadi and M. Ordonez, "Extreme start-up response of LLC converters using average geometric control," 2016 IEEE Energy Conversion Congress and Exposition (ECCE), 2016, pp. 1-7.

[7] M. Mohammadi and M. Ordonez, "Inrush Current Limit or Extreme Startup Response for LLC Converters Using Average Geometric Control," in IEEE Transactions on Power Electronics, vol. 33, no. 1, pp. 777-792, Jan. 2018.

[8] I. Lee and G. Moon, "The k-Q Analysis for an LLC Series Resonant Converter," in IEEE Transactions on Power Electronics, vol. 29, no. 1, pp. 13-16, Jan. 2014.

[9] H. Park, M. Kim and J. Jung, "Spread Spectrum Technique to Reduce EMI Emission for an LLC Resonant Converter Using a Hybrid Modulation Method," in IEEE Transactions on Power Electronics, vol. 33, no. 5, pp. 3717-3721, May 2018.

[10] H. Nguyen, D. Maksimovic and R. Zane, "On/off control of a modular DC-DC converter based on active-clamp LLC modules," 2013 IEEE 14th Workshop on Control and Modeling for Power Electronics (COMPEL), 2013, pp. 1-6, doi: 10.1109/COMPEL.2013.6626413.

[11] Z. Fang, J. Wang, S. Duan, K. Liu and T. Cai, "Control of an LLC Resonant Converter Using Load Feedback Linearization," in IEEE Transactions on Power Electronics, vol. 33, no. 1, pp. 887-898, Jan. 2018.

[12] W. Feng, F. C. Lee and P. Mattavelli, "Optimal Trajectory Control of Burst Mode for LLC Resonant Converter," in IEEE Transactions on Power Electronics, vol. 28, no. 1, pp. 457-466, Jan. 2013.

[13] Hong Huang, "Designing an LLC Resonant Half-Bridge Power Converter," Texas Instruments (TI) Power Supply Design Seminar, SEM1900, 2010, TI Literature No. SLUP263.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 21,2022 at 07:10:17 UTC from IEEE Xplore. Restrictions apply.

iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40594

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Reconfigurable Manufacturing System: A Review

Vipin Kumar[1], Sreeraj Ramesan[2], Vinay Kumar[3], Dr. A. K Madan[4]
[1, 2, 3]*Research Scholar, Department of Mechanical Engineering, Delhi Technological University, Delhi*
[4]*Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi*

Abstract: ***Manufacturing firms/companies in the twenty-first century might be compelled to manage unpredictably high-rate, high-risk market changes energized by overall market competitiveness. To remain competitive in the market, these businesses should put resources into flexible manufacturing solutions and new kinds of manufacturing systems that are both cost-effective and touchy to market changes without forfeiting product and production quality. Re-configurability is a recent fad in designing technology that spotlights cost-effective, fast responses to market changes. Reconfigurable manufacturing systems (RMS) are another manufacturing paradigm that accepts a framework that is more delicate to changing market demands, including components like reconfigurable machines and reconfigurable controllers, as well as approaches for their deliberate plan and fast increase. This paper gives a basic overview of reconfigurable manufacturing systems (RMS) and the objectives that satisfy this approach.***

Keywords: RMS (Reconfigurable Manufacturing System) and FMS (Flexible Manufacturing System).

I. INTRODUCTION

Globalization of economies, fast progress in process technology, and the client-driven market have all contributed to extreme competition and dynamic business climate during the most recent couple of decades. To remain competitive and profitable, the area should adjust to a quickly changing environment wherein growing economies, new business concepts and philosophies, technological breakthroughs, and consistently changing client requirements generally offer severe threats to its existence.

Designing and running a cost-effective manufacturing framework in an internationally competitive context is a critical competitive issue. Manufacturing systems have changed over the course of time to satisfy the needs of a more dynamic and worldwide market that demands greater flexibility and responsiveness. Fast technological advancements have prepared the way for another production framework paradigm known as the reconfigurable manufacturing framework (RMS), which fulfills new manufacturing objectives with high responsiveness. An RMS, as indicated by Liles and Huff, is a framework able to do dynamically adjust the plan of a manufacturing framework to match the production request forced on it. Tsukune et al. characterize 'modular manufacturing,' which is analogous to the reconfigurable manufacturing framework. Later in 1996, the University of Michigan, Ann Arbor, established the designing research place for reconfigurable manufacturing systems (ERC-RMS) to plan and convey reconfigurable manufacturing systems. The core of the reconfigurable manufacturing framework (RMS) is the reconfigurable machine tool (RMT), which gives RMS its characterizing qualities, like customizable functionality and adjustable capacity, on account of its changeable structure. Reconfigurable machine tools are modular machines comprised of a few basic and supplementary modules.

Different combinations of basic and auxiliary modules are utilized to configure the RMTs. The primary modules, like the base, columns, and slideways, are structural; however, the auxiliary modules, for example, spindle heads, tool changers, spacers, indexing units, adapter plates, and angle structure, are kinematical or motion supplying. In comparison to the core modules, the auxiliary modules are smaller, lighter, and more affordable. As a result, they can be replaced more economically and rapidly with less exertion. By preserving the basic modules and replacing the auxiliary modules, the RMTs might be immediately different into a variety of different configurations with variable functionality and capacity.

By and large, RMTs can conduct a wide scope of operations in every one of their current configurations, which can be handled by checking out the operational capability and the configurations into which an RMT might be further reconfigured by adjusting its auxiliary modules, which add to its reconfigurability. The RMTs' responsiveness is portrayed by their working capability and reconfigurability, which additionally influences the responsiveness of the reconfigurable manufacturing framework.

A. Why RMS?

Manufacturers are increasingly confronted with fast and unexpected market shifts as a result of worldwide competitiveness, for example, the quick introduction of new products and continually changing product interest, a saturated market, and quick advancements in process technology. Companies should plan production systems that not just create high-quality products at a modest cost yet additionally consider the fast response to market changes and consumer needs, are effectively upgradeable, and can undoubtedly integrate new advances and functions.

© IJRASET: All Rights are Reserved |

The speed with which a plant might respond to changing business objectives and generate new product models is referred to as responsiveness. Manufacturing systems that are responsive may rapidly send off new goods on existing systems and respond rapidly and cost-effectively to market changes, product modifications, for example, current product updates and new product introductions, and framework faults.

Flexible manufacturing systems (FMS) can conform to product changes, yet they aren't intended to respond to structural changes. In this way, they can't respond to market swings like shifting interest or catastrophic equipment breakdowns. It features fixed hardware and fixed software that can be programmed. Changes to the framework structure are impractical with this architecture. As a result, FMSs have restricted upgrade, add-on, and customization capabilities. FMS are additionally costly in light of the fact that they regularly incorporate more functions than are required, utilize inadequate framework software since creating user-specified software is restrictively expensive, are unreliable, and are dependent upon oldness because of technological advances and their fixed framework software/hardware.

One of the most concerning issues for manufacturers is the high risk of expensive flexible production systems becoming old. Computers, information, processing, controls, optics, high-speed motors, linear drives, and materials advance in six-month cycles, so even the present most effective production framework may become wasteful before long.

Dedicated manufacturing systems, or DMSs, on the other hand, are intended to create a solitary product at a high rate for a minimal price. Manufacturing another product with a similar framework is unimaginable, and changing the manufacturing volume in little stages is unthinkable. Each new product or change in production volume requires the arrangement of another production line. As a result, DMSs are not adequately versatile to market changes.

Future manufacturing systems technology should fulfill the accompanying objectives, which go beyond mass, lean, and flexible manufacturing, to address difficulties, for example, reduced lead time for sending off new manufacturing systems and reconfiguring existing ones and the fast upgrade of existing systems, and the speedy integration of new process technology and capabilities.

The Reconfigurable Manufacturing System (RMS) has been highlighted as the main concern for future manufacturing, with capacity and functionality that can be altered on a case-by-case basis. Since these systems take into consideration the simple replacement of framework components and the inclusion of application-explicit software modules, they won't become outdated.

These systems won't become out of date since they will consider the quick replacement of framework components that can be continuously improved by coordinating new technology, as well as the fast addition of application-explicit software and hardware modules that can be reconfigured rather than rejected and replaced to oblige future products and changes in product interest.

II. LITERATURE REVIEW

Shabaka and Elmaraghy [1] proposed a technique for adjusting machine setups by adjusting the tool and workpiece axes' motion as indicated by the desired component operations and characteristics. A kinematic chain structure is utilized to depict the machine tool. The creators recommended that the framework be made more flexible by producing numerous potential configurations for every action. Framework convertibility was described by Maier-Speredelozzi et al. [2] as a framework's capacity to adjust production functionality. They recommended machine convertibility based on qualitative criteria like tool magazine capacity, fixturing flexibility, programmed tool changer availability, and modular hardware/software component availability. Gumasta et al. [3] proposed reconfigurability of the RMS on a framework level, considering the framework's basic properties like modularity, scalability, convertibility, and diagnosability.

Abdi and Labib [4,5] concocted a conventional RMS model in which every configuration may not just produce a wide scope of goods gathered into a family yet, in addition, respond emphatically to new products presented inside that family. They split the planning stage into strategic and tactical levels, and while addressing the strategic level, they fostered a relationship among market and manufacturing to arrange items into families. To sort out items into families based on operational similitudes and in this manner relegate them to the proper manufacturing processes all through configuration stages, a reconfiguration connection among market and production is offered.

Spicer and Carlo [6] utilized a dynamic programming method to represent different periods of RMS by considering cost and reconfiguration as execution criteria. Notwithstanding, they only treated bases and spindles as two RMT modules in their research, which is a long way from functional.

Son [7] depicted a hereditary algorithm procedure for delivering machining machine setups consequently. For homogenous paralleling flow lines, Son presented a capacity scalability strategy. Son created a couple of framework configurations that satisfied the need for all request times, and afterward, he created numerous other configuration courses involving single-period configurations as seeds.

© IJRASET: All Rights are Reserved |

All of the proposed configuration pathways were surveyed based on the reconfiguration similarity metric and cost, and the best configurations for all request times were chosen to reduce reconfiguration costs. This study addresses a solitary part request situation, and that implies that variations in functionality are not considered.

Goyal et al. [8] gave ways of surveying the reconfigurable machine tool's working competence and reconfigurability (RMT). Machine assignment was based on the created performance index as well as cost.

Reconfigurable manufacturing systems were separated by Makino and Trai [9] into two classifications: statically reconfigurable manufacturing systems and dynamically reconfigurable manufacturing systems. In a static reconfigurable manufacturing framework, the stations are intended to be effortlessly moved around, whereas, in a dynamically reconfigurable manufacturing framework, reconfigurability is accomplished by utilizing progressed material handling systems like AGVs or voyaging robots rather than traditional conveyor systems.

Xia et al. [10] explored a technique for generating a strategy for RMS reconfiguration and maintenance. By utilizing a reconfigurable maintenance time window, they proposed an algorithm for constructing an arrangement that considers each machine's dependability and failure, as well as the costs of replacement and failure (RMTW).

Kumar et al. [11] zeroed in on the RMS gathering framework reconfiguration and presented the production planning algorithm for the get-together booked with minimal measure of reconfiguration cost and the greatest measure of reconfiguration gains. They recommended an algorithm that considered the cost and postponed the ideal opportunity for reconfiguration, task balance, due date, and different variables.

III. RMS (RECONFIGURABLE MANUFACTURING SYSTEM)

Reconfigurable manufacturing systems are extremely dynamic and advancing systems that are intended to manage evolving conditions unpredictably. Re-configurability is characterized in this context as the ability to change a framework's conduct by changing its configuration. Hard reconfiguration is cultivated through hardware adjustments and alterations (e.g., evolving spindles), while delicate reconfiguration is refined through software or organizational changes (e.g., additional movements). RMS's fast adaptability in terms of hard-and software, as well as their structure, is one of their essential characteristics (e.g., adding, removing, and modifying machine tools). These adjustments are put forth with minimal attempt in terms of both time and money. Their functionality and capacity can be adjusted and reconfigured on a case-by-case basis to respond to changes in market conditions on account of this principle property. Modularity, integrability, customization, convertibility, and diagnosis ability are significant aspects of RMS. Modularity, integrability, and analytical ability are critical empowering agents for decreasing reconfiguration time and costs. Customization and convertibility, on the other hand, permit these systems' operational costs to be reduced. A reconfigurable manufacturing framework is another type of manufacturing framework that can undoubtedly and right away change its capacity and functionality on a case-by-case basis.

Flexible manufacturing systems, on the other hand, are constructed with all suitable flexibility and functionality setups and can only change inside their a priori specified flexibility with fluctuated adaptabilities (e.g., process and product flexibility). This incorporates only creating a predetermined product spectrum and amount inside their flexibility range. Adaptations that go beyond this flexibility require a lot of work.

To sum up, FMS has the capacity and functionality limitations. Adaptations inside their flexibility can be overseen rapidly and cheaply. In any case, one negative is the huge initial capital expenditure required to exploit the inherent and sometimes underutilized flexibility. The adaptability and scalability of systems, especially in terms of capacity and flexibility, are significant contrasts among FMS and RMS. Reconfigurable manufacturing systems empower on-request redid flexibility through scalability to incrementally realize additional capabilities and capacities, permitting them to fulfill market demands. Flexible manufacturing systems, on the other hand, have an overall a priori-fixed flexibility and can modify inside their intrinsic flexibility. RMS is created for a particular product portfolio (for instance, product A toward the start and B+C toward the finish of the expansion stage), whereas FMS can manufacture numerous products at the hour of installation. The capacity and functionality of the reconfigurable manufacturing framework (RMS) are actually required.

RMS can be basically altered with new process technology and can react to evolving demands. RMS is a crossover framework that joins the advantages of dedicated and flexible systems. Dedicated manufacturing systems have formed into reconfigurable manufacturing systems. The concept of reconfiguration originates from the concept of utilizing a modular machine. However, it isn't restricted to modular machines. Reconfiguration utilizing material handling systems, relocation, reconfiguration process plan, and different concepts have been proposed by a few researchers.

© IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. *RMS Features*

RMS, as a relatively new sort of manufacturing framework, have a flexible structure in both hardware and software architecture, and they join the accompanying six basic features:

1) Modularity, which is the division of operational functions into units that might be moved around among different production plans to accomplish the best results;
2) Integrability, the ability to rapidly and precisely link components utilizing a bunch of mechanical, informational, and control interfaces that work with integration and communication;
3) Diagnosability, or the ability of a framework to self-read its current condition to distinguish and analyze the fundamental reasons for product faults and repair them quickly;
4) Convertibility, or the capacity to rapidly adjust the functioning of current systems and machines to meet changing production and market demands;
5) Customization, which limits framework and machine flexibility to a particular product family, resulting in modified flexibility;
6) Scalability refers to the ability to change the production capacity of a framework by adding or pulling out resources and modifying framework components.

RMS' dynamic systems have the capacity and functionality to respond to market changes as a result of these features. Furthermore, when compared to conventional manufacturing techniques, RMSs empower the creation of more extensive scope of redone goods. Since the link between the features of the production framework and market expectations is basic, a wide definition of RMS, enveloping and expanding such aspects, is required.

'RMS is a manufacturing framework created to satisfy the needs of a dynamic market that demands high-quality products in differing amounts at a fair cost.

RMS has a flexible hardware and software structure that permits it to adjust production capacity and functionality to accomplish a high throughput rate while keeping up with flexibility and a cell organization design.' The dynamism of RMSs, as well as their connections to both the market and traditional industrial systems, are highlighted in this description. RMS research has detonated in recent years, crossing a wide scope of themes. Most of the examinations distributed up to this point have zeroed in on ways for fusing a portion of the new characteristics into current manufacturing systems, with less accentuation on approaches for effectively designing new RMSs

IV. COMPARISON OF RECONFIGURATION AND RAMP-UP TIME

The time allotment it takes to take on a framework to another product is known as reconfiguration time. This time covers the time spent redesigning the framework, assembling the machines, rearranging the equipment, etc. How much time it takes for a manufacturing framework to ramp up to the suitable quality and production rate after it has been reconfigured is known as ramp-up time.

Pre-manufacturing and change time are incorporated.

1) *DMS:* Since DMS is inflexible, when the workpiece changes, the entire framework should be redesigned, manufactured, and implemented. This procedure can take anything from 18 to 30 months. The ramp-up of an ideal opportunity for a DMS is roughly 2-4 weeks.
2) *FMS:* A FMS is intended to generate a wide number of different forms and sizes of workpieces. This requires a fast reconfiguration and ramp-up period. The reconfiguration process requires 1-3 days, while the ramp-up process requires 1-2 days.
3) *RMS:* The framework and the machines ought to be rebuilt and rearranged at whatever point the workpiece is changed. As a result, RMS might take longer than FMS; however, fundamentally less time than DMS. The reconfiguration procedure requires 1-2 weeks, while the ramp-up process requires 2-4 weeks.

We can deduce from the above focuses that FMS has the most limited reconfiguration and ramp-up time followed by RMS. The time taken for something very similar by a DMS is a lot bigger than the other two.

© IJRASET: All Rights are Reserved |

A. Comparison of Life Cycle Cost

The costs of designing, implementing, reconfiguring, maintaining, ramping up, and disposing of a manufacturing system are included in the system's life cycle cost.

1) *DMS:* In comparison to other systems, DMS is the simplest because it only deals with a single fixed part. However, because of a lack of appropriate design methodologies, the special purpose machines employed are more challenging to develop. DMS has cheap manufacturing costs due to its simple design, except for the SPMs. In this situation, reconfiguration involves the construction of a new system, which is quite costly. Because ramp-up costs are proportional to ramp-up time, DMSs have a relatively high ramp-up cost. Because there are no spare components to manage, a DMS requires the least amount of maintenance. Almost none of the components and machines in a DMS can be reused after reconfiguration, resulting in substantial disposal costs.
2) *FMS:* When compared to other types of production systems, FMS has the most complicated structure, resulting in the highest design cost. Because of its complexity, this design has the highest manufacturing cost. FMS can produce a huge number of different workpieces, and changing them is simple; therefore, the cost of reconfiguration is small. The lowest ramp-up cost is due to the shortest ramp-up time. The administration and maintenance costs of maintaining an FMS are the highest due to the need for highly qualified personnel. An FMS's CNC machines can be reused in a new system, resulting in a minimal disposal cost.
3) *RMS:* Despite the fact that an RMS system is more sophisticated than a DMS, RMS design expenses are lower due to the good design methods available to designers. Because of the modular design, manufacturing costs are lower than those of a DMS. Because an RMS is designed to be quickly reconfigured, it has a cheap cost of reconfiguration. Due to the lengthier ramp-up period, ramp-up expenses are lower than DMSs but higher than FMSs. RMS requires a large number of spare components to ensure adjustability and reconfigurability, resulting in increased maintenance costs. An RMS's components and machines are all built to be highly reconfigurable. That is, they can be reused multiple times, and the cost of remanufacturing must be below.

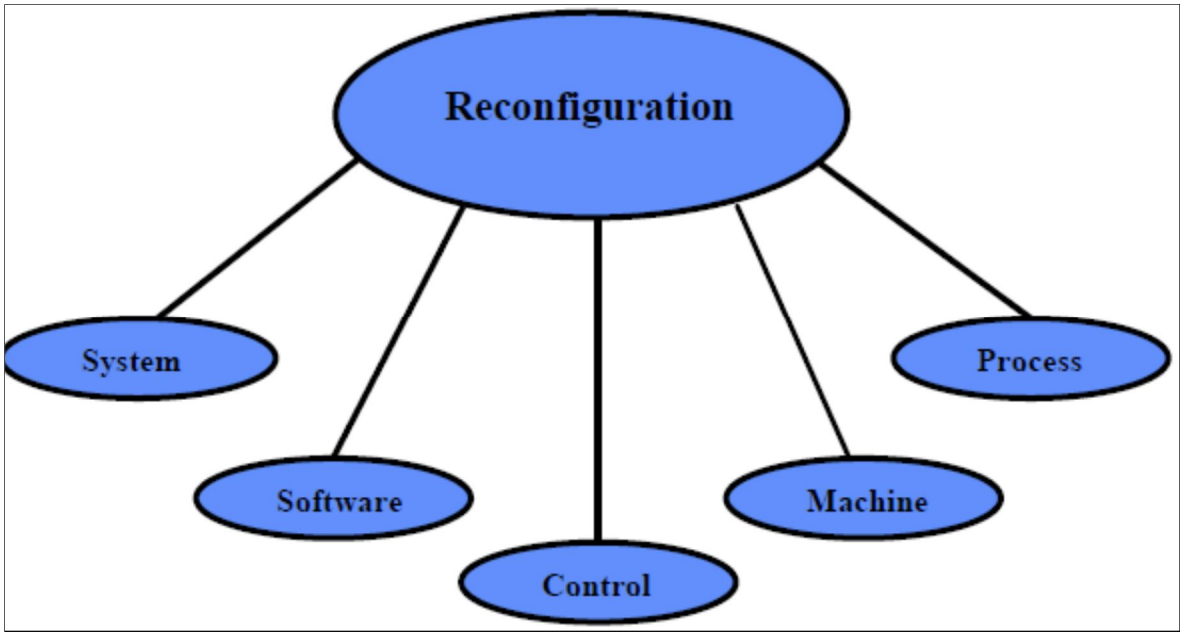
We can conclude that RMSs have the lowest lifecycle costs among all the three manufacturing systems, followed by DMS, while FMS is the most expensive one.

The results from all the comparisons made point towards the fact that RMSs are the future of manufacturing.

V. COMPONENTS OF RECONFIGURABLE MANUFACTURING SYSTEM

The reconfigurable manufacturing systems have two important components:

1) Reconfigurable machine tool
2) Reconfigurable Controller



Reconfigurable Components

A. Reconfigurable Machine Tool

The upside of a reconfigurable manufacturing system is that the system's structure, as well as its machines and controls, might be immediately adjusted in response to market changes. The reconfigurable machine tools are a significant part of RMS (RMT). In contrast to customary CNC machines, which are general-purpose machines, RMT machines are produced for a specific range of operations and might be cost-really changed over as the needs change. Machine tools can be reconfigured to make a specified set of characteristics in a specific cycle time range. Some activity requirements will stay consistent all through the machining system's lifetime. The primary objective of a reconfigurable machine tool is to adjust to changes in the products or parts being made. The following possible modifications must be considered.

1) The size of the workpiece
2) The geometry and intricacy of the parts
3) Production volume and rates
4) Processes that must be followed
5) Geometrical accuracy and surface quality requirements
6) Material properties such as type of material, hardness, and so on

RMT controllers have various new issues as a result of the demand for reconfiguration. The first boundary is the controller architecture's reconfigurability, which is necessary when physical machine tools are reconfigured, or new innovation is added. Controlling RMT with numerous tools working autonomously, as well as RMT with axes in non-orthogonal configurations, is another test. The combination of diverse software and equipment components is another issue. Machine design, machine tool design, kinematics modeling, and dynamics analysis are totally expected for RMT design. There is no finished philosophy or design method that can be applied directly to the RMT. Fixture design, assembly system design, and reconfigurable robotics are generally disciplines where the idea of reconfiguration has been utilized. Coming up next are some examples of normal movement types.

B. RMT Tool Machining Motion (TMM)

All reconfigurable machine tool movements that are occupied with the same machining highlight (an opening, slot, and so on) are assembled as RMT (Tool machining movement). All Tool point motions. have the same directions and execute the same machining activity.

C. RMT Tool Machining Motion Family (TMMF):

An RMT TMMF is an assortment of machining motions shared by several work parts that may have the same movement characteristics.

Tools for creating reconfigurable machine tool control systems comprise four primary parts, as indicated by reconfigurable machines.

1) Automatic Part Transfer System module.
2) Automatic Part Clamping Rotating System module.
3) Automatic Part Lifting System module.
4) Automatic Tool Changing System module.

D. Reconfigurable Controller

The reconfigurable controller contains the machine-specific functions or classes that are expected to work a machine. Reconfigurable Controller is ready to control both a 3-axis desktop factory and a 5-axis plant. Some improvements and adjustments to the reconfigurable controller are expected to permit reconfiguration of the movement organizer and servo controller, which is expected for controlling various mechanisms.

The configuration system is directly connected with the control interface, permitting it to acknowledge configuration commands from it. The configuration system will do one of two processes based on these various configuration instructions. It will either design a legitimate machine working boundary, such as joint machine restrictions, or it will not.

The Reconfigurable controller is progressively reconfigured for a specific mechanism when the configuration system completes these configuration operations. The manufacturing process instructions may then be sent from CAD/CAM software to the Reconfigurable controller for direct machining.

© IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VI. DESIGN CONSIDERATIONS IN RMSS

The advantages associated with RMS are numerous. However, some critical design decisions are required for all types of RMSs. These are present in its architecture design, configuration design, and control design.

A. Architecture Design

A reconfigurable equipment system and a reconfigurable software system make up the system architecture. Reconfigurable machining systems, reconfigurable fixturing systems, reconfigurable assembly systems, and reconfigurable material-handling systems are all part of the equipment system.

B. Configuration Design

Configuration design determines system configuration under given system architecture for a specific task. A configuration is a gathering of modules that have been chosen. A configuration can do the task to its maximum capacity: the various types and assembly possibilities of the system are still up in the air by the system architecture. As a result, a system's architecture dictates which configuration variants it can produce. A system configuration is represented by design variables, the number of modules, and inner configurable parameters inside every module. Design analysis and design synthesis are both engaged with configuration design.

1) *Design Analysis:* The mappings between design variables and design constraints, as well as between design variables and design objectives, are established through design analysis.
2) *Design Synthesis:* From all of the possible configurations, design synthesis is used to select the best one.

C. Control Design

Control design identifies the process variables that must be used for a configuration to perform satisfactorily. Reconfigurable variables and process variables are the two types of variables that can be changed in an RMS system.

1) *Reconfigurable Variables:* These modify the system's configuration. These are established in the configuration design phase.
2) *Process Variables:* The development and motions directed by the machines in accomplishing the task in a configuration are controlled by process variables.

VII. DESIGN REQUIREMENTS

RMS control systems must meet the following requisites:

1) The control system must be self-contained. Because a system objective is deconstructed into module goals, every one of these modules has a singular controller to accomplish its objective. The control system must have the option to amalgamate and synchronize components to accomplish system goals.
2) Since system components are decentralized and topographically scattered, the control system should be distributed and modularized.
3) The control system should be modifiable and updatable so that controlling components can be refreshed. Controlling elements can be made using an assortment of working systems, languages, networks, databases, and protocols, as well as by various manufacturers.
4) Because of usefulness, ability, or empowering technologies being alterable, expansion, expulsion, and up-gradation of equipment components is essential. Consequently, the control system should be scalable and upgradeable.
5) Considering system configurations might change frequently, the control system should be self-arranging, and the connected control system should have the option to self-reconfigure rapidly as well.

VIII. FUTURE RESEARCH TRENDS

RMS provides devoted and versatile production solutions in a single bundle. To manage the issues of designing part families, elective processing routes, planning product features with processing modules, creating part programs for operations designed. Thus, for algorithms to rapidly make new machine configurations and their physical creations, it is necessary to foster a perspective and new computer technologies. Reconfigurable machine tools are a reasonable choice for production scenarios where processing needs change over the machine tool's lifetime. Notwithstanding, procedures for structural approval and diagnosability for quality concerns must be created. To accomplish synergistic benefits, research activities must be multidisciplinary.

© IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Manufacturing firms are under more aggressive pressure today than they were 20 years prior when RMS was sent off because of increased globalization. Volatile market trends, short product life cycles, increased product diversity, lower production costs, and stricter ecological restrictions have all become harder to survive. Re-configurability becomes substantially more basic to keep up with "sustainable competitiveness." Higher reconfigurability can add to the better natural impact, notwithstanding faster response and decreased costs. Besides, RMS and present-day manufacturing systems are approaching another age, based on late innovations in "Industry 4.0." We offer numerous prospective future research trends in this section, as well as discuss how ongoing mechanical improvements can work on the design and activity of RMS.

A. *Concurrent Plan of Product, Manufacturing Systems, and Business Strategies*

A manufacturing organization's product improvement process remembers decisions for the accompanying three aspects:

1) The product's features are constrained by cost and engineering constraints;

2) The product's manufacturing systems, including the framework configuration, picked machines (e.g., functionality, power, accuracy, ranges, and number of axes), and process parameters (e.g., task precedence, task type, access direction, dimension, accuracy, and power required to do the responsibility); and

3) The field-tested strategy or marketing strategy incorporates things like forecasting deals, deciding product valuing, and choosing when to send off another product. The objectives of these many parts are typically in a state of harmony; however, they can sometimes conflict. For instance, it is preferred to reduce product intricacy from a manufacturing point of view; yet, this reduction might result in a product that is less desired in the marketplace.

To resolve the compromise, a shared consideration is required, and concurrent plan approaches ought to be laid out. Most of the current literature centers around only one component of the planned issue, with a couple looking at cooperative decision-production in the product and manufacturing framework improvement. In any case, more research on fusing commercial objectives into the specialized decision-production process is required. Michale c z et al. took a gander at the concurrent plan issue, which included adjusting marketing and manufacturing objectives in a production line. The planning challenge was separated into a few sub-issues. Not exclusively should the best plan solution layout the best manufacturing systems, yet it ought to likewise foster a product evolution strategy and a long-term marketing strategy.

Note that concurrent plan difficulties are hard to tackle since various components should be assessed simultaneously. Regardless of whether only the production framework is considered, convoluted difficulties should be grown, for example, joint process planning and line balance [5,88] or joint process planning and scheduling . Such troublesome plan optimization challenges are ordinarily presented as numerical programming issues, which require the utilization of heuristics to track down a solution. Renzi et al. analyzed non-careful meta-heuristic and artificial intelligence approaches to settling RMS plan challenges. More progressed comprehensive plan approaches that join commercial objectives into engineering ought to be created in the future. An overall plan process that synthesizes the different plan features is additionally still required.

Beneath, we go more than two subjects that require inside and out research in product-framework business plan strategies.

1) The significance of cooperation in the improvement of product-framework business strategies. Cooperate culture hugely affects framework configuration; it's another field we're calling "social engineering." When these components are considered, the product-framework business configuration challenge turns out to be extremely troublesome. For instance, a conflict might emerge between a manufacturing company's marketing and manufacturing offices: the previous require items that are more marketable, while the last option desires to assemble less refined products, which will bring down their cost and make them more reasonable to purchasers. Future research on the relationship between the undertaking management structure and the company's functioning culture ought to be performed to defeat this conundrum. The plan and operation procedures of a framework can be impacted by the helpful culture. Koren et al., for instance, investigated how corporate culture (e.g., response time to critical maintenance requests) may impact framework configuration selection and concluded that a different corporate culture could clarify why the United States and Japan prefer different frameworks configurations.

2) The opportunities for equipment and operations to progress. Because of its ability to create a few generations of goods, RMSs can reach high framework sustainability. Be that as it may, to increase the proficiency of a framework over its entire lifecycle, it is important to analyze present cutting-edge technology as well as how technology might develop in the future. More reliable machinery, novel control innovations, and new sensors, for instance, may affect the best framework plan (e.g., framework configuration). An SLP setup offers a greater throughput than an RMS course of action because of higher machine availability.

B. Improving the Effectiveness and Proficiency of Real-time Operational Decision-Production.

In comparison to the production planning challenge, the literature on real-time operational decision-production for RMSs is scant. The key issue is the framework's intricacy, as well as the requirement that the proposed strategy/algorithm be adequately proficient to be sent in real-time. As previously expressed, maintenance decision-production is troublesome; it turns out to be considerably more troublesome when maintenance and production scheduling issues are consolidated. For instance, how to ideally embrace maintenance based on the product requirement and the health status of machines while likewise picking production courses utilizing the machines that are not under repair? Such difficulties are extremely hard to address, particularly when a decision should be made rapidly. Most standard analytical and decision support tools are either unequipped for managing such intricacy or can only arrange with it wastefully. Wise manufacturing approaches, for example, multi-specialist systems cloud manufacturing, digital manufacturing and cyber-physical systems, can increase the adequacy and effectiveness of real-time decision-production. He et al. , for instance, created an interesting procedure that permits manufacturing resources to self-coordinate cost-effectively inside the structural restrictions of a specially made manufacturing framework for fulfilling client orders. Then, at that point, we'll discuss how large information and cyber-physical manufacturing framework strategies can assist with the RMS plan. More effective and savvy maintenance and production scheduling decisions might be made in real-time utilizing progressed monitoring and examination capabilities.

1) Manufacturing systems' prognostics and diagnostics capabilities have been considerably increased as a result of the fast advancement of sensor innovations and information investigation approaches, which has fuelled the systems' continual improvement. More information is received from machines and processes in the present production systems, which is examined to give further developed knowledge of the framework's condition, including both online and offline characterization. Furthermore, the registering capabilities helped by cloud-based approaches are quickly moving along. Such capabilities consider the turn of events and arrangement of more proficient prognostics and diagnostics approaches in industrial systems, for example, online monitoring, remote monitoring , irregularity identification, and remaining valuable life prediction. Future research should zero in on the advancement of flexible or reconfigurable condition monitoring systems, which can consolidate a variety of decision-support tools, for example, information gathering, feature selection, and sensor allocation.

2) Cyber-physical manufacturing advances, which have recently been created, can address these issues. With different contextual investigations, Monostori et al. assessed the vital approaches in cyber-physical systems (CPSs) and demonstrated how they could be utilized in the manufacturing context. A CPS is comprised of two interrelated components: physical and cyber. A "digital twin" of the real framework is sent by the cyber framework, which might be considered a mirrored version of the real machinery and operations. While the real framework runs on a physical platform, the digital twin suddenly spikes in demand for a cloud platform, mimicking the health of every individual machine in the framework and continuing recording and following machine conditions, energy utilization, product quality, and a variety of different metrics and information. Information-driven models and algorithms can be created, which can then be joined with physical knowledge and integrated into the simulation model. The cyber framework will actually want to give further developed admittance to machine status for plant management because of the universal connectivity given by cloud processing advancements. More essentially, utilizing the digital twin to reproduce will take into consideration better decision-production, for example, contrasting different maintenance and production scheduling strategies. Once the cyber framework has observed ideal or close ideal solutions, they will be implemented in the physical framework to work on its operation.

C. From Mass Customization to Mass Individualization

Clients' needs are turning out to be more changed and individualized nowadays. Personalized products can be made cost-productively as a result of advancements in additive manufacturing and 3D printing, which shift the manufacturing paradigm from mass customization to mass individualization. Open-architecture goods will be created under this new paradigm. Countless different modules that should be joined into a perplexing product are a significant issue for mass individualization. As a result, the manufacturing framework (for this situation, an assembly framework) ought to have the option to manufacture an enormous number of different models and be versatile (RAS). The plan and operation of a RAS like this are more troublesome than a normal RMS for machining. Countless different modules that should be consolidated into an intricate product are a significant issue for mass individualization. As a result, the manufacturing framework (for this situation, an assembly framework) ought to have the option to manufacture an immense number of different models and be versatile (RAS). The RMS characteristics and thoughts examined in Sections 2 and 3 can be utilized to work on the plan and operation of RAS. Nonetheless, the plan and operation of a RAS like this are more troublesome than a commonplace RMS for machining.

© IJRASET: All Rights are Reserved |

The RAS, for example, ought to be expandable to fulfill more need variations and convertible to permit numerous variations and new items. Among the plan topics of a RAS are displaying systems that wed modern relationships among changed machines, line equilibrium, and production scheduling, among different things. New measures to gauge the intricacy of framework configurations and products are likewise required. Another framework structure for the present RAS should be planned to further develop framework effectiveness and limit operational costs.

IX. CONCLUSION

Manufacturing flexibility and reconfigurability are broadly acknowledged as fundamental components for acquiring a competitive edge in the marketplace. For manufacturing businesses, these are the best properties. This record gathers a comprehensive appraisal of the literature on manufacturing flexibility and reconfigurability. The objectives of an ideal manufacturing framework are talked about in this study. This report likewise recognizes the concepts of flexible manufacturing systems and reconfigurable manufacturing systems. Both flexibility and reconfigurability have been discovered to be significant. Notwithstanding, having a little flexibility and reconfigurability is preferable to having a ton of flexibility. Flexibility and reconfigurability are both costly. As a result, what should be the degree of flexibility and reconfigurability in a production framework turns into a research point.

REFERENCES

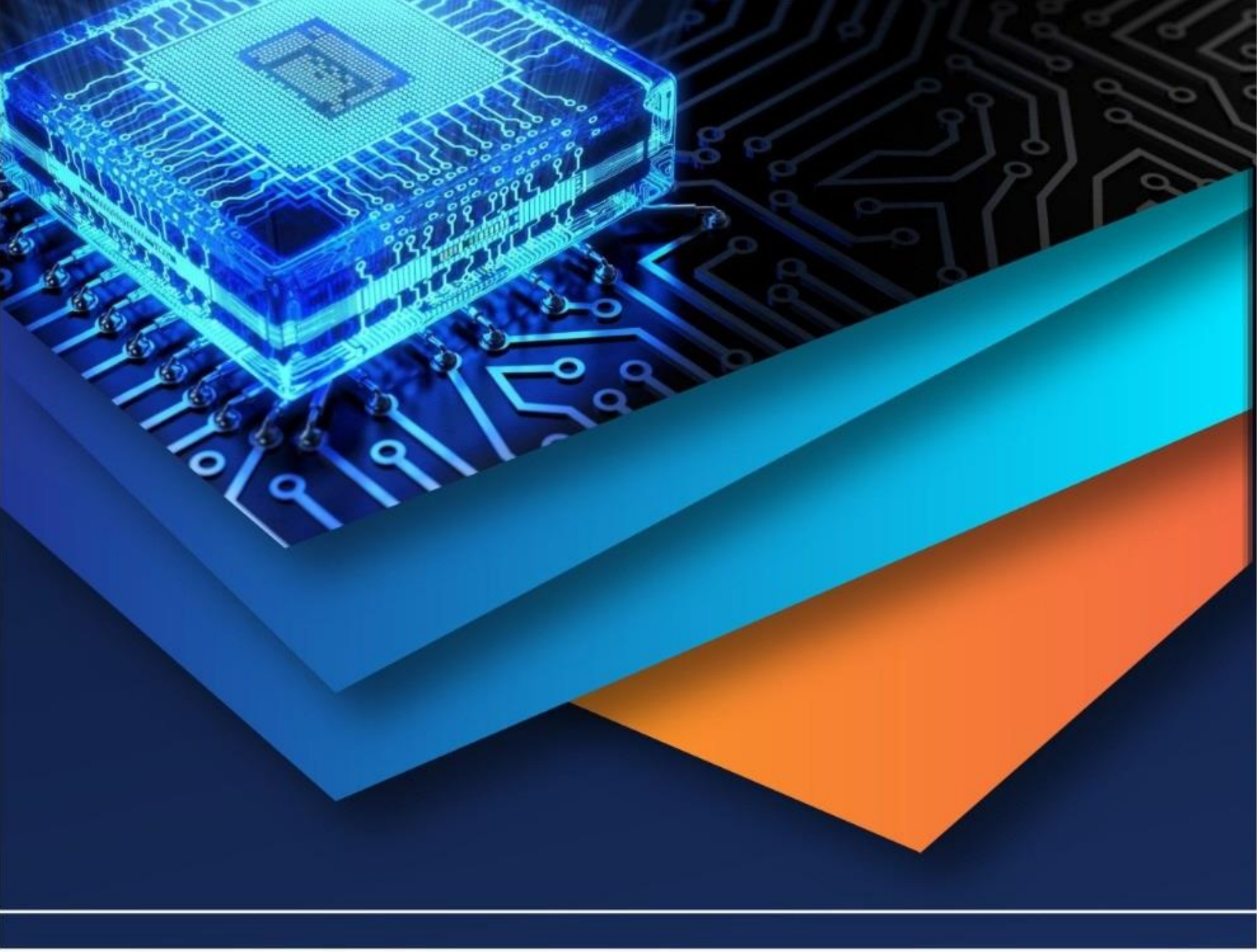
[1] Rheault M, Drolet JR, Abdulnour G. Physically reconfigurable virtual cells: a dynamic model for highly dynamic environment. Computers and Industrial Engineering 1995;29(1–4):221–5.
[2] Koren Y, Hiesel U, Jovane F, Moriwaki T, Pritschow G, Ulsoy G, et al. Reconfigurable manufacturing systems. Annals of the CIRP 1999;48(2):527–40.
[3] Galan R, Racero J, Eguia I, Canca D. A methodology for facilitating reconfiguration in manufacturing: the move towards reconfigurable manufacturing systems. International Journal of Advanced Manufacturing Technology 2007;33(3/4):345–53.
[4] Galan R, Racero J, Eguia I, Garcia JM. A systematic approach for product families formation in reconfigurable manufacturing systems. Robotics and Computer-Integrated Manufacturing 2007;23(5):489–502.
[5] Koren Y, Shpitalni M. Design of reconfigurable manufacturing systems. Journal of Manufacturing Systems 2010;29(4):130–41.
[6] Liles DH, Huff BL. A computer based computer scheduling architecture suitable for driving a reconfigurable manufacturing system. Computers and Industrial Engineering 1990;19(1–4):1–5.
[7] Tsukune H, Tsukamoto M, Matsushita T, Tomita F, Okada K, Ogasawara T, et al. Modular manufacturing. Journal of Intelligent Manufacturing 1993;4(2):163–81.
[8] Landers RG. A new paradigm in machine tools: reconfigurable machine tools. In: Japan–USA Symposium on Flexible Automation. 2000.
[9] Moon YM, Kota S. Design of reconfigurable machine tools. Journal of Manufacturing Science and Engineering 2002;124(2):480–3.
[10] Pattanaik LN, Jain PK, Mehta NK. Cell formation in the presence of reconfigurable machines. International Journal of Advanced Manufacturing Technology 2007;34(3/4):335–45.
[11] Landers RG, Min BK, Koren Y. Reconfigurable machine tools. CIRP Annals Manufacturing Technology 2001;50(1):269–74.
[12] Bi, Z. M., Lang, S. Y., Shen, W., & Wang, L. (2008). Reconfigurable manufacturing systems: The state of the art. International Journal of Production Research, 46(4), 967–992. https://doi.org/10.1080/00207540600905646
[13] Koren, Y., Gu, X., & Guo, W. (2017). Reconfigurable Manufacturing Systems: Principles, design, and future trends. Frontiers of Mechanical Engineering, 13(2), 121–136. https://doi.org/10.1007/s11465-018-0483-0
[14] Kumar, G., Kumar Goyal, K., & Batra, N. K. (2019). Evolution, principles and recent trends in reconfigurable manufacturing system. Journal of Physics: Conference Series, 1240(1), 012161. https://doi.org/10.1088/1742-6596/1240/1/012161
[15] Zhang, G., Liu, R., Gong, L., & Huang, Q. An analytical comparison on cost and performance among DMS, AMS, FMS and RMS. Reconfigurable Manufacturing Systems and Transformable Factories, 659–673. https://doi.org/10.1007/3-540-29397-3_33
[16] Koren, Y., & Shpitalni, M. (2010). Design of reconfigurable manufacturing systems. Journal of Manufacturing Systems, 29(4), 130–141. https://doi.org/10.1016/j.jmsy.2011.01.001
[17] Mehrabi, M. G., Ulsoy, A. G., & Koren, Y. Reconfigurable manufacturing systems: Key to future manufacturing. Journal of Intelligent Manufacturing (2000), 11, 403–419. https://doi.org/DOI: 10.1023/A:1008930403506
[18] Mehrabi, M. G., Ulsoy, A. G., Koren, Y., & Heytler, P. (2002). Trends and perspectives in flexible and reconfigurable manufacturing systems. Journal of Intelligent Manufacturing, 13, 135–146. https://doi.org/10.1023/A:1014536330551
[19] Lameche, K., Najid, N. M., Castagna, P., & Kouiss, K. (2017). Modularity in the design of reconfigurable manufacturing systems. IFAC-PapersOnLine, 50(1), 3511–3516. https://doi.org/10.1016/j.ifacol.2017.08.939
[20] Singh, A., Kumar, P., & Singh, S. (2013). Vision, Principles and Impact of Reconfigurable Manufacturing System. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, 3(1).
[21] Galan, R., Racero, J., Eguia, I., & Canca, D. (2006). A methodology for facilitating reconfiguration in manufacturing: The move towards reconfigurable manufacturing systems. The International Journal of Advanced Manufacturing Technology, 33(3-4), 345–353. https://doi.org/10.1007/s00170-006-0461-2
[22] Koren, Y. (2004). Reconfigurable machine tools. Autonome Produktion, 523–534. https://doi.org/10.1007/978-3-642-18523-6_36
[23] Prasad, D., & Jayswal, S. C. (2019). Assessment of a reconfigurable manufacturing system. Benchmarking: An International Journal, 28(5), 1558–1575. https://doi.org/10.1108/bij-06-2018-0147
[24] Shabaka AI, Elmaraghy HA. Generation of machine configurations based on product features. International Journal of Computer Integrated Manufacturing 2007;20(4):355–69.
[25] Maier-Speredelozzi V, Koren Y, Hu SJ. Convertibility measures for manufacturing systems. Annals of the CIRP 2003;52(1):367–70.

[26] Gumasta K, Gupta SK, Benyoucef L, Tiwari MK. Developing a reconfigurability index using multi-attribute utility theory. International Journal of Production Research 2011;49(6):1669–83.
[27] Abdi MR, Labib AW. A design strategy for reconfigurable manufacturing systems (RMSs) using analytical hierarchical process (AHP): a case study. International Journal of Production Research 2003;41(10):2273–99.
[28] Abdi MR, Labib AW. Grouping and selecting products: the design key of reconfigurable manufacturing systems (RMSs). International Journal of Production Research 2004;42(3):521–46.
[29] Spicer P, Carlo HJ. Integrating reconfiguration cost into the design of multi-period scalable reconfigurable manufacturing systems. Journal of Manufacturing Science and Engineering 2007;129(1):202–10.
[30] Son SY. Design principles and methodologies for reconfigurable machining systems'. Ann Arbor: University of Michigan; 2000 [PhD thesis].
[31] Goyal KK, Jain P, Jain M (2012) Optimal configuration selection for reconfigurable manufacturing system using nsga ii and topsis. International Journal of Production Research 50(15):4175–4191
[32] H. Makino and T. Arai, "New Developments in Assembly Systems," CIRP Annals - Manufacturing Technology, vol. 43, no. 22, pp. 501-512, 1994.
[33] Xia, T.; Xi, L.; Pan, E.; Ni, J. Reconfiguration-oriented opportunistic maintenance policy for reconfigurable manufacturing systems. Reliab. Eng. Syst. Saf. 2017, 166, 87–98.
[34] Kumar, A.; Pattanaik, L.N.; Agrawal, R. Multi-objective Scheduling Model for Reconfigurable Assembly Systems. In Innovations in Soft Computing and Information Technology; Springer: Singapore, 2019; pp. 209–217.
[35] Prasad D, Jayswal SC (2017) Reconfigurability consideration and scheduling of products in a manufacturing industry. International Journal of Production Research https://doi.org/10.1080/00207543.2017.1334979
[36] Prasad D, Jayswal SC (2017) Scheduling of products for reconfiguration effort in reconfigurable manufacturing system. In: Singh SK (ed) 7th International Conference of Materials Processing and Charac- terization (ICMPC 2017), GRIET Hyderabad, India.
[37] Goyal KK, Jain PK, Jain M (2013) A novel methodology to measure the responsiveness of RMTs in reconfigurable manufacturing system. Journal of Manufacturing Systems
[38] Koren Y, Ulsoy A (1997) Reconfigurable manufacturing systems, engineering research center for reconfigurable machining systems (ERC/RMS) report# 1, the university of michigan. Ann Arbor
[39] Koren Y, Heisel U, Jovane F, Moriwaki T, Pritschow G, Ulsoy G, Van Brussel H. Reconfigurable Manufacturing Systems. Annals of the CIRP 1999;48: 527-540.
[40] ElMaraghy HA. Flexible and reconfigurable manufacturing systems paradigms. International Journal of Manufacturing Systems 2006; 17: 261-276.
[41] Wiendahl H-H. Adaptive Production Planning and Control – Elements and Enablers of Changeability. In: ElMaraghy HA, editor. Changeable and Reconfigurable Manufacturing Systems. Berlin: Springer; 2009. p. 197-212.
[42] Schuh G, Westkaemper E, Wiendahl H-H. Liefertreue im Maschinen- und Anlagenbau. Stand - Potenziale - Trends. Aachen: 2006.
[43] Hu SJ. Paradigms of manufacturing – a panel discussion. 3rd Conference on Reconfigurable Manufacturing. Ann Arbor: 2005.
[44] Koren, Y. What are the differences between FMS & RMS. Paradigms of manufacturing – a panel discussion. 3rd Conference on Reconfigurable Manufacturing. Ann Arbor: 2005.
[45] ElMaraghy HA. Flexible and reconfigurable manufacturing systems. Paradigms of manufacturing – a panel discussion. 3rd Conference on Reconfigurable Manufacturing. Ann Arbor: 2005.
[46] Tolio T, Valente A. An Approach to Design the Flexibility Degree in Flexible Manufacturing Systems. In: Proceedings of flexible automation and intelligent manufacturing conference. Limerik: 2006. p. 1229-1236.
[47] Hitomi K. (1994) International Journal of Manufacturing System Design, 2(1), 1-7.
[48] Souza D.E., Williams F.P. (2000) Journal of Operational Analysis, 18(1), 577-593.
[49] Fujii S., Morita H., Kakino Y., Ihara Y.,Takata Y., Murakami D., Miki T., Tatsuta Y. (2000) Proceeding of Pacific Conference Manufacturing, 2(1),970– 980.
[50] Maraghy H.A, Kuzgunkaya O., Urbanic J. (2007) Annals of CIRP, 5(1),445-450.
[51] Brettel M, Klein M, Friederichsen N. The relevance of manufacturing flexibility in the context of Industrie 4.0. Procedia CIRP, 2016, 41: 105–110
[52] Dubey R, Gunasekaran A, Helo P, et al. Explaining the impact of reconfigurable manufacturing systems on environmental performance: The role of top management and organizational culture. Journal of Cleaner Production, 2017, 141: 56–66
[53] Michalek J J, Ceryan O, Papalambros P Y, et al. Balancing marketing and manufacturing objectives in product line design. Journal of Mechanical Design, 2006, 128(6): 1196–1204
[54] Tang L, Yip-Hoi D M, Wang W, et al. Concurrent line-balancing, equipment selection and throughput analysis for multi-part optimal line design. Journal for Manufacturing Science and Production, 2004, 6(1–2): 71–82
[55] Ausaf M F, Gao L, Li X. Optimization of multi-objective integrated process planning and scheduling problem using a priority based optimization algorithm. Frontiers of Mechanical Engineering, 2015, 10(4): 392–404
[56] Wang B, Guan Z, Chen Y, et al. An assemble-to-order production planning with the integration of order scheduling and mixed-model sequencing. Frontiers of Mechanical Engineering, 2013, 8(2): 137–145
[57] Renzi C, Leali F, Cavazzuti M, et al. A review on artificial intelligence applications to the optimal design of dedicated and reconfigurable manufacturing systems. International Journal of Advanced Manufacturing Technology, 2014, 72(1–4): 403–418
[58] Koren Y, Gu X, Freiheit T. The impact of corporate culture on manufacturing system design. CIRP Annals-Manufacturing Technology, 2016, 65(1): 413–416
[59] He N, Zhang D Z, Li Q. Agent-based hierarchical production planning and scheduling in make-to-order manufacturing system. International Journal of Production Economics, 2014, 149: 117– 130 94. Gao R, Wang L, Teti R, et al. Cloud-enabled prognosis for manufacturing. CIRP Annals-Manufacturing Technology, 2015, 64(2): 749–772
[60] Xiong Y, Yin Z. Digital manufacturing—The development direction of the manufacturing technology in the 21st century. Frontiers of Mechanical Engineering in China, 2006, 1(2): 125– 130 96. Monostori L, Kádár B, Bauernhansl T, et al. Cyber-physical systems in manufacturing. CIRP Annals-Manufacturing Technology, 2016, 65(2): 621–641

© IJRASET: All Rights are Reserved |

[61] Guo W, Chen R, Jin J. On-line eccentricity monitoring of seamless tubes in cross-roll piercing mill. ASME Journal Manufacturing Science and Engineering, 2015, 137(2): 021007
[62] Guo W, Shao C, Kim T H, et al. Online process monitoring with near-zero misdetection for ultrasonic welding of Lithium-ion batteries. Journal of Manufacturing Systems, 2016, 38(1): 141–150
[63] Wang S, Chen T, Sun J. Design and realization of a remote monitoring and diagnosis and prediction system for large rotating machinery. Frontiers of Mechanical Engineering in China, 2010, 5 (2): 165–170
[64] Li X, Jiang J, Su H, et al. Identification of abnormal operating conditions and intelligent decision system. Frontiers of Mechanical Engineering in China, 2011, 6(4): 456–462
[65] Xu X, Deng S. Trend prediction technology of condition maintenance for large water injection units. Frontiers of Mechanical Engineering, 2010, 5(2): 171–175
[66] Hu Y, Yang S, Du R. Distributed flexible reconfigurable condition monitoring and diagnosis technology. Frontiers of Mechanical Engineering in China, 2006, 1(3): 276–281
[67] Lee J, Lapira E, Bagheri B, et al. Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters, 2013, 1(1): 38–41
[68] Guo W, Guo S, Wang H, et al. A data-driven diagnostic system utilizing manufacturing data mining and analytics. SAE International Journal of Materials and Manufacturing, 2017, 10(3): 01632923
[69] Guo N, Leu M C. Additive manufacturing: Technology, applications and research needs. Frontiers of Mechanical Engineering, 2013, 8(3): 215–243
[70] Koren Y, Hu S J, Gu P, et al. Open architecture products. CIRP Annals-Manufacturing Technology, 2013, 62(2): 719–729
[71] Hu S J, Ko J, Weyand L, et al. Assembly system design and operations for product variety. CIRP Annals-Manufacturing Technology, 2011, 60(2): 715–733
[72] Koren Y, Hill R. US Patent 6920973, 2004-07-26
[73] Cherubini A, Passama R, Crosnier A, et al. Collaborative manufacturing with physical human-robot interaction. Robotics and Computer-Integrated Manufacturing, 2016, 40: 1–13
[74] Pellegrinelli S, Moro F L, Pedrocchi N, et al. A probabilistic approach to workspace sharing for human-robot cooperation in assembly tasks. CIRP Annals-Manufacturing Technology, 2016, 65(1): 57–60
[75] Wang X V, Kemény Z, Váncza J, et al. Human-robot collaborative assembly in cyber-physical production: Classification framework and implementation. CIRP Annals-Manufacturing Technology, 2017, 66(1): 5.

© IJRASET: All Rights are Reserved |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

symmetry

MDPI

Article

Reversible Data Hiding Using an Improved Pixel Value Ordering and Complementary Strategy

Rajeev Kumar [1], Neeraj Kumar [2] and Ki-Hyun Jung [3,*]

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi 110042, India
[2] Department of Electronics & Communication Engineering, Jamia Millia Islamia University, Delhi 110025, India
[3] Department of Software Convergence, Andong National University, Andong 36729, Republic of Korea
* Correspondence: kingjung@anu.ac.kr; Tel.: +82-54-820-7968

Abstract: Reversible data hiding (RDH) schemes based on pixel value ordering have gained significant popularity due to their unique capability of providing high-quality marked images with a decent embedding capacity, while also enabling secret information extraction and the lossless recovery of the original images at the receiving side. However, the marked image quality may be distorted severely when the pixel value ordering (PVO) method is employed in a layer-wise manner to increase the embedding capacity. In this paper, a new high-capacity RDH scheme using a complementary strategy is introduced to overcome the limitation of the image quality in the case of layer-wise embedding. The proposed RDH scheme first divides the cover image into non-overlapping blocks of 2×2 pixels uniformly and then sorts the pixels of each block according to their intensity values. The secret data are then embedded into two layers. In the first layer, the minimum value of the block is decreased and the maximum value is increased by either 1 or 2 to embed the secret data bits. The second layer is used as a complement to the first layer and is in symmetry with the first layer. In the second layer, the proposed RDH scheme increases the minimum valued pixel and decreases the maximum valued pixel in order to minimize the distortion resulting from the first layer embedding and to embed an additional amount of the secret data. As a result, the proposed RDH scheme significantly increases the embedding capacity, which is clearly evident from the conducted experimental results.

Keywords: complementary strategy; IPVO; PVO-RDH; reversible data hiding

check for updates

Citation: Kumar, R.; Kumar, N.; Jung, K.-H. Reversible Data Hiding Using an Improved Pixel Value Ordering and Complementary Strategy. *Symmetry* **2022**, *14*, 2477. https://doi.org/10.3390/sym14122477

Academic Editor: Jeng-Shyang Pan

Received: 23 September 2022
Accepted: 8 November 2022
Published: 22 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

The internet of things (IoT) is one of the trending technologies in the current technological era. It connects millions of physical devices through the internet network with a high speed and reliable data transfer to implement automatization. Various types of sensors attached to the IoT devices facilitate in the automatization as they collect and share environmental information in digital forms such as text, image, audio, video, and control and measurement signals. IoT networks deployed in the medical and defense fields [1] are sensitive to data integrity and protection. Hence, for the security reasons, digital content is pre-processed before it is shared the network. Pre-processing is performed using one of the two well-known techniques, i.e., encryption and steganography [2]. The encryption technique converts the digital content into a mysterious form using the popular symmetric and/or asymmetric key-based encryption methods. Mysterious content is unrecognizable for the intruders and is not decodable without the corresponding public/private key. However, in some circumstances, this may raise doubts about suspicious activity among outsiders, as the contents (though encoded) are visible to everyone, whereas the steganography technique reduces such risks by covertly embedding confidential messages in some trivial media, known as cover/host media, which can be of any type, such as text, image, audio, video, etc. At times, when both the cover media and embedding contents

are required in their original form at the receiving end, the reversible data hiding (RDH) technique is adopted. The RDH technique guarantees the reversibility of both the cover media and embedding contents at the receiving end [3].

In the early phase, RDH techniques were proposed based on their capacity for lossless compression [3–5] (which usually creates room inside the cover media by compressing its trivial elements to embed the secret message), but these techniques provide a very limited embedding capacity (EC). Thereafter, RDH techniques advanced in three major directions, which are difference expansion (DE) [6], histogram shifting [7,8] and prediction-error expansion (PEE) [9,10]. The difference expansion based the RDH scheme exploits inter-pixel differences to embed the data. In contrast, histogram-shifting-based RDH methods create a histogram of the image and then expand the peak bins while shifting the others. In 2007, Thodi et al. [9] presented a novel high EC with a least distortion RDH technique using prediction error expansion (PEE). The PEE strategy makes use of a predictor to predict a reference pixel based on the surrounding pixels and then modifies the reference pixel to embed the bits of the secret message based on the identified error between the reference pixel and the predicted value. To modify the pixel value, a prediction error histogram (PEH) is plotted, and its peak bins are utilized to embed the secret bits by expanding them and shifting the others to ensure the reversibility. The PEE seeks to embed data in all the pixels and makes only small changes in magnitude ($+1/-1$) per pixel. Thus, it improves both the EC and marked image quality.

A noteworthy improvement of PEE was introduced by Li et al. [10]. Li et al. proposed a pixel value ordering (PVO) technique which provides high-fidelity stego-images with a decent EC. The PVO technique divides the image into uniform-sized blocks and then sorts the pixels of each block. Next, the PVO generates prediction errors by taking the pixel intensity difference between the first and second pixels located at the extreme ends of the block. Thus, the prediction error histogram (PEH) generated from the difference between sorted pixels is usually sharper than the one produced by conventional PEE, which allows for embedding in one of the peak bins while shifting the other bins to ensure the reversibility. Thus, it provides high-fidelity stego-image. However, it has a limited EC, as only one of the peak bins is utilized for embedding the secret data. To overcome this limitation, Peng et al. introduced a value-added extension to PVO, which is popularly known as improved PVO (I-PVO) [11]. I-PVO computes the prediction errors in a different way, which considers the relative locations of the pixels inside the original block so that two peak bins, i.e., '0' and '1', can be expanded to increase the EC, and at the same time, the number of shifted pixels can be reduced to increase the quality of the stego-image. Thus, the IPVO method provides a higher data hiding capacity and stego-image of a better visual quality. Since reversible data hiding can be employed in layer-wise embedding to match the demand for a high capacity, the severity of the distortion of the stego-image is quite high in the case of two-layered IPVO. To address this concern, the proposed work introduces a new RDH scheme which embeds the secret data in two layers, where the second layer is complementary to the first layer. This means that modification performed in the first layer is nullified to a great extent in the second layer of embedding. Thus, the proposed RDH scheme significantly increases the data hiding capacity while maintaining the marked image quality. The rest of the paper is organized as follows. Section 2 discusses the related work. In Section 3, the proposed RDH scheme is explained in detail. The experimental results are discussed in Section 4. In Section 5, the conclusion of the work is provided.

2. Related Works

In this section, pixel-value-ordering-based RDH schemes are briefly reviewed. As discussed in Section 1, the PVO-based RDH scheme was first introduced by Li et al. in 2012 [10]. The PVO-RDH [10] provides high-fidelity stego-images with a decent embedding capacity. The work of [10] was extended by Peng et al. [11] in 2013, proposing an improved PVO-based RDH scheme with the aim of ensuring a higher embedding capacity with less distortion. The IPVO scheme makes use of '0' and '1' prediction errors (which are

usually the peak bins in the prediction error histograms (PEH)) to hide the secret data. Kumar et al. [12] extended the work of [11] to further increase the embedding capacity. For this purpose, Kumar et al. discussed a new IPVO-RDH scheme using the bin reservation strategy. The bin reservation strategy basically creates room to shift the minimum- and maximum-valued pixels to either of the sided, i.e., the left and right sides, so that a greater number of secret information bits can be embedded. However, this increase in the embedding capacity is achieved at the cost of the image distortion. The IPVO-RDH scheme [11] was further researched and improved in the study reported in [13]. In 2018, Jain and Kasana [13] proposed a new data hiding scheme that can embed the secret data into the already sorted blocks of the original cover image, unlike the method employed in [10], so that the embedding capacity can be increased by reducing the number of non-concealable image blocks. In 2019, Kumar et al. [14] introduced another new IPVO-based reversible data hiding scheme using the block extension strategy. The block extension strategy basically extends the smooth block size by two pixels so that the number of non-concealable pixels is reduced. Thus, the blocks of 2×2 pixels are extended to 2×3 pixels (in the case of extremely smooth blocks), and the embedding in the additional pixels is performed in a pixel-wise manner, as in [15], while the embedding in the original 2×2 block is performed in the manner described in [11]. In another noteworthy improvement, Weng et al. [16] introduced a novel PVO-RDH scheme which flexibly modifies the pixels in terms of the block-wise complexity. The scheme conceals the bits of secret information based on the block smoothness, which leads to the optimal utilization of the block redundancy. As a result, a significant enhancement in the embedding performance is achieved.

In addition to the aforementioned schemes, many other schemes, such as those described in [17,18], that use dynamic block size were introduced in the literature to improve the embedding performance of the schemes described in [10,11]. The scheme introduced by Di et al. [18] makes use of a quadtree for efficient embedding by PVO-RDH. The quadtree approach adaptively generates blocks of different sizes based on the correlations of image region, and then the secret data are optimally embedded. By introducing a new direction for efficient embedding, Zhao et al. [19] improved the PVO-RDH schemes using the adaptive threshold strategy for efficiently categorizing noisy blocks. The standard deviation is utilized to efficiently identify the noisy blocks, instead of the difference between the second and third pixels of the sorted blocks. The adaptive-threshold-based strategy helps to achieve a higher-quality stego-image [19,20].

Lee et al. [21] developed a "star-shaped" PVO-based reversible data hiding scheme. The scheme is different from other contemporary schemes, such as Li et al.'s method [10], as the proposed scheme takes a fixed 3×3 block to embed the secret data efficiently and utilizes the prevalent correlations between the pixel elements of the block for the preservation of the image quality and to ensure a high embedding capacity. Furthermore, it enables the sender/hider to perform multi-layer embedding, so that embedding capacity can be increased. He et al. [22] discussed a multi-pass PVO-based RDH scheme which takes a defined number of maximum- and minimum-valued pixels within the block as independent units to embed the secret data. The method increases the data hiding capacity using optimally combined embedding. Weng et al. [23] further extended the work on K-pass pixel value ordering to enhance the embedding performance. Weng et al.'s scheme exploits the traditionally excluded pixels to ensure a more accurate prediction of the local complexity, so that efficient embedding could be performed. Ou et al. [24] proposed a PVO-based data embedding technique in a two-dimensional space. The technique utilizes the correlations between the prediction errors in a block so that the redundancy of the block can be exploited effectively. The authors further discussed a dynamic selection method that can be used to treat different regions of the images differently based on their correlation level so as to enhance the embedding efficiency. He et al. [25] further extended this work by proposing a new RDH method based on optimally manipulating a 2D-PEH. The method flexibly predicts and modifies the maximum and minimum pixels, unlike those described in [10,11], to increase the embedding capacity. Kumar et al. [26] discussed a new enhanced

pairwise IPVO-based RDH scheme which embeds the secret data in not only a block-wise manner but also a pixel-wise manner, so that the pixels can be optimally utilized to embed the secret information. Additionally, the scheme embeds the secret data using the enhanced pairwise IPVO strategy in the case of block-wise embedding to limit the caused distortion. Furthermore, layer-wise embedding is also performed, which helps to further increase the embedding capacity. Thus, the scheme provides a high data hiding capacity with a sustained image quality. The work of [26] was further developed by Kaur et al. [27] in 2021 to further enhance the embedding performance. A comprehensive review of PVO-based RDH schemes can be further consulted in [28].

Based on the analysis of the aforementioned PVO-RDH schemes, it can be observed that none of the existing schemes aim to improve the IPVO scheme for layer-wise embedding. Thus, in this paper, a new RDH scheme using a complementary embedding strategy is proposed. The proposed scheme extended the work described in [11] to increase the embedding capacity. The detailed algorithm of the proposed complementary scheme is outlined in Section 4. The next section briefly reviews Peng at al.'s scheme.

3. The Proposed RDH Scheme

In this section, the proposed IPVO-based reversible data hiding scheme using a complementary embedding strategy is presented. Since it was proved in [11] that partitioning the image in 2×2 blocks provides the highest data hiding capacity, the proposed RDH scheme also partitions the host image into blocks of 2×2 pixels. The pixels of each block are first arranged in ascending order, and then the minimum- and maximum-valued pixels are prudently modified to conceal the secret information bits. The secret information embedding is performed in two layers, where the second layer is complementary to the first layer's work and is almost in symmetry with the first layer. The proposed scheme modifies the pixels in the second layer in such a way that the modification performed in the first layer can be minimized to a great extent. In the next subsections, the proposed embedding algorithm, and extraction and image recovery algorithm are described.

3.1. Embedding Algorithm

In this subsection, the detailed embedding algorithm as described at Algorithm 1 is presented.

Algorithm 1. Embedding Algorithm

1. Partition the host image (H) into blocks of 2×2 pixels $(P_1, \ldots P_4)$ and arrange the pixels of each block in ascending order $(P_{\pi(1),}\ P_{\pi(2),}\ P_{\pi(3),}\ P_{\pi(4)}\ )$, where $P_{\pi(1)} \leq \cdots \leq P_{\pi(4)}$, $\pi(i) < \pi(j)$ if $P_{\pi(i)} = P$ and $i < j$, as in [11].
2. Design a location map (LM) to represent each image block (B_i) by either '0' or '1'. The blocks which have $P_{\pi(1),} < 2$, or $P_{\pi(4),} > 253$ are represented by '0' in the LM. All the other blocks are represented by '1'. LM is compressed using arithmetic encoding to obtain a compressed location map (CLM) of the length $LCLM$, where $LCLM$ represents the length of the compressed location map.
3. Classify block (B_i) into three categories, namely the probable overflow/underflow, complex, and smooth, based on following ruleset:

If B_i is represented by '0' in LM, **then**
B_i is assumed to be a probable overflow/underflow block, which is not considered for embedding S.
else B_i is represented by '1' in LM, **then**
compute noise level (NL_i) such that $NL_i = P_{\pi(3)} - P_{\pi(2)}$, and
if $NL_i > thr$, **then**
B_i is assumed to be a complex block, which, again, is not considered for embedding S.
else B_i is assumed to be a smooth block, which is considered for embedding S.
end if
end if

4. In the case of smooth blocks:

Compute the prediction differences PD_{max} and PD_{min} using Equations (1) and (2), respectively, which are defined as follows:

$$PD_{max} = P_u - P_v \tag{1}$$

$$PD_{min} = P_s - P_t \tag{2}$$

where $u = min(\pi(4),\ \pi(3), v = max(\pi(4), \pi(3)), s = min(\pi(2), \pi(1))$, and $t = \max(\pi(2),\ \pi(1))$.
Embed the bits (b) from S into the maximum-valued pixel $P_{\pi(4)}$ and minimum-valued pixel $P_{\pi(1)}$ using Equations (3) and (4), respectively, which are defined as follows:

$$P\prime_{\pi(4)} = \begin{cases} P_{\pi(4)} + (b+1),\ if\ PD_{max} = 1, \\ P_{\pi(4)} + 2,\ if\ PD_{max} > 1, \\ P_{\pi(4)} + (b+1),\ if\ PD_{max} = 0, \\ P_{\pi(4)} + 2,\ if\ PD_{max} < 0 \end{cases} \tag{3}$$

$$P\prime_{\pi(1)} = \begin{cases} P_{\pi(1)} - (b+1),\ if\ PD_{min} = 1, \\ P_{\pi(1)} - 2,\ if\ PD_{min} > 1, \\ P_{\pi(1)} - (b+1),\ if\ PD_{min} = 0, \\ P_{\pi(1)} - 2,\ if\ PD_{min} < 0 \end{cases} \tag{4}$$

Compute the new prediction differences $PD\prime_{max}$ and $PD\prime_{min}$ between the newly obtained first maximum $(P\prime_{\pi(4)})$ and second maximum pixels $(P_{\pi(3)})$ and the first minimum $(P\prime_{\pi(1)})$ and second minimum pixels $(P_{\pi(2)})$ using Equations (1) and (2), respectively.
Embed the bits (b) from S into the newly obtained maximum-valued pixel $P\prime_{\pi(4)}$ and newly obtained minimum-valued pixel $P\prime_{\pi(1)}$ using Equations (5) and (6), respectively, which are defined as follows.

$$P\prime_{\pi(4)} = \begin{cases} P\prime_{\pi(4)} - b,\ if\ PD'_{max} = -3\ or\ 3, \\ P\prime_{\pi(4)},\ if\ PD\prime_{max} > 3, \\ P\prime_{\pi(4)},\ if\ PD\prime_{max} < -3, \\ P\prime_{\pi(4)} - 1,\ else \end{cases} \tag{5}$$

$$P\prime\prime_{\pi(1)} = \begin{cases} P\prime_{\pi(1)} + b,\ if\ PD'_{min} = -3 \text{ or } 3, \\ P\prime_{\pi(1)},\ if\ PD\prime_{min} > 3, \\ P\prime_{\pi(1)},\ if\ PD\prime_{min} < -3, \\ P\prime_{\pi(1)} + 1,\ else \end{cases} \tag{6}$$

The process of secret information embedding is halted in a case where the all the bits from S are embedded or the last information carrying block B_{end} is reached.

5. Concatenate all the auxiliary information, such as the length of the compressed location map $(LCLM)$, compressed location map (CLM), thr, and B_{end} , which is required for blind decoding to obtain a list Aux.
6. Extract and conceal the least significant bit (LSB) of the first $LAux$ pixels (where $LAux$ represents the length of Aux) in the remaining blocks using *Step 4*. Then, use the simple LSB substitution method to substitute the LSB of first $LAux$ pixels with the bits of Aux.

Thus, all the bits of the secret information bitstream (S) are concealed, and a stego-image H′ is obtained. Next, an illustrative example of the proposed complementary embedding method is discussed.

3.2. Illustrative Example of Embedding Algorithm

In this subsection, an illustrative example of the embedding process of the proposed complementary embedding method is discussed in detail using Figure 1. Here, a host image block of 2 × 2 pixels ($P_1,\ P_2,\ P_3,\ P_4$) = (105, 104, 100, 99) is considered. The pixels of the block are then arranged in ascending order, such as ($P_{\pi(1)},\ P_{\pi(2)},\ P_{\pi(3)},\ P_{\pi(4)}$) = (99, 100,

104, 105). It should be noted that the example does not take the location map and noise level threshold into consideration, so that the focus can be maintained purely on the embedding procedure. Next, two prediction differences (PD_{max} and PD_{min}) are computed, such that $PD_{max} = (105 - 104) = 1$ and $PD_{min} = (100 - 99) = 1$. Suppose that three bits of secret data ("100") are to be embedded into the block. The values of PD_{min} and PD_{max}, are both 1, meaning the values $P_{\pi(1)} = 99$ and $P_{\pi(4)} = 105$ will be decreased and increased by $(b + 1)$, respectively. Thus, the minimum and maximum pixels after the first layer of data embedding become $(99 - (1 + 1) = 97)$ and $(105 + (0 + 1) = 106)$. For the second layer of data embedding, two different values, namely ($PD\prime_{max}$ and $PD\prime_{min}$), are calculated using the new maximum and minimum pixels, such that $PD\prime_{max} = (106 - 104) = 2$ and $PD\prime_{min} = (100 - 97) = 3$. In the second layer, the information is embedded only when the difference is equal to -3 or 3 in Step 5d of the embedding algorithm. In this example, $PD\prime_{min}$ is 3, which means that a new minimum value after the embedding of the secret data bit, i.e., '0', will be $P''_{\pi(1)} = (97 + 0) = 97$. The maximum pixel will not be used for the secret data embedding, as $PD\prime_{max}$ is 2, but its value will be decreased by 1, so that the host image can be restored. Thus, the final stego-block is =(105, 104, 100, 97).

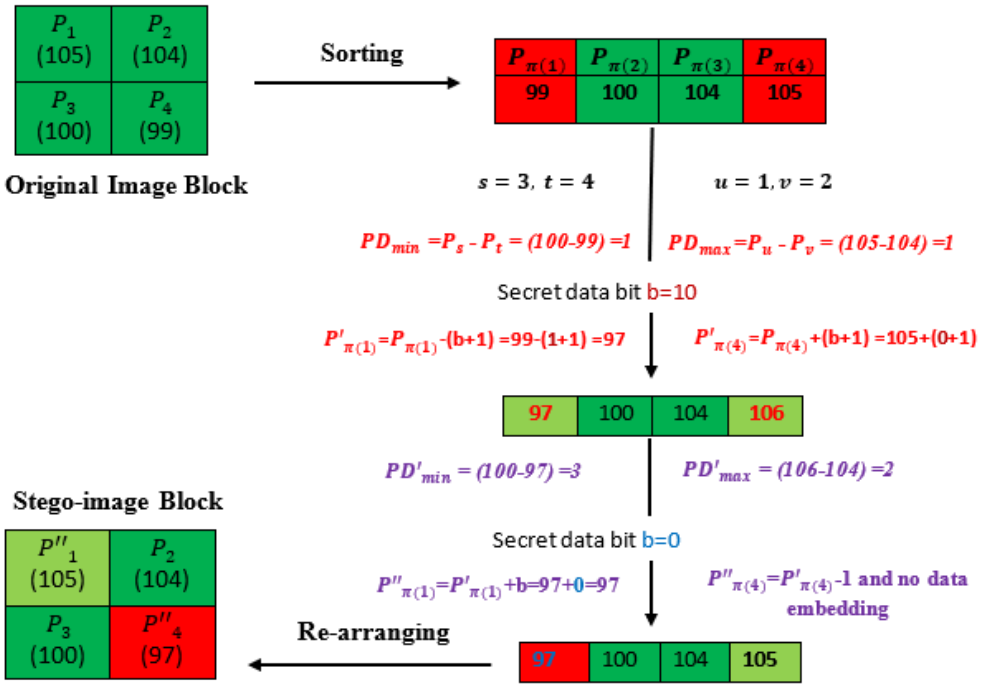


Figure 1. An illustrative example of the proposed embedding procedure.

3.3. Extraction and Image Recovery Algorithm

In this subsection, the proposed extraction and image recovery algorithm as Algorithm 2 is presented.

Algorithm 2. Extraction and Image Recovery Algorithm

1. Read the LSBs of some pre-defined pixels to obtain the length of the auxiliary information Aux. Next, extract and process the entire auxiliary information by reading the LSBs of the next $LAux$ pixels.
2. Partition the stego-image $(H\prime)$ into blocks of 2×2 pixels $(P''_1, \ldots P''_4)$ and arrange the pixels of each block in ascending order $(P_{\pi(1)}, P_{\pi(2)}, P_{\pi(3)}, P_{\pi(4)})$, where $P_{\pi(1)} \leq \cdots \leq P_{\pi(4)}$, $\pi(i) < \pi(j)$ if $P_{\pi(i)} = P_{\pi(j)}$ and $i < j$, as in [11].

3. Classify the block (B_i) starting from the (B_{end+1}) into three categories, namely the probable overflow/underflow, complex, and smooth, based on the following ruleset:

if B_i is represented by '0' in LM, **then**
the B_i is assumed to be a probable overflow/underflow block, which does not have any secret information bits.
else B_i is represented by '1' in the LM, **then**
compute the noise level (NL_i) such that $NL_i = P_{\pi(3)} - P_{\pi(2)}$, and
if the $NL_i > thr$, **then**
B_i is assumed to be a complex block, which, again, does not have any secret information bits.
else B_i is assumed to be a smooth block, and proceed to the next step.
End if
end if

4. **if** B_i is a smooth one, **then**

Compute the two prediction differences PD''_{max} and PD''_{min} between the first maximum ($P''_{\pi(4)}$) and second maximum pixels ($P_{\pi(3)}$) and the first minimum ($P''_{\pi(1)}$) and second minimum pixels ($P_{\pi(2)}$) using Equations (1) and (2), respectively.

Extract the concealed bits of secret information and restore the pixel values using following the ruleset:

if $PD''_{max} == -3$ *or* $PD''_{max} == 3$, **then**
b = '0' and $P\prime_{\pi(4)} = P''_{\pi(4)}$.
else if $PD''_{max} == -2$ *or* $PD''_{max} == 2$, **then**
b = '1' and $P\prime_{\pi(4)} = P''_{\pi(4)} + 1$.
else if $PD''_{max} < -3$ *or* $PD''_{max} > 3$, **then**
$P\prime_{\pi(4)} = P''_{\pi(4)}$, and there is no hidden data.
else $P\prime_{\pi(4)} = P''_{\pi(4)} + 1$.
end if

if $PD''_{min} == -3$ or $PD''_{min} == 3$, **then**
b = '0' and $P\prime_{\pi(1)} = P''_{\pi(1)}$.
else if $PD''_{min} == -2$ or $PD''_{min} == 2$, **then**
b = '1' and $P\prime_{\pi(1)} = P''_{\pi(1)} - 1$.
else if $PD''_{min} < -3$ *or* $PD''_{min} > 3$, **then**
$P\prime_{\pi(1)} = P''_{\pi(1)}$, and there is no hidden data.
else $P\prime_{\pi(1)} = P''_{\pi(1)} - 1$.
end if

Compute the two prediction differences $PD\prime_{max}$ and $PD\prime_{min}$ between the first maximum ($P\prime_{\pi(4)}$) and second maximum pixels ($P_{\pi(3)}$) and the first minimum ($P\prime_{\pi(1)}$) and second minimum pixels ($P_{\pi(2)}$) using Equations (1) and (2), respectively.

Extract the concealed bits of secret information and restore the pixel values using the following ruleset:

if $PD\prime_{max} == -2$ or $PD\prime_{max} == 3$, **then**
b = '1' and $P_{\pi(4)} = P\prime_{\pi(4)} - 2$.
else if $PD\prime_{max} == -1$ or $PD\prime_{max} == 2$, **then**
b = '0' and $P_{\pi(4)} = P\prime_{\pi(4)} - 1$.
else if $PD\prime_{max} < -2$ *or* $PD'_{max} > 3$, **then**
$P_{\pi(4)} = P\prime_{\pi(4)} - 2$, and there is no hidden data.
else $P_{\pi(4)} = P\prime_{\pi(4)}$.
end if

if $PD\prime_{min} == -2$ or $PD\prime_{min} == 3$, **then**
b = '1' and $P_{\pi(1)} = P\prime_{\pi(1)} + 2$.
else if $PD\prime_{min} == -1$ or $PD\prime_{min} == 2$, **then**
b = '0' and $P_{\pi(1)} = P\prime_{\pi(1)} + 1$.
else if $PD\prime_{min} < -2$ *or* $PD'_{min} > 3$, **then**
$P_{\pi(1)} = P\prime_{\pi(1)} - 2$, and there is no hidden data.
else $P_{\pi(1)} = P\prime_{\pi(1)}$.
end if
end if

The step is repeated until the $LAux$ bits are retrieved by processing the bocks of the image.

5. Substitute the LSB of the first $LAux$ pixels with the bits obtained from Step 4. Next, iteratively execute Step 2 to Step 4 until the block (B_{end}) is detected, so that the complete hidden bitstream of the secret information (S) and the host image (H) can be obtained.

3.4. Illustrative Example of the Extraction and Image Recovery Algorithm

In this subsection, an illustrative example of the extraction and image recovery process of the proposed complementary method is discussed in detail using Figure 2. Here, a stego-image block of 2 × 2 pixels (P''_1, $P\prime_2$, $P\prime_3$, P''_4) = (105, 104, 100, 97) is considered. The pixels of the block are then arranged in ascending order, such as ($P''_{\pi(1)}$, $P''_{\pi(2)}$, $P''_{\pi(3)}$, $P''_{\pi(4)}$) = (97, 100, 104, 105). Next, the two prediction differences PD''_{max} and PD''_{min} are computed, such that $PD''_{max} = (105 - 104) = 1$ and $PD''_{min} = (100 - 97) = 3$. It should be noted that the secret information is assumed to be embedded if the difference belongs to either −3 or 3 or −2 or 2, as per the Step 5b of the algorithm, in the first layer. In this example, $PD\prime_{min} = (100 - 97) = 3$, meaning there are '0′ bits concealed in $P''_{\pi(1)}$, and nothing is hidden in $P''_{\pi(4)}$, as per the given rules of the extraction algorithm. Then, the pixel values after the extraction of the secret data, as per the given rules of the algorithm, can be recovered. The new values are $P\prime_{\pi(1)} = P''_{\pi(1)} = 97$ and $P\prime_{\pi(2)} = P''_{\pi(2)} + 1 = 105 + 1 = 106$. For the second layer of the data extraction and image recovery, two new prediction differences, $PD\prime_{max}$ and $PD\prime_{min}$, are computed such that $PD\prime_{max} = (106 - 104) = 2$ and $PD\prime_{min} = (100 - 97) = 3$. It should be noted that the secret information is assumed to be embedded if the difference is either −2 or 3 or −1 or 2, as per the Step 5d of the algorithm, in the second layer. Here, $PD\prime_{min}$ *is* 3, meaning there is '1′ bit concealed, and the original pixel value can be restored such that $P_{\pi(1)} = P\prime_{\pi(1)} + 2 = 97 + 2 = 99$. In the case of the $PD\prime_{max}$, the concealed bit of the secret information is '0′, and the original pixel value can be restored such that $P_{\pi(4)} = P\prime_{\pi(4)} - 1 = 106 - 1 = 105$. Thus, the original host image block (105, 104, 100, 99) after the extraction of the secret information '100′, is received.

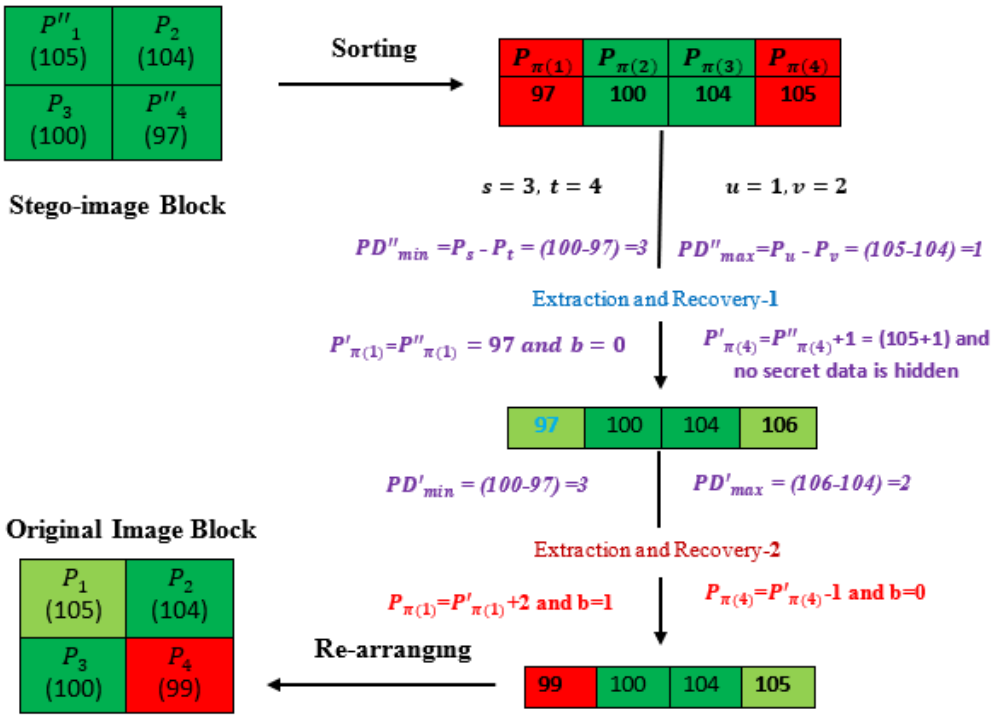


Figure 2. An illustrative example of the proposed extraction and image recovery procedure.

4. Experimental Results and Discussion

This section discusses the experimental performance analysis of our proposed complementary RDH scheme in comparison with the related existing schemes, such as that

of He et al. [25], Weng et al. (K-Pass) [23], that of Weng et al. (DIPVO) [16], Peng et al. (Two-layered) [11] and Di et al.'s method [18]. Peng et al.'s two-layered method is basically a two-time implementation on a given image. The performance of Peng et al.'s two-layered method is evaluated here mainly to enable a comprehensive analysis comparing it with the proposed method, which also makes use of two-layer embedding. However, the second layer embedding of the proposed scheme is basically a complement to the first layer embedding. To cover address the state of the art of pixel-value-ordering-based hiding methods, the results are comparatively evaluated against the results of some of the recently published methods, such as He et al.'s and Di et al.'s methods, along with the parent method, i.e., Peng et al.'s method. The experimental results of Peng et al. [11] and Peng et al.'s two-layered method are taken for the 2×2 pixels blocks, as it was observed from the literature that the maximum embedding capacity is achieved at a block size of 2×2 pixels. The implementation is performed using MATLAB running on an Intel (R) Core (TM) *i*5 processor, 3.20 GHz, with a 4 GB RAM hardware platform. The secret data stream used in our experiment was generated using a random function

The implementation of the complementary RDH scheme along with Peng et al.'s two-layered scheme was performed in MATLAB for experimental purposes. The experimental results of He et al. [25] and Di et al.'s method [18], the Weng et al. (K-Pass) [23] and Weng et al. (DIPVO) [16] method, were taken from their articles. The experiments were performed using standard images, as shown in Figure 3. The performance metrics used for evaluating the existing and the complementary RDH schemes were the embedding capacity and peak signal-to-noise ratio (PSNR). The marked images which were obtained after embedding the secret information through the proposed schemes are shown in Figure 4.

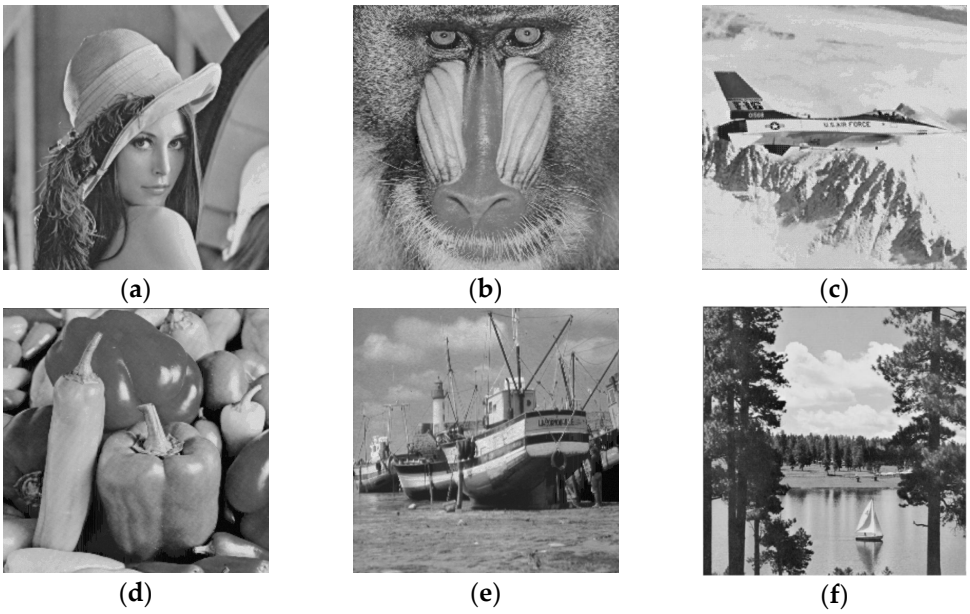


Figure 3. Cover images sized 512×512 pixels. (**a**) Lena. (**b**) Baboon. (**c**) Airplane. (**d**) Peppers. (**e**) Fishing boat. (**f**) Sail boat.

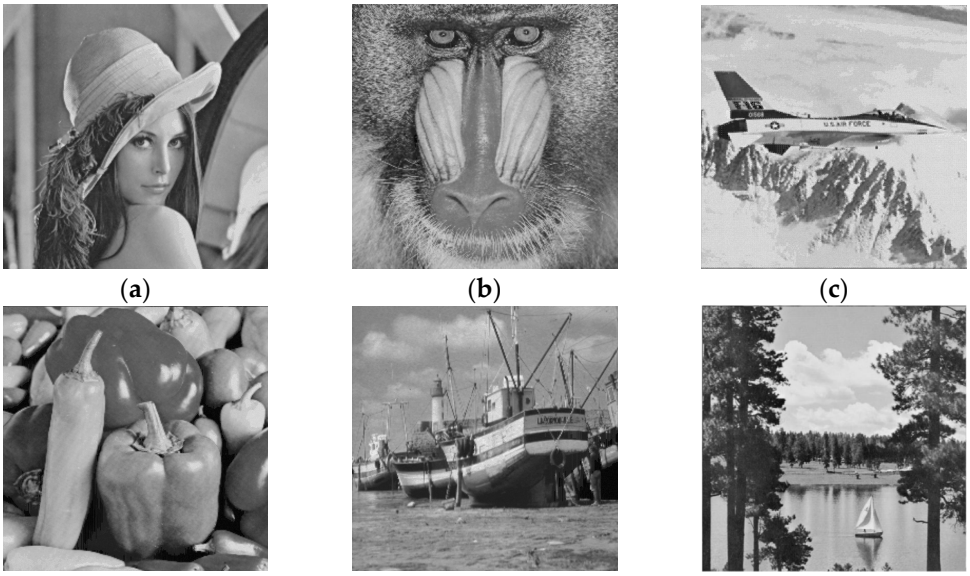


Figure 4. Stego-images sized 512 × 512 pixels. (**a**) Lena. (**b**) Baboon. (**c**) Airplane. (**d**) Peppers. (**e**) Fishing boat. (**f**) Sail boat.

The experimental results and comparison of the proposed complementary RDH scheme with the existing ones, such as that of He et al. [25], Weng et al. (K-Pass) [23], that of Weng et al. (DIPVO) [16], Peng et al.'s [11] two-layered method and Di et al.'s method [18] are illustrated. Figure 5 illustrates the embedding performance of the different existing methods in comparison with the proposed complementary RDH scheme. In our experiments, the noise level thresholds '65′ and '4′ were considered for the baboon, and the remaining test images, respectively, as far as Peng et al. and Peng et al.'s two-layered method and the proposed scheme are concerned. For the illustrative comparison, we varied the embedding capacity from 5000 bits to its maximum value for all the aforementioned schemes. The figure shows that the performance of the complementary RDH scheme has a low PSNR in comparison to the other popular methods, including those of He et al. [25], Weng et al. [16], Peng et al. [11] and Di et al. [18] and the improved K-pass [23]. The main reason for this inferiority is the use of two-layer data embedding, as all of these methods employ only one-layer embedding, which causes a maximum ± 1 change in the pixel value, whereas the proposed scheme employs two-layer embedding, which causes a ± 2 change in the pixel value. Therefore, the methods such as those of He et al. [25], Weng et al. [16], Peng et al. [11] and Di et al. [18] and Weng et al. (K-Pass) [23] achieve a higher PSNR, comparatively. However, the complementary RDH scheme provides superior results in terms of the data embedding capacity in comparison to the existing aforementioned schemes. Furthermore, the quality of the stego-image is also maintained, since second layer-embedding is used to recover the pixel values to some extent. Moreover, it has a significantly higher PSNR and data hiding capacity for all the test images (except the airplane) than Peng et al.'s two-layered method, which is the closest method in terms of the style of the secret data embedding.

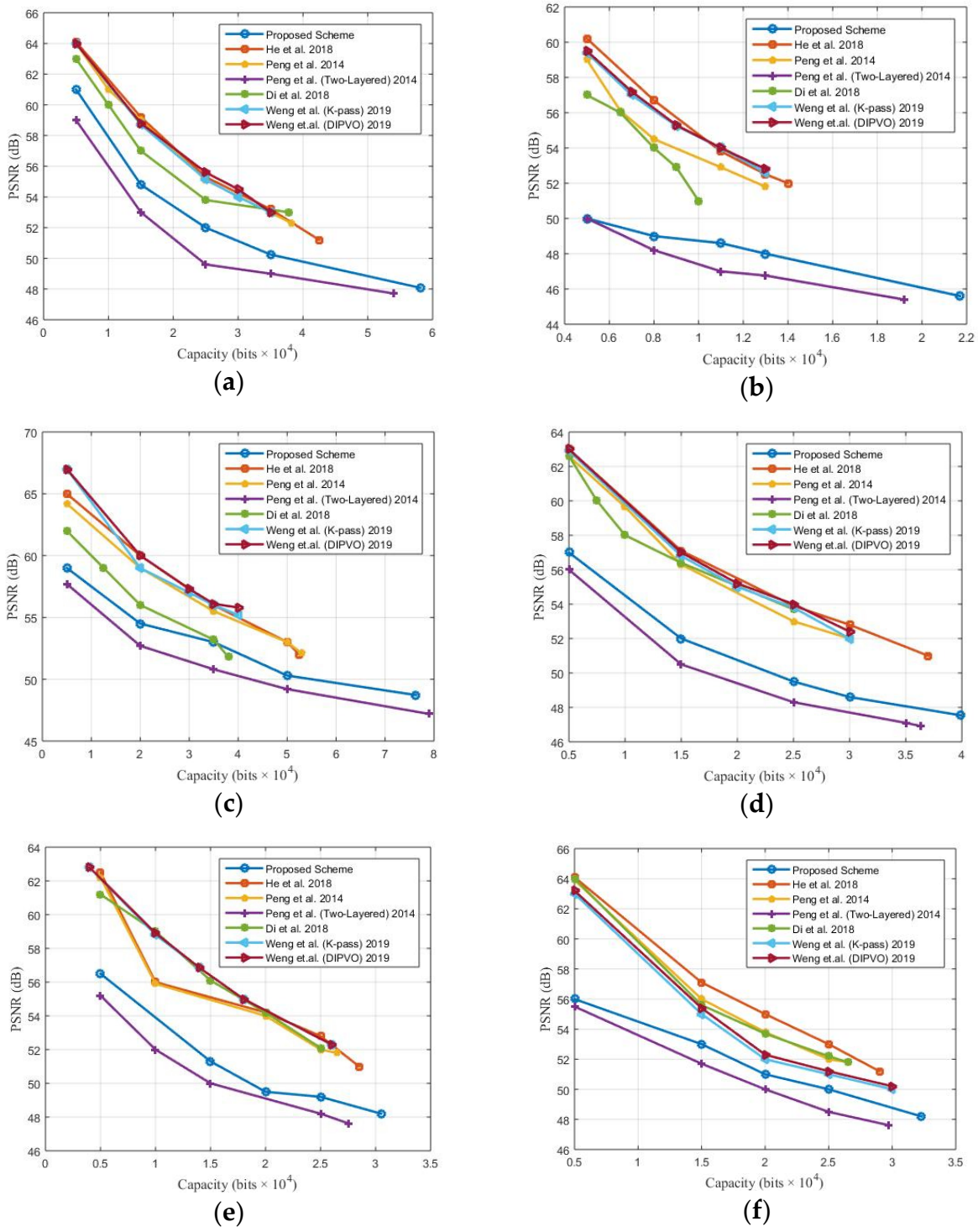


Figure 5. Performance comparisons of PSNR vs. embedding capacity of the proposed scheme with He et al. [25], Peng et al.'s [11], Peng et al. (Two-layered), Di et al.'s method [18], Weng et al. (K-Pass) [23], and Weng et al. (DIPVO) [16]. (**a**) Lena. (**b**) Baboon. (**c**) Airplane. (**d**) Peppers. (**e**) Fishing boat. (**f**) Sail boat.

To succinctly evaluate the performance of the proposed RDH scheme with respect to Peng et al.'s two-layered method at different thresholds (noise level), two cover images, namely the Lena and baboon images, were used, as these represent the extreme directions of digital images as far as standard images are concerned.

Not all the results for all the images are presented due to their similar nature. The experimental results (as shown in Figure 6) were captured by varying the noise level threshold in the ranges of 2 to 20 and 20 to 80 in the case of Lena and the baboon, respectively. It was observed in the experimental analysis that the embedding capacity is usually directly proportional to the noise level threshold, whereas the PSNR is indirectly proportional. The proposed complementary RDH scheme performs significantly better than the Peng et al.'s two-layered method using the same parameters, as evident from Figure 6. Thus, it can easily be stated that the proposed RDH scheme successfully improves the working of the PVO-based RDH schemes as far as the layer-wise embedding is concerned.

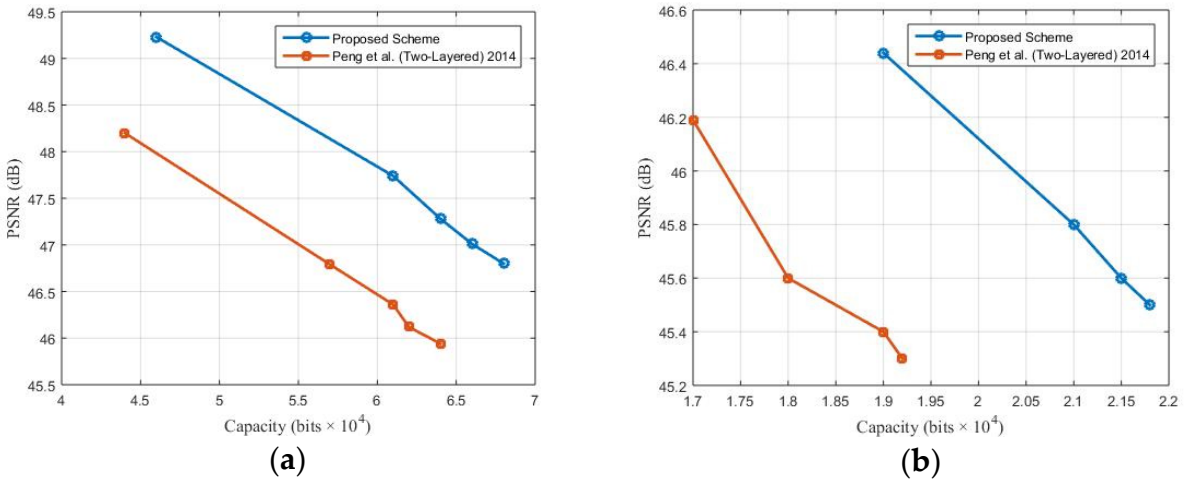


Figure 6. Comparative analysis of the embedding performance at different thresholds of the proposed scheme with the Peng et al. (Two-layered) [11]. (**a**) Lena. (**b**) Baboon.

4.1. Histogram Attack

A histogram analysis of each of the test images was performed to verify the performance of the proposed method against histogram attacks. The performance analysis was accomplished by comparing the histograms of the original images with the stego-images (after embedding the 30,000 bits, as shown in Figure 3), as seen in Figure 7. It is clear that neither the original image histograms nor their associated stego-histograms images underwent any discernible modification. For the exceptions of the baboon and peppers, the histograms also reveal that the majority of the test images lacked pixels valued at 0 or 255, which could result in an overflow or underflow issue and would be displayed clearly. Even in the case of the baboon and peppers, there are very few of these pixels. Thus, it is evident that the compressed location map can be represented by a small number of bits, indicating that the embedding capacity would barely be affected by the supplementary information.

4.2. Statistical Attack

The created stego-images conceal sensitive information using the proposed scheme to safeguard the original information. The marked pixels in stego-images include the original information that we embedded.

The cover image and stego-image correlation coefficients (CC) and standard deviations (SD) are displayed in Table 1 for the sake of comparison, taking into account $thr = 5$ for all the test images. The CC is closer to one, and the difference in SD is closer to zero, which indicates a good concealment according to the data. Thus, the proposed approach is more resistant to a number of attacks. Furthermore, it is possible to correctly recover the original image from the stego-images and obtain the hidden information without any data loss. Due to the scheme's innocence, it is able to thwart potential malevolent attacks.

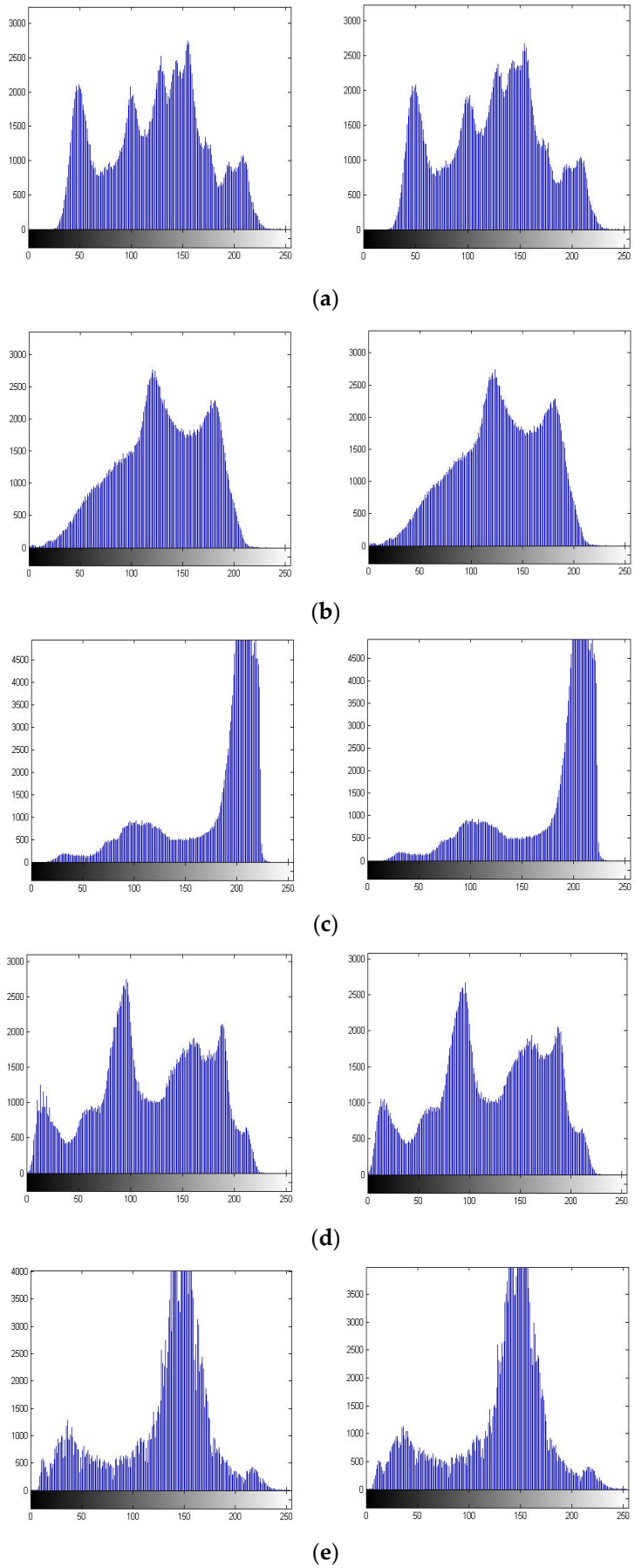
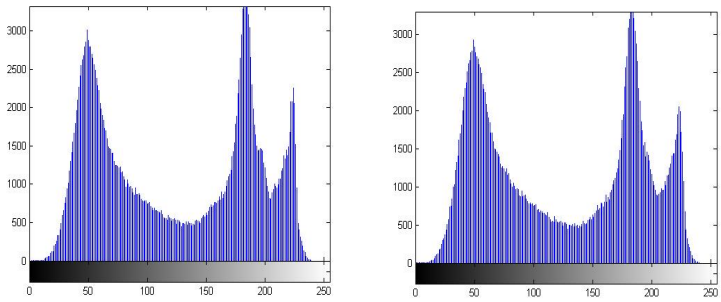


Figure 7. *Cont.*



(**f**)

Original image histogram Stego-image histogram

Figure 7. Histogram distribution of the original and stego-images. (**a**) Lena. (**b**) Baboon. (**c**) Airplane. (**d**) Peppers. (**e**) Fishing boat. (**f**) Sail boat.

Table 1. Standard deviations (SD) and correlation coefficients (CC) of cover images and stego-images.

Test Image	CC	SD (Cover Image)	SD′ (Stego-Image)	Difference (SD′-SD)
Lena	0.9998	47.8557	47.8942	0.0382
Baboon	0.9999	42.3135	42.3845	0.0710
Airplane	0.9999	46.4120	46.4341	0.0220
Peppers	0.9999	53.8794	53.9227	0.0433
Fishing Boat	0.9998	46.6772	46.7309	0.0537
Sail-Boat	0.9999	65.5815	65.6142	0.0328

5. Conclusions

In this paper, a new RDH scheme using the complementary embedding strategy has been proposed. The proposed scheme aimed to preserve the marked image quality in the case of layer-wise embedding. For a high image quality, a cover image was first partitioned uniformly into non-overlapping blocks, and then the secret data were able to be concealed in the maximum and minimum pixel values only in the two layers. In the first layer, the minimum and maximum pixels were decreased and increased by either by '1' or '2', respectively. In the second layer, the minimum pixel value was increased by '0' or '1' and the maximum pixel value was decreased by '0' or '1' to embed the secret data. As a result, the ordering of the pixel values was preserved to ensure the reversibility in the block. The experimental results provided an evidence of the superior embedding capacity of the scheme, with a good PSNR, compared to the previous related works. The proposed scheme could hide 68,000 bits on 46.7 dB and 21,000 bits on 45.5 dB in the Lena and baboon images respectively.

Author Contributions: Conceptualization, R.K. and N.K.; methodology, K.-H.J. and N.K.; visualization, R.K.; writing—original draft, R.K. and N.K.; writing—review and editing, R.K. and K.-H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2021R1I1A3049788), and Brain Pool program, funded by the Ministry of Science and ICT through the National Research Foundation of Korea (2019H1D3A1A01101687, 2021H1D3A2A01099390).

Data Availability Statement: Data is available on the request to any of the author.

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

References

1. Singh, Y.; Singh, S.; Kumar, R. A distributed energy-efficient target tracking protocol for three level heterogeneous sensor networks. *Int. J. Comput. Appl.* **2012**, *51*, 0975–8887. [CrossRef]
2. Kumar, R.; Saini, K.K.; Chand, S. A new steganography technique using snake scan ordering strategy. *Int. J. Image* **2013**, *6*, 25–32. [CrossRef]
3. Kumar, R.; Chand, S. A reversible high capacity data hiding scheme using pixel value adjusting feature. *Multimed. Tools Appl.* **2016**, *75*, 241–259. [CrossRef]
4. Kumar, R.; Chand, S. A reversible data hiding scheme using pixel location. *Int. Arab J. Info. Tech.* **2018**, *15*, 763–768.
5. Malik, A.; Singh, S.; Kumar, R. Recovery based high capacity reversible data hiding scheme using even-odd embedding. *Multimed. Tools Appl.* **2018**, *77*, 15803–15827S. [CrossRef]
6. Tian, J. Reversible data embedding using a difference expansion. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 890–896. [CrossRef]
7. Kumar, R.; Chand, S.; Singh, S. An Improved Histogram-Shifting-Imitated Reversible Data Hiding based on HVS Characteristics. *Multimed. Tools Appl.* **2018**, *77*, 13445–13457. [CrossRef]
8. Kumar, R.; Chand, S. A novel high capacity reversible data hiding scheme based on pixel intensity segmentation. *Multimed. Tools Appl.* **2017**, *76*, 979–996. [CrossRef]
9. Thodi, D.M.; Rodriguez, J.J. Expansion embedding techniques for reversible watermarking. *IEEE Trans. Image Process.* **2007**, *16*, 721–730. [CrossRef]
10. Li, X.; Li, J.; Li, B.; Yang, B. High-fidelity reversible data hiding scheme based on pixel–value-ordering and prediction–error expansion. *Signal Process.* **2013**, *93*, 198–205. [CrossRef]
11. Peng, F.; Li, X.L.; Yang, B. Improved PVO-based Reversible Data Hiding. *Digit. Signal Process.* **2014**, *25*, 255–265. [CrossRef]
12. Kumar, R.; Kumar, N.; Jung, K.H. I-PVO based high capacity reversible data hiding using bin reservation strategy. *Multimed. Tools Appl.* **2020**, *79*, 22635–22651. [CrossRef]
13. Kumar, N.; Kasana, S.S. High-capacity reversible data hiding using modified pixel value ordering approach. *J. Circuits Syst. Comput.* **2018**, *17*, 175–185.
14. Kumar, R.; Kim, D.S.; Lim, S.; Jung, K.H. High-Fidelity Reversible Data Hiding Using Block Extension Strategy. In Proceedings of the 34th International Technical Conference on Circuits/Systems, Jeju, Republic of Korea, 23–26 June 2019; Volume 19, pp. 1–4.
15. Qu, X.; Kim, H.J. Pixel-based pixel value ordering predictor for high-fidelity reversible data hiding. *Signal Process.* **2015**, *111*, 249–260. [CrossRef]
16. Weng, S.; Shi, Y.; Hong, W.; Yao, Y. Dynamic improved pixel value ordering reversible data hiding. *Inf. Sci.* **2019**, *489*, 136–154. [CrossRef]
17. Wang, X.; Ding, J.; Pei, Q. A novel reversible image data hiding scheme based on pixel value ordering and dynamic pixel block partition. *Inf. Sci.* **2015**, *310*, 16–35. [CrossRef]
18. Di, F.; Zhang, M.; Liao, X.; Liu, J. High-fidelity reversible data hiding by quadtree-based pixel value ordering. *Multimed. Tools Appl.* **2018**, *78*, 7125–7141. [CrossRef]
19. Zhao, W.; Yang, B.; Gong, S. A higher efficient reversible data hiding scheme based on pixel value ordering. *J. Inf. Hiding Multimed. Signal Process.* **2018**, *9*, 918–928.
20. Lu, T.C.; Tseng, C.Y.; Huang, S.W.; Vo, T.N. Pixel-value-ordering based reversible information hiding scheme with self-adaptive threshold strategy. *Symmetry* **2018**, *10*, 764. [CrossRef]
21. Lee, C.F.; Shen, J.J.; Kao, Y.C. High-Capacity Reversible Data Hiding Based on Star-Shaped PVO Method. In Proceedings of the 14th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Sendai, Japan, 26–28 November 2018; Volume 109.
22. He, W.; Zhou, K.; Cai, J.; Wang, L.; Xiong, G. Reversible data hiding using multi-pass pixel value ordering and prediction-error expansion. *J. Vis. Commun. Image Represent.* **2017**, *49*, 351–360. [CrossRef]
23. Weng, S.; Chen, Y.; Ou, B.; Chang, C.; Zhang, C. Improved K-Pass Pixel Value Ordering Based Data Hiding. *IEEE Access* **2019**, *7*, 34570–34582. [CrossRef]
24. Ou, B.; Li, X.; Wang, J. High-fidelity reversible data hiding based on pixel-value-ordering and pairwise prediction-error expansion. *J. Vis. Commun. Image Represent.* **2016**, *39*, 12–23. [CrossRef]
25. He, W.; Xiong, G.; Weng, S.; Cai, J.; Wang, Y. Reversible data hiding using multi-pass pixel-value-ordering and pairwise prediction-error expansion. *Inf. Sci.* **2018**, *467*, 784–799. [CrossRef]
26. Kumar, R.; Jung, K.H. Enhanced pairwise IPVO-based reversible data hiding scheme using rhombus context. *Inf. Sci.* **2020**, *536*, 101–119. [CrossRef]
27. Kaur, G.; Singh, S.; Rani, R.; Kumar, R.; Malik, A. High-quality reversible data hiding scheme using sorting and enhanced pairwise PEE. *IET Image Process.* **2022**, *16*, 1096–1110. [CrossRef]
28. Kaur, G.; Singh, S.; Rani, R.; Kumar, R. A comprehensive study of reversible data hiding (RDH) schemes based on pixel value ordering (PVO). *Arch. Comput. Methods Eng.* **2021**, *28*, 3517–3568. [CrossRef]

International Journal of Production Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tprs20

Sales and operations planning for delivery date setting in engineer-to-order manufacturing: a research synthesis and framework

Swapnil Bhalla, Erlend Alfnes, Hans-Henrik Hvolby & Olumide Oluyisola

To cite this article: Swapnil Bhalla, Erlend Alfnes, Hans-Henrik Hvolby & Olumide Oluyisola (2022): Sales and operations planning for delivery date setting in engineer-to-order manufacturing: a research synthesis and framework, International Journal of Production Research, DOI: 10.1080/00207543.2022.2148010

To link to this article: https://doi.org/10.1080/00207543.2022.2148010

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 25 Nov 2022.

Submit your article to this journal

Article views: 57

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tprs20

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
https://doi.org/10.1080/00207543.2022.2148010

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Sales and operations planning for delivery date setting in engineer-to-order manufacturing: a research synthesis and framework

Swapnil Bhalla [a], Erlend Alfnes [a], Hans-Henrik Hvolby [b] and Olumide Oluyisola [a]

[a]Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway; [b]Department of Materials and Production, Aalborg University, Aalborg, Denmark

ABSTRACT

Sales and operations planning (S&OP) has emerged as a planning approach that integrates tactical level decisions across functions and supply chains while aligning day-to-day operations with long-term strategy through these decisions. The extant knowledge on S&OP has evolved primarily based on the needs of mass production contexts, and applications of S&OP in engineer-to-order (ETO) contexts have not been explored by previous research. Arguing that the cross-functionally coordinated planning enabled by S&OP can improve the effectiveness of the challenging and competitively critical tendering process, this paper develops an S&OP framework for the tactical planning process design to support delivery date setting in ETO contexts. The paper adopts a systematic literature review approach for identifying the main tactical planning activities managers in ETO companies should consider while designing the S&OP process and the information inputs required for performing and coordinating these planning activities. The identified planning activities and planning inputs are synthesised to develop the proposed S&OP framework for delivery date setting in ETO contexts. The proposed framework can support managers in assessing which tactical planning activities are strategically essential in their respective companies and redesigning or reconfiguring existing planning processes to address the planning needs of their environment.

ARTICLE HISTORY
Received 30 March 2022
Accepted 8 November 2022

KEYWORDS
Engineer-to-order; sales and operations planning; tactical planning; delivery date setting; customer enquiry management

1. Introduction

Sales and operations planning (S&OP) is an approach for tactical level planning that has received growing interest from academics and practitioners over the last three decades (Kreuter et al. 2022). S&OP emerged in the 1980s as an extension of aggregate production planning to address problems arising from planning and decision-making in functional silos (Danese, Molinaro, and Romano 2018; Stentoft, Freytag, and Mikkelsen 2020). S&OP emphasises integrating or coordinating the tactical planning activities and planning objectives across the various supply chain functions, e.g. procurement, production, sales, etc., for effectively balancing demand and supply at the tactical level while also aligning the day-to-day operations with long-term strategic plans and competitive priorities (Grimson and Pyke 2007; Pereira, Oliveira, and Carravilla 2020; Thomé et al. 2012b).

Since its conception, S&OP has been adopted in a variety of industrial contexts (Kristensen and Jonsson 2018), and various studies have reported the positive impacts of S&OP adoption on companies' performance (Feng, D'Amours, and Beauregard 2008; Oliva and Watson 2011; Thomé et al. 2012a; Thomé, Sousa, and do Carmo 2014). The adoption of S&OP in different industrial contexts has allowed researchers to observe how the design of the S&OP process is adapted across contexts to achieve the intended performance outcomes (Ivert et al. 2015; Kreuter et al. 2021; Kreuter et al. 2022; Kristensen and Jonsson 2018; Tuomikangas and Kaipia 2014). The principle that planning processes should be designed to fit the characteristics and requirements of specific industrial contexts has been widely emphasised in the planning and control literature (Berry and Hill 1992; Buer et al. 2018b; Jonsson and Mattsson 2003; Newman and Sridharan 1995) and is based on the assumptions of the wider-scoped contingency theory (Donaldson 2001; Ivert et al. 2015; Kristensen and Jonsson 2018; Lawrence and Lorsch 1969). Therefore, contextualising or adjusting the design of the S&OP process according to different industrial characteristics and requirements has been one of the main research streams within S&OP literature (Jonsson, Kaipia, and Barratt 2021; Kreuter et al. 2021; Kreuter et al. 2022).

The recent state-of-the-art reviews on S&OP by Kristensen and Jonsson (2018) and Kreuter et al. (2022) indicate that one of the major research gaps within the extant

CONTACT Swapnil Bhalla swapnil.bhalla@ntnu.no

© 2022 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.

S&OP literature is the contextualisation of S&OP design in production contexts operating with the engineer-to-order (ETO) strategy. ETO production contexts, i.e. companies operating with an ETO strategy, produce customised products based on individual customers' requirements, and the adoption of the strategy has been observed across industries supplying high-value, complex products such as industrial machinery, agricultural machinery, ships, ship equipment, etc. (Bertolini et al. 2022; Cannas and Gosling 2021; Gosling and Naim 2009; Zennaro et al. 2019). Due to customer-specific design, engineering, procurement, and production activities, tactical planning is characterised by high complexity and high uncertainty in ETO production contexts (Adrodegari et al. 2015; Alfnes et al. 2021; Løkkegaard et al. 2022; Sylla et al. 2021). The high complexity and uncertainty of tactical planning in ETO contexts amplify the need for coordinated planning across functions, which can be addressed by S&OP (Kreuter et al. 2022; Shurrab, Jonsson, and Johansson 2020b). However, despite these needs of ETO production contexts, the issue of contextualising S&OP design for ETO production has not been investigated in the extant research (Kreuter et al. 2022).

In a step towards addressing the lack of research on S&OP in ETO contexts, the authors' recent case study of an ETO maritime equipment manufacturer identifies customer enquiry management or delivery date setting as one of the main decision areas for S&OP in an ETO manufacturing context (Bhalla et al. 2021). Setting delivery dates in ETO contexts entails (1) estimating the delivery dates to be quoted while tendering for new customer orders or responding to customer enquiries; and (2) assessing the feasibility of delivery dates imposed by customers for potential orders (Carvalho, Oliveira, and Scavarda 2015; Zijm 2000; Zorzini, Stevenson, and Hendry 2012). Setting delivery dates in ETO contexts is a particularly challenging task, and its effectiveness is essential for ETO companies to maintain competitive delivery performance (Amaro, Hendry, and Kingsman 1999; Cannas et al. 2020; Hicks, McGovern, and Earl 2000). Therefore, this paper further explores the contextualisation of S&OP design in ETO contexts, focussing on the tactical planning task of delivery date setting, and addresses the following research question.

> RQ: How should engineer-to-order manufacturers contextualise the design of the sales and operations planning process for effective delivery date setting?

The paper addresses this research question by developing an S&OP reference framework for setting delivery dates while tendering for new customer orders in ETO manufacturing contexts. Delivery date setting is a topic of general relevance for various ETO contexts and has motivated many research contributions supporting practitioners in executing the task effectively. One of these contributions is highlighting that coordinated planning across supply chain functions can help ETO companies in managing the complexity and uncertainty characterising the task of setting delivery dates (Hicks, McGovern, and Earl 2000; Shurrab, Jonsson, and Johansson 2020a, 2020b; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012). However, the extant literature on delivery date setting lacks guidance on designing a coordinated planning process for setting delivery dates, and developing a process reference framework is one of the main research needs in this area (Bhalla, Alfnes, and Hvolby 2022). The proposed S&OP reference framework identifies the main planning activities of different supply chain functions for setting delivery dates in ETO contexts and the information flows required for coordinating these planning activities.

This paper uses a systematic literature review methodology for developing the S&OP reference framework and answering the research question presented above. The remainder of this paper is organised as follows. Section 2 provides an overview of existing frameworks from the research streams on S&OP and delivery date setting to elaborate on the research gaps motivating this study. Section 3 describes the literature review methodology adopted for developing the S&OP reference framework. Section 4 synthesises the relevant literature for developing the framework. Section 5 discusses potential applications of the proposed framework and identifies future research needs. Section 6 concludes the paper by summarising the paper's contributions.

2. Overview of the extant research

The growing interest in S&OP and the evolving knowledge on the topic have motivated various systematic reviews of the literature on the topic over the last decade, albeit with different focuses (Kreuter et al. 2022; Kristensen and Jonsson 2018; Noroozi and Wikner 2017; Pereira, Oliveira, and Carravilla 2020; Thomé et al. 2012a, 2012b; Tuomikangas and Kaipia 2014). These reviews provide overviews and syntheses of the extant research on S&OP from different perspectives (Jonsson, Kaipia, and Barratt 2021). Among these, the reviews considering the effect of the production strategy on S&OP design, i.e. Kreuter et al. (2022) and Kristensen and Jonsson (2018), find that the extant S&OP research has been contextualised in make-to-stock (MTS) and make-to-order (MTO) production contexts. Due to the differences in the contextual characteristics and planning needs of ETO, MTO, and MTS production, the design requirements for the S&OP process are different across these contexts

Table 1. Tactical S&OP attributes required in different production contexts.

Tactical S&OP attribute	MTS production	MTO production	ETO production
Planning strategy	Level [10]	Chase [10]	Chase [10]
Aggregation or planning object	Product families or individual products [6, 7, 9, 10, 12]		Individual products [5, 8, 11]
Demand-input for planning	Forecasts [6, 7, 11, 12]		Tenders, confirmed orders or projects [1, 2, 11]
Main planning outputs	Production volumes [10], inventory targets, promotion timing, price changes [12]	Production volumes, sales targets, inventory & backorder targets [6, 7]	Delivery dates for tenders, production plans for confirmed orders or projects [3, 4, 5, 8]

[1] Adrodegari et al. (2015); [2] Alfnes and Hvolby (2019); [3] Alfieri, Tolio, and Urgo (2011); [4] Alfieri, Tolio, and Urgo (2012); [5] Carvalho, Oliveira, and Scavarda (2015); [6] Feng, D'Amours, and Beauregard (2008); [7] Gansterer (2015); [8] Ghiyasinasab et al. (2021); [9] Grimson and Pyke (2007); [10] Olhager (2013); [11] Olhager, Rudberg, and Wikner (2001); [12] Pereira, Oliveira, and Carravilla (2020).

(Bhalla et al. 2021; Buer et al. 2018b; Kreuter et al. 2022; Kristensen and Jonsson 2018; Olhager, Rudberg, and Wikner 2001). For instance, S&OP has primarily been considered a forecast-driven planning process in MTS and MTO contexts, while tactical planning in ETO contexts is primarily driven by tenders or customer enquiries and confirmed orders or projects (Adrodegari et al. 2015; Alfieri, Tolio, and Urgo 2011; Feng, D'Amours, and Beauregard 2008; Gansterer 2015; Ghiyasinasab et al. 2021; Hans et al. 2007; Olhager, Rudberg, and Wikner 2001). Table 1 highlights the main differences between the required attributes for tactical S&OP in the different types of production contexts.

Within the topic of S&OP and the broader area of planning and control, conceptual and reference frameworks are valuable artefacts with utility for applications in research as well as practice, e.g. for unifying fragmented knowledge on conceptually related topics (Kreuter et al. 2022; Pereira, Oliveira, and Carravilla 2020; Tuomikangas and Kaipia 2014), for identifying and establishing industry-wide best practices (Adrodegari et al. 2015), for investigating and explaining the impact of context on process design and performance (Kristensen and Jonsson 2018; Thomé et al. 2012b; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012), for assessing process maturity (Grimson and Pyke 2007), for guiding improvements of process maturity (Danese, Molinaro, and Romano 2018), for mapping, analysing, and designing or redesigning contextually fitting managerial and planning processes (Adrodegari et al. 2015; Shurrab, Jonsson, and Johansson 2020b), etc. Perhaps the most widely cited framework for S&OP is the five-step process model, which lists the main steps to be implemented in companies' S&OP process, namely product portfolio review, forecasting and demand planning, supply planning, pre-S&OP meeting and executive S&OP meeting (Grimson and Pyke 2007; Jacobs et al. 2011; Kristensen and Jonsson 2018; Thomé et al. 2012b; Wallace 2004; Wallace and Stahl 2008). The main activities within different versions of this five-step process model focus on forecast-driven tactical planning. Wing and Perry (2001); Lapide (2005); Grimson and Pyke (2007); Wagner, Ullrich, and Transchel (2014); Goh and Eldridge (2015); Pedroso et al. (2017); Vereecke et al. (2018); and Danese, Molinaro, and Romano (2018) propose S&OP maturity models for assessing the maturity of companies' S&OP processes, and for identifying measures for improving S&OP maturity. Despite the abundance of research on S&OP maturity models, none of the listed studies investigates the applicability of their proposed maturity models in ETO contexts. Other higher-level frameworks, such as the literature synthesis framework developed by Thomé et al. (2012b); the coordination framework proposed by Tuomikangas and Kaipia (2014); and the contingency framework developed by Kristensen and Jonsson (2018), are sufficiently generalisable with dimensions such as organisation, meetings and collaboration, tools and technologies, etc. and these frameworks can provide different theoretical perspectives for studying how S&OP process design in ETO contexts differs from other production environments. More recently, Pereira, Oliveira, and Carravilla (2020) propose a tactical S&OP framework based on the literature on S&OP, aggregate production planning, tactical planning, etc., highlighting the main information flows, decisions and constraints for the tactical S&OP process. Their proposed framework is based on the underlying assumption of a supply chain for standard or non-customised products, rendering various information flows and decisions within the framework irrelevant for S&OP and delivery date setting in ETO manufacturing.

Due to the challenges and complexity of managing ETO operations, planning and control have been among the main research areas within ETO operations and supply chain management literature (Cannas and Gosling 2021; Gosling and Naim 2009; Zennaro et al. 2019). Consequently, numerous planning frameworks have also been proposed in this literature for ETO contexts with different theoretical perspectives underlying these frameworks. In one of the first contributions to planning and control in ETO firms, Bertrand and Muntslag (1993) propose a production control framework to address the

lack of fit between the functionality of manufacturing resource planning (MRP II) systems and the requirements of ETO contexts. Little et al. (2000) propose a planning and scheduling reference model for ETO companies based on a similar premise. Both Bertrand and Muntslag (1993) and Little et al. (2000) take a planning system perspective in developing their frameworks. Nam et al. (2018) take a similar perspective, proposing a supply chain planning matrix for designing an advanced planning and scheduling system for ETO shipbuilding. Adrodegari et al. (2015) take a business process perspective and propose a process reference framework for the software requirements of ETO machinery building companies, which consists of activities across the order-fulfilment process. These high-level frameworks are broadly scoped across the order-fulfilment process in ETO contexts, and while they provide insights on a few planning activities relevant for setting delivery dates and S&OP, they lack focus on the information flows relevant for these activities.

Unlike the high-level, broadly scoped frameworks mentioned above, extant literature also provides frameworks that focus specifically on tactical planning and the task of delivery date setting (Kingsman et al. 1996; Shurrab, Jonsson, and Johansson 2020b; Zorzini, Corti, and Pozzetti 2008a). However, these frameworks also lack some necessary elements. For instance, the frameworks proposed by Kingsman et al. (1996) and Zorzini, Corti, and Pozzetti (2008a) focus on fabrication and assembly capacity planning for delivery date setting but do not address engineering capacity planning and procurement planning. As a result, these frameworks lack an integrated or cross-functional perspective that is essential for S&OP in ETO contexts to ensure that lead times for all order-fulfilment activities are considered and to ensure that tactical plans and delivery dates are based on shared information and functional expertise rather than conflicting assumptions (Grabenstetter and Usher 2014; Hicks, McGovern, and Earl 2000; Shurrab, Jonsson, and Johansson 2020a; Zorzini, Stevenson, and Hendry 2012). The framework proposed by Shurrab, Jonsson, and Johansson (2020b), despite its cross-functional perspective, does not address the information flows supporting the tactical planning decisions outlined in their framework.

The overview of literature presented above suggests that an S&OP framework to support delivery date setting in ETO contexts is a knowledge gap in the extant research. We observe that due to a lack of consideration of ETO contexts in the extant S&OP research, existing S&OP frameworks do not address the unique planning needs of these contexts (Kreuter et al. 2022; Shurrab, Jonsson, and Johansson 2020b). Furthermore, tactical planning frameworks developed for ETO contexts are either broadly scoped across the entire order-fulfilment process and lack a focus on setting delivery dates, or lack a cross-functional planning perspective required in S&OP for delivery date setting (Bhalla, Alfnes, and Hvolby 2022). Based on the knowledge gap outlined above, there is a compelling need for developing a reference framework to map, analyse, and design the S&OP process for effectively setting delivery dates in ETO contexts.

3. Methodology

This paper adopts a systematic literature review (SLR) approach to answer this study's main research question and develop a reference framework for contextualising S&OP design for effective delivery date setting. The extant research supporting delivery date setting in ETO contexts is fragmented, and although many studies have addressed different elements of tactical planning in these contexts, an overarching framework is a persisting research gap (Bhalla, Alfnes, and Hvolby 2022). This study aims to identify the planning activities and information flows required in S&OP for setting delivery dates by analysing and synthesising this fragmented body of knowledge. The SLR approach is particularly suitable for integrating and synthesising knowledge from past research to inform industrial practice due to the approach's emphasis on transparency of the literature review process (Thomé, Scavarda, and Scavarda 2016; Tranfield, Denyer, and Smart 2003; Watson and Webster 2020; Webster and Watson 2002). The following subsections describe the review methodology adopted in this paper for developing the S&OP reference framework. The methodology is divided into the three steps typical for SLRs in operations management (Bhalla, Alfnes, and Hvolby 2022; Cannas and Gosling 2021; Kristensen and Jonsson 2018; Thomé, Scavarda, and Scavarda 2016; Tranfield, Denyer, and Smart 2003) – problem formulation (3.1), literature identification and selection (3.2), and analysis and synthesis (3.3).

3.1. Problem formulation

As introduced in sections 1 and 2, this paper is motivated by the empirical observation from previous research that setting delivery dates is a challenging and competitively critical task for ETO manufacturers (Bhalla et al. 2021; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012), and the lack of frameworks suitable for contextualising the design of the S&OP process for effective delivery date setting in ETO contexts (Bhalla, Alfnes, and Hvolby 2022; Kreuter et al. 2022). The research problem has been translated into the research question for this

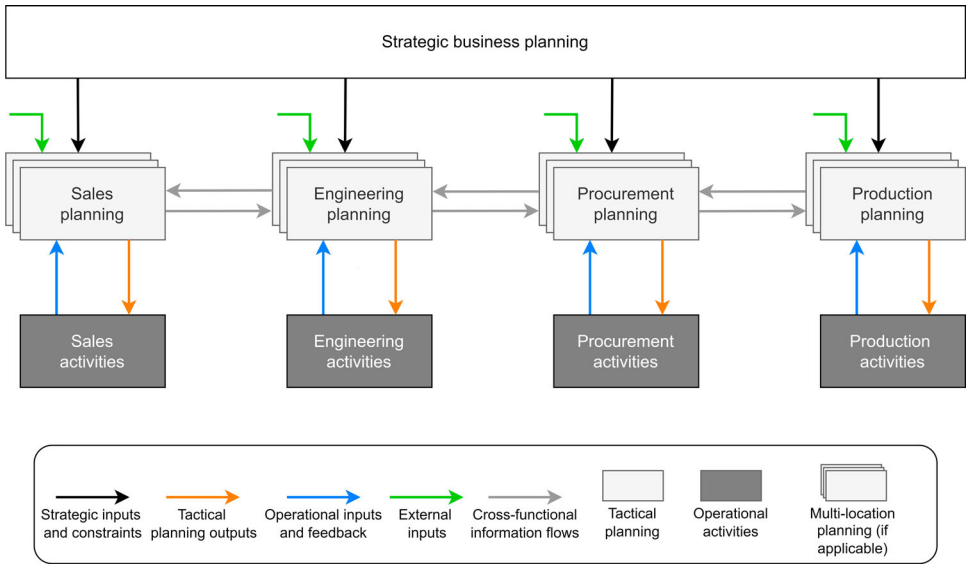


Figure 1. Overall framework for S&OP in ETO contexts (adapted from Pereira, Oliveira, and Carravilla (2020) and Nam et al. (2018)).

paper, as stated in section 1. This problem is addressed in this paper by identifying the planning activities and information flows that ETO companies should consider while designing their S&OP process for effective delivery date setting. The findings are synthesised to propose a reference framework for S&OP in ETO contexts.

The identification and review of literature for identifying the relevant activities and information flows were structured by defining an overall S&OP framework for ETO contexts, as illustrated in Figure 1. This overall framework was defined by adapting the framework from Pereira, Oliveira, and Carravilla (2020), making two main modifications to the original framework. Firstly, the supply chain function of *engineering* was introduced, excluding the *distribution* function, and spatially configuring the supply chain functions of *sales*, *engineering*, *procurement*, and *production* based on the ETO literature (Dekkers 2006; Nam et al. 2018). Secondly, the additional information flow of *operational inputs and feedback* was introduced to account for the uncertain and frequently changing planning environment of ETO contexts, which necessitates considering the current states of operational resources in tactical planning activities (Alfieri, Tolio, and Urgo 2011; Ghiyasinasab et al. 2021). Following Figure 1, the main tactical planning activities or outputs for setting delivery dates were identified under the S&OP subprocesses of *sales planning*, *engineering planning*, *procurement planning*, and *production planning*. As also illustrated in Figure 1, the *information flows* required for these planning activities were identified under the categories of *strategic inputs and constraints*, i.e. outputs of long-term strategic decisions that act as constraints for operations; *external inputs*, i.e. information obtained from actors outside the enterprise, such as suppliers and customers; *operational inputs*, i.e. information about the status and performance of execution and control activities; and *cross-functional information flows*, i.e. information obtained by one planning subprocess or function from another.

3.2. Literature identification and selection

The literature for this review was identified through keyword searches on the Scopus and Web of Science databases and through forward and backward citation searches based on the database search results. The keyword string for searching the two databases was formulated with two blocks. The first block consisted of a wide range of keywords that have been associated in the extant literature with the primary concepts for this study – S&OP and delivery date setting. These included any concepts closely related to, or used synonymously with, either of the primary concepts, e.g. tactical planning, aggregate planning, tactical capacity planning, lead time estimation, etc. The second block of keywords consisted of terms that may be used for referring to ETO contexts, i.e. variations of 'engineer-to-order' and alternative terms, e.g. project manufacturing, customised manufacturing, etc. These included variations of 'make-to-order' for two reasons: (1) some authors use MTO as an umbrella term for collectively referring to all non-MTS contexts, including ETO contexts (Aslan, Stevenson, and Hendry 2012; Kingsman et al. 1996); (2) studies from MTO contexts can also provide insights on planning activities and information flows in ETO contexts, especially for the production function (Adrodegari et al. 2015;

Table 2. Blocks of keywords in the keyword string for literature search.

Block type	Keywords
Concept-related keyword block	'sales and operations planning'; 'sales & operations planning'; 'tactical planning'; 'aggregate planning'; 'capacity planning'; 'production planning'; 'resource planning'; 'lead time estimation'; 'leadtime estimation'; 'delivery date setting'; 'due date setting'; 'customer enquiry management'; 'order acceptance'; 'bid preparation'; 'tendering'; 'delivery date assignment'; 'due date assignment'; 'procurement planning'; 'purchasing planning'; 'sales planning'; 'engineering planning'
Context-related keyword block	'engineer to order'; 'engineer-to-order'; 'engineered to order'; 'engineered-to-order'; 'project manufacturing'; 'project-manufacturing'; 'project production'; 'project-production'; 'project-based production'; 'project based production'; 'project-based manufacturing'; 'project based manufacturing'; 'customized production'; 'customised production'; 'customized manufacturing'; 'customised manufacturing'; 'make to order'; 'made to order'; 'make-to-order'; 'made-to-order'

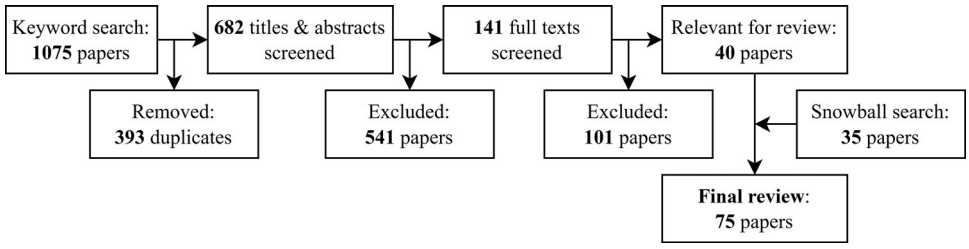


Figure 2. PRISMA flowchart for the process of identifying and selecting literature.

Bhalla, Alfnes, and Hvolby 2022; Sylla et al. 2018). Table 2 shows the keywords in the two blocks of the search string. All keywords in a block were connected by the *or* Boolean operator, while the two blocks were connected by the *and* operator.

The search strings for the databases were formulated to identify papers where the specified keywords appeared in the title, abstract, or author-specified keywords. Searching for publications up to and including May 2022, the searches returned 644 and 431 results on Scopus and Web of Science, respectively, including journal articles, conference papers, and book sections. The citation information (e.g. author(s) of the document, document title, publication year, source type, etc.) for the results from both databases were exported into an EndNote library using RIS (Research Information Systems) format files generated from the databases. After the removal of duplicate results, 682 unique documents remained. The citation information for these documents was exported from the EndNote library into an Excel spreadsheet for record-keeping and documentation of content analysis.

For initial screening, the titles and abstracts of the 682 papers were reviewed to exclude irrelevant papers before the next steps of the review. This screening step led to the exclusion of 541 papers based on one or more of the following criteria: (1) not written in English; (2) not from a peer-reviewed source; (3) research not contextualised or positioned in ETO or MTO production; and (4) research focus on operational or shop-floor level delivery date setting and order-acceptance integrated with detailed scheduling decisions. The next step was to screen the full texts for the remaining 141 papers to assess which of these should be included in the final review. In this full-text assessment, 40 of the 141 papers were found relevant for inclusion in the final review based on two main considerations: (1) the full-text document for the paper is published, and available online; and (2) the paper provides insight into one or more planning activities, or information flows related to delivery date setting, tendering, customer enquiry management, request-for-proposal management, etc.

The final step in identifying relevant literature were the forward and backward citation searches, also known as the snowball search (Thomé, Scavarda, and Scavarda 2016). The reference lists of the 40 papers were screened to identify potentially relevant older papers, and the 'cited by' feature of Google Scholar was used to identify any relevant citations of these papers. Based on this, we identified 35 additional papers that fit the two inclusion criteria presented above, resulting in 75 papers for the final review. Figure 2 illustrates the literature identification process in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart adapted from Buer, Strandhagen, and Chan (2018a); Moher et al. (2009).

3.3. Literature analysis and synthesis

The content of the 75 papers was analysed by coding relevant text from the literature to the main codes or themes outlined in the overall S&OP framework illustrated in Figure 1, i.e. the planning subprocesses and categories of information inputs. First, the relevant quotes from

the papers were coded to one or more of the four planning subprocesses outlined in the framework, i.e. sales-, engineering-, procurement-, and production planning, where each quote identified one or more planning activities for the respective subprocess. Second, quotes identifying one or more planning inputs for a planning activity were coded to these activities and to one of the categories of planning inputs from Figure 1, i.e. strategic constraints, external inputs, operational inputs, or cross-functional inputs.

Among the analysed papers, some of the content contributes explicitly towards answering this study's research question, while others implicitly. Explicit contributions include direct references to specific planning activities, decisions, and planning inputs while describing or discussing case studies, decision-support tools, decision-making methodologies and authors' general observations in the industry. Implicit contributions include quotes that do not mention specific planning activities or planning inputs or mention these without linking them to delivery date setting but allow for logically inferring these. The identified planning activities and inputs were used to populate the overall framework (Figure 1) for creating the final S&OP reference framework.

4. Results

The extant literature provides various insights for contextualising the S&OP design in ETO manufacturing for effective delivery date setting. Based on the content analysis of the reviewed literature, this section identifies the main planning activities and information flows that should be considered in ETO contexts while designing the S&OP process for setting delivery dates. The distribution of the analysed literature across journals and years of publication can be found in Table A1 and Figure A1 in the Appendix.

Table A2 in the Appendix summarises the contributions of the reviewed papers in identifying the planning activities and information inputs for the planning activities for contextualising S&OP design in ETO contexts. As mentioned in subsection 3.3, these activities and information flows are presented in the reviewed literature explicitly or implicitly. Therefore, Table A2 also classifies the contributions of the papers as explicit (E), implicit (I), or partly explicit and partly implicit (E/I). These contributions are further elaborated in the remainder of this section, where each subsection describes the main planning activities and corresponding planning inputs for the four S&OP subprocesses – sales planning (4.1), engineering planning (4.2), procurement planning (4.3), and production planning (4.4).

4.1. Sales planning

Three main tactical planning activities emerge from the reviewed literature for the sales planning subprocess of S&OP for tendering and setting delivery dates in ETO contexts, namely (1) determining which customer enquiries should be pursued, where *customer enquiries* collectively refer to enquiries, tender invitations, sales leads, and requests-for-proposal (RFP) (Aslan, Stevenson, and Hendry 2015; Hans et al. 2007; Hicks, McGovern, and Earl 2000; Shurrab, Jonsson, and Johansson 2020b; Zorzini, Stevenson, and Hendry 2012), (2) setting relative priority levels for customer enquiries (Adrodegari et al. 2015; Ebadian et al. 2009; Shurrab, Jonsson, and Johansson 2020b), and (3) coordinating the preparation of proposals or quotations that are sent to potential customers in response to the enquiries (Carvalho, Oliveira, and Scavarda 2015; Shurrab, Jonsson, and Johansson 2020a). These planning activities and their inputs are described in the following subsections: 4.1.1. Selecting customer enquiries, 4.1.2. Prioritising customer enquiries, and 4.1.3. Responding to customer enquiries. Figure 3 gives an overview of the planning inputs for sales planning, categorising these as strategic inputs, external inputs, operational inputs, and cross-functional inputs using the colour-coding scheme from Figure 1.

4.1.1. Selecting customer enquiries

The reviewed literature emphasises that selecting customer enquiries that a company would pursue is a crucial demand planning decision in ETO contexts, where managers must assess whether it is lucrative to use resources for preparing a proposal (Kingsman et al. 1996; Zorzini et al. 2008b). ETO manufacturing companies typically produce customised products within particular product domains (Adrodegari et al. 2015) with varying degrees of customisation (Alfnes et al. 2021; Cannas et al. 2020). Therefore, a preliminary review of customers' technical and commercial requirements must be conducted to assess if a competitive proposal can be made and its likelihood of success, considering the level of alignment between the company's competitive priorities and the typical order-winning criteria for customers' respective market segments (Adrodegari et al. 2015; Hicks, McGovern, and Earl 2000; Kingsman et al. 1996). This can enable the company's management to determine which orders strategically fit within the context of the company's operations strategy (Amaro, Hendry, and Kingsman 1999; Shurrab, Jonsson, and Johansson 2020b). We identify the main planning inputs for selecting customer enquiries in ETO contexts as the following.

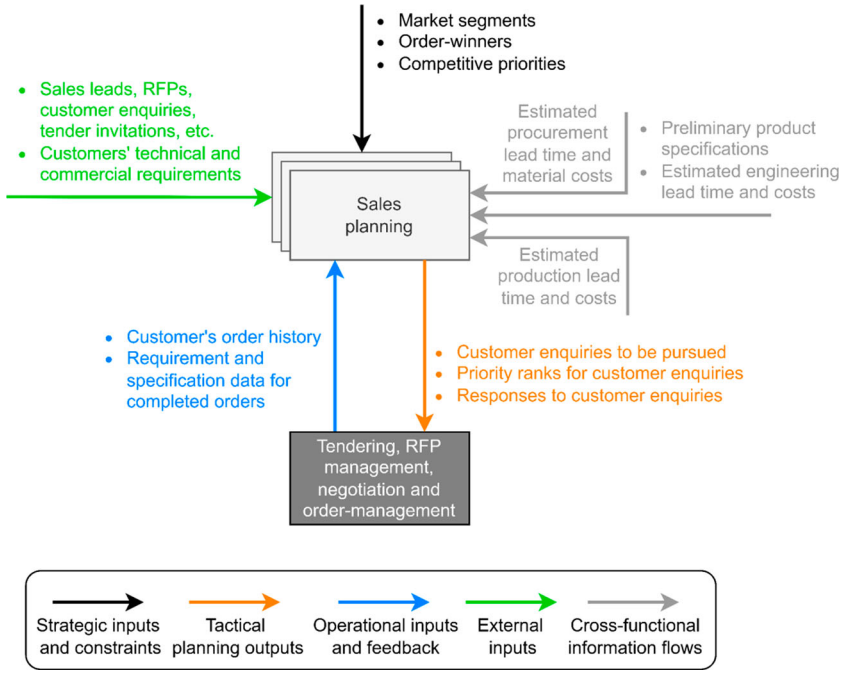


Figure 3. Tactical sales planning – information inputs and planning outputs.

- External inputs – the customer enquiry and customer requirements (Adrodegari et al. 2015; Kingsman et al. 1996; Shurrab, Jonsson, and Johansson 2020b).
- Strategic inputs – the market segmentation strategy, e.g. product-based segmentation, geography-based segmentation, etc.; order-winners, e.g. price, delivery lead time, product features, etc., and the company's competitive priorities (Adrodegari et al. 2015; Amaro, Hendry, and Kingsman 1999; Cannas et al. 2020; Kingsman et al. 1996).

4.1.2. Prioritising customer enquiries

Many ETO companies manage multiple customer enquiries simultaneously, which often compete for the same capacity-constrained managerial resources responsible for coordinating and preparing responses to these enquiries (Adrodegari et al. 2015; Alfnes et al. 2021; Shurrab, Jonsson, and Johansson 2020b). For such instances, ranking the enquiries according to their relative priority level and their level of strategic importance can enable strategic resource allocation to manage these enquiries, where the strategic importance of enquiries may be influenced by factors related to market segmentation, customers' order history, customers' requirements, similarity to previous orders, customer-imposed delivery dates, etc. (Adrodegari et al. 2015; Ebadian et al. 2008; Ebadian et al. 2009; Hans et al. 2007; Kingsman et al. 1996). We identify the following main inputs from the reviewed literature for prioritising customer enquiries.

- External inputs – same as for selecting customer enquiries (4.1.1).
- Strategic inputs – same as for selecting customer enquiries (4.1.1).
- Operational inputs – customer's order history (for assessing strategic relevance of customer) and requirements and specifications for delivered orders (to assess similarity to previous orders) (Adrodegari et al. 2015).

4.1.3. Responding to customer enquiries

The importance of offering competitive product technology, delivery lead times, and prices is widely recognised in the ETO literature (Bertrand and Muntslag 1993; Carvalho, Oliveira, and Scavarda 2015; Cassaigne et al. 1997; Ghiyasinasab et al. 2021; Grabenstetter and Usher 2014; Hans et al. 2007; Zennaro et al. 2019; Zorzini, Corti, and Pozzetti 2008a). Therefore, one of the main sales planning activities in ETO contexts is coordinating the company's response to customer enquiries and providing potential customers with the high-level technical and commercial characteristics of the product, production, and delivery (Adrodegari et al. 2015). The basic technical characteristics are typically based on preliminary engineering (Adrodegari et al. 2015; Sylla et al. 2018; Ulonska and Welo 2016), and the delivery lead time and price can be estimated based on lead time and cost estimates for the main order-fulfilment activities of engineering, procurement, and production functions (Bhalla, Alfnes, and Hvolby 2022; Zorzini et al. 2008b; Zorzini, Stevenson,

and Hendry 2012). Based on the reviewed literature, we identify the following main planning inputs for the sales planning function to respond to customer enquiries in ETO contexts.

- Strategic inputs – market segmentation, order-winning criteria (Amaro, Hendry, and Kingsman 1999; Calosso et al. 2003; Cassaigne et al. 1997; Kingsman and Mercer 1997; Kingsman et al. 1993) for assessing the importance of competitive pricing and lead times for different market segments and customers, such that targeted profit margins and slack for delivery lead time can be decided.
- Cross-functional inputs – preliminary product specifications (Adrodegari et al. 2015; Kingsman et al. 1996), detailed design and engineering activities required for order-specific customisation (Adrodegari et al. 2015), estimated lead time and cost for engineering activities (Ghiyasinasab et al. 2021; Grabenstetter and Usher 2013, 2014), estimated lead time and cost for material and component procurement (Hicks, McGovern, and Earl 2000; Zorzini, Stevenson, and Hendry 2012), estimated lead time and cost for production activities including potential overtime and subcontracting costs (Alfieri, Tolio, and Urgo 2011; Carvalho, Oliveira, and Scavarda 2015).

4.2. Engineering planning

Design and engineering activities are critical sources of competitive advantage for many ETO companies (Amaro, Hendry, and Kingsman 1999). Order-specific product customisation is among the main value-adding activities for many ETO companies (Grabenstetter and Usher 2014), which begins with translating the customer requirements into preliminary product specifications in the tendering phase to win customer orders (Adrodegari et al. 2015). ETO companies must ensure that the correct engineering resources are available at the right time for effectively executing engineering activities after order confirmation (Alfnes et al. 2021; Ghiyasinasab et al. 2021). Based on these needs, four main tactical planning activities for the engineering planning function emerge from the reviewed literature, namely (1) defining the preliminary design, features, and technical characteristics of the product (Bertrand and Muntslag 1993; Nam et al. 2018), (2) determining detailed design and engineering activities and relevant resources required for these activities (Adrodegari et al. 2015; Alfnes et al. 2021), (3) estimating the lead times, feasible due dates, and costs for design and engineering activities (Ghiyasinasab et al. 2021; Grabenstetter and Usher 2013, 2014), and (4) identifying if additional engineering capacity or capabilities are required for a customer order (Alfnes et al. 2021; Brachmann and Kolisch 2021; Gosling, Hewlett, and Naim 2017; Shurrab, Jonsson, and Johansson 2020b). These planning activities and their inputs are described in the following subsections: 4.2.1. Defining preliminary product specifications, 4.2.2. Determining detailed engineering activities and resources, 4.2.3. Estimating lead times and costs and setting due dates, and 4.2.4. Identifying needs for external capabilities and additional capacity. Figure 4 summarises the strategic-, external-, operational-, and cross-functional inputs for engineering planning.

4.2.1. Defining preliminary product specifications

For ETO companies competing on the innovativeness and customisability of their products, effectively defining preliminary product specifications in the tendering phase is crucial for winning orders (Amaro, Hendry, and Kingsman 1999; Cannas et al. 2020). Defining preliminary specifications entails understanding customer requirements and translating them into high-level design, features, and technical characteristics of the product (Adrodegari et al. 2015; Alfnes et al. 2021). The extant research reports that the clarity and preciseness of customer requirements in ETO contexts tend to vary across customers based on their technical and functional knowledge of the product (Cannas et al. 2020; Shurrab, Jonsson, and Johansson 2020b). Consequently, close interaction with potential customers can be beneficial in the tendering phase for ETO companies to clarify requirements (Zorzini et al. 2008b). Moreover, customer feedback on preliminary specifications may also be required before order confirmation for products requiring high degrees of newness and innovation (Adrodegari et al. 2015; Alfnes et al. 2021).

New and innovative product technologies are often developed in ETO contexts as part of order-specific engineering activities (Alfnes et al. 2021; Gosling and Naim 2009; Shurrab, Jonsson, and Johansson 2020a). However, ETO companies may also engage in new product development (NPD) initiatives through strategic, order-independent innovations (Cannas et al. 2019; Fang and Wei 2020; Hicks, McGovern, and Earl 2000), e.g. to expand product capabilities to support digital technology applications such as internet-of-things (IoT), real-time monitoring and control for efficient performance, predictive fault detection and maintenance of critical components, etc. (Oluyisola, Sgarbossa, and Strandhagen 2020; Strandhagen et al. 2020; Zheng et al. 2021). Such innovative product features and capabilities may also be offered to customers as part of the preliminary product specifications. On the other hand, customer

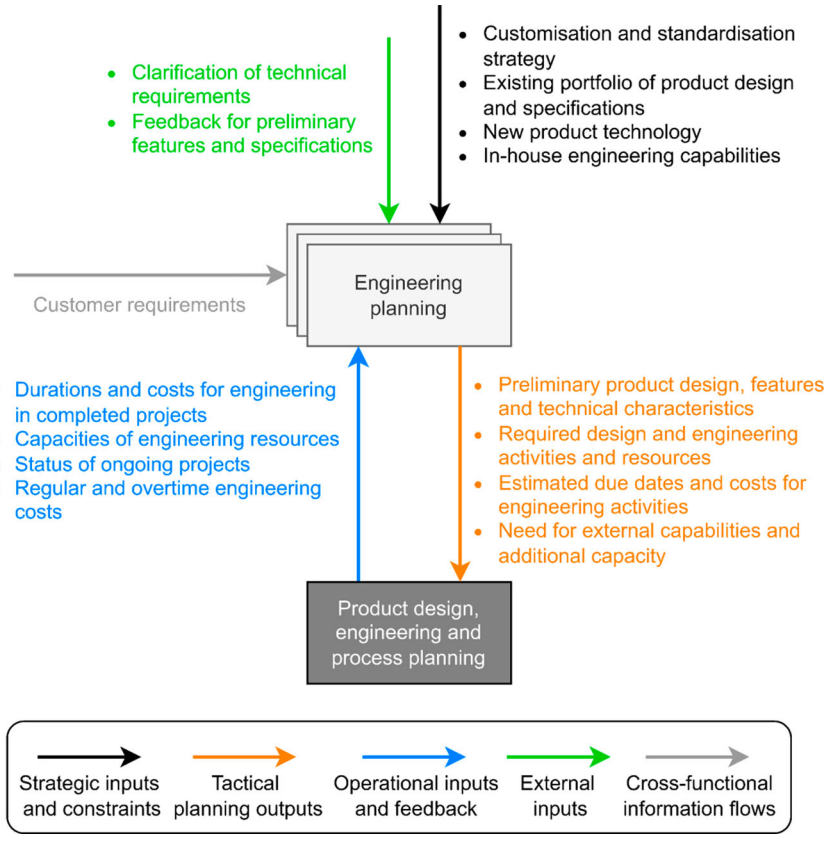


Figure 4. Tactical engineering planning – information inputs and planning outputs.

requirements demanding minimal newness often allow for the reuse of existing design solutions and specifications, which can reduce the time required for defining the preliminary product specifications and increase the reliability of the design offered to the customer (Adrodegari et al. 2015; Grabenstetter and Usher 2013, 2014; Sylla et al. 2018; Ulonska and Welo 2016; Willner, Gosling, and Schönsleben 2016a). As observed in the extant research, the level of order-specific design and customisation required in ETO contexts can vary significantly across market segments and customer orders (Alfnes et al. 2021; Cannas et al. 2020).

Based on the findings from the reviewed literature summarised above, we identify the following main inputs for defining preliminary product specifications.

- Cross-functional input – customer requirements (Adrodegari et al. 2015; Nam et al. 2018) obtained from the sales function.
- Strategic inputs – company's overall strategy for product customisation and standardisation (Fang and Wei 2020; Gosling and Naim 2009; Semini et al. 2014; Willner, Gosling, and Schönsleben 2016a), existing portfolio of product designs and specifications (Adrodegari et al. 2015; Grabenstetter and Usher 2013, 2014; Sylla et al. 2018; Ulonska and Welo 2016; Willner, Gosling, and Schönsleben 2016a), and new product technology (Cannas et al. 2019; Oluyisola, Sgarbossa, and Strandhagen 2020; Strandhagen et al. 2020; Zheng et al. 2021).
- External inputs – clarification of customer requirements and feedback on preliminary product specifications (Alfnes et al. 2021; Shurrab, Jonsson, and Johansson 2020b; Zorzini et al. 2008b).

4.2.2. Determining detailed engineering activities and resources

The order-specific engineering in ETO contexts is a complex transactional and iterative process with substantial uncertainty regarding the specific engineering activities to be undertaken before the detailed product specifications are finalised (Alfnes et al. 2021; Grabenstetter and Usher 2013, 2014). Managers in ETO companies must nevertheless estimate the scope and complexity of these activities and identify the relevant resources for these activities to enable resource and capacity planning for the engineering and design department(s) (Ghiyasinasab et al. 2021; Zijm 2000; Zorzini et al. 2008b). Determining detailed engineering activities and resources entails (1) identifying the design and engineering activities to

be performed after order confirmation (Adrodegari et al. 2015) and (2) identifying the design and engineering capabilities required in different disciplines, e.g. mechanical, hydraulic, electrical, etc., and the workload for these activities in terms of, e.g. personnel, person-hours, etc. (Alfnes et al. 2021).

The preliminary product specifications (4.2.1) defined in the tendering phase can be seen as the primary input for determining the detailed engineering activities, since these activities essentially map the course from the preliminary specifications to the final product specifications. In addition, we identify the following planning inputs from the reviewed literature for determining the detailed engineering activities and resources.

- Cross-functional input – customer requirements (Grabenstetter and Usher 2013, 2014) obtained from the sales function.
- Strategic inputs – the company's overall strategy for product customisation and standardisation (Cannas et al. 2020; Dekkers 2006; Johnsen and Hvam 2019) that constrains the extent of order-specific customisation, e.g. all elements of the product may be customisable, or specific modules of the product may be customisable, etc.; and existing portfolio of product designs and specifications that can be reused to fulfil customer requirements (Grabenstetter and Usher 2013, 2014) and reduce the required order-specific engineering activities.

4.2.3. Estimating lead times and costs and setting due dates

As highlighted in subsection 4.1.3, estimated engineering lead times and costs are essential planning inputs for the sales planning function to estimate the overall delivery lead time and delivery date that should be quoted to customers while responding to enquiries. These engineering lead times are also one of the main sources of planning complexity (Grabenstetter and Usher 2014) and uncertainty (Alfnes et al. 2021) in ETO contexts, and while there is an abundance of planning tools and decision-support models for estimating production lead times and costs, there are few contributions in the literature that propose tools for estimating engineering lead times (Bhalla, Alfnes, and Hvolby 2022). Among the handful of contributions in this area, there are two broad categories of approaches proposed for estimating engineering lead times: (1) estimating engineering lead times solely based on the complexity of engineering activities under an infinite capacity assumption (Cannas et al. 2018; Grabenstetter and Usher 2013, 2014), and (2) estimating engineering lead times based on a tactical capacity planning or tactical resource-loading approach under a finite capacity assumption (Brachmann and Kolisch 2021; Ghiyasinasab et al. 2021). While the infinite-capacity approach focuses only on estimating the lead times for engineering activities and relies on historical cost data for estimating engineering costs, the finite-capacity planning approach can integrate the estimation of engineering lead times and costs.

Based on the reviewed literature, we identify two main sets of planning inputs for estimating engineering lead times and costs in the tendering phase in ETO contexts. First, the planning output of the previous engineering planning activity (4.2.2), i.e. the required detailed engineering activities and resources. Second, the operational inputs required for computing the lead time and cost estimates using infinite- or finite-capacity approaches, as listed below.

- Historical data on duration and costs for engineering activities in completed projects (Grabenstetter and Usher 2013, 2014).
- Capacity of engineering personnel and status of ongoing projects (Brachmann and Kolisch 2021; Ghiyasinasab et al. 2021) for estimating lead times based on finite loading.
- Costs of regular and overtime engineering capacity (Brachmann and Kolisch 2021; Ghiyasinasab et al. 2021) for estimating costs of engineering activities.

4.2.4. Identifying needs for external capabilities and additional capacity

Design and engineering capabilities are a vital source of competitive advantage for many ETO manufacturers, especially in contexts where customers value the innovativeness and customisability of products (Alfnes et al. 2021; Amaro, Hendry, and Kingsman 1999; Cannas et al. 2020). For such ETO companies, an important consideration in responding to customer enquiries is the availability of required capabilities in different engineering disciplines to perform the detailed engineering activities necessary to fulfil the customer requirements (Alfnes et al. 2021; Aslan, Stevenson, and Hendry 2015; Gosling, Hewlett, and Naim 2017). Moreover, ETO contexts requiring frequent innovations in product technology necessitate a continual reassessment of in-house capabilities and expertise in engineering to maintain the competitive advantage in product features and innovativeness (Cannas et al. 2020; Gosling, Hewlett, and Naim 2017). Based on these factors, in addition to the detailed engineering activities and resources (i.e. planning output from 4.2.2), we identify the in-house engineering capabilities as the main strategic input for identifying the needs for external engineering capabilities.

Determining the need for additional engineering capacity is an extension of the planning activity of estimating engineering lead times and costs (4.2.3). ETO companies are multi-project contexts where the same capacity-constrained resources execute multiple projects simultaneously (Adrodegari et al. 2015; Barbosa and Azevedo 2019; Hans et al. 2007; Shurrab, Jonsson, and Johansson 2020b). As a result, the regular capacity of in-house engineering resources may not be sufficient for meeting customer-imposed engineering due dates because of these resources being allocated to other ongoing projects, which can necessitate the use of non-regular capacity alternatives such as overtime (Brachmann and Kolisch 2021; Ghiyasinasab et al. 2021). Identifying such non-regular capacity needs as early as in the tendering phase can enable managers in better capacity planning for the engineering function.

4.3. Procurement planning

ETO companies use combinations of standard and non-standard or customer-specific components and modules for producing the final product (Johnsen and Hvam 2019; Zennaro et al. 2019; Zorzini et al. 2008b). Fabrication and assembly activities for various components and modules are outsourced by ETO companies to varying extents, ranging from highly vertically integrated manufacturers that only procure raw materials and basic components to highly vertically disintegrated companies with entirely outsourced production activities (Alfnes et al. 2021; Hicks, McGovern, and Earl 2000; Hicks, McGovern, and Earl 2001; Zorzini, Stevenson, and Hendry 2012). Nevertheless, almost all ETO companies depend on their suppliers for some stages of the order-fulfilment process, which underlines the importance of procurement planning before order confirmation (Hicks, McGovern, and Earl 2000). The primary role of procurement planning in the tendering phase is the early identification of components and sub-assemblies to be procured if an order is confirmed and identifying relevant suppliers and supplier-related constraints for the order-fulfilment process (Dekkers, Chang, and Kreutzfeldt 2013; Hicks, McGovern, and Earl 2000; Shishank and Dekkers 2013). From the reviewed literature, we identify three main tactical planning activities for the procurement planning function in S&OP, namely (1) identifying critical items for a potential customer order, (2) identifying potential suppliers for the critical items, and (3) determining lead time and cost-related constraints for critical items. These planning activities and their planning inputs are described in the following subsections: 4.3.1. Identifying critical items, 4.3.2. Selecting potential suppliers, and 4.3.3. Determining procurement lead times and prices. Figure 5 summarises the strategic-, external-, operational-, and cross-functional inputs for procurement planning.

4.3.1. Identifying critical items

Products produced with an ETO strategy are typically complex, large-sized, and characterised by deep and wide product structures (Zennaro et al. 2019) that consist of components and subsystems with a wide range of characteristics, e.g. some are used in low volumes while others are used in medium to large quantities, some are highly customised while others are standardised, some are technologically advanced while others are not, etc. (Hicks and Braiden 2000; Hicks, McGovern, and Earl 2000). The typically long delivery lead times in ETO contexts allow for externally sourced items to be procured during the order-fulfilment process, i.e. after order confirmation. Nevertheless, factors such as long supplier lead times, the geographical distance of suppliers, few or no alternate suppliers, low flexibility of suppliers, customisation, etc., render some items critical for timely order-fulfilment and delivery precision in ETO contexts (Emblemsvåg 2014; Mwesiumo, Nujen, and Kvadsheim 2021; Shlopak, Rød, and Oterhals 2016; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012). Some of these factors may also be correlated, e.g. higher levels of customisation are usually associated with higher costs and longer, more uncertain lead times (Alfnes et al. 2021; Hicks, McGovern, and Earl 2000). Identifying such critical components already in the tendering phase can enable managers and planners (1) to consider the supply-related constraints for these items while estimating and quoting delivery dates and prices to customers and (2) to closely monitor the procurement of these items after order confirmation (Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012).

From the reviewed literature, the following main planning inputs for identifying supply-critical items emerge.

- Strategic input – overall outsourcing and offshoring strategy (Hicks, McGovern, and Earl 2000; Sabri, Micheli, and Cagno 2020; Zorzini, Stevenson, and Hendry 2012) that constrains which components and subsystems are sourced from suppliers, and the geographical preferences vis-à-vis suppliers, i.e. localisation versus globalisation of supply.
- Cross-functional inputs – preliminary product specifications (4.2.1) and detailed engineering activities (4.2.2) (Alfnes et al. 2021; Emblemsvåg 2014; Hicks, McGovern, and Earl 2000; Zorzini, Stevenson, and Hendry 2012) that identify subsystems and components that will require order-specific customisation.
- Operational inputs – historical procurement lead times (Hicks, McGovern, and Earl 2000; Zorzini et al.

Figure 5. Tactical procurement planning – information inputs and planning outputs.

2008b; Zorzini, Stevenson, and Hendry 2012) that allow for identifying items with potentially long and variable procurement lead time items.

4.3.2. Selecting potential suppliers

The complex structure of ETO products and a large number of sourced components and sub-assemblies used for assembling these products necessitate effective supplier coordination in ETO contexts (Hicks, McGovern, and Earl 2000; Zorzini, Stevenson, and Hendry 2012). Moreover, the suppliers from whom raw materials, components, and sub-assemblies are procured may change in ETO contexts from one customer order or project to another (Alfnes et al. 2021; Mwesiumo, Nujen, and Kvadsheim 2021). For items that are considered critical from a supply planning perspective (4.3.1), potential suppliers must already be identified in the tendering phase such that realistic lead time and price constraints, which are essential planning inputs for estimating and quoting delivery dates and price (4.1.3), can be identified by contacting the potential suppliers (Hicks, McGovern, and Earl 2000; Mello et al. 2017; Zorzini et al. 2008b). Using internally estimated procurement lead times based on historical data and managerial assumptions can expose ETO companies to a significant risk of delays and cost overruns due to the uncertainty of supplier lead times (Alfnes et al. 2021; Hicks, McGovern, and Earl 2000; Shurrab, Jonsson, and Johansson 2020b).

In addition to the identified critical items (4.3.1), we identify the following planning inputs for identifying the potential suppliers based on the reviewed literature.

- Strategic inputs – supplier or vendor list (Mwesiumo, Nujen, and Kvadsheim 2021; Reid, Bamford, and Ismail 2019; Sabri, Micheli, and Cagno 2020) that identifies the approved suppliers for various materials, components, and subsystems; and any strategic suppliers with whom the company has long-term alliances or partnerships for specific items, e.g. due to a supplier's technological expertise, unique product features, etc. (Hicks, McGovern, and Earl 2000; Hicks, McGovern, and Earl 2001; Mello et al. 2017; Mwesiumo, Nujen, and Kvadsheim 2021; Saghiri and Hill 2014).
- External input – customer's preferred supplier(s) for specific items or customer requirements that can exclusively be fulfilled by particular suppliers (Hicks, McGovern, and Earl 2000), which constrain the choice of suppliers.

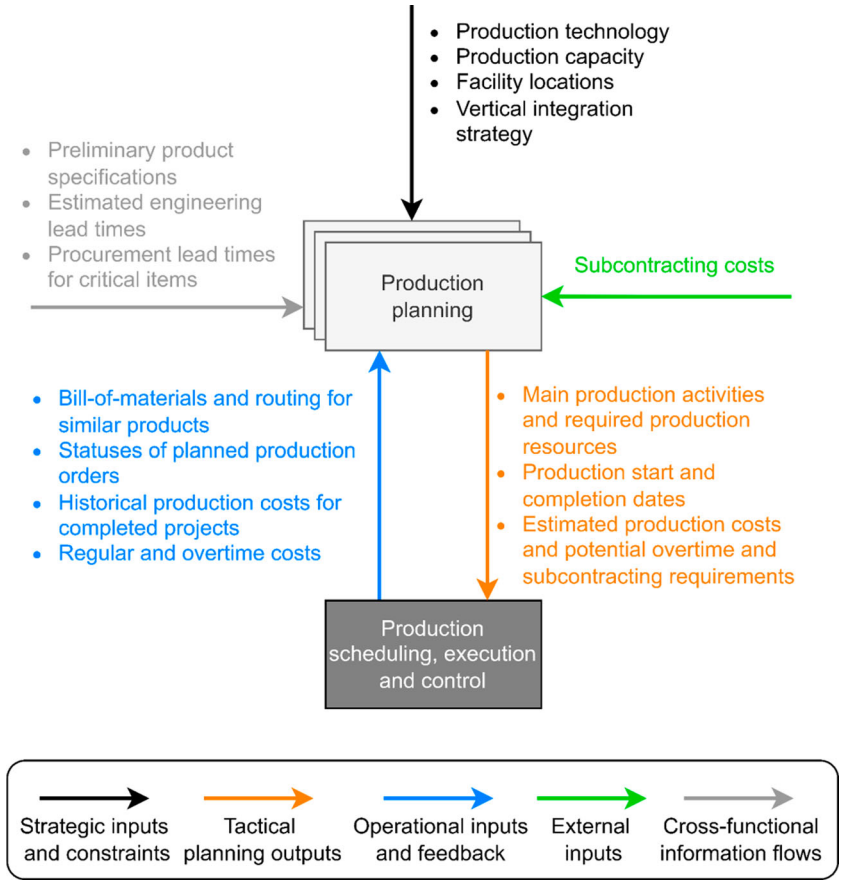


Figure 6. Tactical production planning – information inputs and planning outputs.

4.3.3. Determining procurement lead times and prices

In ETO contexts, procurement lead times and the costs of procured items are usually significant elements of the overall delivery lead time and the overall cost of the product, respectively (Adrodegari et al. 2015; Alfnes et al. 2021; Gourdon and Steidl 2019; Zorzini, Stevenson, and Hendry 2012). Consequently, procurement lead times and costs are indispensable inputs for estimating delivery dates and prices quoted to customers to ensure that products can be delivered to customers within the promised delivery lead times and profitably (Hicks, McGovern, and Earl 2000; Mello et al. 2017). Three main approaches emerge from the extant literature for identifying procurement lead times and costs in the tendering phase emerge, namely (1) estimation based on historical data, (2) identifying lead times and prices from long-term supplier agreements, and (3) identifying lead times and prices by active coordination with suppliers (Calosso et al. 2003; Hicks, McGovern, and Earl 2000; Zorzini et al. 2008b). Based on the reviewed literature, we identify the following planning inputs for determining the procurement lead times and prices.

- Strategic inputs – long-term supplier agreements for lead times and prices (Hicks, McGovern, and Earl 2000; Olhager 2010; Shurrab, Jonsson, and Johansson 2020b), if any.
- Cross-functional inputs – preliminary product specifications (4.2.1) and required detailed engineering activities (4.2.2) for communicating the expected characteristics of customised items to potential suppliers, and the estimated engineering lead times (4.2.3) for communicating the anticipated timeline for availability of the detailed specifications (Dekkers, Chang, and Kreutzfeldt 2013; Hicks, McGovern, and Earl 2000; Shishank and Dekkers 2013; Zorzini, Stevenson, and Hendry 2012).
- External inputs – lead times and prices quoted by suppliers (Alfnes et al. 2021; Calosso et al. 2003; Hicks, McGovern, and Earl 2000; Shishank and Dekkers 2013; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012).
- Operational inputs – historical data for suppliers' lead times and prices (Hicks, McGovern, and Earl 2000; Shishank and Dekkers 2013; Shlopak, Rød, and Oterhals 2016; Zorzini, Stevenson, and Hendry 2012).

4.4. Production planning

Products that are typically engineered and produced for specific customer orders are usually high-value and heavy-duty electromechanical systems consisting of various subsystems (Cannas and Gosling 2021; Gosling and Naim 2009; Zennaro et al. 2019). Consequently, facilities that manufacture these products require diverse equipment and manual expertise for fabricating the components and assembling the subsystems that comprise these products. Furthermore, since ETO manufacturing contexts are multi-project environments, many customer orders for complex products are simultaneously processed by the specialised, capacity-constrained production resources in these contexts (Adrodegari et al. 2015; Alfnes et al. 2021; Hans et al. 2007). Therefore, capacity constraints for production resources must be considered by managers in ETO contexts while quoting delivery dates and prices for new customer orders to ensure that the customer order can be produced within the promised duration without incurring unanticipated costs due to capacity shortfalls (Alfieri, Tolio, and Urgo 2011; Carvalho, Oliveira, and Scavarda 2015; Shurrab, Jonsson, and Johansson 2020b; Wullink et al. 2004). Based on this need, we identify three main tactical planning tasks for the production planning function within S&OP in the tendering phase, namely (1) identifying the main production activities and resource requirements for a potential customer order, (2) identifying the feasible start and finish dates for production activities or stages, and (3) estimating production costs and non-regular capacity (overtime and subcontracting) requirements. These planning activities and their planning inputs are described in the following subsections: 4.4.1. Identifying the main production activities and resource requirements, 4.4.2. Identifying feasible production start and end dates, and 4.4.3. Estimating production costs and non-regular capacity requirements. Figure 6 summarises the strategic-, external-, operational-, and cross-functional inputs for these production planning activities.

4.4.1. Identifying the main production activities and resource requirements

The tendering phase for customer orders in ETO contexts is characterised by substantial uncertainty regarding the product and process specifications since the detailed product engineering and process planning activities are performed after order confirmation (Adrodegari et al. 2015; Alfieri, Tolio, and Urgo 2012; Alfnes et al. 2021; Carvalho, Oliveira, and Scavarda 2015, 2016; Hans et al. 2007; Reid, Bamford, and Ismail 2019; Shurrab, Jonsson, and Johansson 2020b; Wullink et al. 2004). Nevertheless, managers and planners must plan and tentatively allocate resources and capacity for potential customer orders in the tendering phase to ensure the availability of these resources later, to estimate feasible production due dates, and for timely execution of order-fulfilment activities. To enable this planning or capacity allocation, it is an essential planning activity to identify the main production activities for a potential customer order, e.g. cutting, stamping, machining, welding, assembly, testing, packaging, etc., and the resource requirements for performing them, i.e. personnel and equipment (Adrodegari et al. 2015; Carvalho, Oliveira, and Scavarda 2015; De Boer, Schutten, and Zijm 1997; Reid, Bamford, and Ismail 2019) albeit with high-level, aggregated production stages, workloads, resources, and time-buckets (Adrodegari et al. 2015; Aslan, Stevenson, and Hendry 2012; Zorzini, Corti, and Pozzetti 2008a). The level of aggregation and scope for identifying these resource requirements can vary based on contextual factors such as product complexity, degree of customisation, resource flexibility, etc. (Zorzini, Corti, and Pozzetti 2008a; Zorzini et al. 2008b). For instance, production contexts with fixed bottlenecks may focus on capacity requirements for bottleneck resources, while contexts with varying bottlenecks must consider a broader set of resources and corresponding capacities (Alfnes and Hvolby 2019; Park et al. 1999; Ruben and Mahmoodi 2000; Zorzini, Corti, and Pozzetti 2008a). Similarly, factors such as resource capabilities, capacity flexibility, target resource utilisation, etc., may influence the level of aggregation of resources, capacity, and time-buckets (Ebben, Hans, and Weghuis 2005; Robinson and Moses 2006; Zorzini, Corti, and Pozzetti 2008a). Based on the reviewed literature, we identify the following main planning inputs for identifying the main production activities and resource requirements in the tendering phase.

- Strategic input – current manufacturing process technology, which governs if the fabrication activities will be performed using the same techniques as previous customer orders (Adrodegari et al. 2015; Alfieri, Tolio, and Urgo 2011, 2012) or if new process technology alternatives, e.g. additive manufacturing, have been implemented (Eyers et al. 2021).
- Cross-functional input – preliminary product specifications (4.2.1) (Adrodegari et al. 2015; Alfieri, Tolio, and Urgo 2012; Alfnes and Hvolby 2019; De Boer, Schutten, and Zijm 1997; Nam et al. 2018), based on which the required macro-level production processes can be identified.
- Operational input – existing bill-of-materials (BOM) and production routing from previous orders for similar products (Carvalho, Oliveira, and Scavarda 2015,

2016; De Boer, Schutten, and Zijm 1997; Zorzini et al. 2008b) for estimating the workload or resource requirements for various production resources.

4.4.2. Identifying feasible production start and end dates

Similar to the lead times for engineering and procurement, production lead times are an essential input for the sales planning function to estimate the overall delivery lead time (4.1.3) for customer orders in ETO production contexts (Alfieri, Tolio, and Urgo 2011; Carvalho, Oliveira, and Scavarda 2015; Ghiyasinasab et al. 2021). Estimating the lead time for production essentially entails determining feasible production start and end dates while considering finite-capacity constraints and the availability of resources, materials, and product and process specifications (Ebadian et al. 2008; Hicks and Braiden 2000; Wikner and Rudberg 2005; Zorzini, Stevenson, and Hendry 2012). Among the tactical planning activities for S&OP in ETO contexts that are identified in this review, estimation of production lead times is perhaps the planning activity on which the majority of the extant research has focussed, especially the development of planning and decision-support tools for this activity (Bhalla, Alfnes, and Hvolby 2022). In addition to the identified production activities and resource requirements (4.4.1), the following planning inputs for identifying feasible start and end dates for production activities emerge from the reviewed literature.

- Strategic inputs – vertical integration strategy (Hicks, McGovern, and Earl 2000; Hicks, McGovern, and Earl 2001) that constrains which production activities will be performed in-house, production capacity for in-house production resources (Alfieri, Tolio, and Urgo 2011, 2012; Barbosa and Azevedo 2019; Ebben, Hans, and Weghuis 2005; Micale et al. 2021; Park et al. 1999; Ruben and Mahmoodi 2000; Thürer et al. 2012; Wullink et al. 2004; Zorzini, Corti, and Pozzetti 2008a), and the facility locations for companies with multiple production sites (Yang and Fung 2014) for considering the inter-facility transportation times for components or sub-assemblies.
- Cross-functional inputs – estimated engineering lead times (Ghiyasinasab et al. 2021; Grabenstetter and Usher 2014; Wikner and Rudberg 2005; Zorzini et al. 2008b) that govern the availability of the detailed product and process specifications, and estimated procurement lead times (Alfnes et al. 2021; Hicks, McGovern, and Earl 2000; Zorzini, Stevenson, and Hendry 2012) that govern the availability of raw materials and components.
- Operational inputs – status of planned production orders (Barbosa and Azevedo 2019; Carvalho, Oliveira, and Scavarda 2015; Hans et al. 2007; Thürer et al. 2012) for estimating queueing delays, i.e. durations that production orders must wait for required resources to become available; and existing BOMs and production routing from previous orders for similar products (Adam et al. 1993; Burggraf et al. 2021; Thürer et al. 2012) for estimating the processing times and staging delays, i.e. times when components and subassemblies are waiting for other components to be ready for assembly since complex structures of ETO products contain multiple levels of assemblies (Adrodegari et al. 2015; Hicks and Braiden 2000; Zennaro et al. 2019).

4.4.3. Estimating production costs and non-regular capacity requirements

Estimated production costs are an essential input for the sales planning function to determine the overall product price that must be quoted to customers (4.1.3) in the tendering phase. Based on the reviewed literature, two broad approaches for estimating production costs can be identified. First, these production costs can be estimated based on archived historical data on production costs from previous customer orders for similar products (Adrodegari et al. 2015; Kingsman et al. 1996). Second, these production costs can be estimated based on finite-capacity allocation approaches as an extension of estimating the production lead times (Carvalho, Oliveira, and Scavarda 2015; Ghiyasinasab et al. 2021). Finite-capacity approaches entail explicit consideration of the capacity of production resources, as well as any potential non-regular capacity alternatives, e.g. overtime or subcontracting, that might be required for expediting production to meet customer-imposed delivery dates or for quoting short and competitive delivery dates (Amaro, Hendry, and Kingsman 1999; Carvalho, Oliveira, and Scavarda 2015; Ghiyasinasab et al. 2021; Wullink et al. 2004; Zorzini, Corti, and Pozzetti 2008a). Therefore, when using finite-capacity approaches for estimating production lead times and costs, managers must also identify any non-regular capacity requirements associated with the estimated lead times. Based on the reviewed literature, we identify the following planning inputs for estimating production costs based on historical production costs or estimating these costs and non-regular capacity requirements as an extension of estimating production start and end dates (4.4.2).

- Strategic inputs – vertical integration strategy (Hicks, McGovern, and Earl 2000; Hicks, McGovern, and Earl

2001) that constrains which production activities can be outsourced or subcontracted.
- External inputs – subcontracting costs from potential subcontractors (Carvalho, Oliveira, and Scavarda 2015, 2016; Ebadian et al. 2008; Ghiyasinasab et al. 2021; Wullink et al. 2004) for estimating the potential costs of subcontracted production activities.
- Operational inputs – historical production costs for completed projects (Adrodegari et al. 2015; Kingsman et al. 1996), and costs for regular and overtime capacity of production resources (Carvalho, Oliveira, and Scavarda 2015, 2016; Ghiyasinasab et al. 2021).

4.5. S&OP reference framework for ETO contexts

The tactical planning activities and planning inputs identified in the preceding subsections can be synthesised into an S&OP reference framework for delivery date setting in ETO contexts, as illustrated in Figure 7. The planning activities and planning inputs shown in Figure 7 are identified based on the references summarised in Table A2, and are illustrated separately for sales planning (4.1), engineering planning (4.2), procurement planning (4.3), and production planning (4.4) in Figures 3–6, respectively. The S&OP reference framework shown in Figure 7 synthesises the findings from the literature review for the individual S&OP functional subprocesses into a holistic framework for S&OP in ETO contexts, which is constructed by populating Figure 1 with the relevant planning activities and planning inputs, adapting the presentation methodology from Pereira, Oliveira, and Carravilla (2020).

5. Discussion of the framework and future research needs

Based on a systematic review of the extant literature, this paper has identified the main tactical planning activities that ETO companies should consider in contextualising the design of the S&OP process for effectively setting delivery dates while tendering for new customer orders, and the flow of planning information required for performing and coordinating these activities. The findings have been synthesised into an S&OP reference framework for ETO contexts. The remainder of this section discusses the proposed framework's potential applications or usage areas and future research needs for better supporting practitioners in designing and conducting S&OP in ETO contexts.

5.1. Applications of the proposed framework

The high planning complexity characterising the task of setting delivery dates in ETO contexts has been repeatedly emphasised in the extant literature (Shurrab, Jonsson, and Johansson 2020b; Zorzini, Stevenson, and Hendry 2012). For managing this planning complexity, previous research has underlined the need for cross-functionally coordinated tactical planning in ETO contexts, such that the relevant planning factors are considered while tendering, and the in-house tacit knowledge and expertise of managers are utilised for effective planning (Shurrab, Jonsson, and Johansson 2020a, 2020b; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012). Despite the emphasis on coordinated planning for effective delivery date setting in previous research, there are no planning frameworks in the extant literature to support ETO practitioners in designing cross-functionally coordinated tactical planning processes, neither in the research stream on delivery date setting (Bhalla, Alfnes, and Hvolby 2022) nor in the stream on tactical S&OP (Kreuter et al. 2022; Kristensen and Jonsson 2018). Therefore, the proposed framework can serve as a *common reference framework* for S&OP and delivery date setting in ETO contexts. Managers in ETO companies could use the framework as a reference for assessing (1) which planning activities are critical or essential for their planning environment, (2) whether their existing tactical planning process addresses those activities, and (3) if there are necessary mechanisms for making the planning inputs available for those activities. Furthermore, given the fragmented nature of the extant research on delivery date setting in ETO contexts (Bhalla, Alfnes, and Hvolby 2022), the use of the proposed framework as a common reference among researchers can help position and scope the future research contributions supporting delivery date setting, similar to the application of their framework demonstrated by Pereira, Oliveira, and Carravilla (2020).

Despite the framework's potential to serve as an S&OP reference for delivery date setting in ETO contexts, some contextual contingencies and limitations must be considered in such a generalisation of the framework. Firstly, the framework is developed with the manufacturing industry as the primary target context. However, the ETO strategy is also adopted in non-manufacturing contexts, such as the construction industry (Cannas and Gosling 2021). Due to the underlying assumptions specific to manufacturing, some framework elements may not be applicable in non-manufacturing ETO contexts. Secondly, the strategic relevance of the elements of the framework is expected to vary across ETO manufacturing contexts based on the characteristics of their planning environments and corresponding planning needs (Buer et al. 2018b; Kristensen and Jonsson 2018; Zorzini, Corti, and Pozzetti 2008a; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012). Factors such as the degree of product

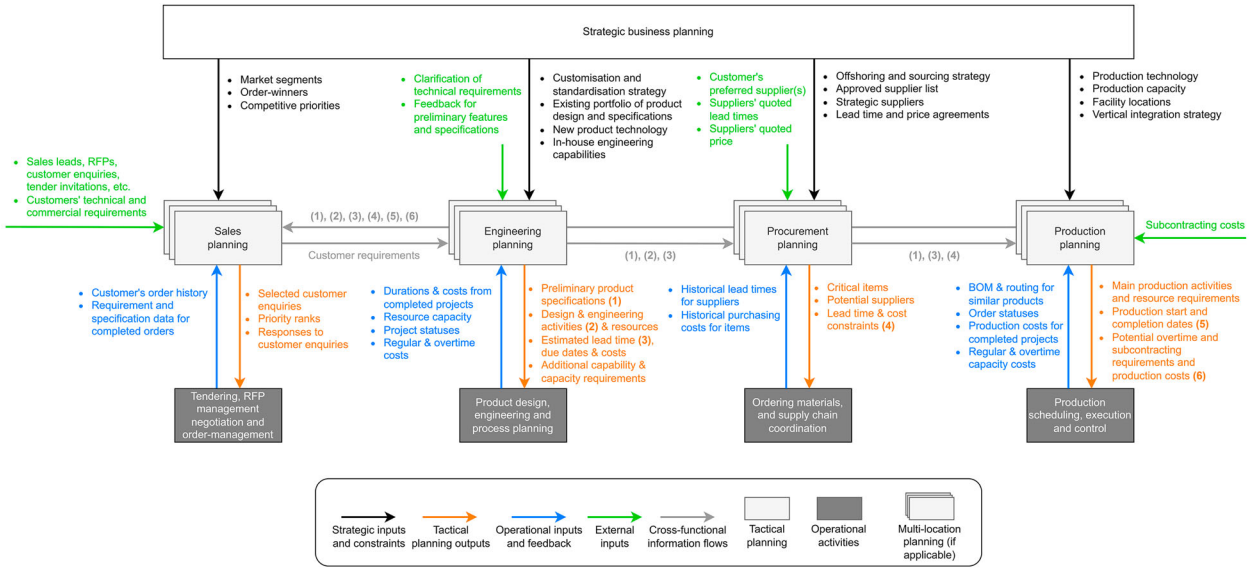


Figure 7. Proposed S&OP reference framework for setting delivery dates in ETO contexts.

customisation, product complexity, order volumes, etc., may amplify the need for cross-functional coordination in tactical planning, while flexibility in capacities of production and engineering functions and the level of technical knowledge of the product and production system across functions may reduce the need for this coordination (Bhalla, Alfnes, and Hvolby 2022; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012). Companies with high levels of vertical integration may require more emphasis on planning activities for production and engineering functions, while companies with low levels of vertical integration may focus more on planning activities for the procurement function (Hicks, McGovern, and Earl 2000; Hicks, McGovern, and Earl 2001; Zorzini, Stevenson, and Hendry 2012). Furthermore, the context of companies' business- and competitive strategy might influence how specific planning activities and decisions are handled. For instance, companies focusing on expanding their market share might prioritise enquiries from new customers, while other companies might prioritise enquiries from existing customers to sustain existing long-term relationships with customers (Ebadian et al. 2008; Ebadian et al. 2009). In some companies, pre-defined product platforms or templates may be used for expedited or automated specification of preliminary product characteristics in the tendering phase (Fang and Wei 2020; Ulonska and Welo 2016). Due to the variety in characteristics of ETO manufacturing companies found in practice (Alfnes et al. 2021; Amaro, Hendry, and Kingsman 1999; Cannas and Gosling 2021; Gosling and Naim 2009; Hicks, McGovern, and Earl 2001; Willner et al. 2016b), and as exemplified above, the contingency between the contextual factors, the planning needs, and the planning process design are essential to consider while applying or generalising the framework. Moreover, the order-fulfilment process in ETO contexts is influenced by requirements related to compliance with design codes, standards, and product certification practices, which are typically industry-specific and are therefore excluded from the framework. For high-value, complex, and technologically advanced ETO products such as power generation equipment, manufacturing machinery, offshore oil and gas production platforms, etc., the complexity of design and engineering activities is typically managed by using design codes and standards established by different professional societies and national or international standard organisations (Gosling, Hewlett, and Naim 2017; Shapiro 1997) such as ISO (International Organisation for Standardisation), IEC (International Electrotechnical Commission), CEN (European Committee for Standardisation), CENELEC (European Committee for Electrotechnical Standardisation), ANSI (American National Standards Institute), etc. Consequently, many of these ETO products undergo post-production inspection or testing procedures to demonstrate and certify their adherence to the relevant standards, where the relevance of different standards may also depend on the geographical context. Such industry-specific certification procedures and their influence on the order-fulfilment process must be considered while designing the S&OP process in specific contexts. For instance, design and production of ships and ship equipment are governed by the rules and standards established by classification societies such as DNV (Det Norske Veritas), Lloyd's register, etc., and the class certification procedures of these societies impose precedence

constraints on shipbuilding projects and ship equipment manufacturing (Alfnes et al. 2021; Emblemsvåg 2014). Therefore, industry-specific regulatory considerations and their impact on S&OP must be incorporated into the framework based on the particular ETO industry or context of application.

The variations in the planning environment characteristics across ETO contexts also suggest another potential usage of the framework for comparative case studies of delivery date setting practices across companies within and across industry sectors. In the extant literature, the handful of multi-case studies on delivery date setting practices are exploratory studies from machinery building companies from a limited set of geographical contexts (Zorzini, Corti, and Pozzetti 2008a; Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012). Despite the valuable contributions of these studies, case research establishing the current status of delivery date setting practices in other industrial and geographical contexts is a gap in the extant research (Bhalla, Alfnes, and Hvolby 2022). The proposed framework can support future case studies on delivery date setting as a *tool for mapping* the tactical planning process for setting delivery dates in ETO companies, identifying variations in companies' focus on planning activities and cross-functional information sharing and the contextual factors influencing these variations. The contingency frameworks for delivery date setting proposed by previous studies of Zorzini et al. (2008b) and Zorzini, Stevenson, and Hendry (2012) identify four high-level design variables for the delivery date setting process, namely delivery date monitoring support, delivery date setting responsibility, coordination, and formalisation. The planning activities and inputs identified in our proposed S&OP framework (Figure 7) provide a more granular set of variables for the design and assessment of the delivery date setting process.

The proposed framework can also be utilised as a basis for *developing maturity models* for delivery date setting practices in ETO contexts. As suggested by numerous contributions within S&OP research and within operations management in general, maturity models are valuable tools for mapping and assessing the current state of business processes and industry practices and for planning strategic process improvements in the studied contexts (Danese, Molinaro, and Romano 2018; Goh and Eldridge 2015; Grimson and Pyke 2007; Pedroso et al. 2017; Vereecke et al. 2018; Wagire et al. 2021; Willner, Gosling, and Schönsleben 2016a). Despite the strategic and competitive importance of the planning task of setting delivery dates (Zorzini et al. 2008b; Zorzini, Stevenson, and Hendry 2012), there are no existing maturity models for delivery date setting practices in ETO contexts (Bhalla, Alfnes, and Hvolby 2022). Furthermore, the existing maturity models for the S&OP process have been developed based on research contextualised in MTS and MTO production contexts (Danese, Molinaro, and Romano 2018; Goh and Eldridge 2015; Grimson and Pyke 2007; Pedroso et al. 2017; Vereecke et al. 2018), and do not consider the unique planning needs for S&OP in ETO production contexts that are highlighted by the framework developed in this paper. The proposed framework can support future research aimed at addressing this gap.

5.2. Research gaps and future research needs

The systematic review of literature conducted to answer this paper's main research question also highlighted several research gaps in the reviewed literature. Based on these research gaps, this subsection suggests research needs that future research should address for supporting practitioners in designing and conducting the S&OP process for effective delivery date setting in ETO contexts.

The first research gap identified in the review concerns the sales planning activities of selecting and prioritising customer enquiries. Over the last three decades, multiple authors and their studies in diverse ETO contexts have highlighted the importance of selecting and prioritising customer enquiries for tendering based on strategic factors (Adrodegari et al. 2015; Hans et al. 2007; Hicks, McGovern, and Earl 2000; Kingsman et al. 1996; Shurrab, Jonsson, and Johansson 2020b; Zorzini et al. 2008b). However, the majority of references for these activities in the extant literature only provide high-level descriptions of these activities without exhaustive accounts of the factors that are or should be considered for these activities. While some of the factors to be considered for selecting and prioritising customer enquiries have been identified or inferred in this paper based on a few references from ETO contexts (Adrodegari et al. 2015; Amaro, Hendry, and Kingsman 1999; Cannas et al. 2020; Hans et al. 2007; Hicks, McGovern, and Earl 2000; Kingsman et al. 1996; Shurrab, Jonsson, and Johansson 2020b; Zorzini et al. 2008b) and MTO contexts (Ebadian et al. 2008; Ebadian et al. 2009), future studies should investigate whether any additional factors should be considered for these activities. Furthermore, the references for these activities in the extant ETO literature do not provide any formal decision-making methodologies for these planning activities. Future research should explore if, similar to MTO contexts (Ebadian et al. 2008; Ebadian et al. 2009), decision models for selecting and prioritising customer enquiries can be developed to support managers in ETO contexts.

The second research gap in the literature review relates to lead time estimation and capacity planning

for engineering activities. Among the few quantitative contributions in this area, Grabenstetter and Usher (2013, 2014) use an infinite-capacity approach for estimating engineering lead times by estimating engineering complexity based on historical data and the characteristics of a new engineering project; while Brachmann and Kolisch (2021); Ghiyasinasab et al. (2021) use finite-capacity planning-based approaches for estimating the engineering lead times. While the regression-based infinite-capacity approach from Grabenstetter and Usher (2013, 2014) may be useful for ETO companies with surplus engineering capacity, such an approach may not fulfil the planning needs of companies where the level of engineering capacity utilisation is high. On the other hand, the approaches proposed by Brachmann and Kolisch (2021); Ghiyasinasab et al. (2021) can be useful references for companies with capacity-constrained engineering resources, however, the authors demonstrate and test the proposed approaches using historical data without explicating how the characteristics of a new engineering project are considered in estimating the duration of individual engineering activities and the workloads for individual engineering resources. Future research should explore how hybrid approaches can be developed for estimating the lead times of engineering activities considering both, the capacity constraints for engineering resources and the characteristics of new engineering projects.

The third research gap observed in the literature concerns procurement planning, which is tasked with coordinating the upstream supply chain actors during the tendering phase. Activities such as supplier selection and determining the type of strategically fitting relationships for different suppliers have been traditionally seen as long-term strategic decisions, based on the needs of high-volume production environments (Ellram 1990; Hesping and Schiele 2015, 2016; Kraljic 1983). Recently proposed approaches for selecting suppliers and supplier relationship types in ETO contexts also adopt this view (Sabri, Micheli, and Cagno 2020; Shlopak, Rød, and Oterhals 2016). However, ETO contexts usually procure items order by order, in low volumes (Adrodegari et al. 2015; Buer et al. 2018b; Jonsson and Mattsson 2003), and the suppliers for components of similar products may vary across customer orders (Alfnes et al. 2021; Mwesiumo, Nujen, and Kvadsheim 2021). As a result, ETO companies require tactical-level managerial approaches for periodically or dynamically reassessing procurement strategies and supplier relationships, and prescriptive guidance on the sourcing levers and coordination mechanisms that should be used for different categories of suppliers (Hesping and Schiele 2016). These are knowledge gaps in the extant sourcing literature where research has been primarily motivated by the needs of mass production contexts (Hesping and Schiele 2015). Studies focussing on addressing the tactical-level procurement and supplier coordination needs of ETO companies are required in future research to support managers' procurement planning activities within S&OP and delivery date setting in ETO contexts.

The final set of knowledge gaps and research needs identified in the review are related to lead time estimation and capacity planning for in-house production activities. The majority of the quantitative decision-support tools and models for tactical planning activities in ETO contexts focus on the mutually linked planning activities of tentatively allocating production capacity, estimating production lead times, assessing the feasibility of completing production activities within customer-imposed due dates, etc. (Alfieri, Tolio, and Urgo 2011, 2012; Carvalho, Oliveira, and Scavarda 2015, 2016; Ghiyasinasab et al. 2021; Micale et al. 2021). Despite the variety of tools proposed for these planning activities, there are gaps and shortcomings in the extant planning and decision-support tools that should be addressed to improve their managerial utility in practice (Bhalla, Alfnes, and Hvolby 2022). Planning tools based on formally specified optimisation models and exact solution techniques are valuable for smaller problem instances but become computationally intractable and practically unusable for industrial applications as the product complexity, the number of unique resources, or the required granularity or detail in planning increase (Alfieri, Tolio, and Urgo 2011, 2012; Carvalho, Oliveira, and Scavarda 2015). Efficient heuristic planning methods have been proposed for MTO contexts (Thürer et al. 2012) that are also potentially useful for estimating production lead times in ETO contexts. However, further development and testing are required for such methods to be viable in practice for complex ETO products with multi-level product structures and multiple subsystems that may be fabricated and assembled in parallel (Bhalla, Alfnes, and Hvolby 2022). Therefore, future research on production planning in ETO contexts should focus on developing effective and efficient heuristic planning methods and decision-support tools for addressing the industrial planning needs within tactical-level lead time estimation and capacity planning for production activities.

6. Conclusion

This paper has investigated the research question: How should engineer-to-order manufacturers contextualise the design of the sales and operations planning process for effective delivery date setting? The paper proposes that based on their specific planning environments,

ETO companies identify their planning needs for effectively setting delivery dates while tendering for new customer orders, and design or redesign their S&OP process focussing on the activities and information flows that address the identified planning needs. To support this in practice, the paper (1) develops an S&OP reference framework that identifies the planning activities and planning inputs that should be considered in ETO companies for contextualising the S&OP process design for effective delivery date setting, and (2) discusses the industrial application of the framework for designing and analysing the S&OP process. The framework is developed by systematically reviewing literature. The paper also (1) discusses applications of the framework in future research, (2) highlights the research gaps within tactical planning in ETO contexts, and (3) suggests the future research needs to address these gaps and to better support ETO practitioners in designing and conducting the S&OP process. The proposed framework contributes to the literature on two, currently isolated streams of research – the research stream on S&OP that has lacked contextual consideration of ETO production (Kreuter et al. 2022; Kristensen and Jonsson 2018) and the research stream on tactical planning and delivery date setting that has lacked an overall framework for cross-functional coordination and coordinated planning (Bhalla, Alfnes, and Hvolby 2022), as highlighted below.

Previous research on S&OP has primarily been contextualised in MTS contexts in food production, consumer electronics production, automotive manufacturing, production of medical products, cardboard production, process industry, etc. (Danese, Molinaro, and Romano 2018; Grimson and Pyke 2007; Noroozi and Wikner 2017; Oliva and Watson 2011), with some contributions addressing MTO contexts in the electrical and electronics industries (Feng, D'Amours, and Beauregard 2008; Grimson and Pyke 2007) and the automotive supplier industry (Gansterer 2015). Consequently, the extant frameworks developed for supporting S&OP research and practice have been implicitly targeted towards the planning needs of MTS and/or MTO contexts, and do not address the planning needs of ETO contexts. Conversely, the framework proposed in this paper is specifically designed to address the needs of ETO contexts. For instance, the proposed framework is structured based on the ETO supply chain matrix proposed by Nam et al. (2018), as opposed to the general supply chain matrix from Stadtler and Kilger (2008) that was used by Pereira, Oliveira, and Carravilla (2020) to develop their S&OP framework. Furthermore, this paper adopts a customer order or customer enquiry-oriented planning perspective for developing the proposed S&OP framework, which is typical for ETO contexts due to low volumes of customer orders, customer-specific order-fulfilment activities, and relatively long delivery lead times (Adrodegari et al. 2015). This is in contrast to MTO and MTS production contexts, where statistical demand forecasts, contractual sales volumes, and backlogged sales volumes are often essential inputs for S&OP (Feng, D'Amours, and Beauregard 2008; Gansterer 2015; Pereira, Oliveira, and Carravilla 2020), and the planning perspective typically adopted at this level concerns volumes of sales, production, distribution, and procurement, while most order-specific activities are planned and controlled with shorter planning horizons at the operational level (Pereira, Oliveira, and Carravilla 2020). Finally, the framework proposed in this paper includes operational inputs as a planning input category which is absent in the S&OP framework from Pereira, Oliveira, and Carravilla (2020). This inclusion is necessitated by the frequently changing planning environment of ETO production contexts where the current and planned states of operational resources must be considered in tactical planning to increase the feasibility of these plans (Alfieri, Tolio, and Urgo 2011, 2012; Alfnes et al. 2021; Carvalho, Oliveira, and Scavarda 2015, 2016; Ghiyasinasab et al. 2021; Wullink et al. 2004; Zorzini et al. 2008b).

The findings of this study and the proposed framework also contribute to the research stream on tactical planning and delivery date setting in ETO contexts, where an overall framework for cross-functional coordination and coordinated planning has been a gap in the extant literature (Bhalla, Alfnes, and Hvolby 2022). The framework addresses the shortcomings of existing frameworks in this literature stream, underlined in section 2 (Adrodegari et al. 2015; Bertrand and Muntslag 1993; Kingsman et al. 1996; Little et al. 2000; Nam et al. 2018; Shurrab, Jonsson, and Johansson 2020b; Zorzini, Corti, and Pozzetti 2008a), by exhaustively identifying S&OP activities and planning inputs for delivery date setting with a cross-functional perspective.

Based on the research gaps highlighted in recent reviews on S&OP (Kreuter et al. 2022; Kristensen and Jonsson 2018) and delivery date setting (Bhalla, Alfnes, and Hvolby 2022), and to the best of our knowledge, the proposed framework is the first to address the design of S&OP and tactical planning processes in ETO contexts focussing on delivery date setting and tendering from a cross-functional perspective. The development and focus of the framework on tendering are partly motivated by empirical observations from the maritime industry reported by the authors in a previous study (Bhalla et al. 2021), and our subsequent research will focus on demonstrating the application of the framework in industrial cases. In the authors' view, the primary potential for industrial application of the framework lies in its use

for supporting the managerial assessment of which planning activities are strategically essential for a company and reconfiguring the design of those planning activities for improving their effectiveness. Future studies can also utilise the framework for defining requirements for decision-support systems and planning functionalities for enterprise planning systems that better address the needs of ETO companies than existing systems (Aslan, Stevenson, and Hendry 2012, 2015). Finally, since the proposed framework focuses on designing S&OP to support delivery date setting, future extensions of the framework can also consider integrating other planning tasks in ETO contexts, e.g. multi-project planning after order confirmation (Adrodegari et al. 2015; Hans et al. 2007), replanning due to engineering changes (Iakymenko et al. 2020), etc., into the scope of the framework.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Norwegian Research Council: [Grant Number 282270].

Notes on contributors

Swapnil Bhalla is a PhD candidate at the Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. He holds an MSc in Global Manufacturing Management from NTNU and a BEng in Mechanical Engineering from Delhi Technological University in New Delhi, India. He has three years of experience in manufacturing process analysis from the automotive industry in India. In his PhD, he is working on solutions for effective tactical sales and operations planning and delivery date quotation in engineer-to-order manufacturing environments focusing on maritime equipment suppliers. His research interests include information systems, data science and analytics, and decision-support in manufacturing supply chains.

Erlend Alfnes is an associate professor at the Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology in Trondheim, Norway. He holds a PhD in Manufacturing Logistics. He has 15 years of experience as a manager of national and international research projects. His main research interests include manufacturing planning and control, enterprise resource planning systems, manufacturing strategy, lean manufacturing, and the integration of lean manufacturing and Industry 4.0, and his research activities focus on engineer-to-order industries.

Hans-Henrik Hvolby is a Full Professor at the Centre for Logistics, Aalborg University in Denmark and a visiting professor at the University of Tasmania, Australia. He has previously held visiting professorships at NTNU in Norway and UNISA in Australia. He holds a PhD in Manufacturing Information Systems, has published more than 170 peer-reviewed papers, and serves as an editorial board member of several international journals. For over two decades, he has been the project manager or partner in large research projects raising national and European research funding. His research areas include manufacturing and supply chain planning, manufacturing information systems, logistics and supply chain integration.

Olumide Emmanuel Oluyisola is a cloud operations professional in the logistics-focused software-as-a-service (SaaS) industry. He holds an MSc in Global Production Management and a PhD in Production and Quality Engineering from the Department of Mechanical and Industrial Engineering, NTNU in Trondheim, Norway. He has five years of industrial experience in various roles as a business process analyst, project manager, and technology consultant in the oil & gas and software industries. He has also collaborated on several national research projects. His current research interests include manufacturing planning and control, smart manufacturing systems design, industrial internet-of-things, applied machine learning, and software security.

Data availability statement

The papers used for developing the framework are cited in the manuscript and included in the list of references. Any additional data about the papers can be made available by the authors upon request.

ORCID

Swapnil Bhalla http://orcid.org/0000-0003-4905-2488
Erlend Alfnes http://orcid.org/0000-0002-9892-3916
Hans-Henrik Hvolby http://orcid.org/0000-0002-5574-5216
Olumide Oluyisola http://orcid.org/0000-0001-7944-3776

References

Adam, Nabil R, J Will M Bertrand, Diane C Morehead, and Julius Surkis. 1993. "Due Date Assignment Procedures with Dynamically Updated Coefficients for Multi-Level Assembly job Shops." *European Journal of Operational Research* 68 (2): 212–227. doi:10.1016/0377-2217(93)90304-6.

Adrodegari, F., A. Bacchetti, R. Pinto, F. Pirola, and M. Zanardini. 2015. "Engineer-to-order (ETO) Production Planning and Control: An Empirical Framework for Machinery-Building Companies." *Production Planning and Control* 26 (11): 910–932. doi:10.1080/09537287.2014.1001808.

Alfieri, Arianna, Tullio Tolio, and Marcello Urgo. 2011. "A Project Scheduling Approach to Production Planning with Feeding Precedence Relations." *International Journal of*

Production Research 49 (4): 995–1020. doi:10.1080/00207541003604844.

Alfieri, Arianna, Tullio Tolio, and Marcello Urgo. 2012. "A two-Stage Stochastic Programming Project Scheduling Approach to Production Planning." *International Journal of Advanced Manufacturing Technology* 62 (1): 279–290. doi:10.1007/s00170-011-3794-4.

Alfnes, Erlend, Jonathan Gosling, Mohamed Naim, and Heidi C. Dreyer. 2021. "Exploring Systemic Factors Creating Uncertainty in Complex Engineer-to-Order Supply Chains: Case Studies from Norwegian Shipbuilding First Tier Suppliers." *International Journal of Production Economics* 108211. doi:10.1016/j.ijpe.2021.108211.

Alfnes, E., and H. H. Hvolby. 2019. "APS Feasibility in an Engineer to Order Environment." In *Advances in Production Management Systems. Production Management for the Factory of the Future. APMS 2019*, edited by Farhad Ameri, Kathryn E. Stecke, Gregor von Cieminski, and Dimitris Kiritsis, 604–611. Cham: Springer.

Amaro, Graça, Linda Hendry, and Brian Kingsman. 1999. "Competitive Advantage, Customisation and a new Taxonomy for non Make-to-Stock Companies." *International Journal of Operations & Production Management* 19 (4): 349–371. doi:10.1108/01443579910254213.

Aslan, B., M. Stevenson, and L. C. Hendry. 2012. "Enterprise Resource Planning Systems: An Assessment of Applicability to Make-To-Order Companies." *Computers in Industry* 63 (7): 692–705. doi:10.1016/j.compind.2012.05.003.

Aslan, B., M. Stevenson, and L. C. Hendry. 2015. "The Applicability and Impact of Enterprise Resource Planning (ERP) Systems: Results from a Mixed Method Study on Make-To-Order (MTO) Companies." *Computers in Industry* 70: 127–143. doi:10.1016/j.compind.2014.10.003.

Barbosa, Cátia, and Américo Azevedo. 2019. "Assessing the Impact of Performance Determinants in Complex MTO/ETO Supply Chains Through an Extended Hybrid Modelling Approach." *International Journal of Production Research* 57 (11): 3577–3597. doi:10.1080/00207543.2018.1543970.

Berry, William L, and Terry Hill. 1992. "Linking Systems to Strategy." *International Journal of Operations & Production Management* 12 (10): 3–15. doi:10.1108/01443579210017204.

Bertolini, Massimo, Marcello Braglia, Leonardo Marrazzini, and Mattia Neroni. 2022. "Project Time Deployment: A new Lean Tool for Losses Analysis in Engineer-to-Order Production Environments." *International Journal of Production Research* 60 (10): 3129–3146. doi:10.1080/00207543.2021.1912428.

Bertrand, J. Will M., and Dennis R. Muntslag. 1993. "Production Control in Engineer-to-Order Firms." *International Journal of Production Economics* 30: 3–22. doi:10.1016/0925-5273(93)90077-X.

Bhalla, Swapnil, Erlend Alfnes, and Hans-Henrik Hvolby. 2022. "Tools and Practices for Tactical Delivery Date Setting in Engineer-to-Order Environments: A Systematic Literature Review." *International Journal of Production Research*, 1–33. doi:10.1080/00207543.2022.2057256.

Bhalla, Swapnil, Erlend Alfnes, Hans-Henrik Hvolby, and Olumide Emmanuel Oluyisola. 2021. "Requirements for Sales and Operations Planning in an Engineer-to-Order Manufacturing Environment." In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*, edited by Alexandre Dolgui, Alain Bernard, David Lemoine, Gregor von Cieminski, and David Romero, 371–380. Cham: Springer.

Brachmann, R., and R. Kolisch. 2021. "The Impact of Flexibility on Engineer-to-Order Production Planning." *International Journal of Production Economics* 239. doi:10.1016/j.ijpe.2021.108183.

Buer, Sven-Vegard, Jan Ola Strandhagen, and Felix TS Chan. 2018a. "The Link Between Industry 4.0 and Lean Manufacturing: Mapping Current Research and Establishing a Research Agenda." *International Journal of Production Research* 56 (8): 2924–2940. doi:10.1080/00207543.2018.1442945.

Buer, Sven-Vegard, Jo Wessel Strandhagen, Jan Ola Strandhagen, and Erlend Alfnes. 2018b. "Strategic fit of Planning Environments: Towards an Integrated Framework." In *Information Systems, Logistics and Supply Chain: 6th International Conference, ILS 2016*, edited by Cecilia Temponi, and Nico Vandaele, 77–92. Cham: Springer.

Burggraf, P., J. Wagner, B. Heinbach, and F. Steinberg. 2021. "Machine Learning-Based Prediction of Missing Components for Assembly - a Case Study at an Engineer-to-Order Manufacturer." *IEEE Access*. doi:10.1109/ACCESS.2021.3075620.

Calosso, T., M. Cantamessa, D. Vu, and A. Villa. 2003. "Production Planning and Order Acceptance in Business to Business Electronic Commerce." *International Journal of Production Economics* 85 (2): 233–249. doi:10.1016/S0925-5273(03)00112-9.

Cannas, Violetta Giada, and Jonathan Gosling. 2021. "A Decade of Engineering-to-Order (2010–2020): Progress and Emerging Themes." *International Journal of Production Economics* 241: 108274. doi:10.1016/j.ijpe.2021.108274.

Cannas, Violetta G, Jonathan Gosling, Margherita Pero, and Tommaso Rossi. 2019. "Engineering and Production Decoupling Configurations: An Empirical Study in the Machinery Industry." *International Journal of Production Economics* 216: 173–189. doi:10.1016/j.ijpe.2019.04.025.

Cannas, Violetta G, Jonathan Gosling, Margherita Pero, and Tommaso Rossi. 2020. "Determinants for Order-Fulfilment Strategies in Engineer-to-Order Companies: Insights from the Machinery Industry." *International Journal of Production Economics* 228: 107743. doi:10.1016/j.ijpe.2020.107743.

Cannas, V. G., M. Pero, R. Pozzi, and T. Rossi. 2018. "An Empirical Application of Lean Management Techniques to Support ETO Design and Production Planning." *IFAC-PapersOnLine* 51 (11): 134–139. doi:10.1016/j.ifacol.2018.08.247.

Carvalho, A. N., F. Oliveira, and L. F. Scavarda. 2015. "Tactical Capacity Planning in a Real-World ETO Industry Case: An Action Research." *International Journal of Production Economics* 167: 187–203. doi:10.1016/j.ijpe.2015.05.032.

Carvalho, A. N., F. Oliveira, and L. F. Scavarda. 2016. "Tactical Capacity Planning in a Real-World ETO Industry Case: A Robust Optimization Approach." *International Journal of Production Economics* 180: 158–171. doi:10.1016/j.ijpe.2016.07.019.

Cassaigne, Nathalie, M. Kromker, Madan G Singh, and S. Wurst. 1997. Decision Support for effective Bidding in a competitive Business Environment. Paper presented at the IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation.

Danese, Pamela, Margherita Molinaro, and Pietro Romano. 2018. "Managing Evolutionary Paths in Sales and Operations Planning: Key Dimensions and Sequences of Implementation." *International Journal of Production Research* 56 (5): 2036–2053. doi:10.1080/00207543.2017.1355119.

De Boer, R., J. M. J. Schutten, and W. H. M. Zijm. 1997. "A Decision Support System for Ship Maintenance Capacity Planning." *CIRP Annals - Manufacturing Technology* 46 (1): 391–396. doi:10.1016/s0007-8506(07)60850-6.

Dekkers, Rob. 2006. "Engineering Management and the Order Entry Point." *International Journal of Production Research* 44 (18–19): 4011–4025. doi:10.1080/00207540600696328.

Dekkers, Rob, C. M. Chang, and Jochen Kreutzfeldt. 2013. "The Interface Between "Product Design and Engineering" and Manufacturing: A Review of the Literature and Empirical Evidence." *International Journal of Production Economics* 144 (1): 316–333. doi:10.1016/j.ijpe.2013.02.020.

Donaldson, Lex. 2001. *The Contingency Theory of Organizations*. Thousand Oaks: Sage.

Easton, F. F., and D. R. Moodie. 1999. "Pricing and Lead Time Decisions for Make-to-Order Firms with Contingent Orders." *European Journal of Operational Research* 116 (2): 305–318. doi:10.1016/s0377-2217(98)00101-5.

Ebadian, M., M. Rabbani, F. Jolai, S. A. Torabi, and R. Tavakkoli-Moghaddam. 2008. "A new Decision-Making Structure for the Order Entry Stage in Make-to-Order Environments." *International Journal of Production Economics* 111 (2): 351–367. doi:10.1016/j.ijpe.2007.01.004.

Ebadian, M., M. Rabbani, S. A. Torabi, and F. Jolai. 2009. "Hierarchical Production Planning and Scheduling in Make-to-Order Environments: Reaching Short and Reliable Delivery Dates." *International Journal of Production Research* 47 (20): 5761–5789. doi:10.1080/00207540802010799.

Ebben, Mark JR, Erwin W Hans, and F. M. Olde Weghuis. 2005. "Workload Based Order Acceptance in job Shop Environments." *OR Spectrum* 27: 107–122. doi:10.1007/s00291-004-0171-9.

Ellram, Lisa M. 1990. "The Supplier Selection Decision in Strategic Partnerships." *Journal of Purchasing and Materials Management* 26 (4): 8–14. doi:10.1111/j.1745-493X.1990.tb00515.x.

Emblemsvåg, Jan. 2014. "Lean Project Planning in Shipbuilding." *Journal of Ship Production and Design* 30 (02): 79–88. doi:10.5957/jspd.2014.30.2.79.

Eyers, Daniel R, Andrew T Potter, Jonathan Gosling, and Mohamed M Naim. 2021. "The Impact of Additive Manufacturing on the Product-Process Matrix." *Production Planning & Control*, 1–17. doi:10.1080/09537287.2021.1876940.

Fang, Jun, and Xing Wei. 2020. "A Knowledge Support Approach for the Preliminary Design of Platform-Based Products in Engineering-To-Order Manufacturing." *Advanced Engineering Informatics* 46: 101196. doi:10.1016/j.aei.2020.101196.

Feng, Y., S. D'Amours, and R. Beauregard. 2008. "The Value of Sales and Operations Planning in Oriented Strand Board Industry with Make-to-Order Manufacturing System: Cross Functional Integration Under Deterministic Demand and Spot Market Recourse." *International Journal of Production Economics* 115 (1): 189–209. doi:10.1016/j.ijpe.2008.06.002.

Gansterer, M. 2015. "Aggregate Planning and Forecasting in Make-to-Order Production Systems." *International Journal of Production Economics* 170: 521–528. doi:10.1016/j.ijpe.2015.06.001.

Ghiyasinasab, M., N. Lehoux, S. Ménard, and C. Cloutier. 2021. "Production Planning and Project Scheduling for Engineer-to-Order Systems- Case Study for Engineered Wood Production." *International Journal of Production Research* 59 (4): 1068–1087. doi:10.1080/00207543.2020.1717009.

Goh, Shao Hung, and Stephen Eldridge. 2015. "New Product Introduction and Supplier Integration in Sales and Operations Planning: Evidence from the Asia Pacific Region." *International Journal of Physical Distribution & Logistics Management* 45 (9/10): 861–886. doi:10.1108/IJPDLM-08-2014-0215.

Gosling, Jonathan, Bill Hewlett, and Mohamed M Naim. 2017. "Extending Customer Order Penetration Concepts to Engineering Designs." *International Journal of Operations & Production Management* 37 (4): 402–422. doi:10.1108/IJOPM-07-2015-0453.

Gosling, Jonathan, and Mohamed M Naim. 2009. "Engineer-to-order Supply Chain Management: A Literature Review and Research Agenda." *International Journal of Production Economics* 122 (2): 741–754. doi:10.1016/j.ijpe.2009.07.002.

Gourdon, Karin, and Christian Steidl. 2019. "Global Value Chains and the Shipbuilding Industry." *OECD Science, Technology and Industry Working Papers* 2019 (08): 1–35. doi:10.1787/7e94709a-en.

Grabenstetter, Douglas H, and John M Usher. 2013. "Determining job Complexity in an Engineer to Order Environment for due Date Estimation Using a Proposed Framework." *International Journal of Production Research* 51 (19): 5728–5740. doi:10.1080/00207543.2013.787169.

Grabenstetter, Douglas H, and John M Usher. 2014. "Developing due Dates in an Engineer-to-Order Engineering Environment." *International Journal of Production Research* 52 (21): 6349–6361. doi:10.1080/00207543.2014.940072.

Grimson, J Andrew, and David F Pyke. 2007. "Sales and Operations Planning: An Exploratory Study and Framework." *The International Journal of Logistics Management* 18 (3): 322–346. doi:10.1108/09574090710835093.

Hans, Erwin W, Willy Herroelen, Roel Leus, and Gerhard Wullink. 2007. "A Hierarchical Approach to Multi-Project Planning Under Uncertainty." *Omega* 35 (5): 563–577. doi:10.1016/j.omega.2005.10.004.

Hesping, Frank Henrik, and Holger Schiele. 2015. "Purchasing Strategy Development: A Multi-Level Review." *Journal of Purchasing and Supply Management* 21 (2): 138–150. doi:10.1016/j.pursup.2014.12.005.

Hesping, Frank Henrik, and Holger Schiele. 2016. "Matching Tactical Sourcing Levers with the Kraljič Matrix: Empirical Evidence on Purchasing Portfolios." *International Journal of Production Economics* 177: 101–117. doi:10.1016/j.ijpe.2016.04.011.

Hicks, C., and P. M. Braiden. 2000. "Computer-aided Production Management Issues in the Engineer-to-Order Production of Complex Capital Goods Explored Using a Simulation Approach." *International Journal of Production Research* 38 (18): 4783–4810. doi:10.1080/00207540010001019.

Hicks, C., T. McGovern, and C. F. Earl. 2000. "Supply Chain Management: A Strategic Issue in Engineer to Order Manufacturing." *International Journal of Production Economics* 65 (2): 179–190. doi:10.1016/S0925-5273(99)00026-2.

Hicks, Chris, Tom McGovern, and Chris F Earl. 2001. "A Typology of UK Engineer-to-Order Companies." *International Journal of Logistics* 4 (1): 43–56. doi:10.1080/13675560110038068.

Iakymenko, Natalia, Anita Romsdal, Erlend Alfnes, Marco Semini, and Jan Ola Strandhagen. 2020. "Status of Engineering Change Management in the Engineer-to-Order Production Environment: Insights from a Multiple Case Study." *International Journal of Production Research* 58 (15): 4506–4528. doi:10.1080/00207543.2020.1759836.

Ivert, Linea Kjellsdotter, Iskra Dukovska-Popovska, Anna Fredriksson, Heidi C Dreyer, and Riikka Kaipia. 2015. "Contingency Between S & OP Design and Planning Environment." *International Journal of Physical Distribution & Logistics Management* 45 (8): 747–773. doi:10.1108/IJPDLM-04-2014-0088.

Jacobs, F Robert, William L Berry, D Clay Whybark, and Thomas E Vollmann. 2011. *Manufacturing Planning and Control for Supply Chain Management: APICS/CPIM Certification Edition*. New York: McGraw-Hill Education.

Johnsen, Sara Markworth, and Lars Hvam. 2019. "Understanding the Impact of non-Standard Customisations in an Engineer-to-Order Context: A Case Study." *International Journal of Production Research* 57 (21): 6780–6794. doi:10.1080/00207543.2018.1471239.

Jonsson, Patrik, Riikka Kaipia, and Mark Barratt. 2021. "Guest Editorial: The Future of S&OP: Dynamic Complexity, Ecosystems and Resilience." *International Journal of Physical Distribution & Logistics Management* 51 (6): 553–565. doi:10.1108/IJPDLM-07-2021-452.

Jonsson, Patrik, and Stig-Arne Mattsson. 2003. "The Implications of fit Between Planning Environments and Manufacturing Planning and Control Methods." *International Journal of Operations & Production Management* 23 (8): 872–900. doi:10.1108/01443570310486338.

Kingsman, B., L. Hendry, A. Mercer, and A. De Souza. 1996. "Responding to Customer Enquiries in Make-to-Order Companies Problems and Solutions." *International Journal of Production Economics* 46-47: 219–231. doi:10.1016/0925-5273(95)00199-9.

Kingsman, B. G., and A. Mercer. 1997. "Strike Rate Matrices for Integrating Marketing and Production During the Tendering Process in Make-to-Order Subcontractors." *International Transactions in Operational Research* 4 (4): 251–257. doi:10.1111/j.1475-3995.1997.tb00081.x.

Kingsman, B. G., I. P. Tatsiopoulos, and L. C. Hendry. 1989. "A Structural Methodology for Managing Manufacturing Lead Times in Make-to-Order Companies." *European Journal of Operational Research* 40 (2): 196–209. doi:10.1016/0377-2217(89)90330-5.

Kingsman, B., L. Worden, L. Hendry, A. Mercer, and E. Wilson. 1993. "Integrating Marketing and Production Planning in Make-to-Order Companies." *International Journal of Production Economics* 30–31: 53–66. doi:10.1016/0925-5273(93)90081-U.

Kraljic, Peter. 1983. "Purchasing Must Become Supply Management." *Harvard Business Review* 61 (5): 109–117.

Kreuter, Tobias, Christian Kalla, Luiz Felipe Scavarda, Antônio Márcio Tavares Thomé, and Bernd Hellingrath. 2021. "Developing and Implementing Contextualised S&OP Designs–an Enterprise Architecture Management Approach." *International Journal of Physical Distribution & Logistics Management* 51 (6): 634–655. doi:10.1108/IJPDLM-06-2019-0199.

Kreuter, Tobias, Luiz Felipe Scavarda, Antonio Márcio Tavares Thomé, Bernd Hellingrath, and Marcelo Xavier Seeling. 2022. "Empirical and Theoretical Perspectives in Sales and Operations Planning." *Review of Managerial Science* 16: 319–354. doi:10.1007/s11846-021-00455-y.

Kristensen, Jesper, and Patrik Jonsson. 2018. "Context-based Sales and Operations Planning (S&OP) Research." *International Journal of Physical Distribution & Logistics Management* 48 (1): 19–46. doi:10.1108/IJPDLM-11-2017-0352.

Lapide, Larry. 2005. "Sales and Operations Planning Part III: A Diagnostic Model." *The Journal of Business Forecasting* 24 (1): 13–16.

Lawrence, Paul R, and Jay W Lorsch. 1969. *Organization and Environment: Managing Differentiation and Integration*. Homewood: Richard D. Irwin, Inc.

Little, David, Ralph Rollins, Matthew Peck, and J. Keith Porter. 2000. "Integrated Planning and Scheduling in the Engineer-to-Order Sector." *International Journal of Computer Integrated Manufacturing* 13 (6): 545–554. doi:10.1080/09511920050195977.

Løkkegaard, Martin, Christian Alexander Bertram, Niels Henrik Mortensen, Lars Hvam, and Anders Haug. 2022. "Identifying Profitable Reference Architectures in an Engineer-to-Order Context." *International Journal of Production Research*, 1–15. doi:10.1080/00207543.2022.2036850.

Mello, Mario Henrique, Jonathan Gosling, Mohamed M Naim, Jan Ola Strandhagen, and Per Olaf Brett. 2017. "Improving Coordination in an Engineer-to-Order Supply Chain Using a Soft Systems Approach." *Production Planning & Control* 28 (2): 89–107. doi:10.1080/09537287.2016.1233471.

Micale, R., C. M. La Fata, M. Enea, and G. La Scalia. 2021. "Regenerative Scheduling Problem in Engineer to Order Manufacturing: An Economic Assessment." *Journal of Intelligent Manufacturing*, doi:10.1007/s10845-020-01728-1.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and Prisma Group. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *PLoS Medicine* 6 (7): e1000097. doi:10.1136/bmj.b2535.

Mwesiumo, Deodat, Bella B Nujen, and Nina Pereira Kvadsheim. 2021. "A Systematic Approach to Implementing Multi-Sourcing Strategy in Engineer-to-Order Production." In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems. APMS 2021*, edited by Alexandre Dolgui, Alain Bernard, David Lemoine, Gregor von Cieminski, and David Romero, 381–389. Cham: Springer.

Nam, S., H. Shen, C. Ryu, and J. G. Shin. 2018. "SCP-Matrix Based Shipyard APS Design: Application to Long-Term Production Plan." *International Journal of Naval Architecture and Ocean Engineering* 10 (6): 741–761. doi:10.1016/j.ijnaoe.2017.10.003.

Newman, W Rocky, and V. Sridharan. 1995. "Linking Manufacturing Planning and Control to the Manufacturing Environment." *Integrated Manufacturing Systems* 6 (4): 36–42. doi:10.1108/09576069510088952.

Noroozi, Sayeh, and Joakim Wikner. 2017. "Sales and Operations Planning in the Process Industry: A Literature Review." *International Journal of Production Economics* 188: 139–155. doi:10.1016/j.ijpe.2017.03.006.

Olhager, Jan. 2010. "The Role of the Customer Order Decoupling Point in Production and Supply Chain Management." *Computers in Industry* 61 (9): 863–868. doi:10.1016/j.compind.2010.07.011.

Olhager, Jan. 2013. "Evolution of Operations Planning and Control: From Production to Supply Chains." *International Journal of Production Research* 51 (23–24): 6836–6843. doi:10.1080/00207543.2012.761363.

Olhager, Jan, Martin Rudberg, and Joakim Wikner. 2001. "Long-term Capacity Management: Linking the Perspectives from Manufacturing Strategy and Sales and Operations Planning." *International Journal of Production Economics* 69 (2): 215–225. doi:10.1016/S0925-5273(99)00098-5.

Oliva, Rogelio, and Noel Watson. 2011. "Cross-Functional Alignment in Supply Chain Planning: A Case Study of Sales and Operations Planning." *Journal of Operations Management* 29 (5): 434–448. doi:10.1016/j.jom.2010.11.012.

Oluyisola, O. E., F. Sgarbossa, and J. O. Strandhagen. 2020. "Smart Production Planning and Control: Concept, use-Cases and Sustainability Implications." *Sustainability (Switzerland)* 12 (9), doi:10.3390/su12093791.

Park, C., J. Song, J. G. Kim, and I. Kim. 1999. "Delivery Date Decision Support System for the Large Scale Make-to-Order Manufacturing Companies: A Korean Electric Motor Company Case." *Production Planning and Control* 10 (6): 585–597. doi:10.1080/095372899232885.

Pedroso, Carolina Belotti, Lucas Daniel Del Rosso Calache, Francisco Rodrigues Lima, Andrea Lago da Silva, and Luiz César Ribeiro Carpinetti. 2017. "Proposal of a Model for Sales and Operations Planning (S&OP) Maturity Evaluation." *Production* 27. doi:10.1590/0103-6513.20170024.

Pereira, Daniel Filipe, José Fernando Oliveira, and Maria Antónia Carravilla. 2020. "Tactical Sales and Operations Planning: A Holistic Framework and a Literature Review of Decision-Making Models." *International Journal of Production Economics* 228: 107695. doi:10.1016/j.ijpe.2020.107695.

Reid, Iain, David Bamford, and Hossam Ismail. 2019. "Reconciling Engineer-to-Order Uncertainty by Supporting Front-end Decision-Making." *International Journal of Production Research* 57 (21): 6856–6874. doi:10.1080/00207543.2018.1552370.

Robinson, K. R., and S. A. Moses. 2006. "Effect of Granularity of Resource Availability on the Accuracy of due Date Assignment." *International Journal of Production Research* 44 (24): 5391–5414. doi:10.1080/00207540600665810.

Ruben, Robert A, and Farzad Mahmoodi. 2000. "Lead Time Prediction in Unbalanced Production Systems." *International Journal of Production Research* 38 (7): 1711–1729. doi:10.1080/002075400188816.

Sabri, Yasmine, Guido JL Micheli, and Enrico Cagno. 2020. "Supplier Selection and Supply Chain Configuration in the Projects Environment." *Production Planning & Control*, 1–19. doi:10.1080/09537287.2020.1853269.

Saghiri, Soroosh, and Alex Hill. 2014. "Supplier Relationship Impacts on Postponement Strategies." *International Journal of Production Research* 52 (7): 2134–2153. doi:10.1080/00207543.2013.857053.

Semini, Marco, Dag E Gotteberg Haartveit, Erlend Alfnes, Emrah Arica, Per Olaf Brett, and Jan Ola Strandhagen. 2014. "Strategies for Customized Shipbuilding with Different Customer Order Decoupling Points." *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment* 228 (4): 362–372. doi:10.1177/2F1475090213493770.

Shapiro, Stuart. 1997. "Degrees of Freedom: The Interaction of Standards of Practice and Engineering Judgment." *Science, Technology, & Human Values* 22 (3): 286–316. doi:10.1177/016224399702200302.

Shishank, Shishank, and Rob Dekkers. 2013. "Outsourcing: Decision-Making Methods and Criteria During Design and Engineering." *Production Planning & Control* 24 (4–5): 318–336. doi:10.1080/09537287.2011.648544.

Shlopak, Mikhail, Espen Rød, and Oddmund Oterhals. 2016. "Developing Supplier Strategies for ETO Companies: A Case Study." In *Advances in Production Management Systems. Initiatives for a Sustainable World. APMS 2016*, edited by Irenilza Nääs, Oduvaldo Vendrametto, João Mendes Reis, Rodrigo Franco Gonçalves, Márcia Terra Silva, Gregor von Cieminski, and Dimitris Kiritsis, 911–918. Cham: Springer.

Shurrab, Hafez, Patrik Jonsson, and Mats I Johansson. 2020a. "Managing Complexity Through Integrative Tactical Planning in Engineer-to-Order Environments: Insights from Four Case Studies." *Production Planning & Control* 33 (9–10): 907–924. doi:10.1080/09537287.2020.1837937.

Shurrab, Hafez, Patrik Jonsson, and Mats I Johansson. 2020b. "A Tactical Demand-Supply Planning Framework to Manage Complexity in Engineer-to-Order Environments: Insights from an in-Depth Case Study." *Production Planning & Control* 33 (5): 462–479. doi:10.1080/09537287.2020.1829147.

Stadtler, Hartmut, and Christoph Kilger. 2008. *Supply Chain Management and Advanced Planning: Concepts, Models, Software, and Case Studies*. Heidelberg: Springer.

Stentoft, Jan, Per Vagn Freytag, and Ole Stegmann Mikkelsen. 2020. "The S&OP Process and the Influence of Personality and key Behavioral Indicators: Insights from a Longitudinal Case Study." *International Journal of Physical Distribution & Logistics Management* 51 (6): 585–606. doi:10.1108/IJPDLM-02-2020-0056.

Strandhagen, Jo Wessel, Sven-Vegard Buer, Marco Semini, Erlend Alfnes, and Jan Ola Strandhagen. 2020. "Sustainability Challenges and how Industry 4.0 Technologies Can Address Them: A Case Study of a Shipbuilding Supply Chain." *Production Planning & Control* 33 (9–10): 995–1010. doi:10.1080/09537287.2020.1837940.

Sylla, Abdourahim, Thierry Coudert, Elise Vareilles, Laurent Geneste, and Michel Aldanondo. 2021. "Possibilistic Pareto-Dominance Approach to Support Technical bid Selection Under Imprecision and Uncertainty in Engineer-to-Order Bidding Process." *International Journal of Production Research* 59 (21): 6361–6381. doi:10.1080/00207543.2020.1812754.

Sylla, A., D. Guillon, E. Vareilles, M. Aldanondo, T. Coudert, and L. Geneste. 2018. "Configuration Knowledge Modeling: How to Extend Configuration from Assemble/Make to Order Towards Engineer to Order for the Bidding Process." *Computers in Industry* 99: 29–41. doi:10.1016/j.compind.2018.03.019.

Thomé, Antônio Márcio Tavares, Luiz Felipe Scavarda, Nicole Suclla Fernandez, and Annibal José Scavarda. 2012a. "Sales and Operations Planning and the Firm Performance." *International Journal of Productivity & Performance Management* 61 (4): 359–381. doi:10.1108/17410401211212643.

Thomé, Antônio Márcio Tavares, Luiz Felipe Scavarda, Nicole Suclla Fernandez, and Annibal José Scavarda. 2012b. "Sales and Operations Planning: A Research Synthesis." *International Journal of Production Economics* 138 (1): 1–13. doi:10.1016/j.ijpe.2011.11.027.

Thomé, Antônio Márcio Tavares, Luiz Felipe Scavarda, and Annibal José Scavarda. 2016. "Conducting Systematic Literature Review in Operations Management." *Production Planning & Control* 27 (5): 408–420. doi:10.1080/09537287.2015.1129464.

Thomé, Antônio Márcio Tavares, Rui Soucasaux Sousa, and Luiz Felipe Roris Rodriguez Scavarda do Carmo. 2014. "The Impact of Sales and Operations Planning Practices on Manufacturing Operational Performance." *International Journal of Production Research* 52 (7): 2108–2121. doi:10.1080/00207543.2013.853889.

Thürer, Matthias, George Huang, Mark Stevenson, Cristovao Silva, and Moacir Godinho Filho. 2012. "The Performance of Due Date Setting Rules in Assembly and Multi-Stage job Shops: An Assessment by Simulation." *International Journal of Production Research* 50 (20): 5949–5965. doi:10.1080/00207543.2011.638942.

Tranfield, David, David Denyer, and Palminder Smart. 2003. "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review." *British Journal of Management* 14 (3): 207–222. doi:10.1111/1467-8551.00375.

Tuomikangas, Nina, and Riikka Kaipia. 2014. "A Coordination Framework for Sales and Operations Planning (S&OP): Synthesis from the Literature." *International Journal of Production Economics* 154: 243–262. doi:10.1016/j.ijpe.2014.04.026.

Ulonska, S., and T. Welo. 2016. "On the use of Product Portfolio and Variant Maps as Visualization Tools to Support Platform-Based Development Strategies." *Concurrent Engineering: Research and Applications* 24 (3): 211–226. doi:10.1177/2F1063293X16654531.

Vereecke, Ann, Karlien Vanderheyden, Philippe Baecke, and Tom Van Steendam. 2018. "Mind the gap–Assessing Maturity of Demand Planning, a Cornerstone of S&OP." *International Journal of Operations & Production Management* 38 (8): 1618–1639. doi:10.1108/IJOPM-11-2016-0698.

Wagire, Aniruddha Anil, Rohit Joshi, Ajay Pal Singh Rathore, and Rakesh Jain. 2021. "Development of Maturity Model for Assessing the Implementation of Industry 4.0: Learning from Theory and Practice." *Production Planning & Control* 32 (8): 603–622. doi:10.1080/09537287.2020.1744763.

Wagner, Stephan M, Kristoph KR Ullrich, and Sandra Transchel. 2014. "The Game Plan for Aligning the Organization." *Business Horizons* 57 (2): 189–201. doi:10.1016/j.bushor.2013.11.002.

Wallace, Thomas F. 2004. *Sales and Operations Planning: The how-to Handbook*. Cincinnati: TF Wallace & Co.

Wallace, Tom, and Bob Stahl. 2008. "The Demand Planning Process in Executive S&OP." *The Journal of Business Forecasting* 27 (3): 19.

Watson, Richard T, and Jane Webster. 2020. "Analysing the Past to Prepare for the Future: Writing a Literature Review a Roadmap for Release 2.0." *Journal of Decision Systems* 29 (3): 129–147. doi:10.1080/12460125.2020.1798591.

Webster, Jane, and Richard T Watson. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review." *MIS Quarterly* 26 (2): xiii–xxiii.

Wikner, Joakim, and Martin Rudberg. 2005. "Integrating Production and Engineering Perspectives on the Customer Order Decoupling Point." *International Journal of Operations & Production Management* 25 (7): 623–641. doi:10.1108/01443570510605072.

Willner, Olga, Jonathan Gosling, and Paul Schönsleben. 2016a. "Establishing a Maturity Model for Design Automation in Sales-Delivery Processes of ETO Products." *Computers in Industry* 82: 57–68. doi:10.1016/j.compind.2016.05.003.

Willner, Olga, Daryl Powell, Markus Gerschberger, and Paul Schönsleben. 2016b. "Exploring the Archetypes of Engineer-to-Order: An Empirical Analysis." *International Journal of Operations & Production Management* 36 (3): 242–264. doi:10.1108/IJOPM-07-2014-0339.

Wing, Larry, and Glynn Perry. 2001. "Toward Twenty-First-Century Pharmaceutical Sales and Operations Planning." *Pharmaceutical Technology* 25 (11): 20–26.

Wullink, Gerhard, A. J. R. M. Gademann, Erwin W Hans, and Aart van Harten. 2004. "Scenario-based Approach for Flexible Resource Loading Under Uncertainty." *International Journal of Production Research* 42 (24): 5079–5098. doi:10.1080/002075410001733887.

Yang, W., and R. Y. K. Fung. 2014. "An Available-to-Promise Decision Support System for a Multi-Site Make-to-Order Production System." *International Journal of Production Research* 52 (14): 4253–4266. doi:10.1080/00207543.2013.877612.

Zennaro, I., S. Finco, D. Battini, and A. Persona. 2019. "Big Size Highly Customised Product Manufacturing Systems: A Literature Review and Future Research Agenda." *International Journal of Production Research* 57 (15–16): 5362–5385. doi:10.1080/00207543.2019.1582819.

Zheng, Ting, Marco Ardolino, Andrea Bacchetti, and Marco Perona. 2021. "The Applications of Industry 4.0 Technologies in Manufacturing Context: A Systematic Literature Review." *International Journal of Production Research* 59 (6): 1922–1954. doi:10.1080/00207543.2020.1824085.

Zijm, W. H. M. 2000. "Towards Intelligent Manufacturing Planning and Control Systems." *OR Spektrum* 22 (3): 313–345. doi:10.1007/s002919900032.

Zorzini, M., D. Corti, and A. Pozzetti. 2008a. "Due Date (DD) Quotation and Capacity Planning in Make-to-Order Companies: Results from an Empirical Analysis." *International Journal of Production Economics* 112 (2): 919–933. doi:10.1016/j.ijpe.2007.08.005.

Zorzini, M., L. Hendry, M. Stevenson, and A. Pozzetti. 2008b. "Customer Enquiry Management and Product Customization: An Empirical Multi-Case Study Analysis in the Italian Capital Goods Sector." *International Journal of Operations & Production Management* 28 (12): 1186–1218. doi:10.1108/01443570810919369.

Zorzini, Marta, Mark Stevenson, and Linda C Hendry. 2012. "Customer Enquiry Management in Global Supply Chains: A Comparative Multi-Case Study Analysis." *European Management Journal* 30 (2): 121–140. doi:10.1016/j.emj.2011.10.006.

Appendix

Table A1 shows the distribution of the 75 reviewed papers across various journals and conference proceedings; Figure A1 shows the distribution of their publication years; Table A2 summarises the papers' contributions to the research question of this study.

Table A1. Distribution of papers across journals and conference proceedings.

Source	Number of papers
International Journal of Production Research	18
International Journal of Production Economics	15
Production Planning and Control	9
Computers in Industry	5
European Journal of Operational Research	3
International Journal of Operations & Production Management	3
IFIP International Conference on Advances in Production Management Systems	3
OR Spectrum	2
Other	17
Total	75

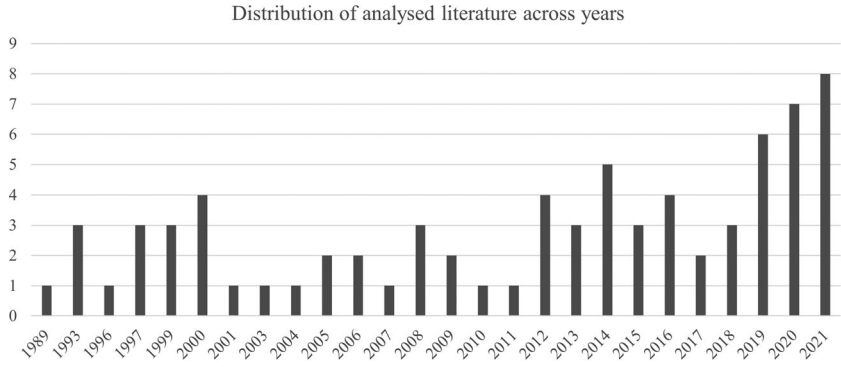


Figure A1. Distribution of papers across years.

Table A2. Overview of the contributions of the analysed literature in addressing the research question.

	Planning activities				Information flow or input				
Reference	Sales	Engineering	Procurement	Production	Strategic	External	Operational	C/F	Type of contribution
Kingsman, Tatsiopoulos, and Hendry (1989)	4.1.3							Sa12	E
Adam et al. (1993)				4.4.2			Pd7		I
Bertrand and Muntslag (1993)	4.1.3	4.2.2, 4.2.3		4.4.1, 4.4.3					E/I
Kingsman et al. (1993)	4.1.3		4.3.3	4.4.3	Sa1, Sa2				E/I
Kingsman et al. (1996)	4.1.1, 4.1.3	4.2.1, 4.2.2	4.3.1	4.4.1, 4.4.3	Sa1, Sa2, Sa3	Sa4, Sa5	Sa7, Pd8	Sa8, Sa12, Sa13	E/I
Cassaigne et al. (1997)	4.1.1, 4.1.3				Sa2	Sa5			I
De Boer, Schutten, and Zijm (1997)				4.4.1, 4.4.2			Pd6, Pd7	Pd10	E/I
Kingsman and Mercer (1997)	4.1.3				Sa1, Sa2				E/I
Amaro, Hendry, and Kingsman (1999)	4.1.1, 4.1.3	4.2.2			Sa1, Sa2, Sa3, En1				I
Easton and Moodie (1999)	4.1.3								E
Park et al. (1999)				4.4.2, 4.4.3	Pd2		Pd7		E
Hicks and Braiden (2000)				4.4.2				Pd12	E
Hicks, McGovern, and Earl (2000)	4.1.1		4.3.1, 4.3.2, 4.3.3		Sa3, Pc1, Pc3, Pc4, En2, En3, Pd4	Sa5, Pc5, Pc6, Pc7	Pc8	Pc10, Pc12	E/I
Ruben and Mahmoodi (2000)				4.4.1, 4.4.2	Pd2		Pd7		E
Zijm (2000)		4.2.1, 4.2.3, 4.2.4		4.4.2, 4.4.3					E
Hicks, McGovern, and Earl (2001)			4.3.2	4.4.1	Pc3, Pd4				I
Calosso et al. (2003)	4.1.3		4.3.3		Sa2	Pc6			I
Wullink et al. (2004)				4.4.2, 4.4.3	Pd2	Pd5			E
Ebben, Hans, and Weghuis (2005)				4.4.1	Pd1, Pd2				I
Wikner and Rudberg (2005)				4.4.2				Pd11	I
Dekkers (2006)		4.2.2			En1				I
Robinson and Moses (2006)				4.4.1	Pd1, Pd2				I
Hans et al. (2007)	4.1.2, 4.1.3	4.2.3		4.4.1, 4.4.2					I
Ebadian et al. (2008)	4.1.2, 4.1.3		4.3.2, 4.3.3	4.4.2, 4.4.3	Sa3, Pc2, Pd2	Sa5, Pd5	Pd7	Sa10, Sa11, Sa12, Sa13, Pd12	E/I
Zorzini, Corti, and Pozzetti (2008a)	4.1.3			4.4.1, 4.4.2, 4.4.3	Sa2, Pd1, Pd2		Pd7		E
Zorzini et al. (2008b)	4.1.1	4.2.1	4.3.1, 4.3.2, 4.3.3	4.4.1, 4.4.2		En5, Pc6	Pd6		I
Ebadian et al. (2009)	4.1.2, 4.1.3				Sa3	Sa5		Sa12, Sa13	E
Gosling and Naim (2009)		4.2.1, 4.2.2			En1				I
Olhager (2010)			4.3.3		Pc4				I
Alfieri, Tolio, and Urgo (2011)				4.4.1, 4.4.2	Pd1, Pd2		Pd6, Pd7	Pd10	E/I
Alfieri, Tolio, and Urgo (2012)				4.4.2	Pd1, Pd2		Pd6, Pd7	Pd10	E/I

(*continued*).

Table A2. Continued.

Reference	Planning activities				Information flow or input				Type of contribution
	Sales	Engineering	Procurement	Production	Strategic	External	Operational	C/F	
Aslan, Stevenson, and Hendry (2012)	4.1.3								E
Thürer et al. (2012)				4.4.2	Pd2		Pd7		E
Zorzini, Stevenson, and Hendry (2012)	4.1.3		4.3.1, 4.3.2, 4.3.3	4.4.2	Pc1	Pc6, Pc7	Pc8, Pc9	Sa11, Pc10, Pc11, Pd12	E/I
Dekkers, Chang, and Kreutzfeldt (2013)			4.3.2, 4.3.3		Pc1			Pc10	I
Grabenstetter and Usher (2013)	4.1.3	4.2.1, 4.2.2, 4.2.3			En1, En2		En7, En8	Sa10, En11	E/I
Shishank and Dekkers (2013)			4.3.2, 4.3.3		Pc1	Pc6, Pc7	Pc8, Pc9	Pc10	E/I
Emblemsvåg (2014)			4.3.1					Pc10, Pc11	I
Grabenstetter and Usher (2014)	4.1.3	4.2.1, 4.2.2, 4.2.3			En1, En2		En7, En8	Sa10, En11	E/I
Saghiri and Hill (2014)			4.3.2		Pc3				E/I
Semini et al. (2014)		4.2.2			En1				I
Yang and Fung (2014)				4.4.2	Pd3				I
Adrodegari et al. (2015)	4.1.1, 4.1.2, 4.1.3	4.2.1, 4.2.2, 4.2.3		4.4.1, 4.4.2, 4.4.3	Sa1, Sa2, Sa3, En2, Pd1, Pd2	Sa4, Sa5	Sa6, Sa7, Pd8	Sa8, Sa9, En11, Pd10	E/I
Aslan, Stevenson, and Hendry (2015)	4.1.3	4.2.4		4.4.1	En4, Pd1, Pd2				E/I
Carvalho, Oliveira, and Scavarda (2015)	4.1.3			4.4.1, 4.4.2, 4.4.3	Pd2	Pd5	Pd6, Pd7, Pd9	Sa12, Sa13	E
Carvalho, Oliveira, and Scavarda (2016)				4.4.1, 4.4.2, 4.4.3	Pd2	Pd5	Pd6, Pd7, Pd9		E
Shlopak, Rød, and Oterhals (2016)			4.3.1, 4.3.2		Pc3		Pc8, Pc9		I
Ulonska and Welo (2016)		4.2.1			En2				E
Willner, Gosling, and Schönsleben (2016a)	4.1.3	4.2.1			En1, En2			Sa10, En11	E
Gosling, Hewlett, and Naim (2017)		4.2.2, 4.2.4			En1, En4				I
Mello et al. (2017)			4.3.2, 4.3.3			Pc6, Pc7			I
Cannas et al. (2018)		4.2.3			En1, En2			En11	E
Nam et al. (2018)		4.2.1		4.4.1				En11, Pd10	E/I
Sylla et al. (2018)	4.1.3	4.2.1			En2	Sa5		En11	E
Alfnes and Hvolby (2019)	4.1.3	4.2.1	4.3.1, 4.3.2	4.4.1, 4.4.2	En2, Pd2	Sa5, Pc6	Pd7	Sa11, Sa12, Pc10, Pd10	E
Barbosa and Azevedo (2019)	4.1.3	4.2.3		4.4.2	Sa1, En1, En2, Pd2	Sa4, Sa5	En8, En9, Pd7		E/I

(*continued*).

Table A2. Continued.

Reference	Planning activities				Information flow or input				Type of contribution
	Sales	Engineering	Procurement	Production	Strategic	External	Operational	C/F	
Cannas et al. (2019)		4.2.1, 4.2.2			En1, En2, En3				I
Johnsen and Hvam (2019)		4.2.2			Sa5, En1			En11	I
Reid, Bamford, and Ismail (2019)			4.3.2		Pc2				E
Zennaro et al. (2019)	4.1.3			4.4.2					E
Cannas et al. (2020)	4.1.1	4.2.1, 4.2.2, 4.2.4			Sa1, Sa2, Sa3, En1, En2				I
Fang and Wei (2020)		4.2.1			En1, En2, En3				I
Oluyisola, Sgarbossa, and Strandhagen (2020)		4.2.1			En3				I
Sabri, Micheli, and Cagno (2020)			4.3.1, 4.3.2		Pc1, Pc2, Pc3, Pc4		Pc8, Pc9		I
Shurrab, Jonsson, and Johansson (2020b)	4.1.1, 4.1.2	4.2.1, 4.2.4	4.3.2, 4.3.3		Sa2, Sa3, En4, Pc2, Pc3	Sa4, Sa5, En5	En8		E/I
Shurrab, Jonsson, and Johansson (2020a)	4.1.3	4.2.1, 4.2.3, 4.2.4	4.3.2, 4.3.3	4.4.1, 4.4.2, 4.4.3	Sa1, Sa2, En1, Pc1, Pd1, Pd2		Sa6, En8, En9, Pc8, Pc9, Pd7		E/I
Strandhagen et al. (2020)		4.2.1			En3				I
Alfnes et al. (2021)	4.1.3	4.2.1, 4.2.3, 4.2.4	4.3.1, 4.3.2, 4.3.3	4.4.1, 4.4.2	En1, En2, En4, Pc1	Sa5, En5, En6, Pc6	En8, En9, Pc8	Sa10, Sa11, Sa12, En11	E/I
Brachmann and Kolisch (2021)		4.2.3, 4.2.4			En4		En8, En9, En10		E
Burggraf et al. (2021)				4.4.2			Pd7		I
Eyers et al. (2021)				4.4.1	Pd1				I
Ghiyasinasab et al. (2021)		4.2.3, 4.2.4		4.4.2, 4.4.3	Pd2	Pd5	En8, En9, En10, Pd7, Pd9		E
Micale et al. (2021)				4.4.2	Pd2		Pd7		E
Mwesiumo, Nujen, and Kvadsheim (2021)			4.3.1, 4.3.2		Pc2				I
Zheng et al. (2021)		4.2.1			En3				I

Information inputs

Sales planning inputs – **Sa1**: Market segmentation; **Sa2**: Order-winning criteria; **Sa3**: Competitive priorities; **Sa4**: Sales leads, requests-for-proposals, customer enquiries, etc.; **Sa5**: Customers' technical and commercial requirements; **Sa6**: Customers' order history; **Sa7**: Requirement and specification data for completed orders; **Sa8**: Preliminary product specifications; **Sa9**: Required detailed design and engineering activities; **Sa10**: Estimated engineering lead time; **Sa11**: Procurement lead times and costs; **Sa12**: Production start and completion dates; **Sa13**: Overtime and subcontracting needs and production costs.

4.1.1: Selecting customer enquiries; **4.1.2**: Prioritising customer enquiries; **4.1.3**: Responding to customer enquiries; **4.2.1**: Defining preliminary product specifications; **4.2.2**: Determining detailed engineering activities and resources; **4.2.3**: Estimating lead times and costs and setting due dates; **4.2.4**: Identifying needs for external capabilities and additional capacity; **4.3.1**: Identifying critical items; **4.3.2**: Selecting potential suppliers; **4.3.3**: Determining procurement lead times and prices; **4.4.1**: Identifying the main production activities and resource requirements; **4.4.2**: Identifying feasible production start and end dates; **4.4.3**: Estimating production costs and non-regular capacity requirements.

Engineering planning inputs – **En1**: Product customisation and standardisation strategy; **En2**: Existing portfolio of product design and specifications; **En3**: New product technology; **En4**: In-house engineering capabilities; **En5**: Clarification of customer's technical requirements; **En6**: Feedback for preliminary features and specifications; **En7**: Durations and cost data from completed projects; **En8**: Capacity of engineering personnel; **En9**: Statuses of ongoing projects; **En10**: Costs of regular and overtime engineering capacity; **En11**: Customer requirements.

Procurement planning inputs – **Pc1**: Offshoring and sourcing strategy; **Pc2**: Approved supplier list; **Pc3**: Strategic suppliers; **Pc4**: Supplier agreements for lead times and prices; **Pc5**: Customer's preferred supplier(s); **Pc6**: Supplier's quoted lead time; **Pc7**: Supplier's quoted price; **Pc8**: Historical lead times for suppliers; **Pc9**: Historical purchasing costs for items; **Pc10**: Preliminary product specifications; **Pc11**: Required detailed design and engineering activities; **Pc12**: Estimated engineering lead time.

Production planning inputs – **Pd1**: Production technology; **Pd2**: Production capacity; **Pd3**: Production facility locations; **Pd4**: Vertical integration strategy; **Pd5**: Subcontracting costs; **Pd6**: Existing bill-of-materials and routing for similar products; **Pd7**: Production status of ongoing projects; **Pd8**: Production cost data for completed projects; **Pd9**: Regular and overtime production capacity costs; **Pd10**: Preliminary product specifications; **Pd11**: Estimated engineering lead time; **Pd12**: Procurement lead times.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model with Convolution Network

Le Hoang Son[1,2], Akshi Kumar[3], Saurabh Raj Sangwan[3], Anshika Arora[3], Anand Nayyar[4], Mohamed Abdel-Basset[5*]

[1] Division of Data Science, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam.
[2] Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam.
Email: lehoangson@tdtu.edu.vn
[3] Department of Computer Science & Engineering, Delhi Technological University, Delhi, India.
Email: akshikumar@dce.ac.in, saurabhsangwan2610@gmail.com, anshika.arora10nov@gmail.com
[4] Graduate School, Duy Tan University, Viet Nam.
Email: anandnayyar@duytan.edu.vn
[5] Department of operations research, Faculty of computers and informatics, Zagazig University, Egypt.
Email: analyst_mohamed@zu.edu.eg
Corresponding author: Mohamed Abdel-Basset (e-mail: analyst_mohamed@zu.edu.eg).

ABSTRACT A large community of research has been developed in the recent years to analyze social media and social networks, with the aim of understanding, discovering insights and exploiting the available information. The focus has shifted from conventional polarity classification to contemporary application-oriented fine-grained aspects such as, emotions, sarcasm, stance, rumor and hate speech detection in the user-generated content. Detecting a sarcastic tone in natural language hinders the performance of sentiment analysis tasks. The majority of the studies on automatic sarcasm detection emphasize on the use of lexical, syntactic, or pragmatic features that are often unequivocally expressed through figurative literary devices such as words, emoticons, and exclamation marks. In this research, we propose a deep learning model called *s*Att-BLSTM convNet that is based on the hybrid of *soft* attention-based bidirectional long short-term memory (*s*Att-BLSTM) and convolution neural network (convNet) applying global vectors for word representation (GLoVe) for building semantic word embeddings. In addition to the feature maps generated by the *s*Att-BLSTM, punctuation-based auxiliary features are also merged into the convNet. The robustness of the proposed model is investigated using balanced (tweets from benchmark SemEval 2015 Task 11) and unbalanced (approximately 40000 random tweets using the Sarcasm Detector tool with 15000 sarcastic and 25000 non-sarcastic messages) datasets. An experimental study using the training- and test-set accuracy metrics is performed to compare the proposed deep neural model with convNet, LSTM, and bidirectional LSTM with/without attention and it is observed that the novel *s*Att-BLSTM convNet model outperforms others with a superior sarcasm-classification accuracy of 97.87% for the Twitter dataset and 93.71% for the random-tweet dataset.

INDEX TERMS Sarcasm, Deep Learning, Attention, Social data.

I. INTRODUCTION

The biggest challenge in sentiment analysis tasks is to accurately determine the veracity of the statement in literal sense so as to classify text on the basis of polarity (positive or negative). Sentiment analysis achieves decent results in the case of literal language as it conveys the expected interpretation. However, the use of figurative language that is inherently emblematic represents something other than the concrete meaning, thereby making sentiment analysis a non-trivial problem. "Sarcasm is defined as a specific type of sentiment where people express their negative feelings using positive or intensified positive words in the text" [1]. It is a rhetoric narrative that presents a communicatively significant discord between the actual situation and the utterance content. For example, a tweet/post, "*It is wonderful feeling to waste hours in traffic jams!"* clearly indicates this discord between the actual situation of "being stuck in traffic jam" and the utterance content "wonderful." This contrast and shift of sentiments in sarcastic expressions validates sarcasm as a special instance of sentiment analysis. Therefore, the

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

detection of sarcastic expressions will enhance the automatic sentiment analysis of huge and diverse social web data.

Automatic sarcasm detection is typically a text-classification problem, which relies on a variety of feature-extraction and learning techniques. In general, the Twitter post size characterizes short text with a character limit of 280, which is insufficient in providing the desired word co-occurrence data. This data sparseness is challenging for conventional learning models based on statistical features, such as term frequency–inverse document frequency (TF-IDF) or co-occurrence. Deep learning architectures have proven capabilities when dealing with natural language, as they can learn by example, offer a flexible set of modeling alternatives while curtailing feature engineering, scale up to big data, exploit unlabeled data to detect trends and identify patterns too complex for humans to notice, and be used for multitask learning [2]. The **primary contribution** of this work includes:

- Improving the performance of the sentiment analysis task by building, training and evaluating a deep learning model, *s*Att-BLSTM convNet, a hybrid of *soft* attention-based bidirectional long short-term memory (*s*Att-BLSTM) and convolution neural network (convNet)
- Feature engineering for automatic sarcasm detection using figurative literary devices such as words, emoticons, and exclamation marks.
- Performance benchmarking by testing different kinds of models and datasets for sarcastic tone classification in real-time.

The proposed deep learning model, *s*Att-BLSTM convNet, has eight layers:

- *Input layer:* The input tweet to the model.
- *Embedding layer:* Each word within tweet is mapped into a low dimension vector using GloVe. This look-up encodes the input into real-valued vectors called embedding vectors.
- *BLSTM layer:* The output of the word-embedding layer is fed into the BLSTM layer. The purpose of this layer is to learn high-level features from the previous step.
- *Attention layer:* A *soft* attention mechanism (based on *softmax*) is used, which generates semantic context vectors. These are merged with the BLSTM output features and auxiliary features to train the convNet.
- *Convolution Layer:* The convolution operation is performed to obtain a convolved feature vector.
- *Activation Layer:* A **Re**ctified **L**inear **U**nit (ReLU) activation function is applied on the feature vectors (output of convNet layer).
- *Down-sampling Layer:* The max-pooling operation is used as a down-sampling strategy in convNets.
- *Representation layer:* A fully connected layer that consists of a linear transformation layer and *softmax* activation function to generate the output predictions.

To build word embeddings, GloVe [3], which generates a word vector table, is used. GloVe is a count-based model of representing words by feature vectors. This model maps all the tokenized words in each tweet to its respective word vector table. Proper padding is performed to unify the feature vector matrix. This matrix is given as an input to the bidirectional long-short-term-memory (BLSTM) layer. The BLSTM networks set up an additional layer to the unidirectional LSTM networks, where the hidden-to-hidden connections flow in opposite temporal order. The network thus exploits information from the past and the future. The output vector representation from the BLSTM is transformed into an attention-weighted sum of these output vectors. This word-level attention hypothesis relies on valuable words within text that contribute to the meaning in context, and thus computes attention probabilities over the input stimuli from the BLSTM. Consequently, the output feature representations from BLSTM are concatenated together as the output of the attention layer, which, combined with auxiliary features, is provided as input to the convNet. The convNet consists of multiple filters with a variable-size window connected to a small part of the input to generate a convolved feature map.

Several filters are used to extract features in a convolution layer and each filter gives an activation map. The next layer is the activation layer, or the ReLU layer, which introduces nonlinearity to the network and generates a rectified feature map. This rectified feature map is fed to the pooling layer to reduce the dimensionality of the feature map. A *k-max* pooling operation is employed, which selects the top *k* features with respect to the various hidden layers and generates a pooled feature map. This pooled feature map is input to the fully connected *softmax* layer, which calculates the probability of any output word and classifies the tweet as sarcastic or non-sarcastic.

- Basic sentiment analysis tools have difficulty distinguishing between positive and negative statements, especially when sarcasm, irony, and mixed feelings are included discretely or in combination. Practical models to predict sarcastic text are thus imperative, and the proposed model combines semantics from the attention-based bidirectional LSTM network with auxiliary pragmatic features to a deep convolution network for enhanced prediction performance.
- The multiple layers involve nonlinearity that aids learning complex representations.

The rest of the paper is structured as follows. Section 2 briefly discusses the pertinent work within the domain of sentiment analysis and deep learning based automatic sarcasm detection followed by Section 3, which illustrates the proposed *s*Att-BLSTM convNet model and its working details. Section 4 explicates the results and provides an

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

analysis of the same. The final section, Section 5, concludes the study and expounds upon the scope of future work.

II. RELATED WORK

The growing pursuit of Internet users in all forms of social media has intensified researchers' interest to intelligently mine the content available, both quantitatively and qualitatively. Twitter, currently the most famous micro-blog, connects people across the globe and has a high level of user involvement. It has gradually emerged as a huge source of sentiment-rich data [4]. The keyword "sentiment analysis" was initially witnessed in the published work [5] in 2003, and since then, both primary [4, 6, 7] and secondary studies have been reported across pertinent literature [8-11]. Furthermore, the literature is well-equipped with studies pertaining to sentiment analysis using machine learning paradigms on specifically textual user-generated online content on social media. Aloufi et al. [12] proposed a model for sentiment analysis of football specific tweets using three classifiers, namely, support vector machine, multinomial Navies Bayes and random forest. Pi et al. [13] put forward a model for prediction of vehicle sales by sentiment analysis of twitter data and stock market values using least squares support vector regression. Research using deep learning models for sentiment analysis has also been reported. Tseng et al. [14] analyzed textual opinions found in teaching evaluation questionnaires and applied the analysis results to assisting the selection of outstanding teaching faculty members using attention-based LSTM. Wu et al. [15] proposed a model having quadratic connections of LSTM capable of capturing complex semantic representations of natural language texts and evaluated on the benchmark dataset, the Stanford Sentiment Treebank. Bouazizi et al. [6] extended the concept of binary or ternary classification and proposed an approach to classify text collected from Twitter into seven sentiment classes. The authors further proposed [16] multi-class sentiment analysis which addresses the identification of the exact sentiment conveyed by the user using the task of quantification that identifies all the existing sentiments within a tweet instead of attributing a single sentiment label to it. Zhang et al. [17] proposed a model combining convolution neural network and bidirectional gated recurrent unit for sentence semantic classification. Han et al. [18] used CNN and BLSTM to design four kinds of memory network models for text sentiment classification. Fu et al. [19] proposed a lexicon-enhanced LSTM model that uses sentiment lexicon as extra information pre-training a word sentiment classifier. Wang et al. [20] proposed a sophisticated algorithm based on deep learning and information geometry for sentiment classification.

Significant studies exemplify sarcasm as a contradiction in polarity [21, 22]. Sarcasm detection has been reported as a non-trivial challenge for sentiment analysis across literature studies. There have been expansive efforts by researchers to investigate sarcasm-detection techniques on various social media channels as a sub-task of sentiment analysis [23]. Bouazizi et al. [24] propose a pattern-based approach to detect sarcasm on Twitter considering four sets of features that cover the different types of sarcasms.

Architectures based on deep learning techniques have gained popularity in natural language programming applications as they are capable of handling data sparseness in unbalanced datasets. Studies on bullying and verbal aggression have been reported deep neural network for short-text sentiment analysis [25]. A few such approaches have been reported for automatic sarcasm detection as well. The authors of [26] proposed a DeepMoji model, which is based on the occurrences of emoji, for detecting the emotional content on Twitter. It uses a variant of LSTM, a 6-layer model that is a hybrid of BLSTM and the attention mechanism for the detection of sarcastic tweets. A new convolution network-based architecture is presented in [27] that learn user embeddings in addition to utterance-based embeddings, which allows the authors to learn user-specific context. The similarity between word embeddings is used by the authors of [28] as features for sarcasm detection. The authors have concatenated features from previous works to these word embedding-based features. The authors of [29] have built a model combining a convNet with a recurrent neural network (RNN, long short-term memory) followed by a deep neural network. It is observed that this approach shows an improvement for the deep learning architecture when compared with recursive SVM. In [30], models based on a pre-trained convNet have been developed that are used to extract emotion, sentiment, and even personality features for sarcasm detection. In another study, [31], bidirectional LSTM is used to capture syntactic and semantic information over tweets, along with a pooling neural network to extract contextual features automatically from history tweets. Performance using neural features is compared with discrete manual features and an improvement is observed. Attention mechanism has been previously used by researchers in the domain of natural language understanding as it selectively focuses on information-rich phrases within the sequence of sentence. Recently, authors Seo at.al [32] and Wang et.al [33] demonstrated the effective use of attention for compression and question-answering.

III. AUTOMATED SARCASM DETECTION

The proposed framework for automatic sarcasm detection has four key components, namely, (*A*) data acquisition, (*B*) pre-processing, (*C*) implementation of the *s*Att-BLSTM convNet model, and (*D*) evaluation using performance measures. These are explained in detail in the following subsections.

A. DATASET ACQUISITION

To evaluate the performance of the proposed model, the following datasets have been used for empirical analysis. The purpose of using two varied datasets was to further validate

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

the results achieved and determine the robustness of the selected features and the model trained on these features.

1) BALANCED DATASET: SEMEVAL 2015 TASK 11

The original SemEval dataset consists of 20000 hashtag-annotated tweets, which were created for the Twitter sentiment analysis task. SemEval 2015 task 11 [34] defined the challenge as a "dataset for sentiment analysis of figurative language on Twitter." The objective of the task is specifically to evaluate the use of figurative language, where the word or phrase described as "figurative" meant that it is being used in a way that was symbolic and represented something other than its concrete meaning. Although figurative language encompasses an extensive variety of literary tools and techniques, such as similes, emoticons, metaphor, personification, idiom, and hyperbole, this benchmark dataset contains tweets with specific figurative types, namely, sarcasm, irony, and metaphor. The sarcasm corpus is thus constructed by collecting self-annotated tweets with the hashtags #irony, #sarcasm, and #not. Although there is a subtle difference between the three, we consider all of them in the same category for performing this empirical analysis. Thus, this study is also indicative of how this generalization influences the obtained results. The experiments used the SemEval 2013 task 2 dataset as the negative class data [35]. Manual perusal of the two datasets revealed that both contain tweets from general topics, thus avoiding any bias in our experimentation. To collect data for the experiments, we tried to retrieve 15961 tweets using a list of the dataset's tweet ids provided, out of which 7994 are sarcastic and 7324 are non-sarcastic.

2) IMBLANCED DATASET: RANDOM TWEETS

In real time, the presence of sarcasm in tweets is less frequent. A randomly sampled dataset containing 15,000 sarcastic and 25,000 non-sarcastic tweets is built using the Sarcasm Detector tool[1].

B. PRE-PROCESSING

After the data acquisition, to intelligently mine the text, the tweets are pre-processed to clean and transform the data for feature extraction [36]. These include the following processes.

- URLs, mentions, hashtags, and numbers in the tweets are replaced with placeholders
- All tweets are tokenized by the Natural Language Toolkit (NLTK) [37]
- All tokens are sampled in lowercase
- All tweets are stemmed to the root word using Porter's Stemmer [38]
- All non-ASCII-English characters were removed, to limit the domain of data to the English language
- All words are replaced by pre-trained GloVe word vectors. Words that do not exist in the pre-trained embeddings are set to a zero vector

Several auxiliary features are used to enhance the classification model. These are primarily pragmatic markers concerned with language use. We consider the following five punctuation-based features that help understand the subtleties of sarcasm in text, as these indicate figurative text and symbolic clues within the text:

- Number of exclamation marks (!) in the tweet
- Number of question marks (?) in the tweet
- Number of periods (.) in the tweet
- Number of capital letters in the tweet
- Number of uses of "or" in the tweet

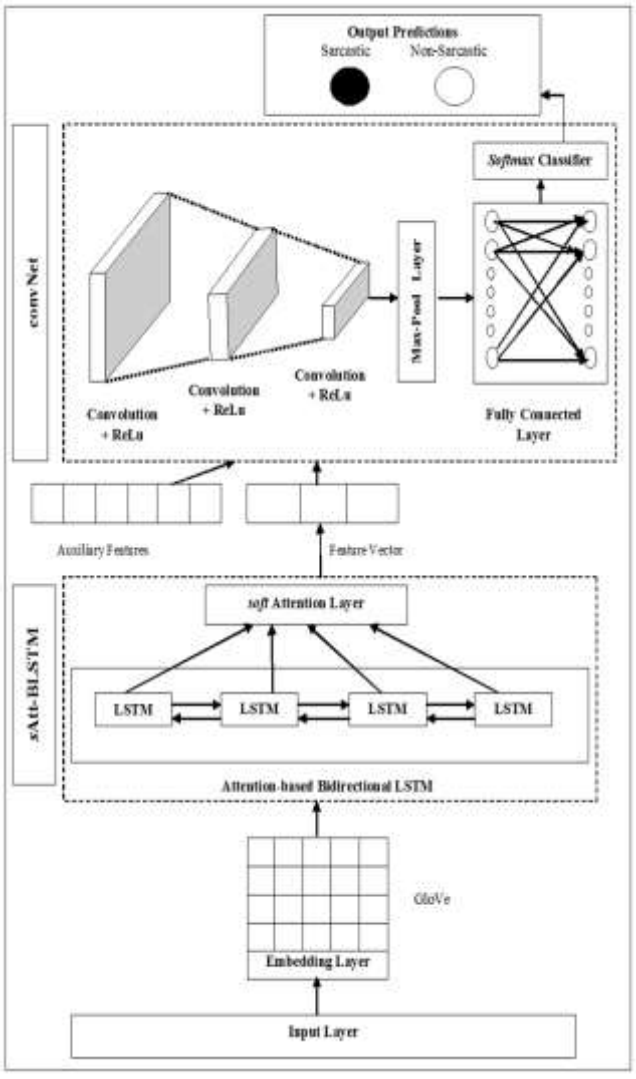


FIGURE 1. System architecture of the proposed sAtt-BLSTM ConvNet.

C. The PROPOSED sATT-BLSTM CONVNET MODEL

The proposed deep learning model uses 8 layers: the input layer, embedding layer, BLSTM layer, attention layer,

[1]www.thesarcasmdetector.com

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

convolution layer, activation and ReLU layer, max pooling layer, and representation layer. Fig. 1 depicts the architecture of the proposed multi-layer model. The details of each layer are provided in the subsequent subsections.

1) INPUT LAYER

The tweets after pre-processing are fed to the input layer. The input layer is connected to the embedding layer, which builds word embeddings using GloVe.

2) EMBEDDING LAYER

The embedding layer maps the input into real-valued vectors using encoding from look-up tables. Word embeddings facilitate learned word representations. The benefits of extracting features based on word embedding to detect sarcasm have been recently reported [28]. In this study, to build word embeddings, GloVe, which generates a word vector table, is used. GloVe is a count-based model of representing words by feature vectors. This log-bilinear model studies the relationship of words by counting the number of times they co-occur. Thus, this model aids in mapping all the tokenized words in each tweet to their respective word vector tables. Proper padding is performed to unify the feature vector matrix. That is, if the total number of given tweets is Z and there is a tweet X with t tokens, generation of a word vector Table V with dimension d of the word vectors is completed using GloVe. Thus, for all Z, each t in X is mapped to its respective V. After this mapping, each X is expressed as a vector of the word embeddings concatenation (E). Thus, the feature vector matrix is obtained as shown in (1).

$$F = U + B + T + P + E, \quad (1)$$

where + is the concatenation operator of the vector. The tweets are of varying length, so to unify the feature vector matrix representation of tweets, the tweets with the maximum length in the given corpus are used as a threshold value. This is done basically to fix the length of the tweet matrix. Hence, for all the tweets that were shorter than this threshold, zero padding was performed. This matrix was finally fed as input (i.e., F) to the BLSTM layer.

3) BLSTM LAYER

LSTM is an RNN that contains special units in the recurrent hidden layer called memory blocks [39]. Self-connections exist within the memory cells. There is an input, output, and forget gate for each memory cell. The hidden layer of LSTM is also called the LSTM cell. LSTM has the capability to plot long-term dependencies by defining each memory cell with a set of gates $\Re^d$, where d is the memory dimension of the hidden state of LSTM [28]. At each iteration t, the LSTM cell has the layer input x_t and the layer output h_t. LSTM contains three gates, which are functions of the current input x_t and hidden state h_{t-1}: input gate i_t, forget gate f_t, and output gate o_t. The cell also takes the cell input state s_t, the cell output state c_t, and the previous cell output state c_{t-1} into account while training and updating parameters.

The input gate, forget gate, output gate, and input cell state can be calculated using equations (2)–(7).

$$i_t = \sigma\left(w_{i_x} x_t + w_{i_h} h_{t-1} + b_i\right) \quad (2)$$

$$f_t = \sigma\left(w_{f_x} x_t + w_{f_h} h_{t-1} + b_f\right) \quad (3)$$

$$o_t = \sigma\left(w_{o_x} x_t + w_{o_h} h_{t-1} + b_o\right) \quad (4)$$

$$s_t = tanh\left(w_{s_x} x_t + w_{s_h} h_{t-1} + b_s\right) \quad (5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot s_t \quad (6)$$

$$h_t = tanh\left(c_t\right) \odot o_t \quad (7)$$

In equations (2)–(7),

- $\odot$ is the element-wise product;
- $w_i\ w_f w_o\ w_s$ are weighing factors used for mapping the hidden layer input to the three gates and the input cell state;
- $b_i b_f b_o b_s$ are bias vectors;
- σ is the gate activation function, which is normally the sigmoid function;
- tanh is the hyperbolic tangent function.

The final output of the LSTM layer is a vector of all the outputs, represented by (8).

$$Y_t = [h_{t-n}, \ldots\ldots, h_{t-1}] \quad (8)$$

The bidirectional LSTM connects two hidden layers to a single output layer. The forward-layer output sequence $h^{\rightarrow}$ is iteratively calculated using inputs in a positive sequence from time t-n to time t-1, while the backward-layer output sequence, $h^{\leftrightarrow}$ is calculated using the reversed inputs from time t-n to t-1 [40]. Both the forward- and backward-layer outputs are calculated by using the standard LSTM updating equations. This creates two copies of the hidden layer, one fit in the input sequences as is and one as a reversed copy of the input sequence [40]. These representations are concatenated together using the attention mechanism. That is, the BLSTM layer generates an output vector in which each element is calculated by using (9).

$$y_t = \sigma(h^{\rightarrow} t, h^{\leftrightarrow} t) \quad (9)$$

where σ is a soft attention function to combine the two output sequences. By this means, our model can capture as many salient words from both directions as possible. Thus, similar to that of the LSTM layer, the final output of the BLSTM layer can be represented by a vector (10).

$$Y_t = [y_{t-n}, \ldots\ldots, y_{t-1}] \quad (10)$$

4) ATTENTION LAYER

The idea of attention [41] has originated by the observation that each word contributes to the meaning of tweet (sentence)

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

differently that is; all words do not contribute equally to the representation of the sentence meaning. The users reading text can intuitively identify important, meaningful parts of the sentence. The attention mechanism adds an additional structure to the network, which is learned as part of the standard model training via back-propagation. A deterministic, differentiable attention mechanism, known as soft attention [42], is used, where the weights on each location are usually given by a *softmax*, and the output of the attention module is a weighted sum of representations at each location. This whole process can be described by a differentiable function, so that training can be performed jointly with the rest of the network using back-propagation.

We use a word-level attention mechanism to focus on the words that have a closer semantic relationship to the sentence meaning. That is, the word-level attention is applied to the feature representations of BLSTM. The output is thus a weighted combination of all the input states, not just the last state. The attention mechanism gives an attention score $e_{i,t}$ to each word t in the sentence i, as given by (11).

$$e_{i,t} = g(Wh_t, c), \tag{11}$$

where g is an activation function.

Then the weight probability $a_{i,t}$ of each h_t is computed by (12).

$$a_{i,t} = \frac{\exp(e_{i,t})}{\sum_{k=1}^{T} \exp(e_{i,k})} \tag{12}$$

- Auxilary Features

Pragmatic features, such as punctuation or repetition of words, aid in understanding the subtleties of sarcasm in text, as these indicate figurative text and symbolic clues within the text. Punctuation is considered ideal for the representation of syntactic features, as the punctuation features contribute to a substantial boost in classification quality [42]. These can be exemplified by the use of exclamation points/marks, repeated letters (especially vowels), or uppercase letters (ALL CAPS) in textual messages such as tweets/posts. A few examples where these punctuation features impact the emotion quotient in text are given below:

- An exclamation point or exclamation mark (!) intensifies the emotional content preserving the semantic orientation [43]. For example: *"The location is great!!!!"* is more intense than *"The location is great."*
- Repeated letters (especially vowels) added to elongate the spelling of a word (e.g., "haaaaaapppyyy") is used to emphasize the word or to express enthusiasm in computer-mediated communication [44].
- Capitalization, such as explicitly using ALL CAPS to emphasize a sentiment-relevant word with other non-capitalized words in the document, increases the intensity of the sentiment preserving semantic orientation [45]. For example, "Simply LOVED the place!" is more intense than "Simply loved the place!"

Thus, sarcasm detection can be made easy by observing the syntactic pattern of tweet to identify punctuation such as:

- Number of contiguous question marks, exclamation points, and combinations thereof
- Whether the last token contains a question mark or exclamation point
- Occurrence of punctuation sequences such as "!!!" and "?!"

In this study, the set of punctuation-based features shown in Table1 are extracted for each tweet.

TABLE I
PUNCTUATION-BASED FEATURES USED

excl	No. of exclamation marks in the post/tweet
ques	No. of question marks in the post/ tweet
dots	No. of periods in the post/tweet
caps	No. of capital letters in the post/tweet
quotes	No. of quotation marks (‘’ or “”) in the post/tweet

5) CONVOLUTION LAYER

ConvNet has two primary parts, namely, feature learning (Conv, ReLU, and Pool) and classification (fully connected and softmax). The output features of the sAtt BLSTM are combined with auxiliary features and provided as input to the convNet. Each convolving filter has a variable window size (s), height, and width. This window slides over to the input matrix and calculates localized feature vector (v_j) for each possible word (window size) using a nonlinear activation function (f) and a bias (b_a).

A transition matrix (T) is generated for each filter. This filter comprises of a layer of weights (weight matrix, W) that are applied to parts of the input matrix to obtain a single unit as the output. This is performed for the bias and weight during the network-learning adjustment, as illustrated in (13).

$$v_j = f\left(T.F_{j:j+s-1} + b_a \right) \tag{13}$$

Here, $F_{j:j+s-1}$shows the local vector from position j to j+s-1 in F. Hence, formation of this v_j together with the completion of the convolution operation generates a new vector, n, as in (14).

$$n = [v_1, v_2, v_3, \ldots, v_{d-s+1}] \tag{14}$$

After the sliding operation, the output from each part is then combined in order to obtain the resultant output, called the transformed feature map, that serves as the input to the next layer: the ReLU Layer.

6) RELU LAYER

The activation or ReLU Layer [46] is applied for dealing with the nonlinearity in the convNet model. It generates

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

a rectified feature map, which is fed to the pooling layer to reduce the dimensionality of the feature map.

7) POOLING LAYER

In this work, max pooling [47] is used, which takes the largest element from the rectified feature map. It reduces the dimensionality of the feature map and helps retain the most significant features. The k-max pooling operation (p) is performed for feature selection that chooses the top k features conforming to various hidden layers, where p is expressed as (15).

$$p = \max\{v_1, v_2, v_3, \ldots, v_{d-s+1}\} \quad (15)$$

8) REPRESNTATION LAYER

The output layer is a fully connected layer that consists of the softmax activation function. The pooled feature map is input to the *fully connected* softmax layer, which calculates the probability of any output word and classifies the tweet as sarcastic or non-sarcastic as an output. The output vector of the *softmax* layer ($O^{(z)}$) is as (16).

$$O^{(z)} = T^{(z)} O^{(z-1)} + b_a{}^{(z)}, \quad (16)$$

where

$O^{(z-1)}$ denotes the output vector the pooling layer

$T^{(z)}$ is the transition matrix

$b_a{}^{(z)}$ represents the bias factor of softmax layer.

Thus, using a fully connected *softmax* layer yields the probability-distribution [48] of the sarcasm detection (P_d), which is denoted using (17).

$$P_d(j|X, \theta) = \frac{[\exp(O_{j[z]})]}{[\sum_{k=1}^{n} \exp(O_{j[z]})]}, \quad (17)$$

where θ is the parameter of the model and j is the class label.

D. RESULTS AND ANALYSIS

For discussing the results, the empirical analysis has been broadly divided into two parts: (i) parameter setting for the proposed model and (ii) comparison with multiple baselines on the basis of classification accuracy.

1) PARAMETER SETTING

Optimal selection of parameters is imperative to achieve superlative performance results. We use the validation data to tune the hyper-parameters so as to obtain the best results. Table 2 lists the values used in this work.

TABLE II
HYPERPARAMETER VALUES

Hyperparameter	Value
Dimension of GloVe vectors	200
Hidden units of LSTMs (Forward, Backward)	500 each
Mini-batch size	10
Number of convNet layers	3
Number of convolution filters	100
Window size	(3,3)
Activation function of convNet	ReLU
Regularization	Dropout Operation
Drop-out rate	0.5: word embedding; 0.2: BLSTM; 0.4: convNet
Learning rate	0.2

The weights of the network are randomly initialized from a Gaussian distribution with variance scaled to 0.01, and the biases are initialized to zero except the forget-gate bias, which is initialized to 1.

2) PERFORMANCE RESULT

The proposed model is evaluated to predict sarcasm in tweets using two datasets, namely SemEval 2015 task 11 and random tweets containing a total of 40,000 tweets. The results have been assessed using key performance indicators (accuracy, recall, precision, and F-measure [49, 50]). Table 3 lists the results of the proposed *s*Att-BLSTM convNet model implemented on the two datasets and Fig. 2 depicts the results graphically.

TABLE III
PERFORMANCE OF THE PROPOSED *S*ATT-BLSTM CONVNET MODEL

*s*Att-BLSTM convNet	SemEval Dataset	Random Tweets
Accuracy	97.87	93.71
Recall	96.83	92.67
Precision	92.14	90.49
F-measure	93.57	88.29

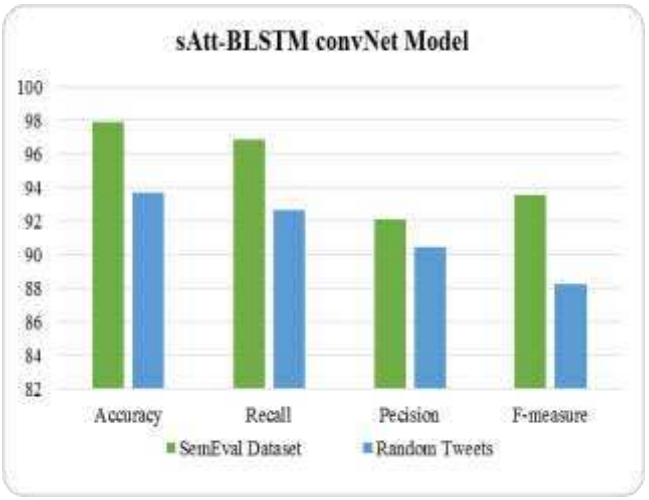


FIGURE 2. sAtt-BLSTM convNet Model Performance on two Datasets

3) COMPARISON WITH BASELINE DEEP LEARNING MODELS RESULT

We compare the results of the proposed model with three other deep learning architectures, namely, LSTM, BLSTM without attention, and convNet. The word embedding was performed using GloVe for each baseline model and the evaluation was made for both datasets using the four key performance indicators. The results obtained for LSTM,

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

BLSTM without attention, BLSTM with attention and convNet are shown in Tables 4, 5, 6 and 7, respectively.

TABLE IV
PERFORMANCE OF LSTM

LSTM model	SemEval Dataset	Random Tweets
Accuracy	84.89	79.75
Recall	83.51	78.01
Precision	86.78	74.98
F-measure	80.39	76.91

TABLE V
PERFORMANCE OF BLSTM WITHOUT ATTENTION

BLSTM without Attention	SemEval Dataset	Random Tweets
Accuracy	86.32	81.03
Recall	84.39	80.27
Precision	81.61	78.76
F-measure	85.25	79.45

TABLE VI
PERFORMANCE OF BLSTM WITH ATTENTION

BLSTM without Attention	SemEval Dataset	Random Tweets
Accuracy	89.03	85.65
Recall	85.51	82.34
Precision	87.76	85.55
F-measure	87.25	84.04

TABLE VII
PERFORMANCE OF CONVNET

ConvNet model	SemEval Dataset	Random Tweets
Accuracy	91.60	88.28
Recall	90.53	87.61
Precision	90.19	86.01
F-measure	88.57	84.72

It can be clearly observed that the proposed *s*Att- BLSTM convNet outperforms the other models with an accuracy of 91.60% achieved for the SemEval dataset and 88.28% for random tweets. LSTM shows the least accuracy of 84.89% and 79.75% for the SemEval and random tweet datasets, respectively. The models, in order from lowest to highest accuracy, are LSTM < BLSTM without Attention < BLSTM with Attention < convNet < *s*Att-BLSTM convNet. The best recall is also observed for the proposed *s*Att-BLSTM convNet for both the datasets. However, the convNet model demonstrates the best precision value of 92.19% for the SemEval dataset, whereas the proposed sAtt-BLSTM convNet model shows the best precision for the random tweets. Table 8 summarizes the comparison of the accuracy results obtained by the above five models. Fig. 3 graphically illustrates these comparative results.

TABLE VIII
ACCURACY COMPARISON OF MODELS

Dataset → / Models ↓	SemEval Dataset	Random Tweets
LSTM	84.89	79.75
BLSTM without Attention	86.32	81.03
BLSTM with Attention	89.03	85.65
convNet	91.60	88.28
*s*Att-BLSTM convNet	97.87	93.71

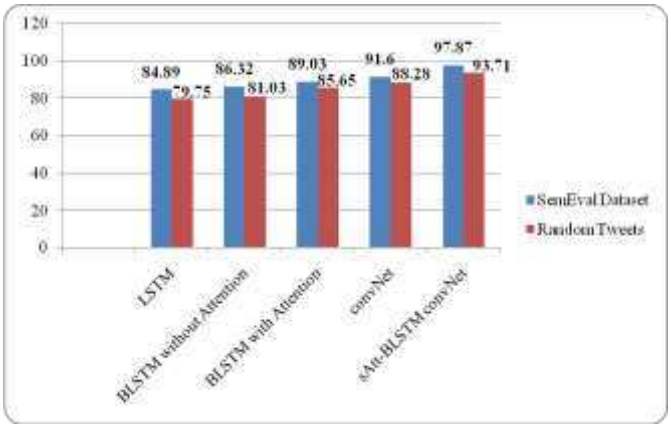


FIGURE 3. Accuracy Comparison of the Proposed Model with Baseline Models

IV. CONCLUSION

Social media is an informal means of communication that includes considerable use of slang, malformed words, short forms, colloquial expressions, mash-up words, etc. All these amplify the ambiguity, vagueness, and imprecision in the available social web content and makes the task of analyzing generic and specific sentiment computationally difficult. Sarcasm is a specific type of sentiment and its automatic detection is a dynamic area of research. In this work, we proposed *s*Att-BLSTM convNet, a hybrid of *soft* attention-based bidirectional long short-term memory (*s*Att-BLSTM) and convolution neural network (convNet), to detect sarcasm in short text (tweets). Semantic word embeddings and pragmatic auxiliary features were used to train the network. The proposed model demonstrates the highest classification accuracy for both the datasets as compared to the baseline models. The use of mash-up languages and novelty in vocabulary with advanced architectures [51-81] add to the challenges of automated detection of sarcasm and characterize some open problems for future research within the domain.

REFERENCES

[1] S.K. Bharti, B. Vachha, R.K. Pradhan, K.S. Babu, S.K. Jena. "Sarcastic sentiment detection in tweets streamed in real time: a big data approach", *Digital Communications and Networks.* vol. 2, no. 3, pp. 108-21, 2016.

[2] S.N. Sivanandam, S.N. Deepa, "Principles of Soft Computing (With CD)", *John Wiley & Sons.* 2007.

[3] J. Pennington, R. Socher, C. Manning, "Glove: Global vectors for word representation" *Proceedings of the 2014 conference on*

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

empirical methods in natural language processing (EMNLP), pp. 1532-1543, 2014.

[4] A. Kumar, A. Jaiswal, "Empirical Study of Twitter and Tumblr for Sentiment Analysis using Soft Computing Techniques", *Proceedings of the World Congress on Engineering and Computer Science,* pp.1: 1-5, 2017.

[5] K. Dave, S. Lawrence, D.M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", *Proceedings of the 12th international conference on World Wide Web. ACM,* pp.519-528, 2003.

[6] Bouazizi, M. and Ohtsuki, T., 2017. A pattern-based approach for multi-class sentiment analysis in twitter. IEEE Access, 5, pp.20617-20639

[7] Shayaa, S., Jaafar, N.I., Bahri, S., Sulaiman, A., Wai, P.S., Chung, Y.W., Piprani, A.Z. and Al-Garadi, M.A., 2018. Sentiment analysis of big data: Methods, applications, and open challenges. *IEEE Access*, *6*, pp.37807-37827.

[8] A. Kumar, T.M. Sebastian, "Sentiment Analysis: A perspective on its Past, Present and Future", *International Journal of Intelligent Systems and Applications,* vol. 4, no. 10, pp. 1-14, 2012.

[9] A. Kumar, T.M. Sebastian, "Sentiment analysis on Twitter*", IJCSI International Journal of Computer Science Issues* vol. 9, no. 3, pp. 372-378, 2012.

[10] Al-Moslmi, T., Omar, N., Abdullah, S. and Albared, M., 2017. Approaches to cross-domain sentiment analysis: a systematic literature review. *IEEE Access*, *5*, pp.16173-16192.

[11] Kumar A, Jaiswal A. Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency Computat Pract Exper. 2019;e5107. https://doi.org/10.1002/cpe.5107

[12] Aloufi, S. and El Saddik, A., 2018. Sentiment Identification in Football-Specific Tweets. *IEEE Access*, *6*, pp.78609-78621.

[13] Pai, P.F. and Liu, C.H., 2018. Predicting Vehicle Sales by Sentiment Analysis of Twitter Data and Stock Market Values. *IEEE Access*, *6*, pp.57655-57662.

[14] Tseng, C.W., Chou, J.J. and Tsai, Y.C., 2018. Text Mining Analysis of Teaching Evaluation Questionnaires for the Selection of Outstanding Teaching Faculty Members. *IEEE Access*, *6*, pp.72870-72879.

[15] Wu, D. and Chi, M., 2017. Long Short-Term Memory with Quadratic Connections in Recursive Neural Networks for Representing Compositional Semantics. IEEE Access, 5, pp.16077-16083.

[16] Bouazizi, M. and Ohtsuki, T., 2018. Multi-Class Sentiment Analysis in Twitter: What if Classification is not the Answer. IEEE Access, 6, pp.64486-64502.

[17] Zhang, D., Tian, L., Hong, M., Han, F., Ren, Y. and Chen, Y., 2018. Combining Convolution Neural Network and Bidirectional Gated Recurrent Unit for Sentence Semantic Classification. *IEEE Access*, *6*, pp.73750-73759.

[18] Han, H., Liu, J. and Liu, G., 2018. Attention-Based Memory Network for Text Sentiment Classification. *IEEE Access*, *6*, pp.68302-68310.

[19] Fu, X., Yang, J., Li, J., Fang, M. and Wang, H., 2018. Lexicon-Enhanced LSTM With Attention for General Sentiment Analysis. IEEE Access, 6, pp.71884-71891.

[20] Wang, M., Ning, Z.H., Xiao, C. and Li, T., 2018. Sentiment Classification Based on Information Geometry and Deep Belief Networks. IEEE Access, 6, pp.35206-35213.

[21] E. Camp, "Sarcasm, pretense, and the semantics/pragmatics distinction. *Noûs* vol. 46, no. 4, pp. 587-634 2012.

[22] E. Riloff, A. Qadir, P. Surve, L.D. Silva, N. Gilbert, R. Huang, "Sarcasm as contrast between a positive sentiment and negative situation", *In EMNLP* vol. 13, pp.704–714, 2013.

[23] E. Cambria, S. Poria, F. Bisio, R. Bajpai, I. Chaturvedi, "The CLSA model: A novel framework for concept-level sentiment analysis", *International Conference on Intelligent Text Processing and Computational Linguistics, Springer,* pp. 3–22, 2015.

[24] Bouazizi, M. and Ohtsuki, T.O., 2016. A pattern-based approach for sarcasm detection on Twitter. *IEEE Access*, *4*, pp.5477-5488.

[25] Junyi, Chen & Yan, Shankai & Wong, Ka-Chun., "Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis". *Neural Computing and Applications,* 2018.

[26] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm", *arXiv preprint arXiv:1708.00524.* 2017.

[27] S. Amir, B.C. Wallace, H. Lyu, P.C.M.J. Silva, "Modelling context with user embeddings for sarcasm detection in social media", *arXiv preprint arXiv:1607.00976* 2016.

[28] A. Joshi, P. Bhattacharyya, M. Carman, J. Saraswati, R. Shukla, "How do cultural differences impact the quality of sarcasm annotation?: A case study of indian annotators and american text.", In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* pp. 95-99, 2016.

[29] A. Ghosh, T. Veale, "Fracking sarcasm using neural network", In Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp. 161-169, 2016.

[30] Felix A. Gers , Jürgen Schmidhuber and Fred Cummins, "Learning to Forget: Continual Prediction with LSTM",1999 *Ninth International Conference on Artificial Neural Networks ICANN 99*. (Conf. Publ. No. 470)

[31] M. Zhang, Y. Zhang, G. Fu, "Tweet sarcasm detection using deep neural network", In Proceedings of COLING 2016, The 26th International Conference on Computational Linguistics: Technical Papers, pp. 2449-2460, 2016.

[32] Seo, M., Kembhavi, A., Farhadi, A., Hajishirzi. H. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603, 2016.*

[33] Wang, W., Yang, N., Wei, F., Chang, B., and Zhou, M. R-NET: Machine Reading Comprehension With Self-Matching Networks, *Technical Report, Natural Language Computing Group, Microsoft Research Asia*, 2017.

[34] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, A. Reyes, "Semeval-2015 task 11: Sentiment analysis of figurative language in twitte", *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* pp. 470-478, 2015.

[35] J. Hltcoe , "Semeval-2013 task 2: Sentiment analysis in Twitter", *Atlanta, Georgia, USA*, 312. 2013.

[36] Guyon, I. and Elisseeff, A., 2006. An introduction to feature extraction. In *Feature extraction* (pp. 1-25). Springer, Berlin, Heidelberg.

[37] E. Loper, S. Bird, "NLTK: The natural language toolkit", Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics, Association for Computational Linguistics vol. 1, pp. 63-70, 2002.

[38] M.F. Porter, 1980, "An algorithm for suffix stripping", Program, 14(3) pp 130−137. Available : https://tartarus.org/martin/PorterStemmer/

[39] H. Salehinejad, J. Baarbe, S. Sankar, J. Barfett, E. Colak, S. Valaee, "Recent Advances in Recurrent Neural Networks", *arXiv preprint arXiv:1801.01078*. 2017.

[40] A. Graves, N. Jaitly, A.R. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM" *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop* pp. 273-278, 2013.

[41] Luong, M.T., Pham, H. and Manning, C.D., 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.

[42] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R. and Bengio, Y., 2015, June. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning* (pp. 2048-2057).

[43] D. Davidov, O. Tsur, A. Rappoport, "Enhanced sentiment learning using twitter hashtags and smileys", Proceedings of the 23rd international conference on computational linguistics:

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

posters, Association for Computational Linguistics pp. 241-249, 2010.

[44] M. Thelwall, K. Buckley, "Topic-based sentiment analysis for the social *web*: The role of mood and issue-related words", *Journal of the American Society for Information Science and Technology,* vol. 64, no.8, pp. 1608-1617, 2013.

[45] C.H.E. Gilbert, A. Vader, "Parsimonious rule-based model for sentiment analysis of social media text*", Eighth International Conference on Weblogs and Social Media (ICWSM-14)* 2014.

[46] Li, Y. and Yuan, Y., 2017. "Convergence analysis of two-layer neural networks with relu activation". In *Advances in Neural Information Processing Systems* (pp. 597-607).

[47] Kim, Y., 2014. "Convolutional neural networks for sentence classification". *arXiv preprint arXiv:1408.5882*.

[48] Jianqiang, Z. and Xiaolin, G., 2018. "Deep Convolution Neural Networks for Twitter Sentiment Analysis". *IEEE Access* (pp. 2169-3536).

[49] Powers, D.M., 2. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", *Journal of Machine Learning Technologies*, 2011

[50] Bhatia MPS., Kumar, A. "A primer on the web information retrieval paradigm", *Journal of Theoretical & Applied Information Technology* 4 (7), 2008.

[51] Hemanth, J., Anitha, J., Naaji, A., Geman, O., Popescu, D., Son, LH. (2018). A Modified Deep Convolutional Neural Network for Abnormal Brain Image Classification. IEEE Access, 7(1), 4275-4283.

[52] Son, LH., Fujita, H. (2019). Neural-fuzzy with representative sets for prediction of student performance. Applied Intelligence, 49 (1), 172-187.

[53] Hemanth, D. J., Anitha, J., Son, LH., & Mittal, M. (2018). Diabetic Retinopathy Diagnosis from Retinal Images Using Modified Hopfield Neural Network. Journal of medical systems, 42(12), 247.

[54] Hemanth, D. J., Anitha, J., Son, LH. (2018). Brain signal based human emotion analysis by circular back propagation and Deep Kohonen Neural Networks. Computers & Electrical Engineering, 68, 170-180.

[55] Giap, CN, Son, LH, Chiclana, F. (2018). Dynamic Structural Neural Network. Journal of Intelligent & Fuzzy Systems 34, 2479 – 2490.

[56] Doss, S., Nayyar, A., Suseendran, G., Tanwar, S., Khanna, A., Son, LH & Thong, P. H. (2018). APD-JFAD: Accurate Prevention and Detection of Jelly Fish Attack in MANET. IEEE ACCESS, 6, 56954-56965.

[57] Hai, D. T., Son, H. Vinh, LT. (2017). Novel fuzzy clustering scheme for 3D wireless sensor networks. Applied Soft Computing, 54, 141-149.

[58] Kapoor, R., Gupta, R., Kumar, R., Son, LH & Jha, S. (2019). New scheme for underwater acoustically wireless transmission using direct sequence code division multiple access in MIMO systems. Wireless Networks, 1-13. DOI: 10.1007/s11276-018-1750-z.

[59] Kapoor, R., Gupta, R., Son, LH, Jha, S., & Kumar, R. (2018). Detection of power quality event using histogram of oriented gradients and support vector machine. Measurement, 120, 52-75.

[60] Kapoor, R., Gupta, R., Son, LH, Jha, S., & Kumar, R. (2018). Boosting performance of power quality event identification with KL Divergence measure and standard deviation, Measurement, 126, 134-142.

[61] Long, H. V., Ali, M., Khan, M., & Tu, D. N. (2019). A novel approach for fuzzy clustering based on neutrosophic association matrix. Computers & Industrial Engineering. DOI: 10.1016/j.cie.2018.11.007.

[62] Phuong, P. T. M., Thong, P. H., Son, LH. (2018). Theoretical analysis of picture fuzzy clustering: Convergence and property. Journal of Computer Science and Cybernetics, 34(1), 17-32.

[63] Robinson, Y. H., Julie, E. G., Saravanan, K., Kumar, R., Son, LH (2019). FD-AOMDV: fault-tolerant disjoint ad-hoc on-demand multipath distance vector routing algorithm in mobile ad-hoc networks. Journal of Ambient Intelligence and Humanized Computing, 1-18. DOI: 10.1007/s12652-018-1126-3.

[64] Saravanan, K., Anusuya, E., Kumar, R., Son, LH (2018). Real-time water quality monitoring using Internet of Things in SCADA. Environmental monitoring and assessment, 190(9), 556.

[65] Saravanan, K., Aswini, S., Kumar, R., Son, LH (2019). How to prevent maritime border collision for fisheries?-A design of Real-Time Automatic Identification System. Earth Science Informatics, 1-12. DOI: 10.1007/s12145-018-0371-5.

[66] Singh, K., Singh, K., Son, LH, Aziz, A. (2018) Congestion control in wireless sensor networks by hybrid multi-objective optimization algorithm, Computer Networks, 138, 90-107.

[67] Singh, N, Son, LH, Chiclana, F, Jean-Pierre, M (2019). A New Fusion of Salp Swarm with Sine Cosine for Optimization of Non-linear Functions. Engineering with Computers. DOI: 10.1007/s00366-018-00696-8.

[68] Son, L. H. (2015). A novel kernel fuzzy clustering algorithm for geo-demographic analysis. Information Sciences—Informatics and Computer Science, Intelligent Systems, Applications: An International Journal, 317(C), 202-223.

[69] Son, L. H. (2016). Generalized picture distance measure and applications to picture fuzzy clustering. Applied Soft Computing, 46(C), 284-295.

[70] Son, LH, Hai, PV. (2016). A novel multiple fuzzy clustering method based on internal clustering validation measures with gradient descent. International Journal of Fuzzy Systems, 18(5), 894-903.

[71] Son, LH, Jha, S., Kumar, R., Chatterjee, J. M., & Khari, M. (2019). Collaborative handshaking approaches between internet of computing and internet of things towards a smart world: a review from 2009–2017. Telecommunication Systems, 1-18. DOI: 10.1007/s11235-018-0481-x.

[72] Son, LH, Tien, N. D. (2017). Tune up fuzzy C-means for big data: some novel hybrid clustering algorithms based on initial selection and incremental clustering. International Journal of Fuzzy Systems, 19(5), 1585-1602.

[73] Son, LH, Tuan, T. M. (2016). A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation. Expert Systems with Applications, 46, 380-393.

[74] Son, LH., Thong, P. H. (2017). Some novel hybrid forecast methods based on picture fuzzy clustering for weather nowcasting from satellite image sequences. Applied Intelligence, 46(1), 1-15.

[75] Son, LH., Tuan, T. M. (2017). Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints. Engineering Applications of Artificial Intelligence, 59, 186-195.

[76] Tam, N. T., Hai, D. T., Son, LH, Vinh, LT. (2018). Improving lifetime and network connections of 3D wireless sensor networks based on fuzzy clustering and particle swarm optimization. Wireless Networks, 24(5), 1477-1490.

[77] Thanh, N. D., Ali, M., Son, LH. (2017). A novel clustering algorithm in a neutrosophic recommender system for medical diagnosis. Cognitive Computation, 9(4), 526-544.

[78] Thong, P. H., Son, LH (2016a). Picture fuzzy clustering: a new computational intelligence method. Soft Computing, 20(9), 3549-3562.

[79] Thong, P. H., Son, LH. (2016b). A novel automatic picture fuzzy clustering method based on particle swarm optimization and picture composite cardinality. Knowledge-Based Systems, 109, 48-60.

[80] Thong, P. H., Son, LH. (2016c). Picture fuzzy clustering for complex data. Engineering Applications of Artificial Intelligence, 56, 121-130.

[81] Tuan, T. M., Ngan, T. T., Son, LH. (2016). A novel semi-supervised fuzzy clustering method based on interactive fuzzy satisficing for dental X-ray image segmentation. Applied Intelligence, 45(2), 402-428.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2899260, IEEE Access

LE HOANG SON obtained the PhD degree on Mathematics – Informatics at VNU University of Science, Vietnam National University (VNU) in 2013. He has been promoted to Associate Professor in Information Technology since 2017. Dr. Son worked as senior researcher and Vice Director at the Center for High Performance Computing, VNU University of Science, Vietnam National University during 2007 - 2018. From August 2018, he is Head of Department of Multimedia and Virtual Reality, VNU Information Technology Institute, VNU. He is also Visiting Professor at Ton Duc Thang University, Vietnam. His major fields include Artificial Intelligence, Data Mining, Soft Computing, Fuzzy Computing, Fuzzy Recommender Systems, and Geographic Information System.

AKSHI KUMAR **is** an Assistant Professor in the Department of Computer Science & Engineering at Delhi Technological University (formerly Delhi College of Engineering). She has been with the university for the past 10 years. She has received her doctorate degree in Computer Engineering in the area of Web Mining, from Faculty of Technology, University of Delhi in 2011. She completed her M.Tech (Master of Technology) with honours in Computer Science & Engineering from Guru Gobind Singh Indraprastha University, Delhi in 2005. She received her BE (Bachelor of Engineering) degree with distinction in Computer Science & Engineering from Maharishi Dayanand University, Rohtak in 2003. Dr. Kumar's research interests are in the area of intelligent systems, user-generated big-data, social media analytics and soft computing.

SAURABH RAJ SANGWAN is a research scholar at the Web Research Group in the Department of Computer Science & Engineering at Delhi Technological University. He has completed his Masters in Software Engineering, [MTech (SWE), 2016-2018] from the Department of Computer Science & Engineering at Delhi Technological University, Delhi, India. He pursued his Bachelors in Computer Science and Engineering from DCRUST, Murthal, India. His research interests include intelligent systems, text mining and social web.

ANSHIKA ARORA has received the Bachelors of Technology (B.Tech) degree in Computer Science and Engineering from University of Delhi, New Delhi, India in 2017. She is currently pursuing her Masters Degree in Software Engineering (M.Tech) in Delhi Technological University, Delhi, India. Her research interests include Social Web, Sentiment Analysis, and Machine Learning.

ANAND NAYYAR is currently working as Professor, Researcher and Scientist in Graduate School, Duy Tan University, Da Nang, Vietnam. Dr. Nayyar received MCA from Punjabi University, Patiala in 2008 with gold medal and distinction. He received his M.Phil in Computer Science in 2009 in 1st Division from V.M.U, Tamil Nadu. He received his M.Tech in Information Technology in 2011 with Distinction cum University Topper. He received his MBA in Information Systems from Sikkim Manipal University with 1st Division. He has done his Ph.D. in Computer Science from Desh Bhagat University, Mandi Gobindgarh in 2017 He has more than 300+ Publications in various National, International Conferences cum International Journals (Scopus/ISI/SCI/SCIe) of high repute. He is currently working in the area of WSN, MANET, Swarm Intelligence, Cloud Computing, Big Data, Network Security, Internet of Things (IoT), Machine Learning, Deep Learning, Network Simulation, 5G Communications.

MOHAMED ABDEL-BASSET received his B.Sc., M.Sc and the Ph.D in Information systems and technology from Faculty of Computers and Informatics, Zagazig University, Egypt. His current research interests are Optimization, Operations Research, Data Mining, Computational Intelligence, Applied Statistics, Decision support systems, Robust Optimization, Engineering Optimization, Multi-objective Optimization, Swarm Intelligence, Evolutionary Algorithms, and Artificial Neural Networks. He is working on the application of multi-objective and robust meta-heuristic optimization techniques. He is also an/a Editor/reviewer in different international journals and conferences. He has published more than 150 articles in international journals and conference proceedings. He holds the program chair in many conferences in the fields of decision making analysis, big data, optimization, complexity and the internet of things, as well as editorial collaboration in some journals of high impact.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-022-01795-y

ORIGINAL ARTICLE

Sensor network driven novel hybrid model based on feature selection and SVR to predict indoor temperature for energy consumption optimisation in smart buildings

Sachin Kumar[1] · Zairu Nisha[2] · Jagvinder Singh[3] · Anuj Kumar Sharma[4]

Received: 3 February 2022 / Revised: 14 August 2022 / Accepted: 21 October 2022
© 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 2022

Abstract Energy is a vital resource for smart cities and smart buildings. Maintaining Indoor temperature to a comfortable level requires energy consumption which comes from several resources renewable or nonrenewable. The indoor temperature depends on several parameters such as environmental attributes, building design and structural attributes, user activity etc. Predictions on energy consumption and indoor temperature have employed few data-driven model and machine learning approaches which require more accuracy, better generalisation and real-time estimation of energy and the indoor temperature. This paper proposed a hybrid model based on feature selection methods such as feature importance and support vector regression (SVR) to predict indoor temperature in a more accurate and efficient manner. The best model SVR gives MAE equals to 0.0174, MSE equals to 0.0006, RMSE equals to 0.0256 which are the least error in the prediction of indoor temperature. Random forest regressor (RFR) model gives MAE equals 0.0820, MSE equals 0.0101, RMSE equals 0.1006. RFR results are the worst in all prediction error matrices. Accuracy and efficiency have been compared with other machine learning models and the proposed model has outperformed them.

Keywords Indoor temperature · SVR · Prediction · Feature Importance · Energy consumption · Smart buildings

✉ Sachin Kumar
skumar@cic.du.ac.in; officialid.sachin@gmail.com

Zairu Nisha
zairunisha@gmail.com

Jagvinder Singh
jagvinder.singh@gmail.com

Anuj Kumar Sharma
anujsharma1920@gmail.com

[1] Cluster Innovation Centre, University of Delhi, Delhi, India

[2] Ramanujan College, University of Delhi, Delhi, India

[3] School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

[4] Department of Mathematics, Shyamlal College, University of Delhi, Delhi, India

1 Introduction

Energy drives modern cities and societies by running the modern equipment, tools, processes and systems to make our life comfortable and sustainable (Allam and Dhunny 2019; Rutherford and Coutard 2014). The demand for energy is increasing day by day in our cities as their systems and tools need energy (Al-Homoud 2001; Perera et al. 2018). In residential and commercial buildings, there are many equipment and processes which are necessary to make our life comfortable and productive (D'Oca et al. 2018; Jensen et al. 2018). These equipment and processes consume energy which is generated from several sources such as fossil-based sources and renewable sources. This energy generation process is also responsible for global warming and climate change (Pelletier et al. 2019). The present pace of development and modernisation in developed and developing countries is responsible for such phenomenons (Sadorsky 2014). This situation expects the stakeholders to use renewable energy resources and use energy efficiently so that energy demand is optimised (Kern et al. 2018). One of the aspects of energy consumption optimization is to model and predict the energy demand and consumption in different sectors. There are several sectors of high energy consumption but the building sector is the major consumer of energy to the

Published online: 09 November 2022

level of 40% of the total energy demand of the countries (D'Agostino et al. 2017). Buildings consume energy in running their equipment, processes and operations. Their operations include, specifically making the indoor environment comfortable to stay in and working by controlling the indoor temperature. It is managed by using heating, ventilation and air conditioning(HVAC) systems (D'Agostino et al. 2017). Other energy consumption items include running of home equipment, security systems, kitchen appliances, user occupancy and activities monitoring (Cao et al. 2016). HVAC systems consume approximately 50% of the energy in residential and commercial buildings (Tian et al. 2019). One report from Spain stated that Spanish household energy consumption is approximately 30 percent of the total energy consumption of Spain (Blázquez et al. 2013; Stoppok et al. 2018). These statistics prepare a case for the importance of energy usage optimization and consequently modelling and prediction of energy usage in buildings. The indoor temperature maintenance is directly related to energy consumption in buildings which can be modelled and predicted. Prediction and modelling of energy through temperature would help in knowing in advance when and how much energy will be required so that its utilisation and management can be done effectively and efficiently. There have been several competitions for the modelling and development of energy-efficient homes and buildings (Perng et al. 2018). As stated that HVAC systems are major contributors, their activation and deactivation if modelled and applied efficiently can save a lot of energy. That depends on the determination of proper comfort level without any negative factors which could disturb the person. There are a large number of parameters and factors such as environmental factors, HVAC systems and user patterns. Having a large number of dynamic factors and their tuning become very difficult and complex. The prediction models developed which are based on mathematical frameworks become rigid and static and unsuitable for real-time temperature prediction and modelling (Širokỳ et al. 2011). There is a need to apply data-driven intelligent approaches for the prediction and modelling of temperature in real-world circumstances. Recent developments in machine learning approaches are data-driven and can adapt to real-world dynamics and complexities with a large number of factors. They have been used in many domains for modelling and prediction, adapting to complex and dynamic environments through parameter tuning using statistical properties (Kumar et al. 2019; Hx and Magoulès 2012). Their applications are in intelligent transportation, smart cities, health and disease prediction, heating and cooling load modelling, classification of objects etc. Several researchers have also applied many mathematical models and statistical rigid models for the prediction of energy consumption and temperature prediction (Ahmad et al. 2017). For instance, Kumar et al. proposed a model to predict the heating and cooling load using buildings structural and design attributes (Kumar et al. 2018b, c). Some of the present-day machine learning models suffer from the local maxima and local minima and inefficiency in parameter tuning which makes them unsuitable for real-time temperature prediction. The proposed research work undertakes a comprehensive study of temperature modelling and prediction using a data set having environmental attributes (outdoor temperature, humidity level, wind speed and sun irradiance), occupancy attributes, equipment related (metres and equipment running) attributes. Such large attributes need to be optimised so that accuracy and efficiency can be improved. The proposed study analyses several factors and their feature importance in the prediction of temperature and energy and quantification of their process of feature importance and correlation. The proposed model with feature importance approaches integrated with support vector machine (SVM) methods give better accuracy and efficiency. The proposed model is compared on accuracy and efficiency with other traditional models and outperforms them.

This paper is organised as follows. Section 2 deals with the recent development and works. Section 3 deals with the mathematical foundations of the machine learning models and feature selection process and its mathematical formulations. Section 4 explains the proposed model and its characteristics. Section 5 discusses the results and critically analyses them. Section 6 concludes the paper.

2 Literature survey

In the energy and building domain, several studies have been conducted which have estimated the energy demand of the buildings based on several parameters and attributes. Machine learning approaches have been applied in many domains such as in indoor temperature forecasting (Kumar et al. 2016), intelligent transportation (Kumar et al. 2018a), energy performance (Kumar et al. 2018b, 2019), occupancy detection (Kumar et al. 2020), news article classification (Kumar et al. 2021, 2022) and depression and mental state detection (Kumar and Nisha 2022). Kumar et al. predicted the energy load of the building based on environment, design and structural related attributes (Kumar et al. 2018c, 2019). Candanedo et al. proposed the data-driven approaches for the prediction model using the appliances data (Candanedo et al. 2017). Zamora-Martínez et al. modelled the indoor temperature using multivariate analysis in which several outdoor and indoor environment variables were recorded and models were developed (Zamora-Martínez et al. 2013). Gupta et al. (2022) used an efficient AR modelling-based electrocardiogram signal analysis for health informatics. Gupta et al also (Gupta et al. 2021) applied ECG signal analysis using CWT, spectrogram and autoregressive technique. Li et al. (2019)

experimented with cloud-terminal-based cyber-physical system architecture for energy efficient machining process optimization. Tomazzoli et al. (2020) used the Internet of things and artificial intelligence enabled energy efficiency. Some of the researchers have used IoT and Fog computing for energy function optimization. Saxena et al. (2022) experimented with reactive power requirement for operating wind-driven Micro Grid in the presence of several proportions and classes of static load. Magalhães et al. conducted a study in Portugal for the prediction of indoor temperature in the residential buildings (Magalhães et al. 2016). Some research studies also developed prediction models of indoor temperature for real-time environment (Afroz et al. 2018). Xu et al. (2019) used the deep neural network structures for better accuracy of the indoor temperatures but such approaches require a large amount of data and computational time (Aznar et al. 2018). Ahmad et al. used machine learning approaches for short term energy demand forecasting using machine learning-based models (Ahmad and Chen 2018) Wang et al. did a comprehensive survey of building energy prediction approaches that were inspired by artificial intelligence (Wang and Srinivasan 2017; Wang et al. 2018). Esen et al. (2008) provided modelling using a ground-coupled heat pump system that used a support vector machine (SVM). Tang et al. (2014) used clustering approaches for short term prediction. Candanedo and Feldheim (2016) also tries to analyse the indoor environment and other attributes for occupancy prediction using statistical learning models. Catalina et al. not only independently used machine learning regression models but also used multiple regression models for improving hearing demand prediction model accuracy (Catalina et al. 2013). Ahmad et al. provided a comprehensive overview of the data-driven large scale forecasting approaches (Ahmad et al. 2018.) Ahmad also demonstrated the comparison of tree-based and neural network-based machine learning approaches for building energy prediction (Ahmad et al. 2017). Yang et al. used the basis for the prediction emerging machine learning approach ELM (Yang et al. 2017). Al-Wadi gave the comprehensive indoor temperatures forecasting models survey (Alawadi et al. 2017). Paul et al. applied the machine learning approaches in the IoT environment for the prediction of building energy prediction (Paul et al. 2018). Foucquier, Aurélie et al. provided the state of art in the building modelling of energy performance (Foucquier et al. 2013). Mba et al. (2016) proposed a model based on ANNs for hourly prediction of indoor temperature and relative humidity in humid regions. Özbalta et al. (2012) demonstrated models for prediction indoor temperature and relative humidity with daily mean. Korolija executed a study specific to the UK and proposed a regression model for predicting building energy consumption based on heating and cooling demands (Korolija et al. 2013). Neto and Fiorelli (2008) provided the comparative analysis of detailed model simulation and artificial neural network for forecasting building energy consumption. Magalhães et al. (2017) proved the relationship between heating energy use and indoor temperatures in residential buildings and used the ANNs with occupant behaviour. Catalina et al. developed and validated regression models to predict monthly heating demand for residential buildings (Catalina et al. 2008). After going through the recent work and literature, it is clear that there is a research gap in the development of more accurate prediction models which are data-driven and efficient.

3 Machine learning approaches

Rigid mathematical approaches have been one of the traditional menthods of modelling and prediction but they suffered from complex rigid structures which made them unsuitable to the dynamic environment modelling and predictions (Flores et al. 2011). On the other hand, recent development in the machine learning domain has presented approaches that are data-driven and provide accurate modelling and prediction in dynamic environments and in the contexts with a large number of variables (Kathirgamanathan et al. 2021). Few of the machine learning approaches which have been used in the experimentation have been described with their mathematical framework and construct. One of the most simple machine learning approaches is linear regression. Mathematically, linear regression tries to fit the line to the data points and predict the unknown points using the trends of the plotted line (Marill 2004). More specifically, It is represented as a linear equation that combines a set of input data values represented as X with the predicted output values represented by Y. In simple mathematical form, it is written as $Y = B_0 + B_1 * X$ which is equivalent to the $Y = mX + c$ in two-dimensional geometry. Here both m and c are parameters that represent the slope of the line and the intercept on the Y-axis respectively. Both parameters/coefficients need tuning according to the data points and different problems to predict and model the trend accurately. Another important machine learning approach is the support vector machine (SVM) which is mostly used for classification types of problems (Suykens and Vandewalle 1999). Its regression or time series forecasting variant is called support vector regression (SVR). The simple concept of SVM is to plot the input data points to the high dimensional hyperplane and try to draw a plan so that the plan maximises the distance between two classes of input data points. This determined hyper plain is called a support vector and let us classify the input data points into classes. A similar concept is applied in SVR for regression and time series forecasting related problems. Mathematically, Suppose, N data points represented as $\{\mathbf{X}_i, Y_i\}\ i = 1, \ldots, n$ where $Y_i \in \{-1, 1\}$ is the class of the point X_i. There are only two classes $Y_i = 1$ and $Y_i = -1$ in

this case although it may work for multiple classes also. To accurately classify and predict the new data points x_i's into one of the classes represented by Y, a hyper plan in hyperspace should have a maximum margin between two classes to avoid misclassification and low accuracy. Let's take the linear scenario of cases. The hyperplane can be written as:

$$\mathbf{x}_i\mathbf{w} + b = 0.$$

Moreover, two parallel hyperplanes must be determined to separate out the data points and their mid distance must be as large as possible. The two parallel hyperplanes can be described as follows

$$\mathbf{x}_i\mathbf{w} + b = +1$$

and

$$\mathbf{x}_i\mathbf{w} + b = -1$$

The distance between the two hyperplanes is measured as $\frac{2}{||w||}$:

$$d_+ + d_- = \frac{|1-b|}{\|w\|} + \frac{|-1-b|}{\|w\|} = \frac{2}{\|w\|}. \tag{1}$$

The model needs minimization of $\|w\|$. Additionally, for $i \in (1, n)$, X_i and Y_i follows the constraints as given below.

$$\mathbf{X}_i\mathbf{w} + b \geq +1,\ Y_i = +1 \tag{2}$$

$$\mathbf{X}_i\mathbf{w} + b \leq -1,\ Y_i = -1 \tag{3}$$

$$\equiv \tag{4}$$

$$Y_i(\mathbf{x}_i\mathbf{w} + b) - 1 \geq 0,\ \forall i \tag{5}$$

In the SVM model, maximisation of the margin is the same as minimising $\|w\|$. For proper classification, optimization problem:

$$\text{Minimize } \frac{\|w\|^2}{2}$$

and

$$\text{s.t } y_i(\mathbf{x}_i\mathbf{w} + b) - 1 \geq 0$$

Here w and b need such tuning so that they can make the above statement true. To solve the optimization problem, Lagrange Multiplier is introduced.

Lagrange multipliers is a strategy for finding the local maxima and minima of a function subject to equality constraints. For the case of only one constraint and only two choice variables, consider the optimization problem: Minimize $f(x, y)$ s.t $g(x, y) = 0$. This looks exactly like the problem that we want to solve and we can switch the problem into a lagrangian problem. Minimize $W_p(\alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^{l} \alpha_i y_i(\mathbf{x}_i\mathbf{w} + b) + \sum_{i=1}^{l} \alpha_i$. We can see this as a convex quadratic programming problem with the dual: Maximize $W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i\alpha_j y_i y_j(\mathbf{x}_i\mathbf{x}_j)$. The reason that we make this transformation is that the solution of the dual problem can be easily approximated with the computer. Although there will always be a small gap between the approximated result and the real answer, it is still very powerful and useful. Another model used in the experiment is Decision Tree (DT) which is primarily based on the induction of decision trees and is used from observations and branches for decision making ensuring the maximum information gained in each decision making (Dietterich and Kong 1995). The decision is taken in the branches or in the internal nodes of the tree and the leaves nodes represent the target values. Decision trees where the target value can take real numbers are called regression trees. Additionally, Random Forest is one of the most robust machine learning models. The random forests method produces better results and works well on large datasets (Biau and Scornet 2016). Moreover, Random forests also work better with missing data values by creating estimates for them. Mathematically, a Random forest is an ensemble process of decision trees machine learning approach. In a random forest, different data points are used to create different random trees and different features are selected for splitting. This way each constructed tree makes its own prediction and several predictions are averaged out to get the single predicted output. The averaging makes a Random Forest regression better than a single Decision Tree hence improving its accuracy and reducing overfitting of the data set. A prediction from the Random Forest Regressor is an average of the predictions produced by the trees in the forest. Random Forest is taken as Random forest = DT (base learner) + bagging (Row sampling with replacement) + feature bagging (column sampling) + aggregation(mean/median, majority vote) to predict the value (Kumar 2019). According to the analysis of the training and run-time complexity. RF training time is $O(log(nd) * k)$. The Running time is $O(depth * k)$ and memory Space is $O(storeeachDT * K)$. Another machine learning model is the Naive Bayes is the most simple algorithm that you can apply to your data. NB makes certain assumptions such as that all the variables in the dataset is "Naive" which means that they are not correlated to each other Rish et al. (2001). For example, mathematically, Here c represents the class eg. ball, cat, bat. X represents features calculated individually. Their probability is determined as P(c|x) represents the posterior probability of the class C given predictor/features. P(c) represents the probability of class. P(x|c) represents the likelihood which is the probability of the predictor given class. P(x) represents the prior probability of the predictor. Another machine learning model is KNN regression

is a non-parametric method that approximates the relationship between independent variables and the output variable by averaging the observations in the same neighbourhood (Song et al. 2017). The size of the neighbourhood needs to be set by the analyst or can be chosen using cross-validation to select the size that minimises the mean-squared error(MSE). While the method is quite appealing, it quickly becomes impractical when the dimension increases, when there are many independent variables. When implementing KNN, the first step is to transform data points into their mathematical values (vectors). The most common way to find the distance between is the Euclidean distance. According to the Euclidean distance formula, the distance between two points in the plane with coordinates $(x, y)and(a, b)$ is given by $dist((x, y), (a, b)) = (x|a) + (y|b)$. One more, the Gradient Boosting algorithm is used to generate an ensemble model by combining the weak learners or weak predictive models. Gradient boosting algorithms can be used to train models for both regression and classification problems. Gradient Boosting Regression (GBR) algorithm is used to fit the model which predicts the continuous value (Bahad and Saxena 2020). Gradient boosting builds an additive mode by using multiple decision trees of fixed size as weak learners or weak predictive models (Mayr et al. 2014). The parameter, $n_estimators$, decides the number of decision trees which will be used in the boosting stages. Gradient boosting differs from AdaBoost in the manner that decision stumps are used in AdaBoost whereas decision trees of fixed size are used in Gradient Boosting (Natekin and Knoll 2013). The process of fitting the model starts with the constant such as the mean value of the target values. In subsequent stages, the decision trees or the estimators are fitted to predict the negative gradients of the samples. The gradients are updated in each iterator (for every subsequent estimator). A learning rate is used to shrink the outcome or the contribution from each subsequent tree or estimator.

4 Research methodology

4.1 Problem formulation

In regression problem, there are two types of attributes:independent attribute and dependent attribute. A time series is a sequence of data points obtained from any process which runs continuously over the period of time. Mathematical formulation of the indoor temperature prediction can be defined as follows: $Y = F(X)$ where Y is dependent matrix of size $N \times 1$ and X is the $N \times M$. Where N represents the number of data points in the research problems data and M represents the independent attributes to predict and model indoor temperature Y. More specifically, It can be described as follows.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1} = \mathbb{F}\begin{bmatrix} x_1 & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_N & \cdots & x_{NM} \end{bmatrix}_{N\times M} \tag{6}$$

Here each of the y_i is obtained by some function of values of $X_i....X_{iM}$. This is simple modelling at a particular time stamp t. It can be modelled on the basis of the values of X on the time stamp t to predict the values of Y at time stamp $t + 1, t + 2...t + s$. Here s is sth time added to the present time stamp value. In this regression problem, we are interested in the value of indoor temperature and this can depend on other independent attributes/values. But we can also model the indoor temperature itself using only a single variable. In that case the regression problem becomes the time series prediction problem. This may require the horizontal expansion of Y to form the X matrix as given in the equation to make a shift of values and predict some point ahead.

4.2 Data and its description

The data includes several types of parameters which can be analysed and used for the modelling of the indoor temperature. The attributes includes time related information such as date and time, weather forecast provided by the authorised weather agency, carbon dioxide information in rooms, relative humidity, lightening status of rain and its quantity, sub dusk, wind speed, sub light, sun irradiance, motors and mechanical equipment, outdoor temperature and also day of the week.There parameters are including the equipment related information, environment related information and the time related information in the data set and has been obtained from SML Experiment available on UCI machine Learning repository (Solar Decathlon Europe Competition 2011). Data set has a recording of all the attributes and variables for 40 days which is sampled every minute and the mean after completion of 15 min is uploaded. Hence in each hour, the mean is determined at 0th min, 15th min, 30th min, and 45th min. More specifically, Data has 24 attributes and a total of 4137 data samples. These samples represent a total 42 days of indoor temperature and associated independent attributes data. There are a total 2764 (28 days) and 1373 (14 days) training and testing samples respectively.The data set can be downloaded from the UCI machine learning repository (Bache 2013).

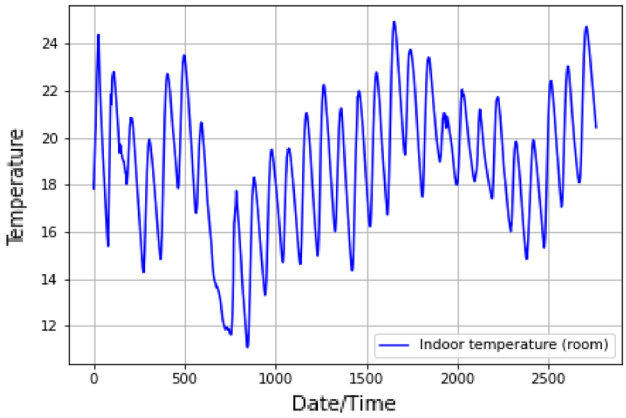


Fig. 1 Plot for indoor temperature room

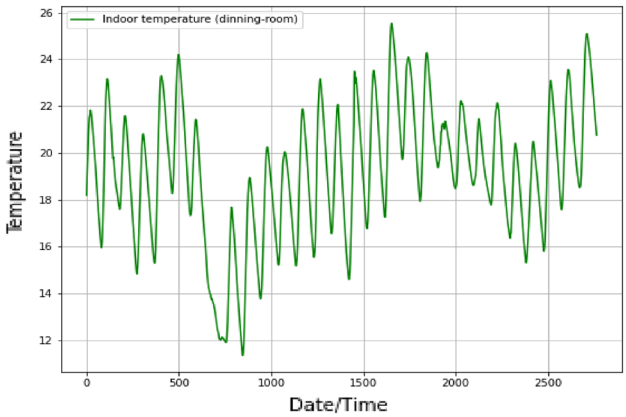


Fig. 2 indoor temperature dining

4.3 Experiment configuration and software tools

In terms of the system configuration and software tools used in the experimentation, the experiment has been conducted on python programming and Google Collab facility. Additionally some python libraries from sklearn have been used in implementing specific ML models and processing modules.

5 Proposed hybrid model

To predict the indoor temperature, data of several attributes obtained by respective sensors and other sources are stored together and combined in structured form so that it can be fed to the proposed model properly. In case, data contains some missing values, they are updated with median values in column and in rows. Data is cleaned and data exploration process is performed to visualise the data and its relationships with other variables/features and indoor temperature itself, our side environment variables which are depicted

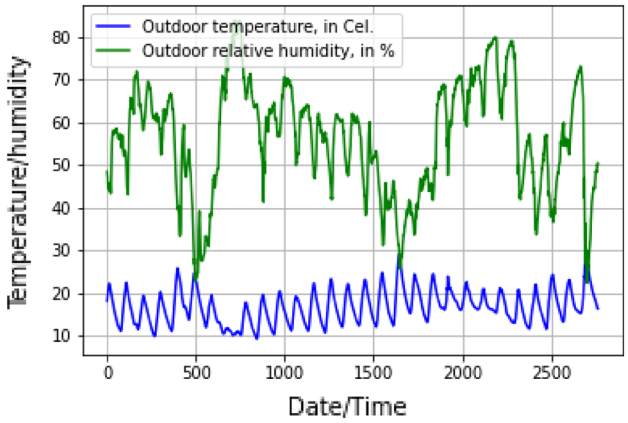


Fig. 3 Outdoor temperature and relative humidity

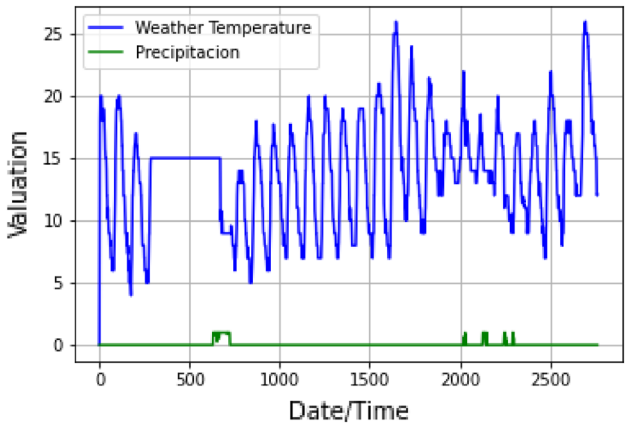


Fig. 4 Weather temperature and precipitation

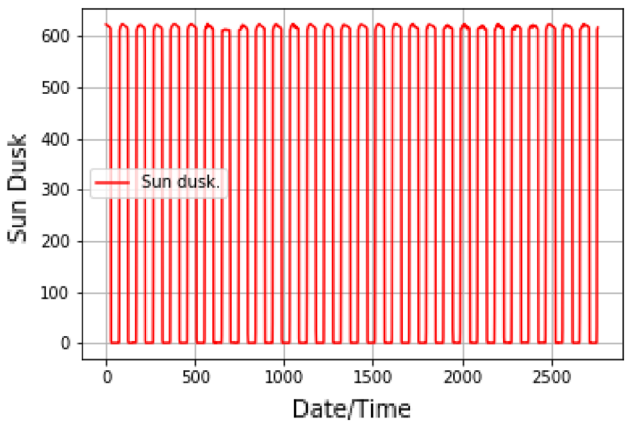


Fig. 5 Plot for sun dusk

in Figs. 1, 2, 3, 4, 5, 6 and 7. This data visualisation and exploration gives some important insights about the relationships of features and indoor temperature. Additionally

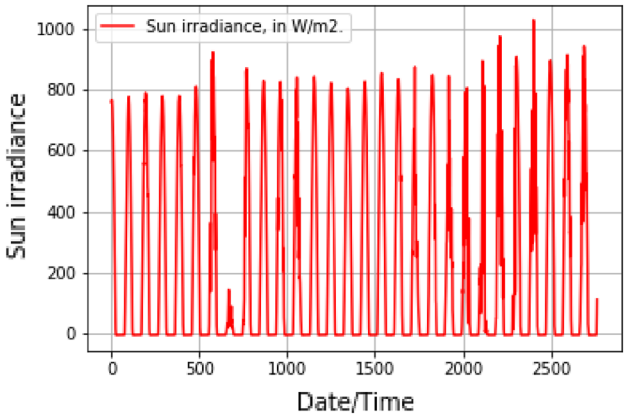


Fig. 6 Plot for sun irradiance

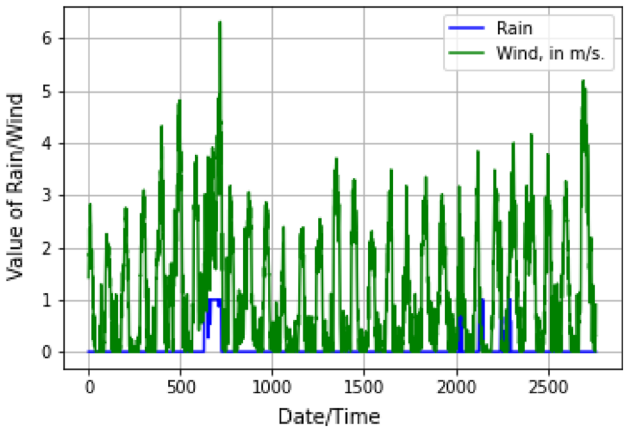


Fig. 7 Plot for rain and Wind

two more feature importance quantification techniques are applied namely, feature importance (FI) and co-relation feature selection. Important features are selected form the data set. The data set is divided into two sets : training data set and testing data set with 70 percent and 30 percent respectively. Selected features with machine learning models are run and results are obtained in terms of error such as MAE and RMSE etc for measuring the accuracy of predicted indoor temperature and actual temperature. The result are organised in tables and discussed in result section. The proposed model is displayed in Fig. 11.

5.1 Evaluation criteria

The proposed model uses error matrices which are mean absolute error (MAE), root mean square error (RMSE), mean squared error(MSE) and R2. Theses are important measure to know the accuracy of regression and time series forecasting model. Mathematical formulation is given as follows.

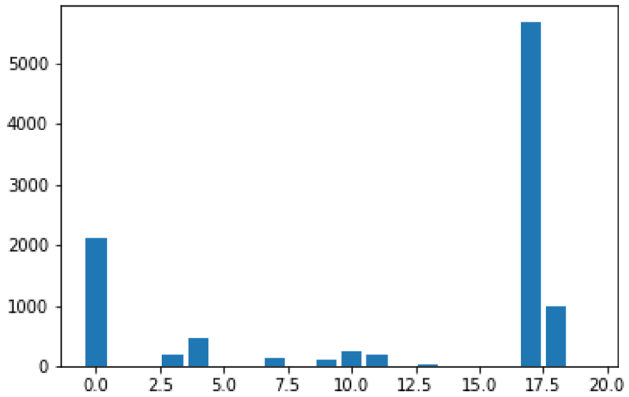


Fig. 8 Plot for correlation feature selection

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|A_i - P_i| \tag{7}$$

$$RMSE = \frac{1}{N}\sum_{i=1}^{N}(|A_i - P_i|)^2 \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(|A_i - P_i|)^2} \tag{9}$$

where N is number of samples, A is actual value and P is predicted value.

5.2 Experiment parameters

The proposed experimentation have used ML methods such as LR, SVR, DTR, RFR, GBR, etc. The associated parameter for each model are as follows. In GBR model, it as loss = 'ls', verbose = 2, criterion = 'mae'. In DTR model it has used "random state"= 0, "max depth" = 100, "max leaf nodes" = 1000, criterion='mae'. In RFR it uses parameters as "max depth" = 2, "max leaf nodes" = 10, "random state = 0", "verbose = 2" , "criterion = 2'mse"'. In SVR the parameters are C = 3.0, epsilon = 0.01.

6 Results and discussion

This section discusses the results and their comparative analysis. Starting from the data and its quantified information. Some of that has already been discussed in exploration section. Table 1 gives details of all features of data set which contains weather variables, equipment variables, indoor environment variables and time variable. Their count, mean values, standard deviation, max values etc

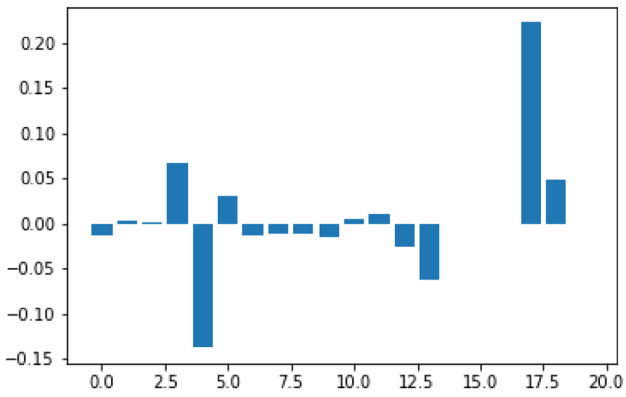


Fig. 9 Plot for feature importance

have been calculated to know the profile of data set. Table 2 and Figs. 8 and 9 provides the data of feature importance quantification of all features in order to their relationship with indoor temperature. Two approaches have been used: Feature importance method and Correlation Feature Selection method. Their values of correlation have been assigned against their name in the table which gives us important measure to identify the features which are importance and will help in better accuracy of models. It is clear that more important features which contribute to indoor temperature is outer water temperature, humidity, precipitation, orientation, outside temperature and outside humidity etc. The proposed hybrid model of feature selection and machine learning approaches is applied and results are obtained in error matrices such as MAE, RMSE, MSE and R2. These are standard error matrices to determine the deviation between the actual values of indoor temperature and predicted value of indoor temperature. Results have been demonstrated in pictorial forms in (Figs. 10, 11). Figure 12 gives the comparative indoor temperature plot of predicted value of indoor temperature by proposed model and actual values of indoor temperature. Figure 13 gives the comparative indoor temperature plot of predicted value of indoor temperature by SVR model and actual values of indoor temperature. Figure 15 gives the comparative indoor temperature plot of predicted value of indoor temperature by RFR model and actual values of indoor temperature.

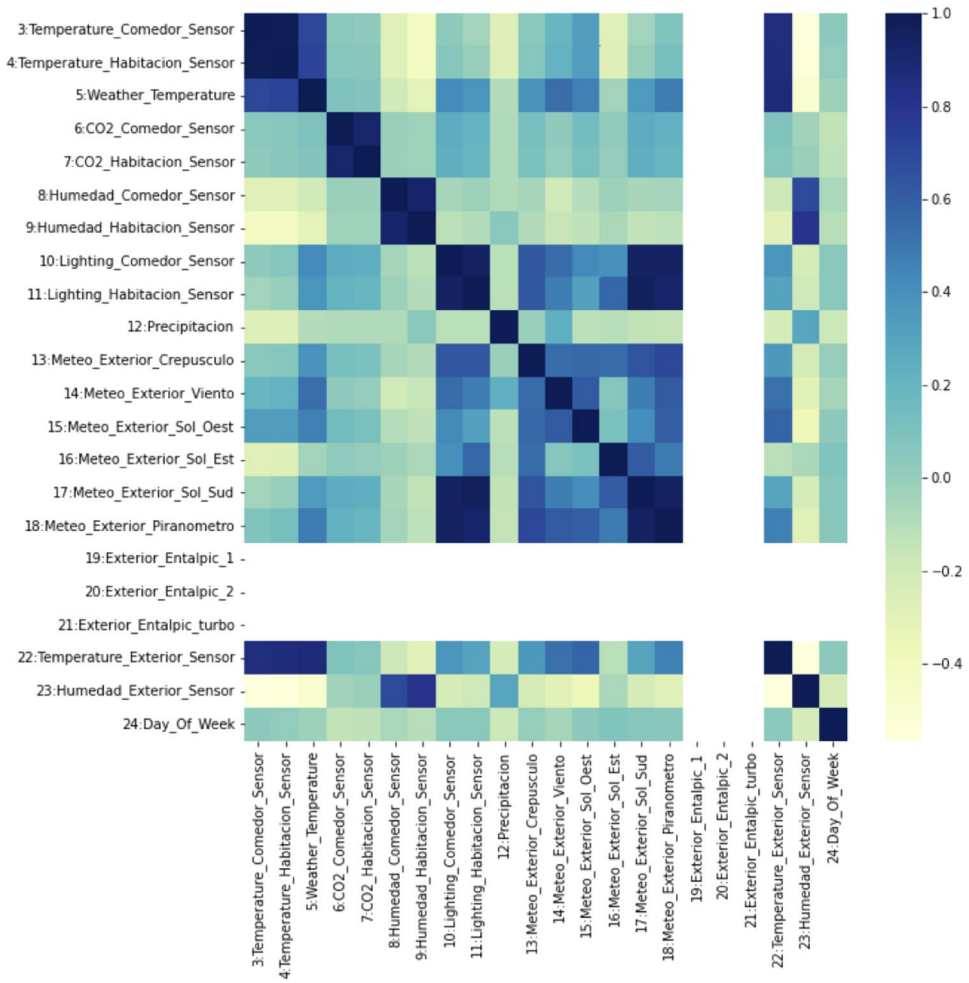


Fig. 10 Heat map to depicted the relationship and quantification of features with indoor temperature

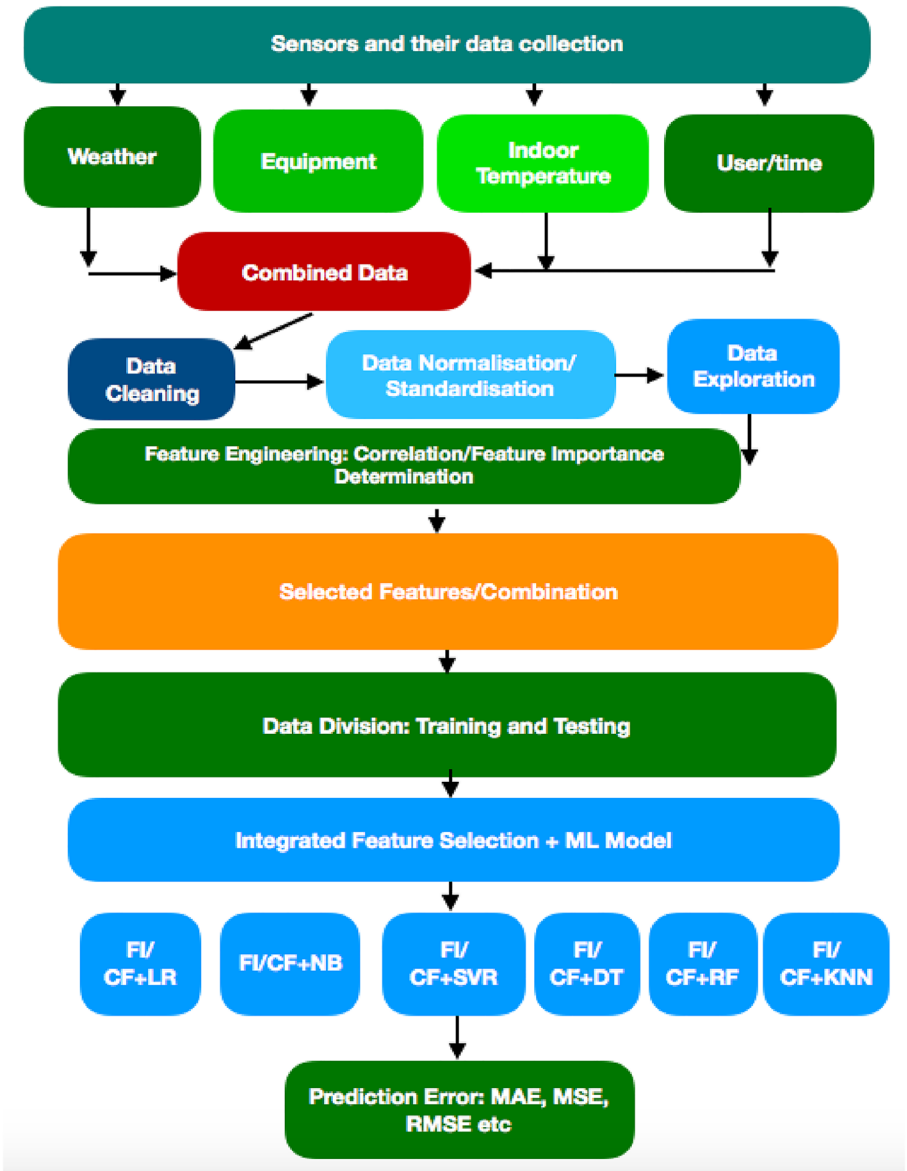


Fig. 11 Proposed hybrid model

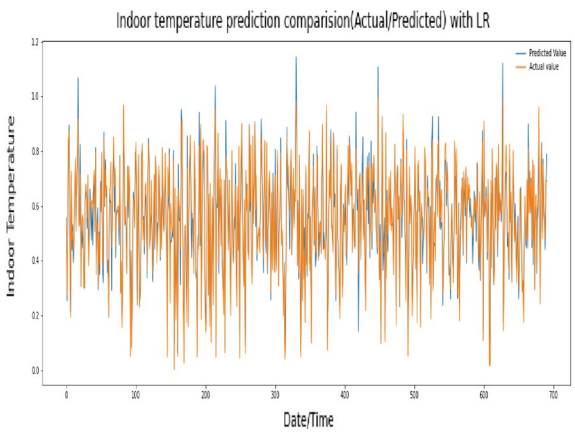


Fig. 12 LR comparison plot for actual and predicted indoor temperature

Figure 17 gives the comparative indoor temperature plot of predicted value of indoor temperature by NB model and actual values of indoor temperature. Figure 14 gives the comparative indoor temperature plot of predicted value of indoor temperature by DT model and actual values of indoor temperature (Fig. 15).

Figure 16 gives the comparative indoor temperature plot of predicted value of indoor temperature by GBR model and actual values of indoor temperature. More clearly, LR model gives MAE equals to 0.04317, MSE equals to 0.0031, RMSE equals to 0.05624, R2 equals to 0.9192. The best model SVR gives MAE equals to 0.0174, MSE equals to 0.0006, RMSE equals to 0.0256 which are the least error in prediction of indoor temperature. RFR model gives MAE equals to 0.0820, MSE equals to 0.0101, RMSE equals to 0.1006. RFR results are worst in all error matrices. DT model gives

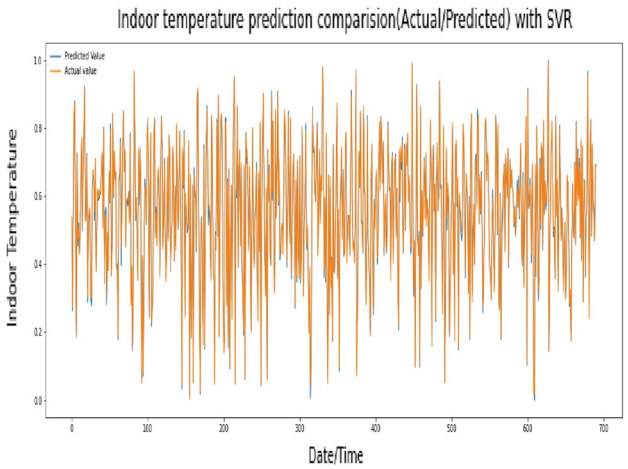


Fig. 13 Plot for SVR comparison plot for actual and predicted indoor temperature

MAE equals to 0.0276, MSE equals to 0.0022, RMSE equals to 0.0469. NB model MAE equal to 0.0431, MSE equal to 0.0031, and RMSE equals to 0.0562. GBR model gives MAE equals to 0.0302, MSE equals to 0.0017, RMSE equals to 0.0420 (Fig. 17). Figure 18 gives comparison of all models with all error matrices. Here the best performing mode is proposed model is feature selection+SVR. In the last, it is visible form the results that the worst performing model is RFR which suffers from comparatively high errors as compared to other models such as Linear regression (LR), GBR, DT, NB. In the case of all error matrices, MEA bars are very high. In terms of the models performing well, the first one is SVR that produces the most accurate indoor temperature prediction with lowest error in MAE equals to 0.01 scale. Second best performing model is GBR followed by DT, NB, LR and RFR. This paper proposed a hybrid model based on a feature selection methods and support vector machine (SVR) to predict indoor temperature prediction in a more accurate and efficient manner. The best model SVR gives MAE equals to 0.0174, MSE equals to 0.0006, RMSE equals to 0.0256 which are the least error in prediction of indoor temperature or energy consumption with environmental and sensor attributes. RFR model gives MAE equals to 0.0820, MSE equals to 0.0101, RMSE equals to 0.1006. RFR results are worst in all prediction error matrices.

7 Conclusion and future directions

Energy consumption optimisation is one of the important requirements of smart cities and smart buildings. Indoor

Table 1 Data description with count, minimum and maximum values

Attribute name	Count	Mean	STD	Min	25%	50%	75%	Max
Temperature_Comedor_Sensor	2764	19.20	2.85	11.35	17.45	19.37	21.23	25.54
Temperature_Habitacion_Sensor	2764	18.82	2.82	11.08	17.06	19.02	20.83	24.94
Weather_Temperature	2764	13.90	4.17	0.00	10.78	15.00	16.67	26.00
CO2_Comedor_Sensor	2764	208.48	27.03	187.34	200.89	207.05	211.25	594.39
CO2_Habitacion_Sensor	2764	211.07	28.47	188.91	202.68	209.41	213.22	609.24
Humedad_Comedor_Sensor	2764	44.88	6.59	27.08	40.35	45.43	49.35	60.96
Humedad_Habitacion_Sensor	2764	47.32	7.56	29.59	42.53	47.53	52.69	62.59
Lighting_Comedor_Sensor	2764	26.75	23.30	10.74	11.59	11.80	31.22	110.69
Lighting_Habitacion_Sensor	2764	40.73	42.33	11.33	13.27	17.69	52.06	162.97
Precipitacion	2764	0.05	0.21	0.00	0.00	0.00	0.00	1.00
Meteo_Exterior_Crepusculo	2764	325.37	305.06	0.61	0.65	611.80	619.21	624.96
Meteo_Exterior_Viento	2764	1.11	1.16	0.00	0.09	0.66	1.97	6.32
Meteo_Exterior_Sol_Oest	2764	14,936.62	25,964.05	0.00	0.00	0.00	15,088.00	95,278.40
Meteo_Exterior_Sol_Est	2764	12,248.00	21,758.55	0.00	0.00	0.00	11,131.27	85,535.40
Meteo_Exterior_Sol_Sud	2764	22,047.53	32,709.39	0.00	0.00	0.00	38,736.58	95,704.40
Meteo_Exterior_Piranometro	2764	215.01	297.23	-4.16	-3.38	3.92	435.43	1,028.27
Exterior_Entalpic_1	2764	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Exterior_Entalpic_2	2764	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Exterior_Entalpic_turbo	2764	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Temperature_Exterior_Sensor	2764	16.76	3.89	9.22	13.66	16.49	19.40	29.91
Humedad_Exterior_Sensor	2764	55.98	13.02	22.26	46.43	57.48	65.65	83.81
Day_Of_Week	2764	3.95	1.99	1.00	2.00	4.00	6.00	7.00

Table 2 Feature importance and correlation feature selection of data set

Feature no.	Feature importance	Correlation feature selection	Feature name
0	– 0.01377	2116.148587	Weather temperature
1	0.00388	4.001909	CO2 Comedor Sensor
2	0.00155	1.06269	CO2 Habitacion Sensor
3	0.06746	203.734634	Humedad Comedor Sensor
4	– 0.13693	473.809254	Humedad Habitacion Sensor
5	0.03082	1.13215	Lighting Comedor Sensor
6	– 0.01378	4.195266	Lighting Habitacion Sensor
7	– 0.01159	133.251723	Precipitacion
8	– 0.01226	5.45774	Meteo Exterior Crepusculo
9	– 0.0153	96.313811	Meteo Exterior Viento
10	0.00478	247.344349	Meteo Exterior Sol Oest
11	0.00987	186.784837	Meteo Exterior Sol Est
12	– 0.02569	4.571113	Meteo Exterior Sol Sud
13	– 0.06245	14.200123	Meteo Exterior Piranometro
14	0.22215	5684.524195	Temperature Exterior Sensor
15	0.04948	994.469056	Humedad Exterior Sensor
16	– 0.0005	2.970419	Day of Week

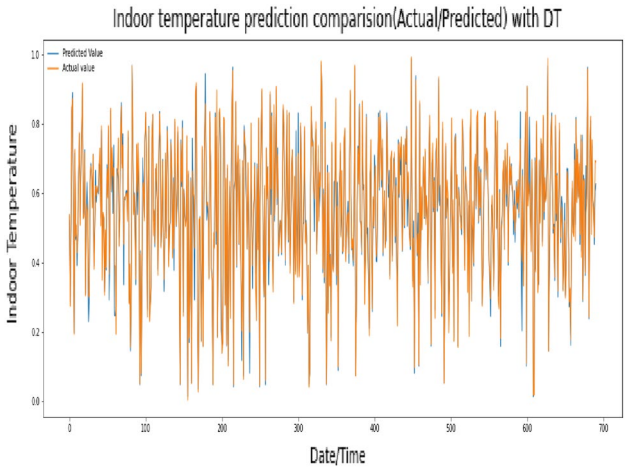


Fig. 14 Plot for DT comparison plot for actual and predicted indoor temperature

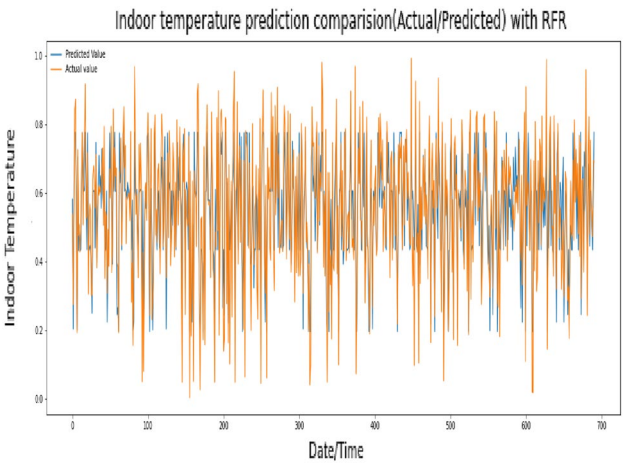


Fig. 15 RFR comparison plot for actual and predicted indoor temperature

temperature and its maintenance to a comfortable level require energy consumption. The indoor temperature depends on several parameters such as environmental attributes, design attributes, user activity in the indoor environment. Presently several studies have been done to predict and model indoor temperature using the data-driven model and machine learning approaches. This domain requires more accuracy, generalisation of prediction ability in accurate prediction of demand of energy to maintain the indoor temperature at a comfortable level. This paper proposed a hybrid model based on a feature selection and support vector machine (SVR) to predict indoor temperature prediction in a more accurate and efficient manner. The best model SVR gives MAE equals to 0.0174, MSE equals to 0.0006, RMSE equals to 0.0256 which are the least error in prediction of indoor temperature. RFR model gives MAE equals to 0.0820, MSE equals to 0.0101, RMSE equals to 0.1006. RFR results are worst in all error matrices. Accuracy and efficiency have been compared with other machine learning models and outperformed them. Future work, deep learning approaches can also be applied and explored.

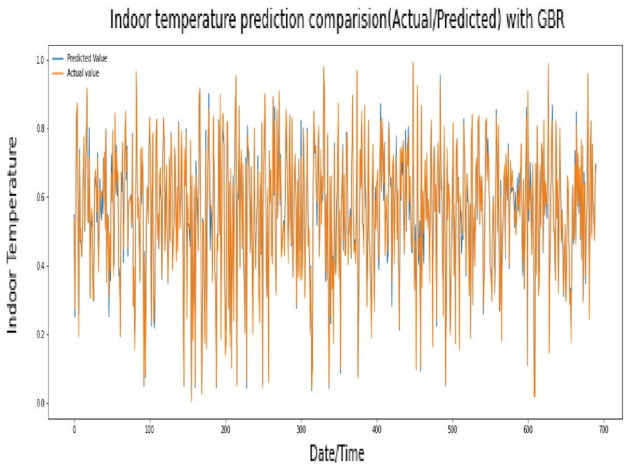


Fig. 16 Plot for GBR comparison plot for actual and predicted indoor temperature

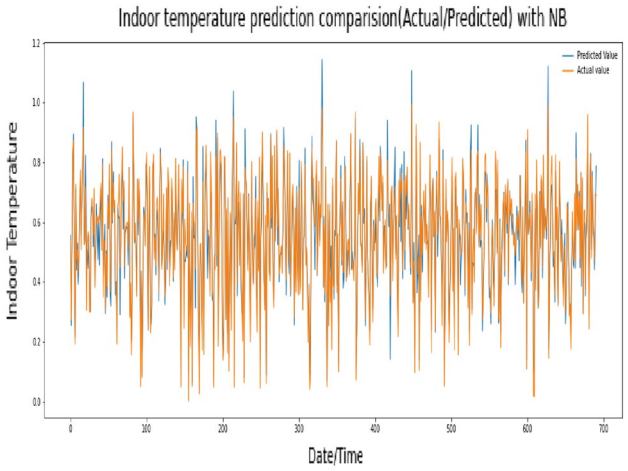


Fig. 17 NB comparison plot for actual and predicted indoor temperature

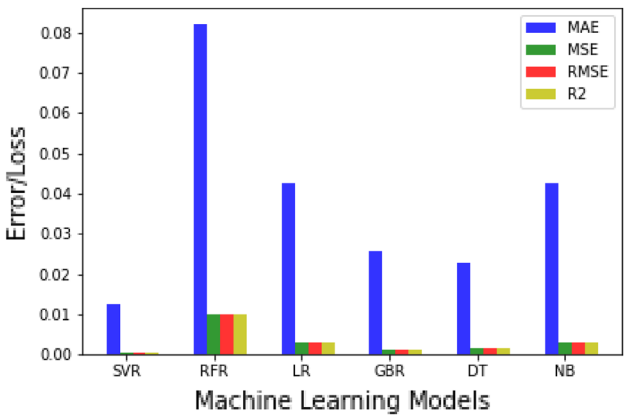


Fig. 18 Indoor temperature error for all machine learning models with all features

Funding No Funding has been used in the Research Work.

Declarations

Conflict of interest There is no potential conflicts of interest. The research does not involve Human Participants and/or Animals. Consent in all other cases have been taken.

References

Afroz Z, Urmee T, Shafiullah G, Higgins G (2018) Real-time prediction model for indoor temperature in a commercial building. Appl Energy 231:29–53

Ahmad T, Chen H (2018) Utility companies strategy for short-term energy demand forecasting using machine learning based models. Sustain Cities Soc 39:401–417

Ahmad MW, Mourshed M, Rezgui Y (2017) Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build 147:77–89

Ahmad T, Chen H, Guo Y, Wang J (2018) A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: a review. Energy Build 165:301–320

Alawadi S, Mera D, Fernández-Delgado M, Taboada JA (2017) Comparative study of artificial neural network models for forecasting the indoor temperature in smart buildings. In: Alba E, Chicano F, Luque G (eds) Smart cities. Smart-CT 2017. Lecture Notes in Computer Science, vol 10268. Springer, Cham, pp 29–38. https://doi.org/10.1007/978-3-319-59513-9_4

Al-Homoud MS (2001) Computer-aided building energy analysis techniques. Build Environ 36(4):421–433

Allam Z, Dhunny ZA (2019) On big data, artificial intelligence and smart cities. Cities 89:80–91

Aznar F, Echarri V, Rizo C, Rizo R (2018) Modelling the thermal behaviour of a building facade using deep learning. PloS One 13(12):e0207616. https://doi.org/10.1371/journal.pone.0207616

Bahad P, Saxena P (2020) Study of adaboost and gradient boosting algorithms for predictive analytics. In: International conference on intelligent computing and smart communication 2019. Springer, pp 235–244

Biau G, Scornet E (2016) A random forest guided tour. TEST 25(2):197–227

Blázquez L, Boogen N, Filippini M (2013) Residential electricity demand in Spain: new empirical evidence using aggregate data. Energy Econ 36:648–657

Candanedo LM, Feldheim V (2016) Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models. Energy Build 112:28–39

Candanedo LM, Feldheim V, Deramaix D (2017) Data driven prediction models of energy use of appliances in a low-energy house. Energy Build 140:81–97

Cao X, Dai X, Liu J (2016) Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. Energy Build 128:198–213

Catalina T, Virgone J, Blanco E (2008) Development and validation of regression models to predict monthly heating demand for residential buildings. Energy Build 40(10):1825–1832

Catalina T, Iordache V, Caracaleanu B (2013) Multiple regression model for fast prediction of the heating energy demand. Energy Build 57:302–312

D'Agostino D, Cuniberti B, Bertoldi P (2017) Energy consumption and efficiency technology measures in European non-residential buildings. Energy Build 153:72–86

Dietterich TG, Kong EB (1995) Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. Technical report, Citeseer

D'Oca S, Hong T, Langevin J (2018) The human dimensions of energy use in buildings: a review. Renew Sustain Energy Rev 81:731–742

Esen H, Inalli M, Sengur A, Esen M (2008) Modeling a ground-coupled heat pump system by a support vector machine. Renew Energy 33(8):1814–1823

Flores P, Leine R, Glocker C (2011) Modeling and analysis of rigid multibody systems with translational clearance joints based on the nonsmooth dynamics approach. In: Arczewski K, Blajer W, Fraczek J, Wojtyra M (eds) Multibody dynamics. Computational methods in applied sciences, vol 23, Springer, Dordrecht. https://doi.org/10.1007/978-90-481-9971-6_6

Foucquier A, Robert S, Suard F, Stéphan L, Jay A (2013) State of the art in building modelling and energy performances prediction: a review. Renew Sustain Energy Rev 23:272–288

Gupta V, Mittal M, Mittal V, Gupta A (2021) ECG signal analysis using CWT, spectrogram and autoregressive technique. Iran J Comput Sci 4(4):265–280

Gupta V, Mittal M, Mittal V, Gupta A (2022) An efficient AR modelling-based electrocardiogram signal analysis for health informatics. Int J Med Eng Inform 14(1):74–89

Jensen PA, Maslesa E, Berg JB, Thuesen C (2018) 10 questions concerning sustainable building renovation. Build Environ 143:130–137

K Bache ML (2013) Uci machine learning repository. University of California http://archive.ics.uci.edu/ml

Kathirgamanathan A, De Rosa M, Mangina E, Finn DP (2021) Data-driven predictive control for unlocking building energy flexibility: a review. Renew Sustain Energy Rev 135:110120

Kern E, Hilty LM, Guldner A, Maksimov YV, Filler A, Gröger J, Naumann S (2018) Sustainable software products-towards assessment criteria for resource and energy efficiency. Futur Gener Comput Syst 86:199–210

Korolija I, Marjanovic-Halburd L, Zhang Y, Hanby VI (2013) Uk office buildings archetypal model as methodological approach in development of regression models for predicting building energy consumption from heating and cooling demands. Energy Build 60:152–162

Kumar D (2019) Tree-based modeling techniques. In: Machine learning techniques for improved business analytics. IGI Global, pp 1–18

Kumar S, Kalia A, Sharma A (2018) Predictive analysis of alertness related features for driver drowsiness detection. Adv Intell Syst Comput 736:368–377

Kumar S, Pal S, Singh R (2018) Intra elm variants ensemble based model to predict energy performance in residential buildings. Sustain Energy Grids Netw 16:177–187

Kumar S, Pal SK, Singh R (2018) A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. Energy Build 176:275–286

Kumar S, Nisha Z (2022) Does social media Feed tell about your mental state? A deep randomised neural network approach. In: 2022 International joint conference on neural networks (IJCNN), pp 1–8. https://doi.org/10.1109/IJCNN55064.2022.9892210

Kumar S, Pal SK, Singh R (2016) Intelligent energy conservation: indoor temperature forecasting with extreme learning machine. In: Proceedings of intelligent systems technologies and applications 2016, vol 2. Springer, Switzerland, pp 977–988

Kumar S, Panwar S, Singh J et al (2022) iCACD: an intelligent deep learning model to categorise current affairs news article for efficient journalistic process. Int J Syst Assur Eng Manag 13:2572–2582. https://doi.org/10.1007/s13198-022-01666-6

Kumar S, Pal SK, Singh R (2019) A novel hybrid model based on particle swarm optimisation and extreme learning machine for short-term temperature prediction using ambient sensors. Sustain Cities Soc 49:101601. https://doi.org/10.1016/j.scs.2019.101601

Kumar S, Sharma A, Reddy BK et al (2022) An intelligent model based on integrated inverse document frequency and multinomial Naive Bayes for current affairs news categorization. Int J Syst Assur Eng Manag 13:1341–1355. https://doi.org/10.1007/s13198-021-01471-7

Kumar S, Singh J, Singh O (2020) Ensemble-based extreme learning machine model for occupancy detection with ambient attributes. Int J Syst Assur Eng Manag 1–11

Li X, He F, Li W (2019) A cloud-terminal-based cyber-physical system architecture for energy efficient machining process optimization. J Ambient Intell Hum Comput 10(3):1049–1064

Magalhães SM, Leal VM, Horta IM (2016) Predicting and characterizing indoor temperatures in residential buildings: results from a monitoring campaign in northern portugal. Energy Build 119:293–308

Magalhães SM, Leal VM, Horta IM (2017) Modelling the relationship between heating energy use and indoor temperatures in residential buildings through artificial neural networks considering occupant behavior. Energy Build 151:332–343

Marill KA (2004) Advanced statistics: linear regression, part II: multiple linear regression. Acad Emerg Med 11(1):94–102

Mayr A, Binder H, Gefeller O, Schmid M (2014) The evolution of boosting algorithms. Methods Inf Med 53(06):419–427

Mba L, Meukam P, Kemajou A (2016) Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region. Energy Build 121:32–42

Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. Front Neurorobot 7:21

Neto AH, Fiorelli FAS (2008) Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy Build 40(12):2169–2176

Özbalta TG, Sezer A, Yıldız Y (2012) Models for prediction of daily mean indoor temperature and relative humidity: education building in Izmir, turkey. Indoor Built Environ 21(6):772–781

Paul D, Chakraborty T, Datta SK, Paul D (2018) Iot and machine learning based prediction of smart building indoor temperature. In: 2018 4th international conference on computer and information sciences (ICCOINS). IEEE, pp 1–6

Pelletier C, Rogaume Y, Dieckhoff L, Bardeau G, Pons MN, Dufour A (2019) Effect of combustion technology and biogenic CO_2 impact factor on global warming potential of wood-to-heat chains. Appl Energy 235:1381–1388

Perera A, Coccolo S, Scartezzini JL, Mauree D (2018) Quantifying the impact of urban climate by extending the boundaries of urban energy system modeling. Appl Energy 222:847–860

Perng SY, Kitchin R, Mac Donncha D (2018) Hackathons, entrepreneurial life and the making of smart cities. Geoforum 97:189–197

Rish I, et al (2001) An empirical study of the naive bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence, vol 3, pp 41–46

Rutherford J, Coutard O (2014) Urban energy transitions: places, processes and politics of socio-technical change. Urban stud 51(7):1353–1377

Sadorsky P (2014) The effect of urbanization and industrialization on energy use in emerging economies: implications for sustainable development. Am J Econ Sociol 73(2):392–409

Saxena NK, Gupta V, Rajput RS, Kumar A, Gupta AR (2022) Reactive power requirement for operating wind-driven micro grid

in the presence of several proportions and classes of static load. In: Kumar A, Srivastava SC, Singh SN (eds) Renewable energy towards smart grid. Lecture Notes in Electrical Engineering, vol 823, Springer, Singapore. https://doi.org/10.1007/978-981-16-7472-3_3

Širokỳ J, Oldewurtel F, Cigler J, Prívara S (2011) Experimental analysis of model predictive control for an energy efficient building heating system. Appl Energy 88(9):3079–3087

Solar Decathlon Europe competition (2011) 2012. United States Department of Energy. http://www.solardecathlon.gov

Song Y, Liang J, Lu J, Zhao X (2017) An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing 251:26–34

Stoppok M, Jess A, Freitag R, Alber E (2018) Of culture, consumption and cost: a comparative analysis of household energy consumption in Kenya, Germany and Spain. Energy Res Soc Sci 40:127–139

Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300

Tang F, Kusiak A, Wei X (2014) Modeling and short-term prediction of HVAC system with a clustering algorithm. Energy Build 82:310–321

Tian Z, Si B, Shi X, Fang Z (2019) An application of Bayesian network approach for selecting energy efficient HVAC systems. J Build Eng 25:100796

Tomazzoli C, Scannapieco S, Cristani M (2020) Internet of Things and artificial intelligence enable energy efficiency. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-02151-3

Wang Z, Srinivasan RS (2017) A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev 75:796–808

Wang Z, Wang Y, Srinivasan RS (2018) A novel ensemble learning approach to support building energy use prediction. Energy Build 159:109–122

Xu C, Chen H, Wang J, Guo Y, Yuan Y (2019) Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method. Build Environ 148:128–135

Yang X, Gao J, Zhang L, Li X, Gu L, Cui J, Tong C (2017) A forecasting method of air conditioning energy consumption based on extreme learning machine algorithm. In: 2017 6th data driven control and learning systems (DDCLS). IEEE, pp 89–93

Zamora-Martínez F, Romeu P, Botella-Rocamora P, Pardo J (2013) Towards energy efficiency: forecasting indoor temperature via multivariate analysis. Energies 6(9):4639–4659

Zhao Hx, Magoulès F (2012) A review on the prediction of building energy consumption. Renew Sustain Energy Rev 16(6):3586–3592

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.

Advanced Production and Industrial Engineering
R.M. Singari and P.K. Kankar (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE220726

Simulation of Compression Molding of Thermoplastics Using CAE Tools: Assessment of Operational Conditions for Mold Design Development

Annu Pawar*[1], Roli Purwar

Department of Applied Chemistry, Delhi Technological University, New Delhi-110042, India

Abstract. This study aims to develop a simulation-based research to evaluate the influence of primary operational conditions on the efficiency of compression molding in processing thermoplastic parts in a mold design with no air venting. Moldflow Insight has been used to simulate the compression molding process to assess the quality of a range of semi-crystalline and amorphous thermoplastic products, to study the performance efficiency of the said mold. Simulation results were obtained by varying the mold surface temperature, melt temperature, charge geometry and mesh element size during processing thermoplastics like polypropylene, polyethylene, polystyrene, and polycarbonate. The findings showcased air entrapment at the corners of the mold cavity, which lowered the process efficiency when the mold operated with no venting under varied processing conditions. Simulation results indicated a need to introduce air vents on the mold's parting line. The findings revealed that the compression cycle time and waste generation could be diminished when air vents were included in the mold design and optimal operating conditions were used. Hence, a compression molding mold design incorporating air vents in areas of high air entrapment was developed.

Keywords. Compression Molding, Thermoplastics, Air Entrapment, Venting

1. Introduction

Compression molding of thermoplastics is highly discussed in the scope of use for composites but is seldom discussed on an individual basis to determine appropriate processing parameters and temperatures. Other than compression molding there are different methods used for making composites including the Hand Lay-Up, Open molding, resin infusion methods and injection molding, but compression molding is the most used and convenient method as it provide better surface finish, high pressure processability and high final product quality. (1-3) Hence, leading to a potential research gap.

With the advancement in computational simulation technology, engineering projects can ensure high precision in manufacturing compression molds for

[1] Annu Pawar, Department of Applied Chemistry, Delhi Technological University, New Delhi, India ; E-mail: annupawar0508@gmail.com

thermoplastics.(4) Computation analysis aided by commercially available software is a widely accepted strategy applied by plastics processing industries before manufacturing compression molds. (5-6) Computer Designing and Computer Simulation can aid in determining the suitability of a mold and product design for mass production upon clearly represents part designs and verifies the stress and deformation values. (7-9) Further, engineers can also work on evaluating alternative materials without physically using any material and compromising the availability of machine time for production. (10) Furthermore, simulations can also be utilized to optimize the processing parameters to produce better quality processing and products.

Henceforth, this study covers the CAE analysis cycle of compression molding of semi-crystalline and amorphous polymers. The following sections cover the methodology and results of the said simulations.

2. Methodology

The methodology for the said study was conducted in two steps. The first step comprised the designing the compression-molded product and the mold used for processing. The second step involved studying the compression molding of semi-crystalline and amorphous polymers using simulation software to identify major areas of air entrapment.

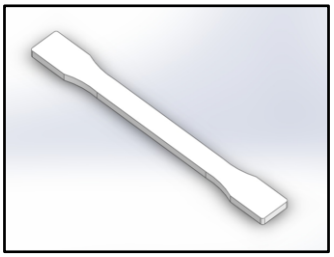


Figure 1. Model of the product design developed in SolidWorks CAD version 2016

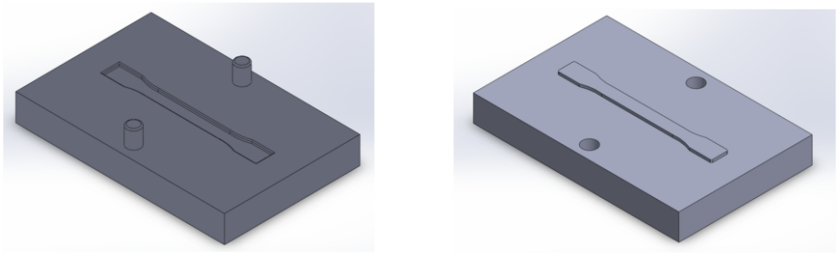


Figure 2. Mold model (a) Cavity and (b) Core developed in SolidWorks CAD version 2016

The simulations were performed for each polymeric sample of the compression-molded product for different operational conditions using the CAE package in Moldflow Insights (version 2019). As per the cavity design in Figure 2a, a three-dimensional compression model was developed to be studied in Moldflow Insight. In the present simulation study, a fill+pack model was used to analyze the behavior of thermoplastics while undergoing compression molding. The polymeric materials were compressed for 100 s during one cycle.

For each polymer, simulations were conducted by varying the initial charge geometry, mesh size, melt temperature, and mold temperature (discussed in the subsequent sections), resulting in 400 simulations. Other parameters that affected processing were maintained constant (Table 1). (11) Based on the simulation results, the masses of the different compression-molded products (M_{comp}) were used to derive the average compression-molded masses to evaluate the process variability, keeping into consideration the standard deviations. Here, the expected theoretical ideal mass (M_{Ideal}) obtained for each polymeric sample was used as a standard to indicate any variability in processing results. The efficiency of mold filling (n) was calculated using the simulated and ideal masses in equation mass. Equation 1 can further be used to evaluate the performance efficiency of the process and identify the specific operating conditions under which the mold gets filled completely without the presence of air bubbles. These results helped investigate the process performance under distinct processing conditions using a single parameter associated with each processing condition. Furthermore, the presence of air traps was also studied using the air traps analysis in Moldflow Insight.

$$\eta\% = \frac{\boldsymbol{M\ Ideal}}{\boldsymbol{MComp}} * \mathbf{100} \tag{1}$$

Table 1: Range of the processing parameters utilized in the simulation study

	Range Variable Parameters			
Polymeric Material	**Mold temperature (°C)**	**Melt temperature (°C)**	**Mesh size (mm)**	**Initial Charge Geometry**
Polypropylene	30,40,50,60,70	200,210,220,230,240	4mm, 2mm	Cuboidal, Cylindrical
Polyethylene	20,30,40,50,60	200,210,220,230,240	4mm, 2mm	Cuboidal, Cylindrical
Polystyrene	30,40,50,60,70	210,220,230,240,250	4mm, 2mm	Cuboidal, Cylindrical
Polycarbonate	80,90,100,110,120	270, 280, 290, 300, 310	4mm, 2mm	Cuboidal, Cylindrical

2.1. Processing Parameters

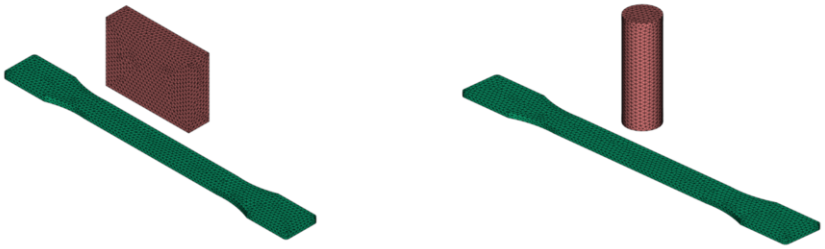
2.1.1. Initial- Charge Geometry

The amount of initial charge used highly impacts the amount of raw material required for a process. Hence, initial charge mass is an important factor that was taken into consideration for the scope of this simulation to determine the cost of the final product for industrial mass production. Equation 2 was used to calculate the initial mass of the charge.

$$i = V\rho + s \tag{2}$$

Where 'i' is the initial mass of the charge, 'V' is the volume of the product, 'ρ' is the density of the polymer and 's' is the percentage shrinkage of the polymer. The percentage shrinkage and required volume of the initial charge of the polymers used. The volumetric shrinkage of polymers under study was taken into account while designing the compression element to ensure that the final volume of the compression-molded product is equivalent to the product volume, i.e., around 7284.24mm^3 after mold filling, cooling, and solidification.

Different charge geometries are known to have different fill times. Hence, for the scope of this study, two types of charge geometry were used; cylindrical and cuboidal. Using the volume derived using equation 2, the dimensions of different charge geometries were calculated. All the charge geometries were designed in Solidworks 2019 and placed above the dumble-shaped product for the scope of the compression molding study (figure 6). The simulation results were then monitored to note the impact of charge geometry on the mold filling efficiency.



(a) **(b)**

Figure 3: (a) Rectangular Charge Geometry (b) Cylindrical Charge Geometry

2.1.2. Meshing

Moldflow Insight generates a two-dimensional or three-dimensional mesh to divide and annotate the domains in a mold or part design in the pre-processing stage. (12-14) In the study, the CAD models of the product and initial charge were converted into tetrahedral solid meshes before performing the required analysis. (15) Different mesh sizes i.e., 4mm and 2mm, were used to obtain results across different processing parameters as indicated in figure 5. The resulting geometry was divided into different numbers of elements which increased with a decrease in mesh size. The simulation

results were then used to study the process efficiency, time taken for simulations, and change in output.

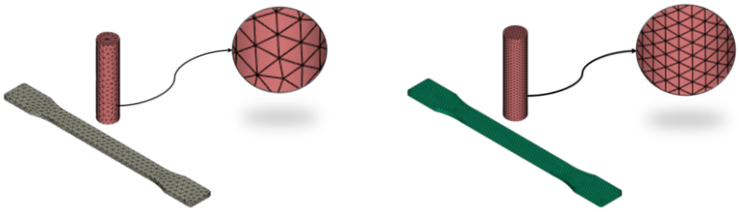


Figure 4. Different sizes of mesh element (a) 4 mm (b) 2 mm

2.1.3. Melt Temperature and Mold Surface Temperature

In this study, a range of mold and melt temperatures were studied to obtain the optimal range of processing. Furthermore, the impact of mold and melt temperature on process efficiency was also studied.

As per the simulations, the mass of the dumble-shaped products was used to evaluate the efficiency of the compression molding process with respect to the expected ideal mass of the product (as per density calculations) The compression-molded mass (M_{ccomp}) estimated in the simulations was compared with the expected ideal mass (M_{ideal}) of the product and the deviations between the data (δ) were calculated using equation 3. This was used to obtain the accuracy of the simulation results to obtain optimal processing conditions.

$$\delta\ \% = \{(\frac{Mideal}{M\ comp})- 1\} * 100 \quad (3)$$

3. Results

3.1. Effect of Processing Parameters

3.1.1. Initial Charge Geometry

In the simulation study, it was observed that cuboidal geometries required less material to completely fill the mold to avoid short shots, but the outputs indicated lower mold filling efficiency and higher degree of deviation from ideal expected product mass. On the other hand, cylindrical geometries offered higher degree of mold filling efficiency and less deviation from ideal expected product mass. Hence, becoming a more favorable initial charge geometry, as observed in figure 5 and figure 6. Further, semi-crystalline polymers (Polypropylene, Polyethylene) showcased higher level of deviations (30-40%) and low mold filling efficiencies than amorphous polymers (Polystyrene, Polycarbonate- 10-20%) when subjected to similar processing conditions and charge geometries. Small air traps were observed around the ends of the product geometry (Figure 7). Here, higher mold filling efficiencies indicate better utilization of raw materials, less energy consumption, better mechanical properties of product, and smaller air traps.

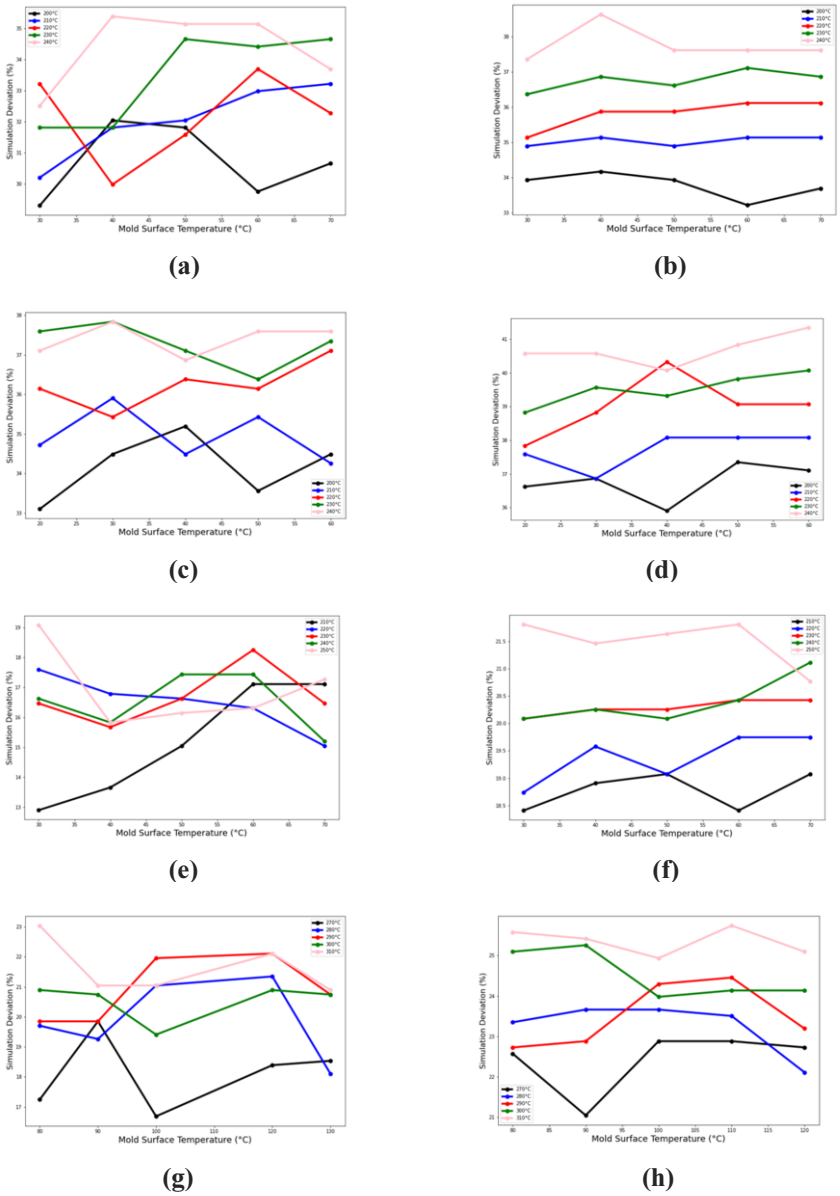


Figure 5: Simulation deviation with varying mold surface temperature and melt temperature with 4mm mesh size for studied polymers using cylindrical charge geometry (a) PP (c) PE (e) PS (g) PC and rectangular charge geometry (b) PP (d) PE (f) PS (h) PC

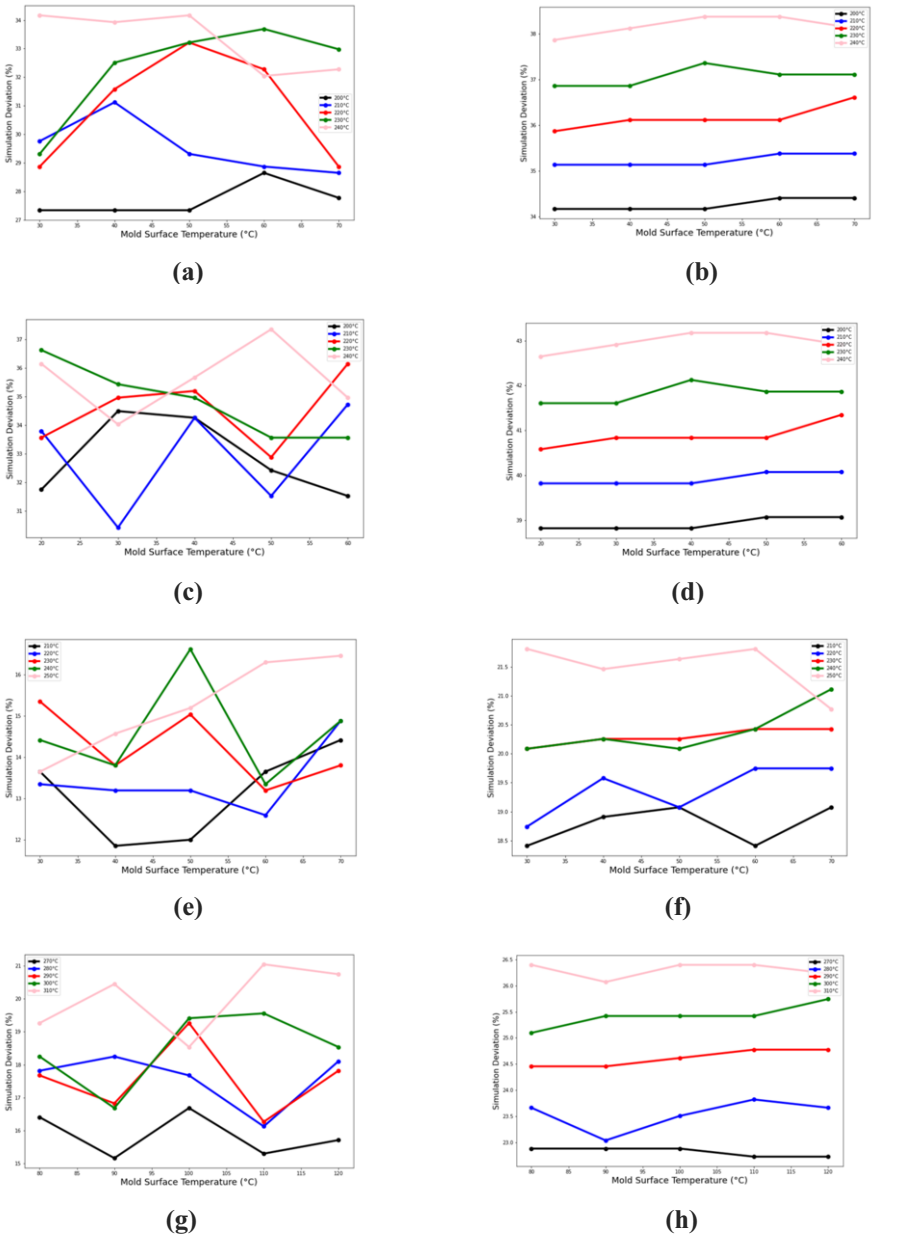


Figure 6: Simulation deviation with varying mold surface temperature and melt temperature with 2mm mesh size for studied polymers using cylindrical charge geometry (a) PP (c) PE (e) PS (g) PC and rectangular charge geometry (b) PP (d) PE (f) PS (h) PC

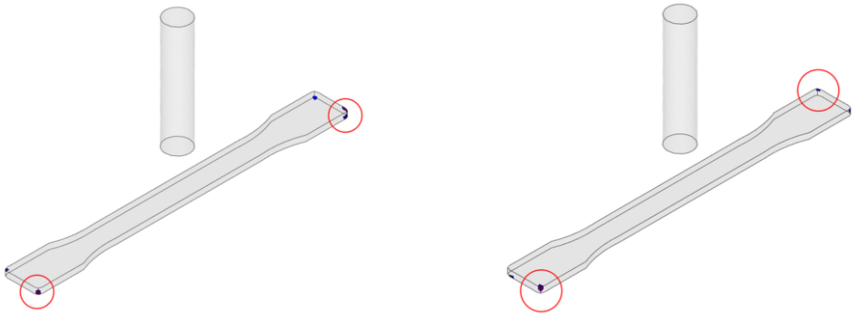


Figure 7: Air entrapments observed during simulation

3.1.2. Mesh Size

The simulation studies indicated that a tetrahedral geometry is able to obtain 97.1 % node refinement in complex geometries, but fails to refine nodes across corners and edges. It was observed that the level of mesh refinement or mesh density only has a very low impact (2-3%) on increasing mold filling efficiency for both semi-crystalline and amorphous polymers (Figure 5). In the cases studied, the mesh with highest level of refinement or highest mesh density was associated with the longest time for computation. Even as the accuracy of results increased slightly with mesh refinement, the time taken computation increased manifolds hence, increasing the processing time for the simulation. It was further observed that higher mesh refinement showcased more warnings during processing due to issues in node refinement across the edges of the products. Upon processing, the presence of air traps was more pronounced in lower mesh densities.

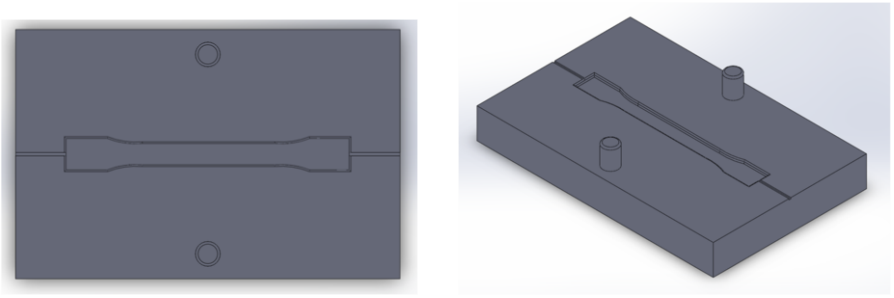
3.1.3. Melt Temperature and Mold Surface Temperature

In the study, it was observed that an increase in melt temperature of polymers around the processing range, increased the deviation percentage (5-10%) from ideal expected product mass (Figure 5 and figure 6). This increase in deviation percentage could be attributed to inferior mold filling efficiencies and increased air traps. It was observed that at lower temperatures, higher masses of samples were obtained, leading to higher average molecular weight in molded parts, which translates well into increase in impact resistance, shorter cycle time, lower energy consumption, and reduced air traps.

Further, the impact of increase in mold temperature was not very profound on mold filling efficiency and process deviation. In the simulations, it was observed that increase in mold surface temperatures in amorphous polymers showcased lower degrees of molded-in stress, which could produce better crack-resistance and fatigue performance. While in semi-crystalline polymers, no significant observations on molded-in were observed.

A mold designs without air vents (venting system) was evaluated in the study to determine an appropriate location for a venting system. Based on the simulation results

on air trap analysis, air traps are concentrations on the corners of the compression molded product, making it a region of concern. To prevent mold degradation which can lead to additional cost of processing and material wastage, a second mold design is presented that accounts for the presence of a venting system which is composed of channels (air vents). These channels are distributed along the end of the mold (female cavity). The virtual project of the air vents of the mold can be visualized in Figure 8. The vents were introduced into the mold with 29.65 mm length, 2 mm width, and 1 mm depth on Solidworks 2016. The design has been developed keeping in mind air trap analysis, mold filling efficiency, and process deviation percentages of semi-crystalline and amorphous thermoplastics. The model predicts that the new mold design with air vents could significantly increase the mold filling efficiency and reduce the presence of air vents.



(a) **(b)**

Figure 8: Mold cavity with air vent developed in SolidWorks CAD version 2016 (a) Top view (b) Isometric view

4. Conclusion

This simulation study through computational analysis has evaluated the effect of various processing parameters on the efficiency of the compression molding process for thermoplastics. Furthermore, the study analyzed the distribution of air traps across the product using Moldflow Insight 2019. As the simulation package in Moldflow Insight successfully located a region of concentration for the presence of air vents, mold modifications were presented to result in higher mold filling efficiency. The impact of charge geometry on ensuring proper raw material utilization was indicated via the simulation results. It was concluded that a cylindrical initial charge geometry offered higher mold filling efficiency with less raw material requirements in comparison to a cuboidal initial charge geometry. The simulations further showcased how higher mesh refinement could increase the mold filling efficiencies but also increase the computation time. The inadequacy of a tetrahedral mesh design along the edges of the product design was observed. The results indicated an decrease in mold filling efficiencies with increase in elt temperature which resulted in partial filling of the mold. Whereas, the impact of increase in mold surface temperature on melt temperature was not very significant. It was observed that all temperature and pressure combinations showed some degree of air entrapment. Using these results and the air traps analysis a new mold design was developed. The mathematical model used by

Moldflow Insight showcased great caliber in predicting the behavior of various polymers undergoing compression molding. Based on the results obtained, it can be possible to optimize the compression cycle and reduce material wastage by introducing a venting system in the mold design and appropriate processing parameters. The findings illustrated that computational simulations can be a powerful tool to aid decision-making and process optimization to be applied in compression molding processes.

5. Acknowledgements

The authors gratefully acknowledge Delhi Technological University, New Delhi, India for providing educational facilities and research infrastructure for the completion of this study.

References

[1] Spasojevic PM. Chapter 15 - Thermal and Rheological Properties of Unsaturated Polyester Resins-Based Composites. In: Thomas S, Hosur M, Chirayil CJ, editors. Unsaturated Polyester Resins [Internet]. Elsevier; 2019 [cited 2022 Apr 12]. p. 367–406. Available from: https://www.sciencedirect.com/science/article/pii/B9780128161296000156.

[2] Mohd Rawi S, Alias S, Wahid S. Design and development mini compression molding for teaching and learning. 2020 Jun 30;

[3] Wulfsberg J, Herrmann A, Ziegmann G, Lonsdorfer G, Stöß N, Fette M. Combination of Carbon Fibre Sheet Moulding Compound and Prepreg Compression Moulding in Aerospace Industry. Procedia Eng. 2014 Jan 1;81:1601–7.

[4] Asproiu S, Str E. INFLUENCE OF MOLD PROPERTIES ON THE QUALITY OF MOLDED PARTS. :8.

[5] Chauhan V, Kärki T, Varis J. Optimization of Compression Molding Process Parameters for NFPC Manufacturing Using Taguchi Design of Experiment and Moldflow Analysis. Processes. 2021 Oct;9(10):1853.

[6] Yadav A. Design and Analysis of Pneumatic Compression Moulding Machine. In 2019.

[7] Romanenko V, Duhovic M, Schommer D, Hausmann J, Eschl J. Advanced process simulation of compression molded carbon fiber sheet molding compound (C-SMC) parts in automotive series applications. Compos Part Appl Sci Manuf. 2022 Jun 1;157:106924.

[8] Sheikh M. DESIGN ANALYSIS FOR COMPONENTS OF PNEUMATIC INJECTION MOULDING MACHINE USING PRO-E NOVATEUR PUBLICATIONS INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]. 2015 Jan 1;

[9] Akinlabi O, Adeiza BA, Emmanuel OO. Development of Laboratory Hot Compression Mounting Press Machine. J Sci Res Rep. 2018 Feb 21;1–12.

[10] Rios A, Davis B, Gramann P. Computer Aided Engineering in Compression Molding [Internet]. [cited 2022 Apr 10]. Available from: https://www.semanticscholar.org/paper/Computer-Aided-Engineering-in-Compression-Molding-Rios-Davis/99aad409f546c033852c6359a90150000f0b94eb

[11] Song Y, Gandhi U, Sekito T, Vaidya UK, Hsu J, Yang A, et al. A Novel CAE Method for Compression Molding Simulation of Carbon Fiber-Reinforced Thermoplastic Composite Sheet Materials. J Compos Sci. 2018 Jun;2(2):33.

[12] Yang J, Xu Y. Warpage Analysis of Injection Molding Based on Mold Flow. Adv Mater Res. 2012 Jun 1;538–541:1192–6.

[13] Nie Y, Zhang H, Niu J. Optimization of the Injection Molding Process Parameters Based on Moldflow and Orthogonal Experiment. Key Eng Mater. 2013 Jul 1;561:239–43.

[14] Kiam TM, Pereira NC. Study of injection-compression molded part using CAE analysis. Polímeros. 2007 Mar;17(1):16–22.

[15] Lai JY, Wu JW, Song PP, Chou TY, Tsai YC, Hsu D. Hybrid mesh generation for the thin shell of thin-shell plastic parts for mold flow analysis. Eng Comput. 2021 Aug 28;

2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA) | 978-1-6654-5834-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCSEA54677.2022.9936302

Simulation Tools and Testbeds for Internet of Things(IoT): "Comparative Insight"

Abhishek Singh
Information Technology
Delhi Technological University
Delhi, India
abhishek.g4plus@gmail.com

Himanshu Nandanwar
Information Technology
Delhi Technological University
Delhi, India
himanshunandanwar9cm0@gmail.com

Anamika Chauhan
Information Technology
Delhi Technological University
Delhi, India
anamika@dtu.ac.in

[1] ***Abstract*—The recent years have seen major technological transformation with IoT in the center. The application of IoT span across multiple areas affecting our day to day life. This has set off major interest into application and associated research in this area. This nessecitates a condusive environment for design, build, test and analysis of IoT-based solutions to existing applications or for creating novel applications. This is provided by variety of simulators and testbeds. This paper provides an extensive analysis of proof-of-concept and real-world experimentations of commonly used simulators and testbeds. This paper provides an exhaustive review of various IoT simulation tools and testbeds that can provide new functionality for designing, modeling and analyzing the IoT problem and determining their solution. This will lead to improvised and optimal solution for improving quality of human life. According to the findings of the study, NS3, an open-access simulator, has been used in numerous studies and is widely used by researchers to solve IoT problems.**

***Index Terms*—Simulator, Testbeds, IoT applications, NS-3, FiT IoT-LAB.**

I. Introduction

Internet of Things(IoT) aims to develop and deploy solutions to build a smart environment, and these smart environments are made up of different components of the IoT ecosystem, having sensors, actuators, objects, and devices having sensing and communicating capability. Nowadays, almost everything we can configure to sense and build having communication capabilities to connect to the internet is associated with the IoT network.

There is almost every sector in which IoT is being used in the present, whether it is the health , education , manufacturing , energy , and agriculture sector each of the sectors is making use of IoT systems to make the process digitized by converting each sector from brick and motor to digital media. Integration of other technologies like Fog computing and Cloud computing can also provide the new dimensions to the IoT technology. It can help in the area of E-Health system, Face recognition system, traffic organization and engery control system [21].

Researchers are investigating more areas or sectors where IoT can be applied or trying to optimize the existing IoT devices by analyzing the models. Due to the diverse nature of IoT systems with various types of sensors and network topologies, implementation and designing a solution is complex. Researchers find it difficult to perform analysis, build and test their models and also, in the initial phase of design, it is difficult to build a prototype when there are multiple devices, each supporting a different protocol.

So to resolve this problem, simulation tools play a crucial role by providing proof-of-concept to the proposed solution of the researchers, where simulation tools help researchers test their hypothesis, perform experiments with changing the parameters and analyze the system performance.

However, only proof-of-concept is not sufficient, but verifying the proposed solution in a real-world environment will provide the basis for their proof-of-concept. So to solve the purpose of real-world experimentation, open testbeds are being used. They have been developed at some sites and have infrastructure support to test the firmware by deploying over the internet. They replicate real-world conditions by providing support for real-world sensor nodes and communication networks.

Researchers in the fields of IoT, while doing their implementation, need to verify their work. So, to demonstrate their proposed solution, they generally use different simulators, emulators, or testbeds. Selecting the appropriate type of simulator, emulator, or testbed that will be best suited for their domain of work will be an additional knowledge they need to have for verifying their proposed solution. So to resolve this issue of an extra burden on researchers, we opted to provide an exhaustive study on the IoT simulator and testbed.

This paper presents an in-depth discussion of frequently used simulation tools and IoT testbeds, covering their advantages and disadvantages. In the later section, we have discussed some of the open challenges which need to be addressed in the future.

A. Internet of Things (IOT):

Internet of Things(IoT), refers to the system of connected objects or groups of objects that have ability to collect data via sensors and transfer those collected data over a wireless network without any need of human intervention.

A "thing" in Internet of Things can be appliances of our home, electric motor, medical diagonastic tool or any device which can be fitted with sensors to receive the environment data and have capability to transmit over the internet to process it and infer some decisions on the gathered data IOT is not just about

[1]978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:35:23 UTC from IEEE Xplore. Restrictions apply.

connecting devices, but also in making decisions so that these devices operate in an optimized way and improve our lives.

B. IoT Characteristics:

Let us discuss some of the prominent and important characteristics of the IoT.

- Data: which is the basis for all the purposes and applications. In IoT domain everything circle around the data as the sensors collect the data and based on the collected data actions are triggered.
- Sensing: is necessary because that is how we get information from the object such as temperature,humidity,moisture etc, as IoT ecosystem will be comprising of different sensors to gather the data.
- Connectivity: is establishing proper infrastructure such that sensed data could be connected to the internet so that data flow can be established .
- Communication: is needed because we need to communicate that gathered information for processing and any decisions that are made out of that processing needs to be communicated back to devices.
- Intelligence: Just like how we need brains to receive information and make decisions, just like we need intelligence in IoT systems and devices with which based on the gathered data, analysis of that data is performed and based on this analysis certain decisions will be made and communicated back to the IoT system.
- Action: is the outcome of processing and analysis of the sensed data which will be in the form of changing the state of our physical environment or providing the insights of that environment .

  Atlast we discuss the IoT ecosystem consists of multiple components that allows different kinds of customers to access their IoT devices.

C. IoT Applications

IoT finds its applications wherever sensors can be embedded and they can connect to the internet for transferring the sensed data.There are different sectors or areas where IoT has already been implemented and plays an important role in optimizing process flow and making life easier and simpler[19].

- Health Care: One of the major sectors is healthcare which is basically using different types of sensors to keep track of the health status of their patients .
- Smart City: is also a major application which is connecting not only citizens of it but also different objects inside the city are getting connected with each other to mitigate energy usage ,waste management,traffic management and a few other things as well.
- Smart homes: are a part of Smart city applications as smart cities need a smart home as they will build a network of these smart homes which in turn will contribute to smart city development.
- Smart Wearables: like watches and different health monitoring devices are present which will be monitoring different health parameters like heart rate, oxygen saturation percentage, blood pressure so these health parameters are being monitored and analyzed with the help of different IoT components.

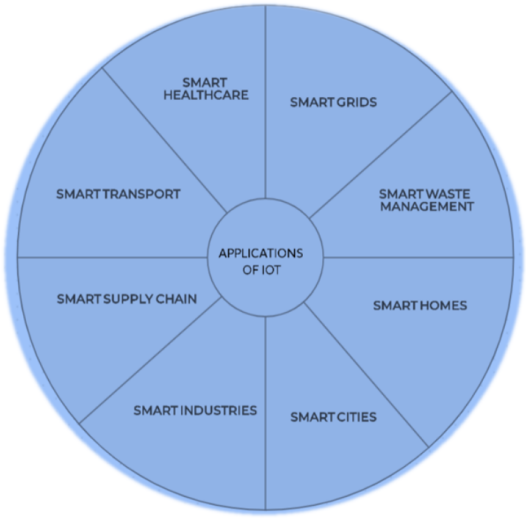


Fig. 1. Applications of IoT

D. IoT Architecture

Talking about IoT architecture, there is not any single design out there which has been agreed universally. But out of different architectures being used 3-Layered and 5-Layered architecture is being used widely.

So the first layer which is common in both layered architectures is the perception layer which connects the physical world with the digital infrastructure.

The next layer in continuation is the Network layer which focuses on building network infrastructures with the use of different existing technologies of communication and algorithms for routing the sensed data from the perception layer to the next layer.

Now the gathered and communicated data needs to be analyzed, filtered, processed and stored so that any useful information could be made out of it, this is the working of the middleware layer. This layer is present in only 5-layered architecture only.

Now the preprocessed data will have to be presented before the consumer or end-user for that purpose application layer is being implemented which will benefit the end-user by showing the functionality for which it was intended to be.

On top of all these layers stands the business layer, any device or product when it is working on a large scale environment, then they need to be analyzed with the help of different visualization techniques in order to improve the end-user experience. This layer is only implemented in 5-layered architecture.

In general designing any idea on paper by assuming different parameters or constraints does not guarantee its successful completion. So implementing these ideas in the real world to validate their concept is quite a tiring and complex task.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:35:23 UTC from IEEE Xplore. Restrictions apply.

So to minimize this implementation part, time, and resources we make use of simulators. In general designing any idea on

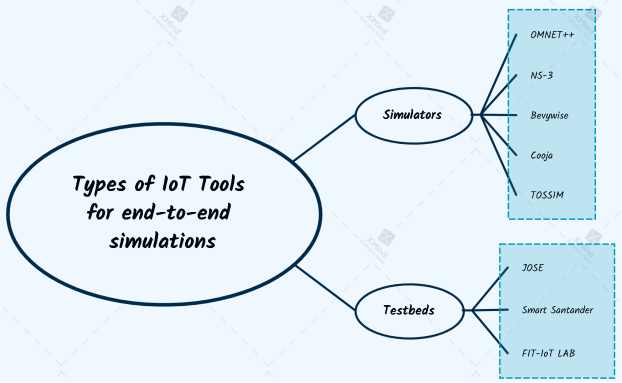


Fig. 2. IoT Tree for IoT Simulators and Testbeds

paper by assuming different parameters or constraints does not guarantee its successful completion. So implementing these ideas in the real world to validate their concept is quite a tiring and complex task. So to minimize this implementation part, time, and resources we make use of simulators.

II. Objective of the Paper

- To gain comprehensive knowledge of various IoT simulation tools and testbeds.
- To compare various simulation tools and testbeds on multiple parameters from the standpoint of functionality.
- To address the various open challenges of IoT simulators and testbeds and provide suggestions to overcome those challenges.

III. End-to-End IoT domain platform

A. IoT Simulators

Wikipedia defines a simulator as: " the imitation of the operation of a real-world process or system over time". Basically, a simulator is a software program designed to perform the task which is meant to be deployed in the real world.
Now by saying "designed to perform the task", we mean the functionality of the implemented model . Simulators make it easy for researchers to visualize the performance or tweak the parameters to optimize the results.
There are a huge number of simulators available today for performing different tasks in the IoT area. These simulators focus on different areas based on different layers of IoT architecture like many of the simulators focus on network layers so models which require new network topologies or want to test new protocols make use of such simulation tools. In Spite of these, there are few which we can call full stack simulators which perform end to end tasks so researchers can make use of such simulation tools.

1) OMNET++: OMNET++ is a sought-after network simulation framework that is non-commercial and also it is being used at a large scale in doing wireless sensor network(WSN) research. Not only this but also it is used free of cost in educational institutions. This is being implemented by using the C++ simulation framework and library.This is of discrete event simulation(DES) type model.It makes use of CSMA, TDMA, and Hybrid routing protocols for simulating different protocols and it has dynamic network topology support.

2) NS-3: NS-3 is a simulator of the NS series which majorly focuses on network simulation and can be used for performing a realistic simulation of wireless sensor networks(WSNs) for duplicating the perceptual layer of IoT architecture. This makes use of different protocols such as FTP, TCP, UDP, and DSR.This makes use of Wifi, Wimax, LR-WPAN routing protocols. This is being implemented in two languages namely C++ and python.

3) Cooja: Cooja is a network-based simulator that is majorly focused on Contiki OS applications and wireless sensor networks (WSN).Cooja simulator having motes have access to most of the standards and protocols which is being implemented by Contiki OS which gives researchers privilege to replicate the real scenarios by incorporating the most used application layer protocols like message queue telemetry transport(MQTT) and constrained application protocol(CoAP).This simulator is being implemented in the Java programming language.This uses IPV4 and IPV6 as the routing protocol and supports dynamic network topology.

4) TOSSIM: TOSSIM is abbreviated as TinyOS Simulator and is used for simulating different TinyOS applications and services.TinyOS is an operating system that is designed keeping in mind the low memory usage and smaller size and low power consumption as there are certain areas or situation which requires such kind of operating systems such as wireless sensor network(WSN). So all the applications designed using this OS can be simulated by the TOSSIM.TOSSIM comes with graphical user interface(GUI) support which helps to monitor, control and debug ongoing simulations.This is being implemented in C and C++ programming languages.It uses CSMA routing protocol and supports static network topology.

5) Bevywise: Bevywise IoT simulator is the graphical user interface(GUI) based data simulator tool that majorly focuses on the message queue telemetry transport(MQTT) and IoT application.BevyWise IoT Simulator is a sophisticated yet user-friendly MQTT simulation tool that allows you to simulate tens of thousands of IoT devices [20]. It helps to analyze the performance level of these applications by creating virtual IoT networks and devices, publishing different events with normal or complex JSON payloads. This simulator is being implemented in C, Python, and Java programming languages. This simulation tool uses MQTT as a routing protocol and supports dynamic network topology.

So as we discussed above different domains of simulation tools we have picked a few simulation tools and performed comparative analysis on certain parameters which are being given in Table I. The parameters on which analysis is done are target area, Supported network band, routing protocol used, network topology support, IoT protocol coverage, Programming Language used, IoT architecture layer, supports GUI, and Licensed or Open.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:35:23 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF DIFFERENT IOT SIMULATION TOOLS

Simulator	Target Area	Supports (2G/3G/4G) Network	Routing Protocol used	N/W Topology Support	IoT Protocol Coverage	Programming Language	IoT Architecture layer	Supports GUI	Licensed/ Open
OMNET++	Network	✓	CSMA, TDMA, Hybrid	Dynamic	Manual Extension	C and C++	Perceptual Network	✓	Licensed
NS-3	Network	✓	Wifi, Wimax,LR-WPAN	Dynamic	802.15.4 LoRa-WAN	C++ and Python	Perceptual Network	✗	Open
Bevywise	Big Data Processing and Data Analysis	✓	MQTT	Dynamic	Real-Time Services	C, Python and Java	Application Network	✓	Licensed
Cooja	Network	✓	IPV4, IPV6	Dynamic	Contiki OS (All the protocols supported by this OS)	Java	Perceptual Network	✓	Licensed
TOSSIM	WSN Protocols and application	✗	CSMA	Static	Tiny OS supported protocols	C and C++	Perceptual Network	✗	Open

B. IoT Testbeds

Simulators provide researchers with a feasibility study of their proposed solution, now there arises a need to perform their proposed solution in the real world by using hardware and networks through experiments. In the past years, researchers have performed these experimentation tasks by implementing their proposed solution. But it required additional resources and knowledge which resulted in extra overhead for them.
However, this problem is now being solved with the availability of a number of open testbeds, now researchers have availability of already deployed hardware and network for performing experiments.Nowadays most researchers are relying on open testbeds for experimentation-based prototyping for verifying their research efforts.
According to, Papadopoulos et al. [8], more than two-thirds of the 596 analyzed studies have verified their research concept using experimentation testbeds after validating it through simulation tools.
There are three mainly testbeds which are being used in research community are:

1) Smart Santander: Smart Santander mainly focuses on the IoT application and services which are targeted at smart city domains. This testbed initially starts from a city in Spain named Santander, but after a few years, later locations across countries namely Serbia, Germany, United Kingdom(UK) were also included. Originally this testbed has a very much mixed set of sensors which includes WSN, parking sensors,RFID, and QR codes. These sensors consist of a mix of nodes in which some nodes are fixed and a few of them are mobile nodes that are being fitted into vehicles like buses and taxis. This testbed is operating with around 20000 sensor nodes. This has been implemented by using Java and Javascript programming languages.

2) FIT IoT-LAB: The FIT IoT-LAB [10,9] which is an acronym for Federated IoT testbed for large scale experimentation, is a large scale and open access testbed that is currently operating from six different sites in France. The platform has a large number of distributed nodes that provide users with a variety of accurate scientific tools for designing, developing, and tuning IoT-related things [22]. This testbed has more than 2500 wireless sensor nodes(WSNs) that perform communication over IEEE 802.15.4 protocol.Programmable robots provide mobility to these testbeds which follow circuit-like mobility models. It has been implemented by using the Java programming language.

3) JOSE: The JOSE which is an acronym for Japan-wide orchestrated sensor environment, this testbed is located in Japan. This provides the functionality to support more than one IoT service at the same time running parallel by providing separate sensor networks and virtual cloud infrastructure for each service. This testbeds perceptual layer supports a large number of sensors that sense data and transmit via a standard protocol such as IEEE 802.14, 3G, and LTE. Currently, this testbed is operating by providing support of 27 IoT services. This has been implemented in C, Java, and Javascript programming languages.

A comparison of three majorly used open testbeds is being given in Table II. The table comprises parameters like Number of nodes, Variety of nodes, Mobility, programming language used ,API integration provided, and focused area.

IV. ADVANTAGES AND DISADVANTAGES

OMNET++ supports a better graphical user interface(GUI), which allows users to easily perform the simulation.One of the major concerns with this simulator is compatibility issues. Whereas NS-3 provides support for a large variety of sensors or wireless sensor nodes(WSN).There is only one internet

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:35:23 UTC from IEEE Xplore. Restrictions apply.

TABLE II
COMPARISON OF DIFFERENT IOT TESTBEDS

Testbed	Number of Sensors	Variety of Node	Mobility	Programming Language	API Integration	Focused Area
FIT-IoT LAB	More than 2,700	Yes(support different types of nodes)	Robot driven	Java	REST	Analysis of protocols and algorithms performance
Smart Santander	Approximately 20,000	Yes(support different types of nodes)	mix of robot and vehicle driven	Java and JavaScript	REST	Smart City application and services
JOSE	Number is huge (exact count not known)	Yes(support different types of nodes)	Not Known	C, Java and JavaScript	SOAP	Real-time execution of multiple IoT services.

protocol i.e IPv4 which it supports. It does not support a graphical user interface(GUI). Cooja provides support for a wide variety of protocols which are supported by wireless sensor nodes(WSN).It is one of the most used simulators by researchers.It requires a higher configuration for performing simulation tasks.

One of the novel advantages of using TOSSIM is that it provides very accurate results.It has the issue of getting erroneous compilation results.Bevywise is one of the widely used simulators for industrial IoT purposes.It has very good support for the graphical user interface(GUI).It is time-consuming, as it takes time while performing analysis.

FIT IoT-LAB provides support for incorporating customized hardware into their experiments.It makes use of heterogeneous sensors/wireless sensor nodes(WSN).Only robots are used for providing mobility which does not provide effective replication of real-world scenarios. Whereas Smart Santander is best suited for areas like Smart Cities, Smart Homes, and Smart grids.It lacks support other than the areas like smart cities and smart homes.

JOSE provides faster computing and access by providing better application layer infrastructure which is spread across five data centers.It only focuses on outdoor environment areas but could not provide better support for projects or areas having laboratory-like experiments.

V. OPEN CHALLENGES

Future of IoT applications and services largely depends on the selection of tools to configure, develop, and test the IoT applications and services. Simulators and Testbeds play an important role in the development of IoT applications. Simulators help to verify the concept of the proposed solution design and testbeds provide experimental proof to the proposed solution. Next, we discuss some of the challenges linked with existing simulators and testbeds which need to be worked upon in the future.

- As discussed in [10], simulators provide varying levels of support for the same layer of IoT architecture. For instance, OMNET++ still does not provide support for IEEE 802.15.4 which is one of the popular protocols for wireless sensor networks (WSN) which is categorized under the perceptual layer of IoT architecture. There are few simulators that provide support for the perceptual layer and media access control(MAC) layer but they fail to implement the support for the application layer.So, there need to be simulators that focus on implementing common IoT standards and protocols.
- Another issue is that the current simulators lack the support for end-to-end simulation service. There are simulators that try to focus on one or two layers and there are some which try to implement whole IoT architecture but they have to make use of abstraction for the layers which they do not make use of. Like Cooja [9], which is simulating WSN, they have implemented the perceptual layer in detail and upto some extent they implemented the network layer as well.So future work will be to provide an end-to-end simulation of IoT applications by maintaining the transparency of interactions across all layers.
- The lack of effective energy measurement and analysis tools is another challenge in the area of IoT simulators. Like in the paper by Al-Roubaiey at el. [2], they have implemented the extension named POWERTOSSIM for TOSSIM simulators for performing energy measurement and analysis tasks. But integrating different extensions will also be overhead for researchers so there need to be specific simulators that will provide end-to-end support with energy measurement.
- Due to the rapid increase in the number of IoT devices, there is an increase in the number of sensors or actuators associated with particular cloud infrastructures. So there is a need to implement the simulation framework which will be able to take loads of huge numbers of sensors and actuators at the same time while performing the computation within the given deadline.
- the IoT devices are used in use cases like smart cities, smart vehicles, energy management, etc. These all use cases require a mix of sensors and actuators. That's why simulation frameworks need to be focusing on this diverse nature of sensor nodes and actuators.

VI. DISCUSSION

We have presented a very detailed review of the simulators and IoT testbeds. As we have discussed in earlier sections,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:35:23 UTC from IEEE Xplore. Restrictions apply.

simulators form state of art for any sort of assumption and hypothesis which is being studied by various researchers.
Nowadays there is an increasing trend among the researcher's community to make use of IoT testbeds for experimentation-based prototyping which means simulators are used for concept validation and testbeds for verifying those validated concepts.
OMNET++ is based on a discrete event simulator(DES) model for wireless sensor networks(WSN). It is generally based on C++ programming language/libraries. Similarly NS-3 based discrete event simulation(DES) model works on C++ programming language, but it uses 802.15.4 LoRa WAN. OMNET++ helps us in building network simulators based on the Eclipse framework. On the other hand, NS-3 supports various protocols like Transmission control protocol(TCP), User datagram protocol(UDP), File transfer protocol(FTP), Dynamic source routing protocol(DSR). OMNET++ is object-oriented and only manual extensions can be converted to the cyber resilience simulator. OMNET++ is assembled into major components and techniques using higher-level language which is based on a large scale of operations.
Cooja is based on universal platform support, it has high scalability/protocol optimization/dynamic network topology. It is based on a multicast network. Cooja is a built-in ContikiOS supported by IoT standards, cyber resilience (custom extensions). Cooja operates on three levels of abstractions.
TOSSIM is a wireless sensor network(WSN)/IoT-based simulator based on sensor observation service. It was developed in the year 2003 by the UC Berkeley TinyOS project group. It supports a graphical user interface(GUI) unlike the NS-3 simulator.
Bevywise has used info based computing environment for the user and a functional testing environment on the cloud. It provides end to end solution for smart cities service domain applications. It supports the MQTT protocol.
NS-3 is widely used for testing different routing protocols and networking topologies whereas Bevywise can be used in industrial applications using IoT technology. TOSSIM is mainly focused on the devices which support TinyOS whereas Cooja is focused on the devices using ContikiOS . Testbeds are based on real-time experimentation. This paper presents three types of testbeds.Testbeds uses real-time dimensions, sensors to interact with the different operating system(OS).
FIT IoT-LAB is based on a different spectrum of small-scale operations, it can be used in broadcasting different objects/things and help to create a large-scale framework based on the 802.15.4 LoRa WAN.
Smart Santander is based on IoT mobile sensing. It provides the facility to experiment and provide an environment to analyze and evaluate IoT applications based on smart cities. It consists of heterogeneous nodes. It supports remotely programming to researchers to develop and deploy new framework architecture based on service provisioning and experiments.

We have plotted line plots in figure 3 and 4 for simulator tools namely OMNET++, NS-3, Cooja, TOSSIM, Bevywise,

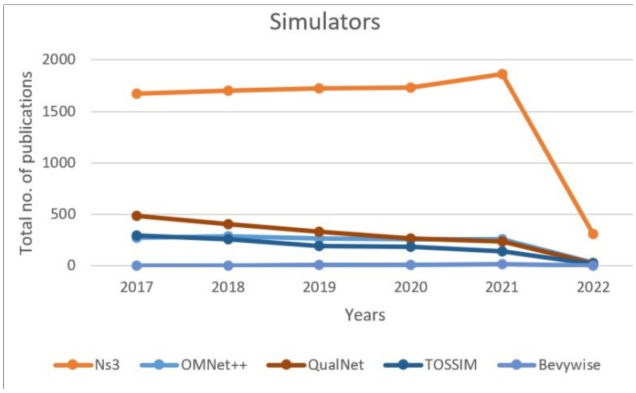


Fig. 3. Year-wise total number of publications for different simulators

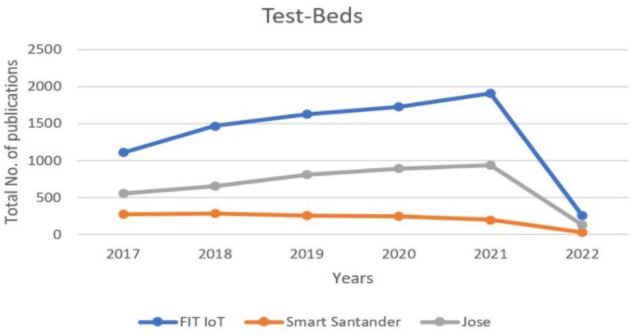


Fig. 4. Year-wise total number of publications for different IoT Testbeds

and IoT Testbeds namely FIT IoT-LAB, Smart Santander, JOSE which shows the data for a number of publications for last six years.

VII. Conclusion

As this study shows, the IoT paradigm has been the focus of major research during the preceding half-decade. Given that the internet of things paradigm is the result of the convergence of various pre-existing technologies, such as wsn, internet, and cloud computing, the path to conducting full-cycle iot research is not always straightforward. The simulation tools used in this study were chosen based on the criteria for locating appropriate components for the design.
This study examined five simulators and three testbeds and found them to be useful for on and IoT. This research will assist researchers in building, constructing, and selecting appropriate simulators and testbeds based on the needs of the IoT ecosystem. The first factor to consider when determining whether to utilize a simulator or testbeds in the IoT environment and whether they are appropriate for all phases of testing under diverse conditions. Lack of support for IoT infrastructure standards, effective iot service simulators, and large-scale interoperable testbeds are all issues that must be addressed in the future, especially given the increasing complexity of emerging IoT implementations. In the future, an ideal algorithm can be employed in the IoT layers architecture to overcome the quality of service issue.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:35:23 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Zeng, Xuezhi, Saurabh Kumar Garg, Peter Strazdins, Prem Prakash Jayaraman, Dimitrios Georgakopoulos, and Rajiv Ranjan. "IOTSim: A simulator for analyzing IoT applications." Journal of Systems Architecture 72 (2017): 93-107

[2] Al-Roubaiey, Anas, and Hamdi Al-Jamimi. "Online Power Tossim Simulator for Wireless Sensor Networks." In 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), pp. 1-5. IEEE, 2019.

[3] Lin, Jenn-Wei, Joseph M. Arul, and Jia-Ting Kao. "A bottom-up tree based storage approach for efficient IoT data analytics in cloud systems." Journal of Grid Computing 19, no. 1 (2021): 1-19.

[4] Adjih, Cedric, Emmanuel Baccelli, Eric Fleury, Gaetan Harter, Nathalie Mitton, Thomas Noel, Roger Pissard-Gibollet et al. "FIT IoT-LAB: A large scale open experimental IoT testbed." In 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), pp. 459-464. IEEE, 2015.

[5] Teranishi, Yuuichi, Yuki Saito, Sakae Murono, and Nozomu Nishinaga. "JOSE: An open testbed for field trials of large-scale IoT services." Journal of the National Institute of Information and Communications Technology 62, no. 2 (2016): 151-159.

[6] https://www.itrelease.com/2018/07/what-is-tinyos-and-its-applications/

[7] https://www.bevywise.com/iot-simulator/help-document.html

[8] Papadopoulos, Georgios Z., Antoine Gallais, Guillaume Schreiner, Emery Jou, and Thomas Noel. "Thorough IoT testbed characterization: From proof-of-concept to repeatable experimentations." Computer Networks 119 (2017): 86-101.

[9] Patel, N. D., B. M. Mehtre, and Rajeev Wankar. "Simulators, emulators, and test-beds for internet of things: A Comparison." In 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 139-145. IEEE, 2019.

[10] Chernyshev, Maxim, Zubair Baig, Oladayo Bello, and Sherali Zeadally. "Internet of things (iot): Research, simulators, and testbeds." IEEE Internet of Things Journal 5, no. 3 (2017): 1637-1647.

[11] Sanchez, Luis, Luis Muñoz, Jose Antonio Galache, Pablo Sotres, Juan R. Santana, Veronica Gutierrez, Rajiv Ramdhany et al. "SmartSantander: IoT experimentation over a smart city testbed." Computer Networks 61 (2014): 217-238.

[12] Salama, Maria, Yehia Elkhatib, and Gordon Blair. "IoTNetSim: A modelling and simulation platform for end-to-end IoT services and networking." In Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing, pp. 251-261. 2019.

[13] Manivannan, T., and P. Radhakrishnan. "A Comprehensive Analysis of Simulation Tools for Internet of Things." Solid State Technology 63, no. 5 (2020): 461-471.

[14] Ojie, Ehizojie, and Ella Pereira. "Simulation tools in internet of things: a review." In Proceedings of the 1st international conference on internet of things and machine learning, pp. 1-7. 2017.

[15] D'Angelo, Gabriele, Stefano Ferretti, and Vittorio Ghini. "Simulation of the Internet of Things." In 2016 International Conference on High Performance Computing Simulation (HPCS), pp. 1-8. IEEE, 2016.

[16] Gomez, Carles, M. Kovatsch, H. Tian, and Z. Cao. "Energy-efficient features of internet of things protocols." draft-ietf-lwigenergy-efficient-06 (trabajo en curso) (2017).

[17] Kumar, R. Praveen, and S. Smys. "A novel report on architecture, protocols and applications in Internet of Things (IoT)." In 2018 2nd International Conference on Inventive Systems and control (ICISC), pp. 1156-1161. IEEE, 2018.

[18] Henderson, Thomas R., Mathieu Lacage, George F. Riley, Craig Dowell, and Joseph Kopena. "Network simulations with the ns-3 simulator." SIGCOMM demonstration 14, no. 14 (2008): 527.

[19] Nandanwar, Himanshu, and Anamika Chauhan. "IOT based Smart Environment Monitoring Systems: A Key To Smart and Clean Urban Living Spaces." In 2021 Asian Conference on Innovation in Technology (ASIANCON), pp. 1-9. IEEE, 2021.

[20] John, A., T. Ananth Kumar, M. Adimoolam, and Angelin Blessy. "Energy management and monitoring using IoT with CupCarbon platform." In Green computing in smart cities: Simulation and techniques, pp. 189-206. Springer, Cham, 2021.

[21] Walia, Agampreet Kaur, Amit Chhabra, and Dakshraj Sharma. "Comparative Analysis of Contemporary Network Simulators." In Innovative Data Communication Technologies and Application, pp. 369-383. Springer, Singapore, 2022.

[22] Shetty, Spoorthi P., and Udaya Kumar K Shenoy. "Performance of RPL Objective Functions Using FIT IoT Lab." In International Conference on Innovative Computing and Communications, pp. 623-630. Springer, Singapore, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:35:23 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/355477063

Solar Panels Crack Detection using Overhead Images

Article · October 2021

DOI: 10.22214/ijraset.2021.38532

CITATIONS
0

READS
56

3 authors, including:

Saksham Checker
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Saksham Checker on 30 October 2021.

The user has requested enhancement of the downloaded file.

iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** X **Month of publication:** October 2021

DOI: https://doi.org/10.22214/ijraset.2021.38532

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

Solar Panels Crack Detection using Overhead Images

Saksham Checker[1], Saksham Gupta[2], J. P. Kesari[3]
[1, 2]*Department of Applied Physics, Delhi Technological University*
[3]*Department of Mechanical Engineering, Delhi Technological University*

Abstract: ***Many countries like India have been working on plans to shift from conventional energy sources to renewable, and solar energy is one of them. The technology of Solar Panels faces defects on a large scale. These defects can be within the cells or on the panel as a whole. Management authorities require automated systems to detect the physical faults on the solar panels. These faults can either be detected by physically looking at the panels or through the energy supply, which requires Machine Learning models. This paper proposes a model with 95.34% accuracy that can be deployed on drones and automatically check for physical damage on the panels. We collected the dataset manually from the internet, cleaned and split it into training and validation parts. Data augmentation was performed to get a better view of the functioning of the model.***
Keywords: Solar Panels, Solar Energy, Renewable Energy, Deep Learning, Machine Learning.

I. INTRODUCTION

There has been a sudden increase in the use of Solar Panels due to the rise in energy prices and a large amount of greenhouse gas released during the production of Energy. With all these issues raised in using non-renewable resources, the awareness of using renewable resources has been increased [1].

Along with other countries, India has tremendous energy potential with the increase in energy needs. The Jawaharlal Nehru National Solar Mission (JNNSM) initiative aims to generate 20GW of power by 2022 [2]. India's perspective in Solar Energy covers both Rural and Urban requirements like Solar Powered Battery well, Solar Powered Vehicles, Solar-powered food preservation facilities, and many more [2].

Looking at the rapidly increasing use of Solar Energy, there is a need to detect faults in solar panels. Many studies cover both Physical as well as statistical methods of fault detection. One of the methods is Therma Image Characterization of solar panel hotspots [3][4]. As in solar panels, the failures occur as hot spots. The poor thermal conductivity of a cell is when the material has reached 100 degrees Celsius of temperature, micro defect in cells with 71.3 degrees Celsius of temperature, and deformation of resins at 136.1 degrees Celsius of temperature [3]. The faults can also be analysed using electricity production monitoring, where produced energy anomalies can conclude to different defects [1]. The Variation in the IV characteristics curve can also help detect PV module degradation [5].

The physical analysis of fault detection includes the visualization of faults in solar cells. Deep learning models like Dense CNNs are used in image classification in the same [5]. In addition to image processing techniques, many pattern recognition methods can be applied to visible defects of the PV modules [5]. Deep Convolutional Neural Networks can also be used to find the surface defects in solar panels.

Our study proposes a Convolutional Neural Network-based model to detect the surface defects of solar panel images. Section 2 covers the related work done by other researchers in solar panels and their defects and faults detection. Section 3 and 4 cover the proposed detection and the experimental results, respectively, followed by the Conclusion and future work in section 5.

II. RELATED WORKS

There have been many studies on Solar Panels and their fault detection due to the gradual increase in the usage of Solar Energy as a replacement of Conventional methods of Energy.

The study [2] covers India's vision and perspective on the use of Solar Energy and concludes the country's capability of solar energy production. It covers the schemes being used by the government and the steps that are important for the better output of solar energy. They also sum up the advancements and encouragements of solar energy generation and usage.

The authors of [1] discussed the methods for detecting malfunctions in solar panels using electricity production monitoring. They proposed a two-stage approach to see the patterns. In the first stage, they selected a representative sample of the production processes in a controlled state. In this stage, they detected the global outliners and local anomalies.

©IJRASET: All Rights are Reserved

Following in stage two, they monitored the PV energy production of a new day d. These energy productions were monitored for several years. They have used six sets of PV panels for three years. The proposed model was able to detect minimal changes in energy output.

Thermal image processing can also be used to analyze the defects of solar panels. The study [4] explains the basic working of solar panels and explains the laws of thermology. They present a detailed report on the identification of defects through visual and infrared thermology. They analyzed the temperature variation in different natural conditions, like with panels within the shades of a tree and without. They give a great visualization using 3D temperature plots with MATLAB image processing. Electroluminescence images can also be used, which provide better analysis, as mentioned in [6]. They used the EL dataset obtained from different solar farms installed in the United Kingdom. Their dataset had a variety of different types of solar panels. They characterized four various faults: Crack A, Crack B, Crack C, and Finger failure. They used Support Vector Machines and random forest algorithms and compared them.

A similar type of ELPV dataset was used by [5] in which they proposed a light CNN architecture in EL image classification. The framework works on automatic PV defect detection in the field. They used VGG structure-based layers to train the model and work on fault detection. They used their experimental setup for testing the framework. The authors of [7] also proposed a CNN model on the dataset having photovoltaic module cells in electromedicine images. Their model predicted the defect probability using a CNN model. They tested different models to test their model.

III.PROPOSED METHODOLOGY

To classify whether the solar panel from the surface, a Convolutional Layers based model is thus proposed. The images are classified into two classes, Normal and Broken. This section explains the dataset collected, the building blocks of the proposed model, and the final model, and the algorithm used.

A. Dataset

Initially, 1314 images were taken from the internet. This set consisted of 632 pictures of broken solar panels and 682 photos of working for solar panels. These images were added to two different folders named broken and regular. It was observed that these images had many photos that had many things other than solar panels., these images were manually cleaned, resulting in 338 standard images and 165 broken solar panel images. All the photos were in jpg format. Figure 1 shows images of broken and figure 2 as typical panels.

Fig. 1 Sample overhead images of Solar Panels without cracks

Fig. 2 Sample overhead images of Solar Panels with cracks

©IJRASET: All Rights are Reserved

B. Building Blocks of the Model

The layers applied in this model are used from the Keras library of TensorFlow.

1) *Convolutional Layer*: The convolution process could be considered an image convolution, and the convolutional layer is the convolution output of the previous layer. Figure 3 shows the process of Convolution acted upon a nxn input layer with mxm filter [8] [9].

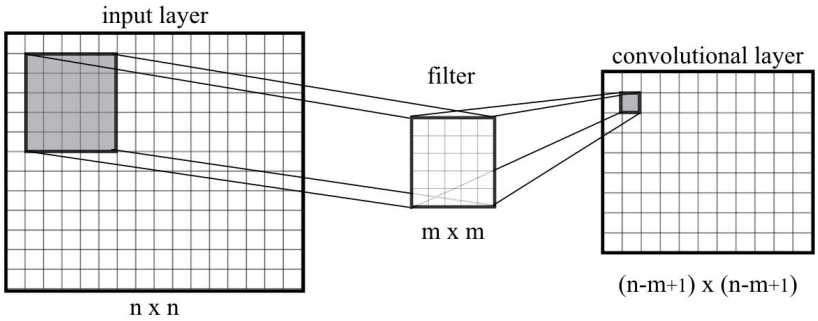


Fig. 3 Convolutional Process

2) *Sampling Layer:* This process is also known as MaxPooling. Pooling can be understood more clearly by the process in figure 4. This layer is applied after a convolutional layer [8] [9].

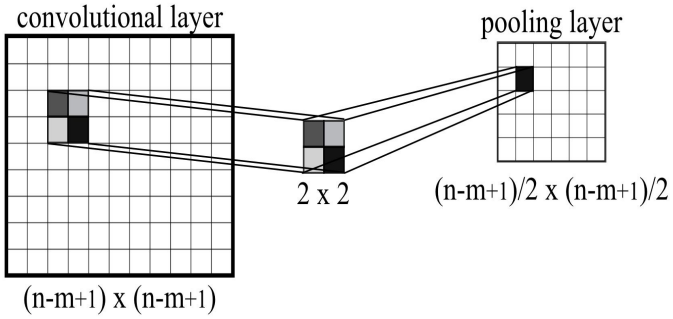


Fig. 4 Sampling Process

3) *Fully Connected Layer:* This layer consists of many perceptrons which individually act as a logistic regression unit. It has input, an activation unit, and output generated after the activation function application on the information. Figure 5 shows the working of a perceptron. *[8]*

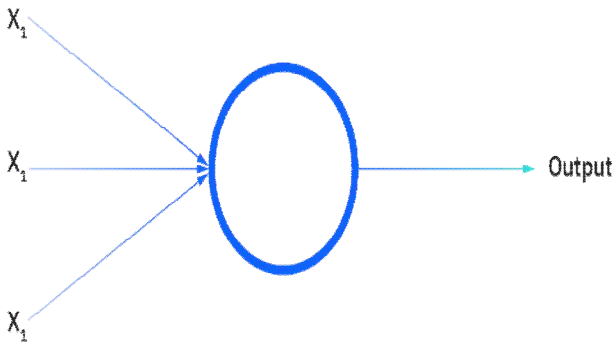


Fig. 5 Perceptron of a Fully Connected layer

4) *Activation Unit:* All Layers of the Network as ReLU as activation units except the last fully connected layer [8]. Equation 1 explains the ReLU function.

$$ReLU: f(x) = \max(0, x) \quad (1)$$

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

C. Data Augmentation

Data augmentation is applied to smaller datasets to generate larger datasets using many parameters like rescaling, rotation, shifting, and flipping [10]. Data augmentation is applied to the cleaned dataset in this paper.

D. Model Visualization and Algorithm

All the Layers mentioned above form the proposed model, as shown in the figure below. The images from the dataset are reshaped to 82 x 82 x 3 to be fitted in this model. Figure 6 shows the structure of the proposed sequential model.
Figure 7 explains the step by step algorithm for the pre-processing and training on the clean dataset.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 80, 80, 24)	672
dropout (Dropout)	(None, 80, 80, 24)	0
max_pooling2d (MaxPooling2D)	(None, 40, 40, 24)	0
conv2d_1 (Conv2D)	(None, 36, 36, 64)	38464
flatten (Flatten)	(None, 82944)	0
dense (Dense)	(None, 1024)	84935680
dense_1 (Dense)	(None, 64)	65600
dense_2 (Dense)	(None, 1)	65

Fig.6 Structure of the proposed model

ALGORITHM 1: Pre-processing and Training on Dataset

INPUT: Images along with their pixel values.
OUTPUT: Trained Model

STEP 1: Loading images from the train folder.
STEP 2: Load labels from the folder.
STEP 3: Perform Data augmentation and split the data into training and validation batches.
STEP 4: Train the formulated model on training batches.
STEP 6: Save the model for the future.

Fig.7 Algorithm 1, Pre-processing and Training on the Dataset.

IV. EXPERIMENTAL RESULTS

The generated model was trained on Kaggle Notebook Kernel having 13 GB of RAM and 16 GB of GPU memory.The data augmentation done was as follows - rescale=1./255, shear_range=0.2, zoom_range=0.2, horizontal_flip=True, validation_split=0.3.
The model was thus compiled using Adam as an optimizer. Metrics used to check the performance were Accuracy (Equation3) and Binary cross-entropy /Log loss (Equation 4).

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3)$$

$$\log loss = -\frac{1}{N}\sum_{i=1}^{N} -(\log(p_i)) \quad (4)$$

Here, TP is True Positive, FP is False Positive, TN is True Negative, and FP is False Negative from equation three, N is the number of data entries, and pi is corrected probability.
The model was trained for 100 epochs. Figures 7 and 8 show the model accuracy and model loss with the increase in epochs, respectively. The model, when tested on the validation set, gave an accuracy of 95.34%.

©IJRASET: All Rights are Reserved

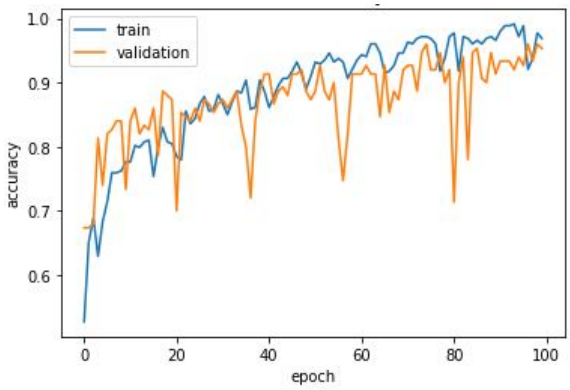


Fig. 8 Model Accuracy vs Epochs

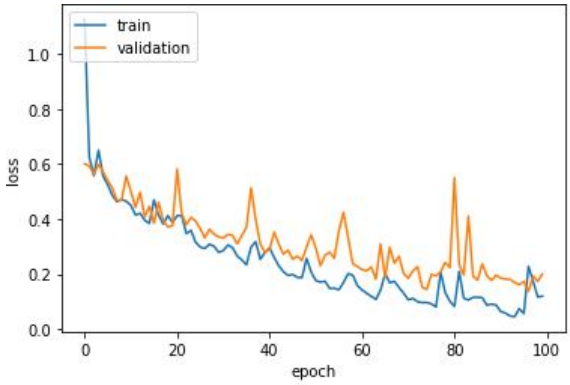


Fig. 9 Model Loss vs Epochs

REFERENCES

[1] F. Mallor, T. León, L. De Boeck, S. Van Gulck, M. Meulders, and B. Van der Meerssche, "A method for detecting malfunctions in PV solar panels based on electricity production monitoring," Sol. Energy, vol. 153, pp. 51–63, 2017, doi: 10.1016/j.solener.2017.05.014.

[2] Y. P. Chandra, A. Singh, V. Kannojiya, and J. P. Kesari, "Solar Energy a Path to India's Prosperity," J. Inst. Eng. Ser. C, vol. 100, no. 3, pp. 539–546, 2019, doi: 10.1007/s40032-018-0454-6.

[3] L. Cardinale-Villalobos, R. Rimolo-Donadio, and C. Meza, "Solar panel failure detection by infrared UAS digital photogrammetry: A case study," Int. J. Renew. Energy Res., vol. 10, no. 3, pp. 1154–1164, 2020.

[4] A. Singh Chaudhary and D. K. Chaturvedi, "Analyzing Defects of Solar Panels under Natural Atmospheric Conditions with Thermal Image Processing," Int. J. Image, Graph. Signal Process., vol. 10, no. 6, pp. 10–21, 2018, doi: 10.5815/ijigsp.2018.06.02.

[5] M. W. Akram et al., "CNN based automatic detection of photovoltaic cell defects in electroluminescence images," Energy, vol. 189, p. 116319, 2019, doi: 10.1016/j.energy.2019.116319.

[6] C. Mantel et al., "Machine learning prediction of defect types for electroluminescence images of photovoltaic panels," p. 1, 2019, doi: 10.1117/12.2528440.

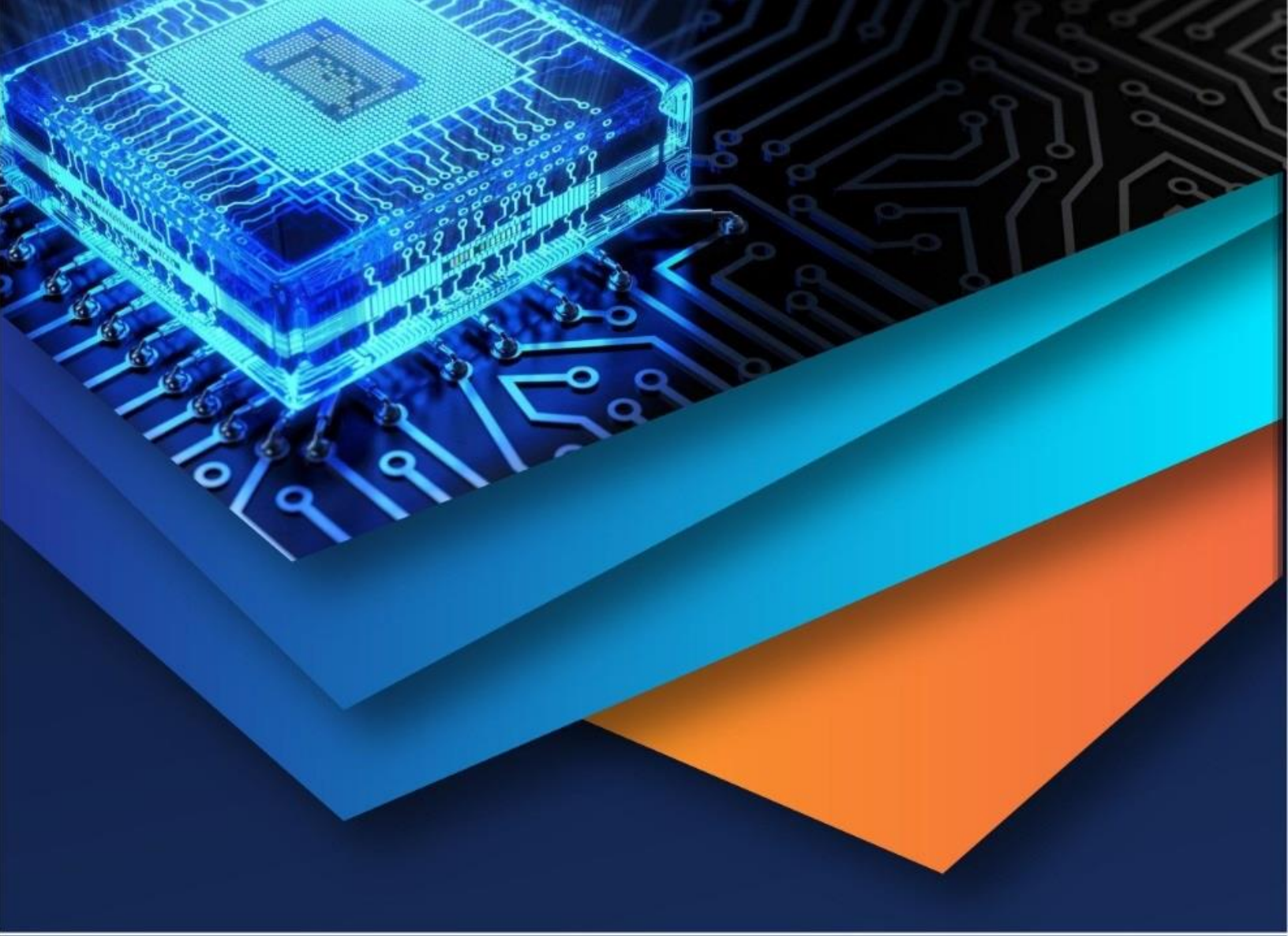
[7] S. Deitsch et al., "Automatic classification of defective photovoltaic module cells in electroluminescence images," Sol. Energy, vol. 185, no. July 2018, pp. 455–468, 2019, doi: 10.1016/j.solener.2019.02.067.

[8] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," Proc. 2017 Int. Conf. Eng. Technol. ICET 2017, vol. 2018-Janua, pp. 1–6, 2018, doi: 10.1109/ICEngTechnol.2017.8308186.

[9] C. Wang and Y. Xi, "Convolutional Neural Network for Image Classification," no. 1969, pp. 1–7, 1997, [Online]. Available: http://www.cs.jhu.edu/~cwang107/files/cnn.pdf.

[10] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537- 019-0197-0.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

View publication stats

Environ. Eng. Res. 2023; 28(4): 220249
https://doi.org/10.4491/eer.2022.249

pISSN 1226-1025
eISSN 2005-968X

Review

Solid Waste Management in India: A State-of-the-Art Review

Anunay A. Gour and S.K. Singh†

Department of Environmental Engineering, Delhi Technological University, Main Bawana Road, Delhi-110042, Delhi, India

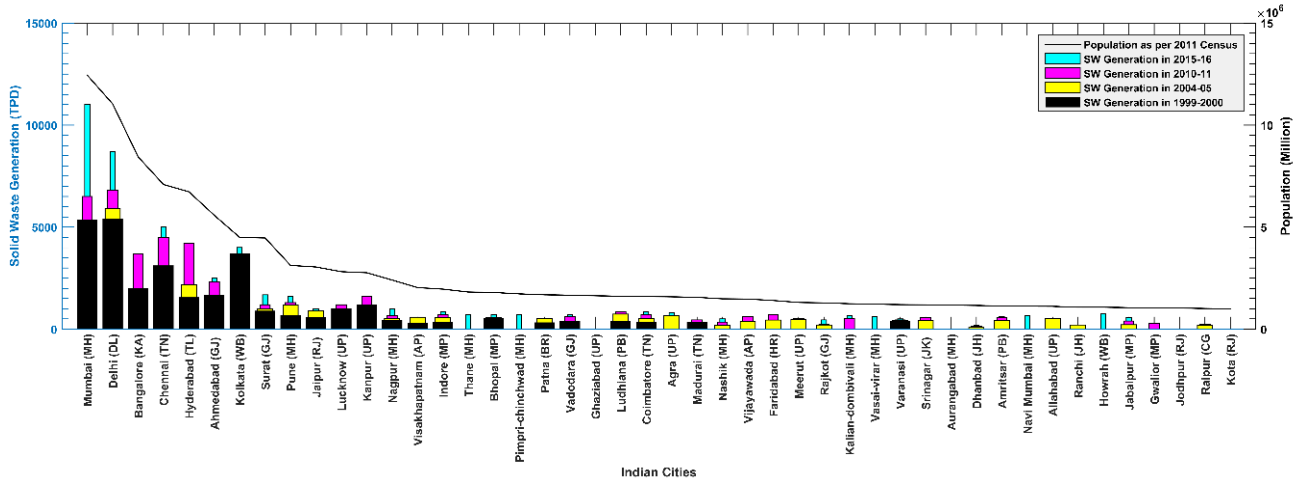
Received June 8, 2022 **Revised** October 5, 2022 **Accepted** October 11, 2022

ABSTRACT

This paper presents the current scenario of solid waste management aspects and its challenges in India, which will benefit developing and low-income countries. The leading cause of waste generation is the growing population and the new lifestyle due to the increased per capita income. Consequently, the magnitude of solid waste is continuously growing along with its compositional diversity. In earlier days, the wastes were organic and could be disposed of in low-lying areas conveniently without causing any adverse impact on the environment. But today, the organic fraction of waste has steeply declined while the inorganic portion has increased manifold. Moreover, wastes from industries, hospitals, construction sites, households, and many other sources severely affect the environment and public health. Also, the chemicals generated from the improper disposal of these wastes enter the air, soil, and water resources, causing hazardous and toxic effects in countries that could not implement the adopted policy framework strictly. A state-of-the-art review is conducted in this paper to further search other primary and prevalent reasons behind the inability of proper waste management and to find a real solution.

Keywords: Challenges, Municipal solid waste, Solid waste management, State-of-the-art review, Waste characterization, Waste generation

Graphical Abstract



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2023 Korean Society of Environmental Engineers

† Corresponding author
E-mail: sksinghdce@gmail.com
Tel: +91 9891 599 903
Fax:
ORCID: 0000-0001-8215-8562

1. Introduction

The improvement in people's living standards because of the industrial revolution and economic development has entirely changed their consumption patterns. But simultaneously, it has created extreme complexity in solid waste management (SWM). This altered lifestyle has brought a tremendous increase in the share of the inorganic constituents in the wastes, due to which the SWM has been severely affected and has become a global problem today [1,2]. Presently, the waste disposal scenario has become a considerable threat compared to earlier times when the waste was primarily organic and could be safely disposed of in low-lying areas, assimilating into the natural biogeochemical cycles [3]. Moreover, the rapid increase in industrially-manufactured materials such as metals, glass, plastics, papers, rags, and polystyrene has enhanced the share of inorganic wastes [4]. It has brought several new solid wastes, such as municipal solid waste from fruit and vegetable markets, gardens and parks, drain silt and street sweepings from the domestic areas, commercial sectors, and common spaces [5-9]; construction and demolition waste from stones and bricks, reinforcement steel, and infrastructure-building activities [10-12]; biomedical waste from used/expired drugs and medical instruments, tissues and organs generated from healthcare and veterinary establishments [13-15]; E-waste generated from the disposal of end-of-life electronic equipment [16]; industrial waste from various industrial manufacturing centers [17-20]; special waste toxic and hazardous wastes from households, and trading zones [21-23]. The significant causes of the rise in these manufactured materials are the expanding urban areas, rising consumerism, and the exponentially growing population and their varying activities. These factors not only affect a particular area or a country but also have made solid waste generation (SWG) a worldwide problem [24-27].

Almost all the countries of the World are undergoing an intense struggle but still have failed to find a proper and effective solution to manage waste. However, a few developed countries like Germany, Italy, Canada, and Australia have effectively addressed this issue. But the rest of the World, including numerous developing and under-developed countries, is still establishing the basic infrastructure, let alone managing solid waste. Therefore, this paper extensively reviews the work done globally in this field and finds a feasible solution to India's SWM condition. A practical solution for managing solid waste in India can lead the developing and under-developed countries of the World, as India shares 17.5% of the World's population and is the 2nd most populous country. Moreover, India has more than 1.4 billion people spread over 3.3 million sq. km, approximately 2.4% of the World's terrestrial area.

In addition to the introduction given in Section 1., the rest of the paper is organized as follows - Section 2. presents a criteria-based state-of-the-art review of the solid waste management processes employed in India along with an outlook of solid waste management practices globally; Section 3. deals with the present scenario of solid waste management including the framework of MSWM, the trend of waste generation in various cities, the composition of waste, it also discusses the status of waste processing undertaken in India; Section 4. discusses the key challenges and tries to find solutions for an efficient solid waste management paradigm in India; Section 5. presents the detailed conclusion, and the References are given at last in the paper.

2. State-of-The-Art Review on Solid Waste Management

The state-of-the-art review presented in this paper tends to address more current matters in solid waste management in contrast to other combined retrospective and contemporary literature along with new perspectives on SWM and points out areas for further research, as depicted in the tables below. In order to have better clarity and to grasp the matter at a glance, the significant contribution of each study is concisely and precisely presented year-wise/criteria-wise/country-wise in Table 1., Table 2., and Table 3., along with a short description of gaps and novelty.

A decadal highlight of literature presented in Table 1. points toward increasing population, urbanization, industrialization, and improvement in the standard of living as factors behind increasing solid waste generation. Recent advancements in solid waste direct administrative attention toward the role of public-private partnerships. It identifies numerous blockades to effective MSW management. It shows that even small steps, such as segregation at source, can significantly determine the feasibility of employing waste-to-energy plants. This may reduce greenhouse gas emissions and provide sustainable logistics and waste transport solutions. Modelling resources and wastes considering their lifecycles is another holistic way to determine their reusability and potential for recycling and material recovery. Implementing a systems approach in MSWM will consider not only the individual components but also their interaction, their effect on the environment, and that of the environment on the MSWM system. Application of geographically and topographically relevant technology based on the existing condition of an area can efficiently optimize, considering the MSW management processes as a system. It may benefit the management agencies to form better policies, solve imminent and long-term issues, and improve the health of the environment.

This points toward increasing population, urbanization, industrialization, and improvement in the standard of living as factors behind increasing solid waste generation. Recent advancements in solid waste direct administrative attention toward the role of the public-private partnerships. It identifies numerous blockades to effective MSW management. It shows that even small steps such as segregation at source can significantly determine the feasibility of employing waste-to-energy plants. This may reduce greenhouse gas emissions and provide sustainable logistics and waste transport solutions. Modelling resources and wastes considering their lifecycles is another holistic way to determine their reusability and potential for recycling and material recovery. Implementing a systems approach in MSWM will consider not only the individual components but also their interaction, their effect on the environment, and that of the environment on the MSWM system. Application of geographically and topographically relevant technology based on the existing condition of an area can efficiently

Table 1. A summary of research paper studies that covered solid waste management

Year	Highlights	Ref.
2022	Proper engagement of the private sector in SWM can create sustainable circumstances. Education and awareness are crucial for household waste management and new approaches.	[28]
2021	The life cycle approach can close the gap between solid waste management in metropolitan and regional cities. And waste-to-energy significantly reduces GHG emissions and particulate matter.	[29]
2020	The fuzzy Delphi method identified 146 barriers to attaining sustainable SWM. The most significant ones are household hazardous waste, inadequate research capital, local architecture, lack of staff capability, and standard processes.	[30]
2019	A strong correlation exists between geographical position and economic status with waste characteristics. It also undertakes LCA models to select appropriate waste management algorithms to evaluate and find a sustainable solution.	[31]
2018	The rapid increase of solid waste is due to population increase, urbanization, rapid industrialization, and economic sustainability.	[32]
2017	The model was generated utilizing TOPSIS in the GIS environment to develop a suitability indicator for siting the units to minimize fixed and transportation costs and maximize the system's suitability.	[33]
2016	Greenhouse gas emissions could be reduced by applying the proper framework of SWM, such as waste-to-energy and material recovery, instead of landfilling.	[34]
2015	Systems engineering approaches such as systems engineering, industrial ecology, integrated solid waste management strategies, integrated systems planning, design and management, and uncertainty analysis shall be applied to SWM.	[35]
2014	Multiple methods for decision-making regarding waste management suggest direction by mass-balance approach, goal-oriented evaluation, and transparent and reproducible presentation of the methodology, data, and results.	[36]
2013	Urbanization, economic growth, inequality, institutional issues, governance, policy, international influences, and cultural and socio-economic aspects are the factors that have complicated the SWM in developing countries.	[37]
2012	All the SWM aspects shall be examined for health and safety, environmental issues, and disaster-prone/affected areas.	[38]

Table 2. A criteria-based summary of recent solid waste management studies

Study Criteria	Objective and key findings
Management	Waste management systems must be continually improved; waste diversion indicators analyze waste management and create a framework and an easy-to-use metric for improving waste management efficiency [39]. Innovation in SWM shall extend to product, organization, and process innovation in the disposal segment [40].
Framework	Creating avenues for the circular economy through recovering and recycling materials from solid waste, similar to the developed countries, may help achieve climate change and sustainable development goals [41]. Literature finds that clear laws, regular public campaigns, and fee methodology are pillars of effective SWM [42]. Also, as involvement in social community activities, education level, and per capita expenditure [43].
Awareness	Public perception and willingness to pay for solid waste management have been highlighted as significant impediments to the effective collection, segregation and processing of solid waste [44,45]. A unique study aimed at raising awareness of SWM through formal education and gauging the gap in knowledge and attitude of students and teachers revealed that students have a positive attitude and high understanding of environmental issues. In contrast, there is a gap in teachers' practical knowledge and education to implement SWM [46].
Generation	The waste generation shall be studied along with analyzing economic, social and population indicators [47]. In addition to the inherent difficulties in MSWM operations, infrastructure, finance, and improper implementation of policies are among the significant challenges in SWM in India [48]. Another study finds a correlation between geographical position and economic status with waste characteristics. It also undertakes LCA-based models to select appropriate waste management algorithms to evaluate and find a sustainable solution [31].
Collection	The relation between distance travelled, working time, and waste collection system can be used to improve the solid waste management system, develop an effective solid waste collection system, and optimize the map of recycling stations in urban settings [49]. Clustering is also one of the potent optimization modelling to enhance solid waste collection [50]. These can provide appropriate collection plan, route map and usage of geo-positional attributes for optimal routing and improved efficiency and offers significant savings in time, labor, and pathways.

Table 2. Continue

Study Criteria	Objective and key findings
Segregation	Multi-criteria decision analysis aimed at assessing and comparing the sustainability of different waste management schemes finds that a door-to-door collection system (efficiency>80%) is far better than point collection (efficiency=35%) in the segregated collection of waste [51]. Internet-of-Things has found its application in detecting the variation in relative humidity in the presence of wet waste, helping to monitor waste bin status regularly [52].
Characterization	Waste broadly includes plastics, paper, diapers, LDPE, plastic bags, glass, pharmaceuticals, and organic material [53]. Organic waste constitutes the primary category (57%) at pilgrimage events, indicating energy conversion potential [54]. Several pieces of literature signify characterizing as the first step to a successful SWM. Artificial intelligence and machine learning approaches can investigate the impact of seasonal variation on the characterization of MSW [55].
Minimization	Life-cycle analysis of hospitality solid waste management can result in cost avoidance, enhance consumer image and employee morale, and reduce environmental impacts [56]. Open dumping of the increasing MSW in Thailand has led to various problems associated with solid waste, including the emission of greenhouse gas. A report intends to ascertain the influencing factors and ways to reduce the per capita generation of MSW [57].
Storage	The location of solid waste storage based on the number of stations, maximum walking distance, and container capacity improves waste storage services [58]. The storage capacity of landfills can be effectively increased by increasing storage height, but it can cause leachate-induced slope instability [59].
Transport	Vehicle routing optimized for travel duration and path strengthens the triple bottom line of sustainability and improves the solid waste management framework [60]. Another study points out that with no marginal increase in cost, the transportation model based on locational planning of MSWM systems shows considerable savings in CO_2 emission [61].
Treatment	Composting is an economical, sustainable, and environmentally benign alternative for treating organic constituents [62]. A study develops a multi-objective optimization model to select optimum technological solutions for MSW; with the help of this decision support system, sludge co-processing can be promoted for managing solid waste [63].
Recycling	Stone-cutting waste can be recycled as a feasible replacement for aggregates in construction, concerning cost reduction in storage and disposal [10]. Several countries are striving hard to inculcate a culture of recycling in the day-to-day life of people and reduce the magnitude of solid waste reaching open dumping sites.
Disposal	High concentrations of heavy metals such as Cu, Mn, Fe, Cr, Pb, Co, Zn, and Ni in groundwater near waste dumping areas. Advances in disposal technology, such as bioreactor landfills and improved liner materials, may help prevent contaminants in the environment [64]. The prevalence of open burning, dumping, management, treatment and disposal of solid waste in Bangladesh has also been documented [65].
Impact	Using artificial intelligence to gauge the temporal variation in waste characteristics can help predict and reduce the impact of MSW by incorporating early sustainable waste management decisions [55]. The presence of odorous compounds causes severe nuisance near and around MSW processing facilities due to the organic wastes' high moisture content and perishability. After that, aromatic and halogenated hydrocarbons have carcinogenic risks [66].
GHG Emission	Sustainable circular waste management systems may contribute considerably to alleviating climatic heating by reducing the emission of GHGs [67]. The Integrated SWM approach results in the maximum reduction of GHG emissions compared to the recycling and incineration approaches [68]. Composting has been the best scenario for the least GHG emission, whereas landfilling generated the most GHG, followed by open burning [69].
COVID-19	The recent COVID-19 pandemic has put massive pressure on solid waste management with its peculiar biomedical waste. An enormous instability has been inducted into healthcare waste handling, recycling, treatment, and disposal [70]. It has been found that the SARs-COV-2 virus, responsible for COVID-19, has more than 72 hrs. of stability on plastics and steel compared to copper or cardboard [71]. The wastes produced during this period are bio-hazardous and highly undesirable.
Waste-to-energy	Estimating methane emissions and technical know-how for initiating landfill gas projects to extract energy needs to be developed along with developing novel technologies for enhancing the energy content of waste. Along similar lines, an article finds a novel approach for WTE by converting solid waste plastics into graphene nanosheets to create high-performance supercapacitors [72]. An assessment links WTE resource potential to an area-based prioritization framework for better performance [73]. Another article proposes the recovery of waste and energy generation and urges the public authorities to invest more in WTE technologies to contribute to sustainable development, economic growth, and ecological and environmental well-being [74].

optimize, considering the MSW management processes as a system. It may benefit the management agencies to form better policies, solve imminent and long-term issues, and improve the health of the environment.

Table 2. presents a criteria-based summary of solid waste management studies undertaken in the past five years. This categorical review informs the objectives and key findings in research areas that require more focus in the present times. It reveals that modern methods of continuous improvement of the management system and holistic innovation through product, organizational, and process innovations may fill the gap in the management of solid waste. Advancing into a circular economy framework, a country benefits by alleviating pressure on the environment and promoting sustainable development. It also increases competition and stimulates innovation in the field. The concept of a circular economy has the potential to boost the economic growth of a nation and save time, resources, and adverse environmental impacts.

Similarly, the inclusion of socio-economic factors such as society, literacy rate, and awareness by public campaigns play a crucial role in improving solid waste management. Yet, few consider the economic and social indicators, along with population indicators. This increases the gap between the theory and practice of MSWM, causing inadequacy of infrastructure and finance. Moreover, the influence of geographical location and economic status on the changing solid waste characteristics remains unknown.

2.1. Solid Waste Management Worldwide Outlook

The various studies [3,25,27,75,76] conducted to determine the trend of SWM worldwide show that the SWG increases following the economic advancement of countries from low-income to high-income. Based on the data from a World Bank Report conducted on 192 countries [76], the statistical correlation of SWG

Table 3. A summary of solid waste management studies done in various countries

Year	Criteria	Issue addressed	Place	Country	Economic Status	Ref.
2022	Awareness	Willingness-to-pay	Hawassa	Ethiopia	LI	[44]
2022	Characterization	Variation of waste characteristics	Johannesburg	South Africa	UMI	[55]
2022	COVID-19	Emerging issues in SWM	Luzon, Visayas, Mindanao	Philippines	LMI	[77]
2022	Impact	Impact assessment using multispectral imagery	Yenagoa, Bayelsa State	Nigeria	LI	[78]
2022	Minimization	Factors for reducing solid waste	Shanghai	China	LMI	[57]
2021	Characterization	MSW Composition Analysis	Nur-Sultan City	Kazakhstan	UMI	[53]
2021	Collection	Efficiency of collection	Mexican Municipalities	Mexico	UMI	[79]
2021	COVID-19	Effect of COVID-19 on SWM	Brazil	Brazil	UMI	[80]
2021	Framework	Life cycle assessment for MSW management	New South Wales	Australia	HIC	[81]
2021	Framework	SWM at Ajmer, India	Ajmer, Rajasthan	India	LMI	[15]
2021	GHG	Air pollution due to MSWM	Accra	Ghana		[82]
2021	Management	Agent-based modelling for eco-effective SWM	Norte Pioneiro	Brazil	UMI	[83]
2021	Management	Input-output indices for SWM quantification	Nova Scotia, Saskatchewan	Canada	HIC	[39]
2021	Management	Implementation analysis of SWM	Ludhiana, Punjab	India	LMI	[84]
2021	Management	Life cycle cost analysis of MSWM	Mumbai, Maharashtra	India	LMI	[85]
2021	Management	Management of SW at dumpsites	Ibadan	Nigeria	LI	[86]
2021	Management	SWM at Papua New Guinea	Lae City	Papua New Guinea	LMI	[20]
2021	Recycling	Residential solid waste recycling	Annaba	Algeria	LMI	[87]
2021	Segregation	Segregation at source and separate collection	Hoi An	Vietnam	LI	[49]
2021	Treatment	Composting of organic waste	Al-Karak	Jordan	LMI	[62]
2021	Waste-to-energy	Solid waste fuel pellets	Newai, Rajasthan	India	LMI	[88]
2020	COVID-19	Effect of SWM on COVID-19 spread	South Africa, Egypt, Morocco, Algeria, Nigeria, Ghana, Cameroon, Guinea, Ivory Coast, Djibouti, Senegal, Tunisia, Niger, Burkina Faso, DR Congo	Africa	UMI, LI	[89]

Table 3. Continue

Year	Criteria	Issue addressed	Place	Country	Economic Status	Ref.
2020	Framework	Community participation for effective SWM	Jakarta	Indonesia	LMI	[43]
2020	Minimization	Cost minimization	Ankara	Turkey	UMI	[61]
2019	Collection	Efficiency of collection	Bharatpur	Nepal	LI	[90]
2019	Framework	Applying sustainability to MSWM	Mata de São João	Brazil	UMI	[91]
2019	Framework	Mathematical modelling for MSWM	Istanbul	Turkey	UMI	[92]
2019	Management	Community-based SWM	Nkulumane Suburb, Bulawayo	Zimbabwe	LI	[93]
2019	Waste-to-energy	WTE at South Africa	Johannesburg	South Africa	UMI	[74]
2018	Collection	Impact of tourism on waste collection	Italian Municipalities	Italy	HIC	[94]
2017	Characterization	Evaluation of SWM	Mohali & Panchkula, Punjab	India	LMI	[95]
2017	Characterization	MSW Composition Analysis	Kerbala	Iraq	UMI	[54]
2017	Characterization	Waste generation & composition	Lagos	Nigeria	LI	[96]
2017	Management	Biomedical waste management	Francistown, Gaborone, Selebi Phikwe, Lobatse, and major villages	Botswana	UMI	[97]
2017	Management	Multi-dimensional modelling for integrated SWM	Tehran	Iran	LMI	[33]
2016	GHG	Greenhouse gas emissions from landfills	Delhi	India	LMI	[98]
2016	Waste-to-energy	WTE and recycling value	Lahore	Pakistan	LI	[99]
2015	Generation	Public perception hazards of MSWM	West Bank, Gaza Strip	Palestine	LMI	[100]
2014	Disposal	Impact of landfilling on groundwater	Kolhapur, Maharashtra	India	LMI	[101]
2014	Framework	MSWM system in a developing country	Dhaka, Chittagong, Khulna, Rajshahi, Barisal, and Sylhet	Bangladesh	LI	[102]
2014	Framework	SWM and Agenda 21 in Brazil	Brazil	Brazil	UMI	[103]
2014	Management	SWM at Nairobi	Nairobi	Kenya	LI	[104]
2013	Awareness	Willingness-to-pay	Kolkata	India	LMI	[105]
2013	Characterization	MSW Composition Analysis	Beijing	China	LMI	[106]
2013	Disposal	Route Optimization for MSW disposal sites	Gondia, Maharashtra	India	LMI	[107]
2013	Management	SWM at Delhi, India	Delhi	India	LMI	[108]
2012	Awareness	Education and Awareness	Universiti Kebangsaan Malaysia	Malaysia	UMI	[109]
2012	Generation	SWM at Bahrain	Kingdom of Bahrain	Bahrain	HIC	[110]
2012	Impact	Landfill fires and their impacts	Lagos	Nigeria	LI	[111]
2012	Recycling	Recycling solid waste	Delhi	India	LMI	[24]
2012	Recycling	Sustainability in SWM	Gianyar	Indonesia	LMI	[112]
2011	Transport	Collection route using GIS	Nagpur, Maharashtra	India	LMI	[113]
2010	Awareness	Health issues of workers	Cairo	Egypt	LMI	[114]
2010	Awareness	Segregation at source	Mercato San Severino	Italy	HIC	[115]
2009	Disposal	Impact of disposal	Ado-Ekiti	Nigeria	LI	[116]
2008	GHG	Greenhouse gas emissions from MSWM	Chennai, Tamil Nadu	India	LMI	[117]
2008	Management	SWM at Karlsruhe City, Germany	Karlsruhe City	Germany	HIC	[118]
2007	Framework	Optimizing Integrated solid waste management in India	Mumbai, Maharashtra	India	LMI	[119]
2007	Minimization	SWM and disposal emission inventory	Bangkok	Thailand	LMI	[120]

was determined with the urban population and the total population of both urban and rural areas. It was found that the statistical correlation of the quantity of waste generated with the entire population was 0.6390, while that with the urban population was 0.8421. It shows that the statistical correlation with the urban population is nearly 1.5 times more than that of the total population. These figures demonstrate a notable trend of SWG, which increases with the countries' affluence. Still, the measures undertaken to improve the SWM are not up to the mark required for properly disposing of the increased SWG. A summary based on different criteria and issues of the SWM adopted by various countries is shown in Table 3. The economic condition of the nations, such as low-income (LI), low-middle-income (LMI), upper-middle-income (UMI), and high-income countries (HIC), is also shown in this table.

3. Solid Waste Management in India

The present rate of SWG in India is 0.34 kg per capita per day, which is expected to reach up to 0.7 kg per day by 2025 [76]. However, this SWG rate is much lower than nearly 90% of countries worldwide. But, due to the massive population, the total waste generated in India is more than 168,403.24 TPD, because of which India ranks seventh globally in SWG [121]. But it is forecasted that by 2025, India may become a more significant waste generator than Germany, Japan, and Brazil, which are now at the 4th, 5th, and 6th rank, respectively [76]. The SWG will still increase in India because the linear projections show that India's population may grow to 2.6 billion in 2051, as presented in Fig. 1. The urban population's present growth rate is 31.45%, which may increase to 36% in 2051. In contrast, the rural population is 17.97% [122]. Moreover, the total number of urban agglomerations (UAs) and towns may rise to more than 10,000 from 6166.

3.1. SWM Framework

The Central Pollution Control Board and the State Pollution Control Boards/Committees impart environmental policy, pollution monitoring, and reporting at the central and state government levels. Municipal agencies/cantonment boards implement plans and rules in the areas under their jurisdiction at the local level. They also develop infrastructure for collection, segregation, transportation, storage, treatment and processing, and solid waste disposal. India's Municipal Solid Waste (Management and Handling) Rules assign citizens the duty to segregate waste at household levels. It also gives strict adherence to avoiding littering streets and delivering waste through the established delivery system notified by the civic bodies [123]. The SWM model followed in India was simplistic and rudimentary. But, the recent integration of ideas such as reduce, reuse, and recycle (3R) and material and energy recovery has significantly improved the model, as observed in developed countries [26,118,120,124-127].

3.2. Solid Waste Generation

Out of 28 states and 8 union territories (UT) of India, about 50% of its solid waste is generated by only five states/UT - Maharashtra, Uttar Pradesh, West Bengal, Gujarat, and NCT of Delhi. Generating more than 23000 TPD of solid waste, Maharashtra stands at the top, followed by Uttar Pradesh and West Bengal, with SWG of about 15,000 TPD each. States such as Andhra Pradesh, Delhi, Gujarat, Karnataka, and Tamil Nadu create more than 10,000 TPD.

Fig. 2. shows the trend of SWG in India. Moreover, the more affluent the economy of a particular state/UT, the more substantial waste is likely to be generated. Maharashtra generates the highest amount of solid waste of all states and collects 99.3 % of its solid waste. It also utilizes more than 53% of the collected waste for composting, recycling, and energy recovery. The remaining 47% of waste is sent to landfills. Arunachal Pradesh, Chhattisgarh, Karnataka, Manipur, Nagaland, and Puducherry collected around

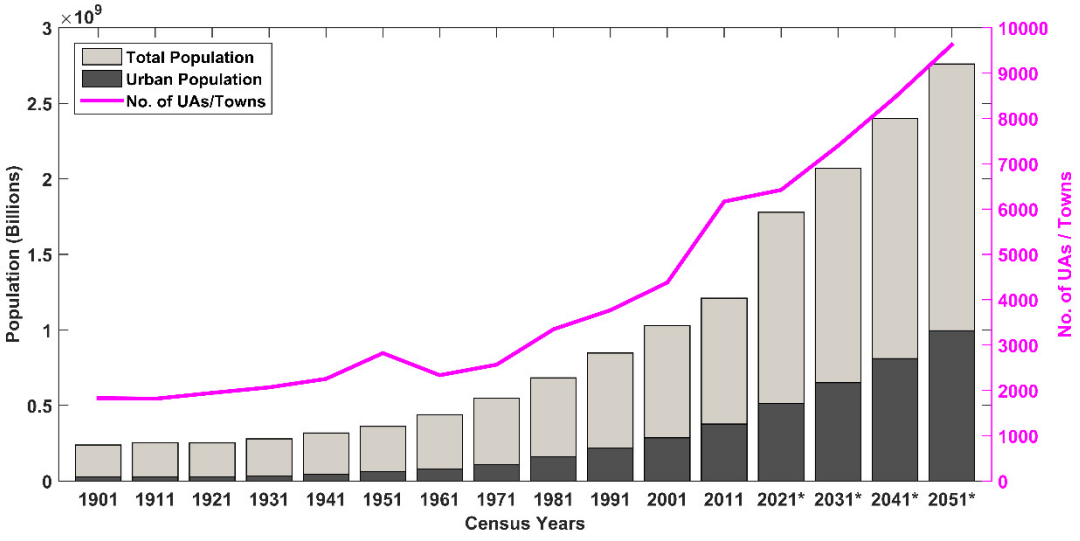


Fig. 1. Trend of growth of UAs/Towns (on the right y-axis) and urban population (on the left y-axis) in India for census years 1901-2011 [122] with linear projection till 2051.

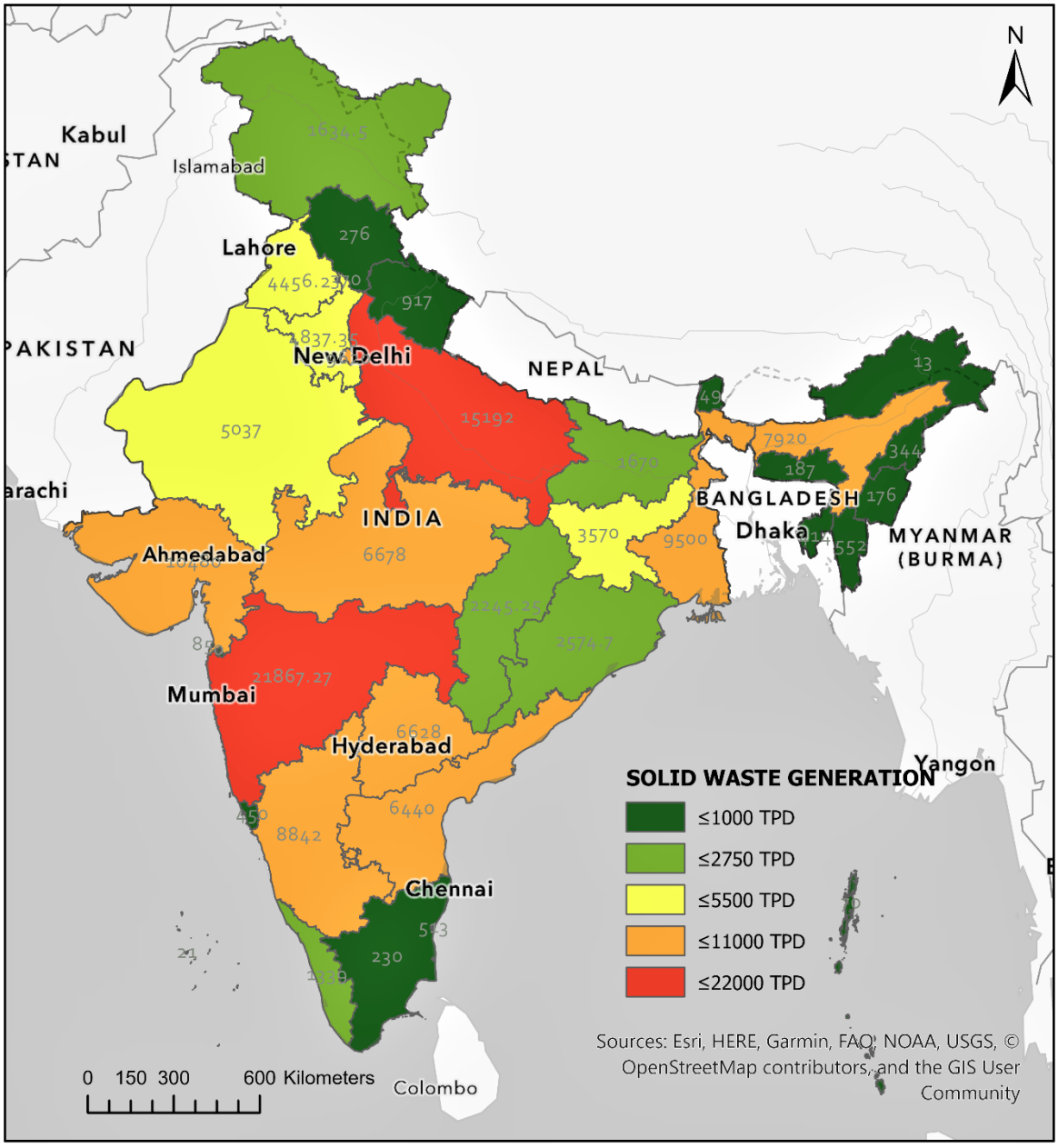


Fig. 2. Solid waste generation trend in India

80% of the solid waste generated and treated only 25% of waste. Mumbai and Delhi produce more than 11,000 TPD of solid waste at the city level. Only about 50% of solid waste is treated, and the remaining 50% finds its way to landfills. Both mega-urban centers lack sanitary landfill facilities and rely on open dumping of solid waste [108]. From 2009 to 2012, India produced 127485.107 TPD of solid waste at 0.17 to 0.76 kg/capita/day for Tier I, II, and III Indian cities [34,123,128]. According to the World Bank, the average per capita waste generation is approximately 0.35 to 0.4 kg/capita/day [76]. SWG increased by nearly 50% from 2001 to 2011 [129].

Fig. 3. in the grouped column and line chart shows the magnitude of SWG for 1999-2000, 2004-05, 2010-11, and 2015-16 in tons per day (TPD) by urban centers of India on the primary y-axis (on the left side). It also compares the trend with the cities' populations in millions, scaled on the secondary y-axis (on the right). It establishes the relationship between SWG and people. SWG has been increasing consistently since yesteryears, with metropolitans such as Mumbai and Delhi leading in waste generation due to their enormous population, immigration, and varied economic activities. Kolkata and Bengaluru have rigorously controlled their SWM as waste generation has increased marginally from 1999-2000

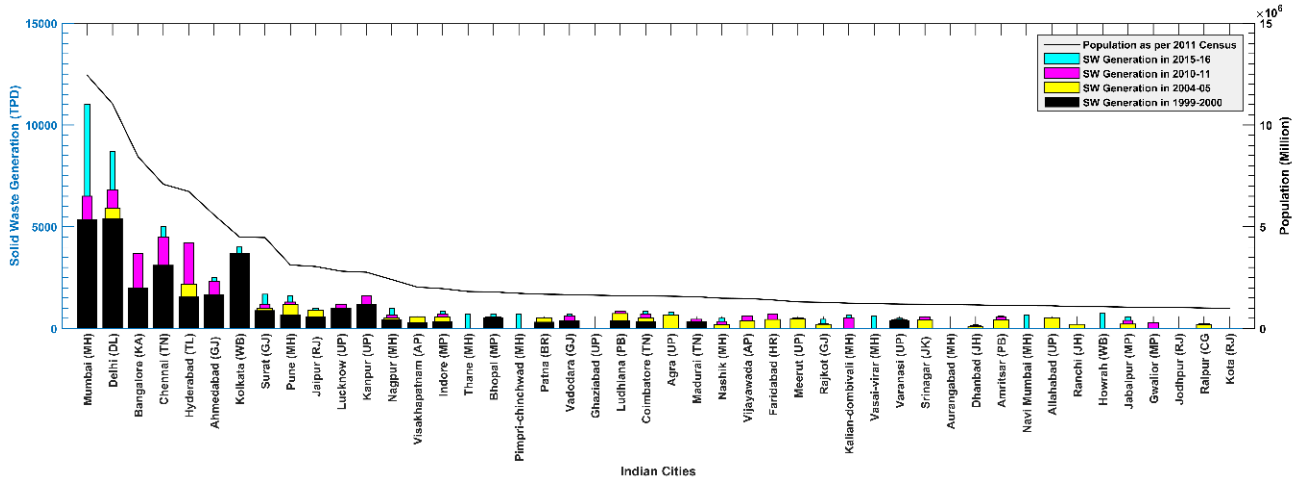


Fig. 3. Solid waste generation by Indian cities [122]

till 2015-16. Summarizing the magnitude of solid waste generated by cities classified based on their economic status as Tier I, II, and III cities present a consistent trend in the solid waste generated by cities of various economic classifications. SWG directly relates to a city's financial state [25]. The city's economic condition: more is the quantity of solid waste generated. Mumbai & Delhi leads the Tier I cities by producing more than 7000 TPD and more than 11500 TPD in 2016. Amongst Tier-II cities, Chandigarh generates the highest quantity of waste. Cities that generate more than 7000 TPD, i.e., Greater Mumbai and Delhi, lack sanitary landfilling facilities [130]. The population growth rate of urban zones is around 36.45%, which is more than twice the population growth in rural areas (17.97%) [122]. Fig. S1. presents the relationship between quantity of solid waste generated by Indian Cities vs their population graphed using curve fitting with 95% confidence bounds. This relation can be developed into a linear model as shown in Eq. (1).

$$Q_{SWG} = 393.2 \times P - 111.2 \qquad (1)$$

where,

Q_{SWG} = quantity of solid waste generated in TPD,
P = population of a city in millions.

3.3. Waste Composition

India's solid waste differs significantly from Western countries in its composition and hazards. The physical composition of India's solid waste on a wet-weight basis includes (a) biodegradable organic fraction (51.3±8.3%), (b) recyclables up to 17.48±5%, (c) the ash content of 30-40%, (d) paper about 3-6%, and (e) inert such as glass, plastic, and metals up to1-6% [128]. Characteristics of waste from 59 Indian cities classified based on population, as shown in Table 4., ascertain the value of the carbon-to-nitrogen ratio (C/N) of 31.91±26.68 low, the moderate calorific value of 1751.20±782 kcal/kg [34,128]. Waste is often attributed to a moisture content of 46.76±13%, making the waste less feasible for energy recovery. High-income cities tend to produce more inorganic waste than organic ones. Solid waste from Tier III cities, i.e., the low per capita income cities, has higher calorific value waste than other towns [34]. Hence, as a town gets economically advanced, the organic share of the waste reduces—consequently, the paper and plastic component increases, which points toward consumerism in the urbanized areas.

3.4. Status of Waste Processing

3.4.1. Waste minimization, recycling, and reuse

In India, an extensive informal and formal recycling network mini-

Table 4. Characteristics of waste from Indian cities classified with respect to their population (in millions)

Population Range (millions)	Number of cities under Observation	Waste Quantity (TPD)	Waste Generation Rate (kg/c/day)	Recyclable (%)	C/N Ratio	Moisture (%)	Compostable (%)	HCV (Kcal/Kg)
< 0.1	8	22.5 ± 23.1	0.4 ± 0.2	21.0 ± 4.9	26.2 ± 7.1	48.6 ± 12.7	51.7 ± 12.5	2161.0 ± 786.8
0.1 - 0.5	11	96.8 ± 65.2	0.4 ± 0.1	19.6 ± 6.6	26.4 ± 6.8	49.8 ± 11.9	51.2 ± 8.0	2162.3 ± 1079.9
0.5 - 1	16	328.1 ± 139.8	0.4 ± 0.2	16.2 ± 4.1	28.9 ± 10.7	47.7 ± 15.3	54.2 ± 9.0	1480.9 ± 560.2
1 - 2	11	480.8 ± 151.3	0.4 ± 0.1	17.2 ± 4.7	27.1 ± 11.4	40.7 ± 13.3	47.4 ± 4.7	1410.9 ± 711.9
2 - 3	6	859.7 ± 301.1	0.4 ± 0.1	13.8 ± 2.3	32.7 ± 9.0	47.0 ± 15.2	51.2 ± 6.8	1685.8 ± 754.9
3 - 10	5	2169.4 ± 705	0.5 ± 0.1	16.7 ± 5.2	30.3 ± 3.4	45.2 ± 8.3	47.8 ± 6.2	1866.0 ± 656.5
> 10	2	5621 ± 425.7	0.5 ± 0.1	16.1 ± 0.8	37.0 ± 2.9	51.5 ± 3.5	58.4 ± 5.7	1794.0 ± 11.3

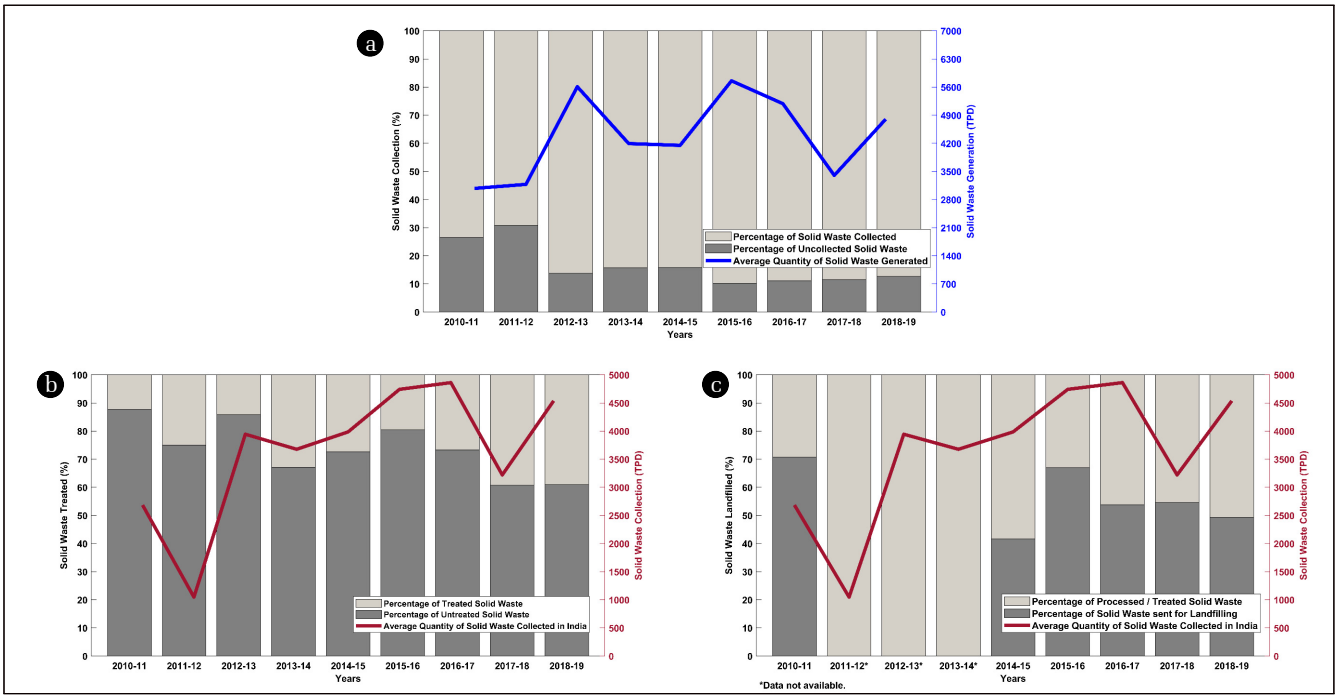


Fig. 4. Percentage of Solid waste collection, treatment, and disposal at landfills w.r.t Solid waste collected in India

mizes solid waste. The rag pickers collect metals, plastics, glass, leather, rubber, textiles, paper/cardboard, and other recyclables. Many households also sell recyclables to itinerant waste buyers. The recyclable materials pass from the rag pickers or itinerant buyers to small waste dealers. After that, it reaches middle-level dealers and finally reaches recycling units. End-users of cleaned and sorted recyclables are the recycling units or factories which use these recyclables as raw materials for manufacturing a new product. On average, about 50.72% of the waste collected in India is minimized, recycled, and reused. Hence, landfills are saved from being piled up with a large volume of garbage. Jharkhand, Uttar Pradesh, Goa, West Bengal, Andhra Pradesh / Telangana, and Chhattisgarh recycle more than 90% of the waste collected. Andaman & Nicobar Islands, Manipur, Tamil Nadu, Maharashtra, and Delhi recycle just over 50% of the waste. The lowest amounts of waste recycling are observed in Haryana, Gujarat, Daman & Diu, Meghalaya, Puducherry, and Odisha. Due to a lack of data, no observation could be made about states/UTs such as Arunachal Pradesh, Assam, Bihar, Jammu & Kashmir, Karnataka, Kerala, Lakshadweep, Mizoram, and Uttarakhand. Major waste-producing states shall coordinate and increase waste minimization, recycling, and reuse. A serious impetus is required to push the recycling sector of the blockages in India. Recycling is a dirty and unhygienic activity; waste is often littered around the recycling units. Although measuring the recycling sector's contribution to environmental conservation is difficult to ascertain, it plays an essential role in saving virgin materials.

3.4.2. Collection, storage, and segregation of waste

Despite the efforts to improve the waste collection mechanism, more than half of the cities still fail to collect even 25% of their solid waste. On average, 17% of all the waste generated in India remains uncollected. Cities such as Mumbai, Delhi, and Hyderabad collected more than 90% of their solid waste and put it into the conventional waste processing method. Maharashtra state can collect 99.3 % of its generated solid waste. Other states have not fared well in collecting their solid waste. Regardless of the trend of waste generation, the segregation of wastes is the bottleneck. In India, waste collection and segregation are carried out in the same instance. The onus lies with the citizen segregating and disposing of waste in the receptacles. However, some municipal agencies also provide door-to-door collection and transport of waste to disposal sites by mini-dumpers. They utilize covered garbage bins of 12 to 16 tons capacity and street dustbins in the societies. Management of open areas for the storage of solid waste requires further efforts. Metal containers with color codes to help MSW segregation (green containers for biodegradable waste and blue containers for non-biodegradable wastes) exist in most municipal areas. Municipal workers collect the waste from domestic and commercial spaces and bring it to the nearest waste receptacles. The cantonment boards also have garbage collection points in civilian and military areas. Street-sweeping in the military area is undertaken by the military, while sweeping in the civilian areas is conducted by the respective Cantonment Boards. A combined community bins and door-to-door collection system is necessary to avoid street littering [131].

3.4.3. Transport and transfer

Waste transportation is a significant part of municipal expenditure. Door-to-door collection vehicles equipped with two color-coded containers increase public awareness of segregation at the source. Municipal agencies across the country maintain a fleet of vehicles

that collect MSW from waste receptacles and transport it to the disposal sites. The primary haulers are refuse removal trucks, tractor-trailers, mini dumpers, loaders in urban areas, and carts in rural areas. Methods of transportation of MSW depend on the type of collection points. The logistical system faces difficulty transporting solid waste and works at low efficiency. Besides, the uncovered truck is a nuisance in the city; they spread the garbage all along their way to landfill sites. The municipal agencies carry out the MSW collection, transportation, disposal, and treatment. Disposal trucks are adequately covered to prevent the spreading of the collected waste on the streets. About three-fourths of expenditure in solid waste management is consumed against the collection and transportation of waste. Approximately Rs. 800 per ton ($10.45/ton) is spent on waste transportation in India. Hence, much focus is required to optimize the activity. Waste disposal routes shall be designed to achieve logistical efficiency and cost optimization. The transportation and logistics system for waste management needs significant overhauling using modern tools. Logistical models based on GIS suggest it improved overall efficiency and optimization of resources [132-135]. States like Delhi, Mumbai, and Ahmedabad are moving forward with a decentralized approach to cutting transportation costs [27].

3.4.4. Treatment and disposal

Almost 90% of landfills in India, including Ghazipur and Bhalswa in Delhi, and Deonar in Mumbai, do not have landfill liners and lack leachate collection or re-circulation system [117,136]. The disposed of MSW is spread evenly and compressed by a single bulldozer movement and covered by soil and silt. Rag pickers also carry out their recycling operations at landfill sites. Animals like vultures, pigs, cows, and dogs are common on the sites [137]. A direct rational relation between the quantity of waste collected and waste sent to landfills is shown in Fig. 4. It also presents an inverse relationship between solid waste treated and solid waste generated. The more solid waste is collected, the more waste is sent to landfills. It points toward the inadequacy of treatment facilities in India. Only 26% of the collected solid waste undergoes treatment. Lack of segregation primarily causes processing plants for palletization and composting to work without adequate bio-degradable waste. Modern, eco-friendly treatment technologies need more emphasis than incineration [138]. As a method of waste disposal, landfilling needs substantial efforts to prevent its threats. An under-drainage system shall be available to provide efficient and adequate cover for the successive landfill layer with ventilation. Moreover, regular monitoring programs check greenhouse gas emissions into the air and leachate into the water environment. Greening and covering the landfill limits the wastes from birds and scavengers and prevent bio-magnification [108].

4. Key Challenges and Solutions to Solid Waste Management in India

Municipal facilities are improving, yet the most significant challenge for effective municipal SWM in India lies with the weak organizational structure; unscientific methods, financial constraints; shortage of low-lying land for landfilling, and, most importantly, lack of public awareness [115,118,124,127,139]. In India, solid waste management faces severe problems as only a fraction of the waste is disposed of appropriately. With numerous efforts, municipal agencies across India have increased waste collection coverage. However, about 18% of the waste remains uncollected. A meagre amount of waste is processed and treated for recycling or energy recovery. Several landfill sites, including larger ones such as Ghazipur, Bhalswa, Okhla in Delhi, and Deonar in Mumbai, have already reached their design life and are overflowing. Developing and operating new landfill sites using the scientific disposal method is challenging. Citizens should be aware of segregating waste at the source, avoiding littering on roadsides, and observing their duties towards the environment [25,131]. The informal sector, including the rag-pickers, manages to recycle in the cities. It is essential to segregate recyclable solid waste components from garbage bins and landfills [131]. Authorities have implemented policy measures to include the unorganized sector in solid waste management. It has increased waste collection and segregation and simultaneously solves logistical problems.

4.1. Waste-to-Energy Conversion

The solid, liquid, and gaseous discharges from industrial processes possess different WTE potential. Urban waste (solid and liquid discharges), cattle farm (solid discharges), distillery (liquid discharges), vegetable raw (solid discharges), poultry (solid discharges), slaughterhouse (liquid discharges), paper (liquid discharges), fruit raw (solid discharges), sugar press mud (solid discharges) are the top sectors in India with more than 5225 MW of WTE potential [140]. Until 2020, biogas generation projects have been set up in Uttar Pradesh, Madhya Pradesh, Andhra Pradesh, and Telangana, generating a cumulative installed capacity of 37900 m^3/day. States such as Maharashtra, Karnataka, and Punjab have established 4 MW power projects based on WTE. Bio-CNG generation has increased with 25731 kg/day cumulative installed capacity. Moreover, 139.80 MW grid-interactive and 114.93 MW off-grid power projects are operational. Presently, 57426 m^3 of biogas is generated per day from industrial units in Madhya Pradesh and Maharashtra from maize and starch processing. Ahmedabad and Surat generate 8600 kg/day of Bio-CNG from vegetable waste, hotel waste, cow dung, and sewage treatment plants. Wastes from gardens, poultry, animal husbandry, slaughterhouses, distillery, sugar manufacture, dairy, oil refining, leather processing, food, and food grain processing shall be employed aggressively for WTE. Self-sufficient industries shall be encouraged with in-house power generation from their wastes using WTE and other renewable energy sources.

4.2. Greenhouse Gas Emission

India is a developing nation improving its technology and infrastructure for waste-to-energy generation. Energy potential here can increase by tapping gases such as methane emitted from landfills. It can quickly and effectively be used as biogas for heat and electricity generation. On the contrary, landfills are the major contributor to methane emissions [98,117,137,141,142]. Landfill gas remains untapped as a potential source of energy. IPCC states

that 30% of the total GHG emission to the environment is from landfills. Delhi contributes to 10% of the total methane emission from MSW [4,34,143]. Further efforts are required to improve sustainability and reduce GHG emissions [124]. IPCC reports the worsening of climate change issues [144].

4.3. Health Issues

The landfills' wastes are mainly polymers, plastics, chemicals, electronic components, biomedical products, construction and demolition waste, household wastes, and many substances. Contaminants such as organic acids and heavy metals are proven allergens, potential carcinogens, and toxic, causing acute and chronic ailments. Cancer, congenital disabilities, tumor formation, reproductive defects, kidney problems, and cardiac, renal, and circulatory issues are known diseases handling solid wastes. There exists evidential proof that operators and manual workers suffer from acute toxicity to the kidney and nervous system [114]. Chronic diseases, such as cadmium poisoning and mercury poisoning, affect the population engaged in solid waste management. Moreover, the health risk is not limited to the landfill but extends to several kilometers from the landfill. Municipal workers and other communities associated with the handling of solid waste and management face respiratory and cardiovascular health issues such as bacterial infection, impaired lung function, low haemoglobin / RBC, altered immunity, allergy, asthma, and inflammation in the nose, throat, lung, and other impairments bear their mark [145].

4.4. Solutions to Key challenges

Awareness is an issue that mars the public and administration's potential. Both need to keep abreast of the present scenario and the fate of SWM. Awareness-related matters can be addressed by assessing public awareness and involvement through knowledge gap analysis, which promotes informed decision-making. Hence, better policy formulation, clear laws, and fee methodology. Improved public awareness plays a pivotal role in deciding the fate of the environment during pilgrimage and religious events. A moral and ethical value system toward solid waste management enhances the quality and quantity of waste collected. Numerous studies have pointed out that a door-to-door collection system is better at segregation than a point collection system. Improvement in segregation primarily impacts waste conversion into recycled goods and energy resources. This helps to obviate the emission of greenhouse gasses. Awareness of the administrative system can be increased by learning from advanced countries. Techniques such as artificial intelligence, machine learning, convolutional neural networks, internet-of-things, computer vision, etc., followed in developed countries, may help the automated identification and segregation of waste.

Waste-to-energy issues can be resolved by exploring newer avenues where untapped resources, such as the provisioning of landfill gas as fuel on a large scale—innovative ways to convert raw wastes into high-energy producing substances. Improved efficiency of collection and segregation is possible by forming a workable collection plan, and optimized route map created using geo-positional attributes. It will invariably reduce time-operation and savings in labor, distance, and pathways. Multi-criteria decision support systems can be used to select an appropriate methodology for waste treatment and save resources. Focus on environment-friendly ways such as composting shall be encouraged.

GHG emissions can be reduced by enabling a circular economy-based framework for industrialized products. The indirect causes shall also be identified and addressed in addition to factors directly affecting GHG emissions. An integrated solid waste management approach is a strategic approach to sustainable management of solid wastes covering all sources and all aspects, covering generation, segregation, transfer, sorting, treatment, recovery, and disposal in an integrated manner, with an emphasis on maximizing resource use efficiency, shall be implemented. Compared to incineration and landfilling, simple and natural methods such as composting are the best at reducing GHG emissions. Employing strategically placed storage areas of waste to facilitate quick loading and disposal also affects the emission of GHGs. Another avenue to reduce the generation of GHG is recycling. Solid waste, whenever possible, shall be reused and recycled. Recycling of construction and demolition minimizes the requirement for virgin materials to a great extent. India possesses a high potential for developing a culture of recycling in day-to-day life. GIS-based optimized routing plan for the transport and transit of wastes proves to have considerable savings in CO2 emission.

Numerous health issues from handling, processing and exposure to solid wastes are avoidable. Measures such as minimizing waste generation by studying and applying life-cycle analysis within the manufacturing process. A product designed for the smallest waste footprint can significantly influence waste generation. Disposal of waste may be kept in check for emission of heavy metals and other toxic chemicals. At present, open burning and dumping are the most prevalent methods of waste disposal Advancement in disposal technology is warranted for environment-friendly disposal methods. Reduction of Impact through an early warning system and decision-making system based on continuous analysis of temporal and spatial variation of waste characteristics is required. Mitigation shall not only focus on immediate impact minimization but also work on a long-term solution for environmental health. Due to the recent COVID-19 pandemic, a dedicated healthcare waste handling system, including measures for recycling, treatment, and disposal, has been established in many countries.

5. Conclusion

The study concludes that prompt action is necessary for data management, training, education, planning, monitoring, strict policy changes with strong legislation, the establishment of a reliable regulatory system, proper and effective management of all waste streams, public awareness, and best housekeeping practices for adequate implementation of solid waste management in the developing countries. This state-of-the-art review examines the literature on management, framework, awareness, generation, collection, segregation, characterization, minimization, storage, transport, treatment, recycling, disposal, impact, GHG emission, COVID-19, and waste-to-energy aspects of solid waste management. It finds

that one of the significant challenges to SWM is the steps to be employed to reverse the impacts caused by the previously used disposal methods. Despite plenty of studies on solid waste management aspects such as exploring and implementing scientific monitoring, treatment, and disposal methods; overhauling the organizational structure; searching for best practices for recycling and recovery from solid waste; tapping the waste-to-energy potential to its fullest extent from landfill-gas-to-energy and waste-to-energy; search for low-cost, and location-based strategies to achieve zero waste disposal at all sources of waste; achieving a superior level of awareness among citizens needs to be the emphasis. Future research in the solid waste management sector shall also focus on the lack of solid waste management infrastructure for rapidly growing urban areas, energy conversion from wastes and tapping landfill gasses, and trapping greenhouse gas emissions from various solid waste processing and disposal. After rigorous review, this paper concludes that all the World's developing and underdeveloped countries face similar solid waste management problems. Therefore, all countries' cumulative efforts for managing solid waste can bring forth excellent results around the globe.

Conflict-of-Interest

The authors declare that they have no conflict of interest.

Author Contributions

A.A.G. (Assistant Professor) conducted the literature review and data analysis and wrote the manuscript. S.K.S. (Professor) revised the manuscript.

References

1. Khajuria A, Matsui T, Machimura T. Economic growth decoupling municipal solid waste loads in terms of environmental kuznets curve: Symptom of the Decoupling in India. *J. Sustain. Dev.* 2011;4:51-8. https://doi.org/10.5539/jsd.v4n3p51.
2. Bhoyar SP, Dusad S, Shrivastava R, Mishra S, Gupta N, Rao AB. Understanding the impact of lifestyle on individual carbon-footprint. *Procedia - Soc. Behav. Sci.* 2014;133:47-60. https://doi.org/10.1016/j.sbspro.2014.04.168.
3. Beede DN, Bloom DE. The economics of municipal solid waste. *World Bank Res. Obs.* 1995;10:113-50. https://doi.org/10.1093/wbro/10.2.113.
4. Mor S, Ravindra K, De Visscher A, Dahiya RP, Chandra A. Municipal solid waste characterization and its assessment for potential methane generation: A case study. *Sci. Total Environ.* 2006;371:1-10. https://doi.org/10.1016/j.scitotenv.2006.04.014.
5. Yoshikawa N, Fujiwara N, Nagata J. Scenario analysis of greenhouse gases reduction by changing consumer's shopping behavior. *Energy Procedia* 2014;61:1532-5. https://doi.org/10.1016/j.egypro.2014.12.163.
6. Dinh C Le, Fujiwara T, Asari M, Duy B, Phu S. Optimization of solid waste collection system in a tourism destination. *Glob. J. Environ. Sci. Manag.* 2022;8:419-36. https://doi.org/ 10.22034/gjesm.2022.03.09.
7. Zeng C, Niu D, Zhao Y. A comprehensive overview of rural solid waste management in China. *Front. Environ. Sci. Eng.* 2015;9:949-61. https://doi.org/10.1007/s11783-015-0816-8.
8. Russo P, Walchhutter S, Slavov T, Santana L. Municipal solid waste management: Analysing the principles of the Brazilian National solid waste policy. *Int. J. Environ. Waste Manag.* 2022;1:1. https://doi.org/10.1504/IJEWM.2022.10035736.
9. Liang Y, Song Q, Wu N, Li J, Zhong Y, Zeng W. Repercussions of COVID-19 pandemic on solid waste generation and management strategies. *Front. Environ. Sci. Eng.* 2021;15:115. https://doi.org/10.1007/s11783-021-1407-5.
10. Al-Zboon K, Masoud T. Recycling of stone cutting waste in the construction sector: A review. *J. Solid Waste Technol. Manag.* 2021;47:56-60. https://doi.org/10.5276/JSWTM/2021.56.
11. Aziz R, Ruslinda Y, Yeni YS, Dewilda Y. Solid waste management system model of village tourism area of Nagari Tuo Pariangan, Indonesia. *IOP Conf. Ser. Mater. Sci. Eng.* 2021;1041:012030. https://doi.org/10.1088/1757-899X/1041/1/012030.
12. Iyamu HO, Anda M, Ho G. A review of municipal solid waste management in the BRIC and high-income countries: A thematic framework for low-income countries. *Habitat Int.* 2020;95:102097. https://doi.org/10.1016/j.habitatint.2019.102097.
13. Lee GB. Solid waste management: Issues and challenges in Asia. *Asian Product. Organ.* 2007:339. Available from: www.apo-tokyo.org
14. Ganguly RK, Chakraborty SK. Integrated approach in municipal solid waste management in COVID-19 pandemic: Perspectives of a developing country like India in a global scenario. *Case Stud. Chem. Environ. Eng.* 2021;3:100087. https://doi.org/10.1016/j.cscee.2021.100087.
15. Jha MK, Sondhi OAK, Pansare M. Solid waste management a case study of Ajmer. *Adv. Environ. Waste Manag. Recycl.* 2021;4:1153-60. https://doi.org/10.33140/AEWMR.04.03.08.
16. Mehta A, Chauhan D, Kumar S, Gour A. Assessing the environmental impacts associated with the life cycle of electronic equipment. *IOSR J. Environ. Sci. Ver. III* 2015;9:2319-99. https://doi.org/10.9790/2402-09634853.
17. Ahmad SFK, Ali UFM, Isa KM, Gopinath SCB. Study on characterization of bio-oil derived from sugarcane bagasse (Saccharum barberi) for application as biofuel. *Clean Energy* 2022;6:297-304. https://doi.org/10.1093/ce/zkac012.
18. Ahmad SFK, Ali UFM, Isa KM, Gopinath SCB. Evaluation of phenol formaldehyde resin synthesized from sugarcane from sugarcane bagasse bio-oil under optimized parameters. *J. Eng. Sci. Technol.* 2022;17:1523-31. Available from: http://jestec.taylors.edu.my/Vol%2017%20Issue%202%20April%20%202022/17_2_49.pdf.
19. Srivastava V, Ismail SA, Singh P, Singh RP. Urban solid waste management in the developing world with emphasis on India: challenges and opportunities. *Rev. Environ. Sci. Bio/Technology* 2015;14:317-37. https://doi.org/10.1007/s11157-014-9352-4.
20. Subramanyam R. Solid waste management in Lae City, Papua New Guinea. *J. Solid Waste Technol. Manag.* 2021;47:371-82. https://doi.org/10.5276/JSWTM/2021.371.
21. Jha R, Dwivedi S, Modhera B. Chapter 4 - Measurement and practices for hazardous waste management. Hazardous Waste

Management: An Overview of Advanced and Cost-Effective Solutions Manag. 2022, p. 89-115. https://doi.org/10.1016/B978-0-12-824344-2.00011-2.
22. Khan S, Anjum R, Raza ST, Ahmed Bazai N, Ihtisham M. Technologies for municipal solid waste management: Current status, challenges, and future perspectives. *Chemosphere* 2022; 288:132403. https://doi.org/10.1016/j.chemosphere.2021.132403.
23. Yousefloo A, Babazadeh R. Designing an integrated municipal solid waste management network: A case study. *J. Clean. Prod.* 2020;244:118824. https://doi.org/10.1016/j.jclepro.2019.118824.
24. Mathur V. Scope of recycling municipal solid waste in Delhi and National Capital Region (NCR). *Integral Rev. - J. Manag.* 2012;5:27-36. Available from: http://iul.ac.in/documents/12052010/IRJM_Paper3_dec2012.pdf.
25. Kaushal RK, Varghese GK, Chabukdhara M. Municipal solid waste management in India-Current state and future challenges: A review. *Int. J. Eng. Sci. Technol.* 2012;4:1473-89. Available from: http://www.ijest.info/docs/IJEST12-04-04-196.pdf.
26. Visvanathan C, Glawe U, Box P, Luang K. Domestic solid waste management in South Asian Countries - A comparative analysis 2. Solid waste generation. In: 3R South Asia Expert Workshop. 30 August - 1 Sept. 2006, Kathmandu. p. 1-14. Available from: https://www.iges.or.jp.
27. Sridhar KS. Solid Waste Management in Asia Pacific. *Public Work. Manag. Policy* 2016;21:53-70. https://doi.org/10.1177/1087724X15570642.
28. Ebekozien A, Aigbavboa C, Nwaole ANC, Aginah IL, Aigbedion M. Sustainable cities through household waste management: an unexplored approach to challenges confronting private solid waste management. *Facilities* 2022:40(5/6):365-379. https://doi.org/10.1108/F-09-2021-0078.
29. Pérez LE, Ziegler-Rodríguez K, Pérez ATE, Vásquez ÓC, Vázquez-Rowe I. Closing the gap in the municipal solid waste management between metropolitan and regional cities from developing countries: A life cycle assessment approach. *Waste Manag.* 2021;124:314-24. https://doi.org/10.1016/j.wasman.2021.02.020.
30. Bui TD, Tsai FM, Tseng M-L, Ali MH. Identifying sustainable solid waste management barriers in practice using the fuzzy Delphi method. *Resour. Conserv. Recycl.* 2020;154:104625. https://doi.org/10.1016/j.resconrec.2019.104625.
31. Das S, Lee S, Kumar P, Kim K-H, Lee SS, Bhattacharya SS. Solid waste management: Scope and the challenge of sustainability. *J. Clean. Prod.* 2019;228:658-78. https://doi.org/10.1016/j.jclepro.2019.04.323.
32. Adipah S, Kwame ON. A novel introduction of municipal solid waste management. *J. Environ. Sci. Public Heal.* 2018;03. https://doi.org/10.26502/jesph.96120055.
33. Asefi H, Lim S. A novel multi-dimensional modeling approach to integrated municipal solid waste management. *J. Clean. Prod.* 2017;166:1131-43. https://doi.org/10.1016/j.jclepro.2017.08.061.
34. Chen Y-C, Lo S-L. Evaluation of greenhouse gas emissions for several municipal solid waste management strategies. *J. Clean. Prod.* 2016;113:606-12. https://doi.org/10.1016/j.jclepro.2015.11.058.
35. Chang N-B, Pires A. Sustainable solid waste management: A systems engineering approach. John Wiley & Sons, Inc.; 2015. p. 1-908 https://doi.org/10.1002/9781119035848.
36. Allesch A, Brunner PH. Assessment methods for solid waste management: A literature review. *Waste Manag. Res.: J. Sustain. Circ. Econ.* 2014;32:461-73. https://doi.org/10.1177/0734242X14535653.
37. Marshall RE, Farahbakhsh K. Systems approaches to integrated solid waste management in developing countries. *Waste Manag.* 2013;33:988-1003. https://doi.org/10.1016/j.wasman.2012.12.023.
38. Chandrappa R, Das DB. Solid Waste Management: Principles and practice. Environmental Science and Engineering Book Series. Springer; 2012. p. 1-414. https://doi.org/10.1007/978-3-642-28681-0.
39. Bolingbroke D, Ng KTW, Vu HL, Richter A. Quantification of solid waste management system efficiency using input-output indices. *J. Mater. Cycles Waste Manag.* 2021;23:1015-25. https://doi.org/10.1007/s10163-021-01187-7.
40. Gaeta GL, Ghinoi S, Silvestri F, Tassinari M. Innovation in the solid waste management industry: Integrating neoclassical and complexity theory perspectives. *Waste Manag.* 2021;120:50-8. https://doi.org/10.1016/j.wasman.2020.11.009.
41. Folarin OS. Achieving sustainable solid waste management in Sub-Saharan Africa: The option of valorisation and circular economy model. In: Yaser AZ, Tajarudin HA, Embrandiri A (eds) Waste Management Processing and Valorisation, Springer; 2022, p. 285-300. https://doi.org/10.1007/978-981-16-7653-6_15.
42. Azevedo BD, Scavarda LF, Caiado RGG, Fuss M. Improving urban household solid waste management in developing countries based on the German experience. *Waste Manag.* 2021;120:772-83. https://doi.org/10.1016/j.wasman.2020.11.001.
43. Brotosusilo A, Nabila SH, Negoro HA, Utari D. The level of individual participation of community in implementing effective solid waste management policies. *Glob. J. Environ. Sci. Manag.* 2020;6:341-54. https://doi.org/10.22034/gjesm.2020.03.05.
44. Kayamo SE. Willingness to pay for solid waste management improvement in Hawassa city, Ethiopia. *J. Environ. Manage.* 2022;302:113973. https://doi.org/10.1016/j.jenvman.2021.113973.
45. Kuya RM, Oindo BO, Long'ora AE. Influence of levels of awareness, willingness to pay, and incentives on solid waste management practices by commercial enterprises in Kakamega Town. *Eur. J. Environ. Earth Sci.* 2022;3:7-13. https://doi.org/10.24018/ejgeo.2022.3.1.236.
46. Debrah JK, Vidal DG, Dinis MAP. Raising Awareness on solid waste management through formal education for sustainability: A developing countries evidence review. *Recycling* 2021;6:1-21. https://doi.org/10.3390/recycling6010006.
47. Liu B, Zhang L, Wang Q. Demand gap analysis of municipal solid waste landfill in Beijing: Based on the municipal solid waste generation. *Waste Manag.* 2021;134:42-51. https://doi.org/10.1016/j.wasman.2021.08.007.
48. Kumar A, Agrawal A. Recent trends in solid waste management status, challenges, and potential for the future Indian cities - A review. *Curr. Res. Environ. Sustain.* 2020;2:100011. https://doi.org/10.1016/j.crsust.2020.100011.
49. Dinh C, Fujiwara T, Phu S. Waste segregation at source and separately collection-The inadequacy in implementation in Vietnam. *Chem. Eng. Trans.* 2021;89:499-504. https://doi.org/10.3303/CET2189084.

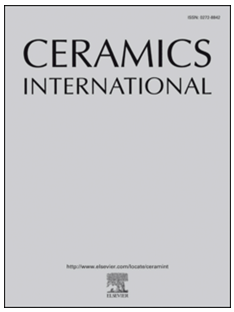
50. Al-Refaie A, Al-Hawadi A, Fraij S. Optimization models for clustering of solid waste collection process. *Eng. Optim.* 2021; 53:2056-69. https://doi.org/10.1080/0305215X.2020.1843165.
51. Gadaleta G, De Gisi S, Todaro F, Campanaro V, Teodosiu C, Notarnicola M. Sustainability assessment of municipal solid waste separate collection and treatment systems in a large metropolitan area. *Sustain. Prod. Consum.* 2022;29:328-40. https://doi.org/10.1016/j.spc.2021.10.023.
52. Das A, Shukla A, Manjunatha R, Lodhi EA. IoT based solid waste segregation using relative humidity values. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE; 04-06 February 2021; India. p. 312-319. https://doi.org/10.1109/ICICV50876.2021.9388611.
53. Abylkhani B, Guney M, Aiymbetov B et al. Detailed municipal solid waste composition analysis for Nur-Sultan City, Kazakhstan with implications for sustainable waste management in Central Asia. *Environ. Sci. Pollut. Res.* 2021;28: 24406-18. https://doi.org/10.1007/s11356-020-08431-x.
54. Abdulredha M, Khaddar RAL, Jordan D, Hashim K. The Development of a Waste Management System in Kerbala during Major Pilgrimage Events: Determination of Solid Waste Composition. *Procedia Eng.* 2017;196:779-84. https://doi.org/10.1016/j.proeng.2017.08.007.
55. Adeleke O, Akinlabi S, Jen T-C, Dunmade I. A machine learning approach for investigating the impact of seasonal variation on physical composition of municipal solid waste. *J. Reliab. Intell. Environ.* 2022. https://doi.org/10.1007/s40860-021-00168-9.
56. Padmavathy A, Anbarashan M. Chapter 17-Potential of solid waste prevention and minimization strategies. Cost effective technologies for solid waste and wastewater treatment: Advances in environmental pollution research, Elsevier; 2022, p. 219-28. https://doi.org/10.1016/B978-0-12-822933-0.00019-X.
57. Pan Y, Li M, Guo H, Li Y, Han J. Influencing factors and reduction of domestic solid waste at university dormitory in Shanghai, China. *Sci. Rep.* 2022;12:570. https://doi.org/10.1038/s41598-021-04582-0.
58. Erfani SMH, Danesh S, Karrabi SM, Gheibi M, Nemati S. Statistical analysis of effective variables on the performance of waste storage service using geographical information system and response surface methodology. *J. Environ. Manag.* 2019; 235:453-62. https://doi.org/10.1016/j.jenvman.2019.01.061.
59. Gao W, Xue-Cheng Bian ;, Xu W-J, Chen Y-M. Storage Capacity and Slope Stability Analysis of Municipal Solid Waste Landfills. *J. Perform. Constr. Facil.* 2018;32:04018036. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001180.
60. Mojtahedi M, Fathollahi-Fard AM, Tavakkoli-Moghaddam R, Newton S. Sustainable vehicle routing problem for coordinated solid waste management. *J. Ind. Inf. Integr.* 2021;23:100220. https://doi.org/10.1016/j.jii.2021.100220.
61. Mohsenizadeh M, Tural MK, Kentel E. Municipal solid waste management with cost minimization and emission control objectives: A case study of Ankara. *Sustain. Cities Soc.* 2020;52: 101807. https://doi.org/10.1016/j.scs.2019.101807.
62. Al-Nawaiseh A, Aljbour S, Al-Hamaiedeh H, El-Hasan T, Hemidat S, Nassour A. Composting of organic waste: A sustainable alternative solution for solid waste management in Jordan. *Jordan J. Civ. Eng.* 2021;15:363-77. https://jjce.just.edu.jo/issues/paper.php?p=6040.pdf.
63. Chen C, Wen Z, Wang Y, Zhang W, Zhang T. Multi-objective optimization of technology solutions in municipal solid waste treatment system coupled with pollutants cross-media metabolism issues. *Sci. Total Environ.* 2022;807:150664. https://doi.org/10.1016/j.scitotenv.2021.150664.
64. Srivastava AN, Chakma S. 19-Bioreactor landfills: sustainable solution for disposal of municipal solid waste. Advanced Organic Waste Management: Sustainable Practices and Approaches, Elsevier; 2022, p. 315-28. https://doi.org/10.1016/B978-0-323-85792-5.00009-5.
65. Alam O, Qiao X. An in-depth review on municipal solid waste management, treatment and disposal in Bangladesh. *Sustain. Cities Soc.* 2020;52:101775. https://doi.org/10.1016/j.scs.2019.101775.
66. Fang W, Huang Y, Ding Y, Qi G, Liu Y, Bi J. Health risks of odorous compounds during the whole process of municipal solid waste collection and treatment in China. *Environ. Int.* 2022;158:106951. https://doi.org/10.1016/j.envint.2021.106951.
67. Gómez-Sanabria A, Kiesewetter G, Klimont Z, Schoepp W, Haberl H. Potential for future reductions of global GHG and air pollutants from circular waste management systems. *Nat. Commun.* 2022;13:106. https://doi.org/10.1038/s41467-021-27624-7.
68. Michel Devadoss PS, Agamuthu P, Mehran SB, Santha C, Fauziah SH. Implications of municipal solid waste management on greenhouse gas emissions in Malaysia and the way forward. *Waste Manag.* 2021;119:135-44. https://doi.org/10.1016/j.wasman.2020.09.038.
69. Kristanto GA, Koven W. Estimating greenhouse gas emissions from municipal solid waste management in Depok, Indonesia. *City Environ. Interact.* 2019;4:100027. https://doi.org/10.1016/j.cacint.2020.100027.
70. Das AK, Islam MN, Billah MM, Sarker A. COVID-19 pandemic and healthcare solid waste management strategy - A mini-review. *Sci. Total Environ.* 2021;778:146220. https://doi.org/10.1016/j.scitotenv.2021.146220.
71. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* 2020;382:1564-7. https://doi.org/10.1056/NEJMc2004973.
72. Karakoti M, Pandey S, Tatrari G, Dhapola PS, Jangra R, Dhali S, et al. An approach towards waste to energy from effective conversion of solid waste plastics into graphene nanosheets by using different catalysts for high performance supercapacitors: A comparative study. *Mater. Adv.* 2022;11. https://doi.org/10.1039/D1MA01136G.
73. Chaianong A, Pharino C. How to design an area-based prioritization of biogas production from organic municipal solid waste? Evidence from Thailand. *Waste Manag.* 2022;138:243-52. https://doi.org/10.1016/j.wasman.2021.11.042.
74. Dlamini S, Simatele MD, Serge Kubanza N. Municipal solid waste management in South Africa: from waste to energy recovery through waste-to-energy technologies in Johannesburg. *Local Environ.* 2019;24:249-57. https://doi.org/10.1080/13549839.2018.1561656.
75. Khajuria A, Matsui T, Machimura T, Morioka T. Decoupling

and environmental kuznets curve for municipal solid waste generation: Evidence from India. *Int. J. Environ. Sci.* 2012;2: 1670-4. https://doi.org/10.6088/ijes.00202030051.
76. Hoornweg D, Bhada-Tata P. What a waste: a global review of solid waste management. World Bank, Washington DC; 2012. Available from: http://hdl.handle.net/10986/17388
77. Limon MR, Vallente JPC, Cajigal ARV, Aquino MU, Aragon JA, Acosta RL. Unmasking emerging issues in solid waste management: Knowledge and self-reported practices on the discarded disposable masks during the COVID-19 pandemic in the Philippines. *Environ. Challenges* 2022;6:100435. https://doi.org/10.1016/j.envc.2021.100435.
78. Rowland ED, Omonefe F. Environmental monitory and impact assessment of solid waste dumpsite using multispectral imagery in Yenagoa, Bayelsa state, Nigeria. *Int. J. Environ. Sci. Technol.* 2022;19:1007-24. https://doi.org/10.1007/s13762-021-03456-2.
79. Salazar-Adams A. The efficiency of municipal solid waste collection in Mexico. *Waste Manag.* 2021;133:71-9. https://doi.org/10.1016/j.wasman.2021.07.008.
80. Penteado CSG, Castro MAS de. COVID-19 effects on municipal solid waste management: What can effectively be done in the Brazilian scenario? *Resour. Conserv. Recycl.* 2021;164:105152. https://doi.org/10.1016/j.resconrec.2020.105152.
81. Dastjerdi B, Strezov V, Kumar R, He J, Behnia M. Comparative life cycle assessment of system solution scenarios for residual municipal solid waste management in NSW, Australia. *Sci. Total Environ.* 2021;767:144355. https://doi.org/10.1016/j.scitotenv.2020.144355.
82. Kanhai G, Fobil JN, Nartey BA, Spadaro J V., Mudu P. Urban Municipal Solid Waste management: Modeling air pollution scenarios and health impacts in the case of Accra, Ghana. *Waste Manag.* 2021;123:15-22. https://doi.org/10.1016/j.wasman.2021.01.005.
83. de Souza VM, Bloemhof J, Borsato M. Assessing the eco-effectiveness of a solid waste management plan using agent-based modelling. *Waste Manag.* 2021;125:235-48. https://doi.org/10.1016/j.wasman.2021.02.019.
84. Mir IS, Cheema PPS, Singh SP. Implementation analysis of solid waste management in Ludhiana city of Punjab. *Environ. Challenges* 2021;2:100023. https://doi.org/10.1016/j.envc.2021.100023.
85. Sharma BK, Chandel MK. Life cycle cost analysis of municipal solid waste management scenarios for Mumbai, India. *Waste Manag.* 2021;124:293-302. https://doi.org/10.1016/j.wasman.2021.02.002.
86. Oladunjoye MA, Adeniran MA. Towards efficient management of solid wastes: An overview of dumpsites in Ibadan Metropolis, Southwestern Nigeria. *J. Solid Waste Technol. Manag.* 2021; 47:271-84. https://doi.org/10.5276/JSWTM/2021.271.
87. Cheniti H, Cheniti M, Brahamia K. Use of GIS and Moran's I to support residential solid waste recycling in the city of Annaba, Algeria. *Environ. Sci. Pollut. Res.* 2021;28:34027-41. https://doi.org/10.1007/s11356-020-10911-z.
88. Mehra A, Pandya J, Singh J, Golhar P, Kamble DS, Bhargava DA. Transformation of Municipal Solid Waste into Fuel Pellets Case Study of Newai, Rajasthan. *Adv. Environ. Waste Manag. Recycl.* 2021;4:43-6. https://doi.org/10.33140/aewmr.04.01.06.
89. Nzediegwu C, Chang SX. Improper solid waste management increases potential for COVID-19 spread in developing countries. *Resour. Conserv. Recycl.* 2020;161:104947. https://doi.org/10.1016/j.resconrec.2020.104947.
90. Rai RK, Nepal M, Khadayat MS, Bhardwaj B. Improving Municipal Solid Waste Collection Services in Developing Countries: A Case of Bharatpur Metropolitan City, Nepal. *Sustainability* 2019;11:3010. https://doi.org/10.3390/su11113010.
91. Pereira TS, Fernandino G. Evaluation of solid waste management sustainability of a coastal municipality from northeastern Brazil. *Ocean Coast. Manag.* 2019;179:104839. https://doi.org/10.1016/j.ocecoaman.2019.104839.
92. Ayvaz-Cavdaroglu N, Coban A, Firtina-Ertis I. Municipal solid waste management via mathematical modeling: A case study in İstanbul, Turkey. *J. Environ. Manage.* 2019;244:362-9. https://doi.org/10.1016/j.jenvman.2019.05.065.
93. Sinthumule N, Mkumbuzi S. Participation in community-based solid waste management in Nkulumane Suburb, Bulawayo, Zimbabwe. *Resources* 2019;8:30. https://doi.org/10.3390/resources8010030.
94. Greco G, Cenciarelli VG, Allegrini M. Tourism's impacts on the costs of municipal solid waste collection: Evidence from Italy. *J. Clean. Prod.* 2018;177:62-8. https://doi.org/10.1016/j.jclepro.2017.12.179.
95. Rana R, Ganguly R, Kumar Gupta A. Evaluation of solid waste management in satellite towns of Mohali and Panchkula-India. *J. Solid Waste Technol. Manag.* 2017;43:280-94. https://doi.org/10.5276/JSWTM.2017.280.
96. Adeniran A, Nubi A, Adelopo AO. Solid waste generation and characterization in the University of Lagos for a sustainable waste management. *Waste Manag.* 2017;67:3-10. https://doi.org/10.1016/j.wasman.2017.05.002.
97. Mmereki D, Baldwin A, Li B, Liu M. Healthcare waste management in Botswana: storage, collection, treatment and disposal system. *J. Mater. Cycles Waste Manag.* 2017;19:351-65. https://doi.org/10.1007/s10163-015-0429-0.
98. Singh SK, Gour AA, Rohit G, Shivangi G, Vipul V. Greenhouse gas emissions from landfills: A case of NCT of Delhi, India. *J. Climatol. Weather Forecast.* 2016;04. https://doi.org/10.4172/2332-2594.1000157.
99. Sadef Y, Nizami AS, Batool SA et al. Waste-to-energy and recycling value for developing integrated solid waste management plan in Lahore. *Energy Sources B: Econ. Plan. Policy.* 2016;11:569-79. https://doi.org/10.1080/15567249.2015.1052595.
100. Al-Khatib IA, Kontogianni S, Abu Nabaa H, Alshami N, Al-Sari MI. Public perception of hazardousness caused by current trends of municipal solid waste management. *Waste Manag.* 2015;36:323-30. https://doi.org/10.1016/j.wasman.2014.10.026.
101. Chavan BL, Zambare NS, Pawar SH. Impact of solid waste dump on ground water quality in the village Kasaba-Bawda, Kolhapur district, Maharashtra, India. *Adv. Appl. Sci. Res.* 2014;5:59-64. Available from: www.pelagiaresearchlibrary.com.
102. Ahsan A, Alamgir M, El-Sergany MM, Shams S, Rowshon MK, Daud NNN. Assessment of Municipal Solid Waste Management System in a Developing Country. *Chinese J. Eng.* 2014;2014:1-11. https://doi.org/10.1155/2014/561935.
103. Chaves GLD, Santos JLJ, Rocha SMS. The challenges for solid

waste management in accordance with Agenda 21: A Brazilian case review. *Waste Manag. Res.: J. Sustain. Circ. Econ.* 2014;32:19-31. https://doi.org/10.1177/0734242X14541987.
104. Njoroge BNK, Kimani M, Ndunge D. Review of Municipal Solid Waste Management: A Case Study of Nairobi. *Res. Inven. Int. J. Eng. Sci.* 2014;4:16-20. Available from: https://www.researchinventy.com.
105. Hazra T, Goel S, Maitra B. Willingness-to-pay for solid waste management service attributes: Kolkata Municipal Corporation area, India, as a case study. *Int. J. Environ. Waste Manag.* 2013;12:406. https://doi.org/10.1504/IJEWM.2013.056627.
106. Wang H, Wang C. Municipal solid waste management in Beijing: characteristics and challenges. *Waste Manag. Res. J. a Sustain. Circ. Econ.* 2013;31:67-72. https://doi.org/10.1177/0734242X12468199.
107. Pandey D, Choudhry BM. Route optimization for municipal solid waste disposal sites in Gondia, Maharashtra State, India. *Int. J. Bioassays* 2013;2:661-6. Available from: https://www.ijbio.com.
108. Kumar MA. Existing situation of municipal solid waste management in NCT of Delhi, India. *Res. Forum Int. J. Soc. Sci.* 2014;1(1):6-17. http://www.ijss.publicationsupport.com/docs/paper/Volume-1/issue_1/IJSS-104.pdf.
109. Desa A, Kadir NBA, Yusooff F. Waste education and awareness strategy: towards solid waste management (SWM) program at UKM. *Procedia - Soc. Behav. Sci.* 2012;59:47-50. https://doi.org/10.1016/j.sbspro.2012.09.244.
110. Ansari A, Saleh M. Municipal solid waste management systems in the Kingdom of Bahrain. *Int. J. Water Resour. Environ. Eng.* 2012;4:150-61. https://doi.org/10.5897/IJWREE1\2.022.
111. Aderemi A, Otitoloju A. An assessment of landfill fires and their potential health effects - A case study of a municipal solid waste landfill in Lagos, Nigeria. *Int. J. Environ. Prot.* 2012;2:22-5. Available from: https://www.ij-ep.org.
112. Zurbrügg C, Gfrerer M, Ashadi H, Brenner W, Küper D. Determinants of sustainability in solid waste management - The Gianyar Waste Recovery Project in Indonesia. *Waste Manag.* 2012;32:2126-33. https://doi.org/10.1016/j.wasman.2012.01.011.
113. Bhambulkar A V, Khedikar IP. Municipal solid waste (MSW) collection route for Laxmi Nagar by geographical information system. *Int. J. Adv. Eng. Technol.* 2011;2:48-53. Available from: https://www.technicaljournalsonline.com.
114. Fahmi W, Sutton K. Cairo's Contested Garbage: Sustainable solid waste management and the Zabaleen's right to the city. *Sustainability* 2010;2:1765-83. https://doi.org/10.3390/su2061765.
115. De Feo G, De Gisi S. Domestic separation and collection of municipal solid waste: Opinion and awareness of citizens and workers. *Sustainability* 2010;2:1297-326. https://doi.org/10.3390/su2051297.
116. Adefemi SO, Awokunmi EE. The impact of municipal solid waste disposal in Ado-Ekiti metropolis, Ekiti-State, Nigeria. *African J. Environ. Sci. Technol.* 2009;3:186-9. https://doi.org/10.5897/AJEST09.075.
117. Jha AK, Sharma C, Singh N, Ramesh R, Purvaja R, Gupta PK. Greenhouse gas emissions from municipal solid waste management in Indian mega-cities: A case study of Chennai landfill sites. *Chemosphere* 2008;71:750-8. https://doi.org/10.1016/j.chemosphere.2007.10.024.
118. Schwarz-Herion O, Practical Action Nepal. A case study on successful municipal solid waste management in industrialized countries by the example of Karlsruhe City, Germany. *Ann. Fac. Eng. Hunedoara: J. Eng.* 2008:VI(3):1-59. Available from: https://annals.fih.upt.ro/pdf-full/2008/ANNALS-2008-3-46.pdf
119. Rathi S. Optimization model for integrated municipal solid waste management in Mumbai, India. *Environ. Dev. Econ.* 2007;12:105-21. https://doi.org/10.1017/S1355770X0600341X.
120. Chiemchaisri C, Juanga JP, Visvanathan C. Municipal solid waste management in Thailand and disposal emission inventory. *Environ. Monit. Assess.* 2007;135:13-20. https://doi.org/10.1007/s10661-007-9707-1.
121. Kaza S, Yao L, Bhada-Tata P, Van Woerden F. What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. 2018. https://doi.org/10.1596/978-1-4648-1329-0.
122. Bose A. Census of India, 2011. *Economic and Political Weekly*. 2011. Available from: www.censusindia.gov.in.
123. Municipal Solid Waste (Management & Handling) Rule, India. Ministry of Environment Forest and Climate Change-Govt. of India; 2016. Available from: https://cpcb.nic.in/uploads/MSW/SWM_2016.pdf.
124. Dhokhikah Y, Trihadiningrum Y. Solid waste management in Asian Developing Countries: Challenges and opportunities. *J. Appl. Environ. Biol. Sci.* 2012;2:329-35. Available from: www.textroad.com.
125. Khajuria A, Yamamoto Y, Morioka T. Estimation of municipal solid waste generation and landfill area in Asian developing countries. *J. Environ. Biol.* 2010;31:649-54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21387916.
126. Chua KH, Jati E, Sahid M, Leong YP. Sustainable Municipal Solid Waste Management and GHG Abatement in Malaysia. In: International Conference on Innovation and Technopreneurship-ST-4: Green Energy Management; 26-28 July 2011; Malaysia. p. 1-8. https://www.academia.edu/download/31052017/Sustainable_Municipal_Solid_Waste_Management_and_GHG_Abatement_in_Malaysia.pdf.
127. Eaton D, Hilhorst T. Opportunities for managing solid waste flows in the peri-urban interface of Bamako and Ouagadougou. *Environ. Urban.* 2003;15:53-63. https://doi.org/10.1630/095624703101286538.
128. Consolidated Annual Report (For the year 2016-2017) on Implementation of Solid Waste Management Rules, 2016. 2016. Available from: https://cpcb.nic.in/uploads/MSW/MSW_AnnualReport_2016-17.pdf.
129. Pariatamby A, Tanaka M. Municipal Solid Waste Management in Asia and the Pacific Islands. 1st ed. Singapore: Springer Singapore; 2014. https://doi.org/10.1007/978-981-4451-73-4.
130. Survey on the Current Status of Municipal Solid Waste Management in Indian Cities and the Potential of Landfill Gas to Energy Projects in India. Federation of Indian Chambers Commerce and Industry. 2009:1-18. https://www.niua.org/csc/assets/pdf/waste/Survey-on-the-Current-Status-of-Municipal-SWM-by-FICCI.pdf.
131. Ramachandra TV, Bachamanda S. Environmental audit of

Municipal Solid Waste Management. *Int. J. Environ. Technol. Manag.* 2007;7:369. https://doi.org/10.1504/IJETM.2007.015152.
132. Ghose MK, Dikshit AK, Sharma SK. A GIS based transportation model for solid waste disposal - A case study on Asansol municipality. *Waste Manag.* 2006;26:1287-93. https://doi.org/10.1016/j.wasman.2005.09.022.
133. Sumathi VR, Natesan U, Sarkar C. GIS-based approach for optimized siting of municipal solid waste landfill. *Waste Manag.* 2008. https://doi.org/10.1016/j.wasman.2007.09.032.
134. Lella J, Mandla VR, Zhu X. Solid waste collection/transport optimization and vegetation land cover estimation using Geographic Information System (GIS): A case study of a proposed smart-city. *Sustain. Cities Soc.* 2017. https://doi.org/10.1016/j.scs.2017.08.023.
135. Singh A. Remote sensing and GIS applications for municipal waste management. *J. Environ. Manage.* 2019. https://doi.org/10.1016/j.jenvman.2019.05.017.
136. Sarkar P. Solid waste management in Delhi - A social vulnerability study. In: Bunch MJ, Suresh VM, Kumaran TV (eds.) Proceedings of the Third International Conference on Environment and Health, 15-17 December, 2003. India. P. 451-464. Available from: https://www.yorku.ca/bunchmj/ICEH/proceedings/Sarkar_P_ICEH_papers_451to464.pdf.
137. Chakraborty M, Sharma C, Pandey J, Singh N, Gupta PK. Methane emission estimation from landfills in Delhi: A comparative assessment of different methodologies. *Atmos. Environ.* 2011;45:7135-42. https://doi.org/10.1016/j.atmosenv.2011.09.015.
138. Li M, Xiang J, Hu S, Sun L-S, Su S, Li P-S, et al. Characterization of solid residues from municipal solid waste incinerator. *Fuel* 2004;83:1397-405. https://doi.org/10.1016/j.fuel.2004.01.005.
139. Wang H, Nie Y. Municipal solid waste characteristics and management in China. *J. Air Waste Manag. Assoc.* 2001;51:250-63. https://doi.org/10.1080/10473289.2001.10464266.
140. Waste to Energy: Overview. Ministry of New and Renewable Energy-Govt. India; 2020. https://mnre.gov.in/waste-to-energy/current-status.
141. Pardo G, Moral R, Aguilera E, Prado A. Gaseous emissions from management of solid waste: a systematic review. *Glob. Chang. Biol.* 2015;21:1313-27. https://doi.org/10.1111/gcb.12806.
142. Kumar A, Sharma MP. Estimation of GHG emission and energy recovery potential from MSW landfill sites. *Sustain. Energy Technol. Assessments.* 2014;5:50-61. https://doi.org/10.1016/j.seta.2013.11.004.
143. Talyan V, Dahiya RP, Anand S, Sreekrishnan TR. Quantification of methane emission from municipal solid waste disposal in Delhi. *Resour. Conserv. Recycl.* 2007;50:240-59. https://doi.org/10.1016/j.resconrec.2006.06.002.
144. Solomon SC, Qin D, Manning MR et al. Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Available from: https://www.ipcc.ch/site/assets/uploads/2018/05/ar4_wg1_full_report-1.pdf.
145. Yannawar VB, Shaikh P, Bhosle AB. The impact of landfill on soil and groundwater quality of the Nanded city, Maharashtra. *Researcher* 2012;4:56-63. Available from: http://www.sciencepub.net/researcher/research0407/007_10305research0407_56_63.pdf.

Journal Pre-proof



Structural and optical studies on Dy^{3+} doped $Gd_2Ti_2O_7$ pyrochlore as white light emission

Umang Berwal, Vinod Singh, Rinku Sharma

PII: S0272-8842(22)04026-3

DOI: https://doi.org/10.1016/j.ceramint.2022.11.045

Reference: CERI 34852

To appear in: *Ceramics International*

Received Date: 19 September 2022

Revised Date: 29 October 2022

Accepted Date: 2 November 2022

Please cite this article as: U. Berwal, V. Singh, R. Sharma, Structural and optical studies on Dy^{3+} doped $Gd_2Ti_2O_7$ pyrochlore as white light emission, *Ceramics International* (2022), doi: https://doi.org/10.1016/j.ceramint.2022.11.045.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Ltd.

Journal Pre-proof

Structural and optical studies on Dy^{3+} doped $Gd_2Ti_2O_7$ pyrochlore as white light emission

Umang Berwal[1], Vinod Singh[1], and Rinku Sharma[1]

[1]Department of Applied Physics, Delhi Technological University, New Delhi-110042, India

E-mail: *vinodsingh@dtu.ac.in*

Abstract

A series of Dy^{3+} trivalent ion embedded into the GTO pyrochlore ($Gd_{2-x}Dy_xTi_2O_7$ (x = 0.5, 1.0, 1.5, 2.0, and 2.5 mol%)) have been synthesized through standard solid state reaction technique. The single phasic nature with Fd3m space group was determined through XRD, Raman spectroscopy, and photoluminescence spectroscopy. Rietveld refinement analysis results confirm that the substitution is not going to disturb the crystallographic structure as well unit cell dimensions. From Raman deconvoluted spectrums, it was found that all the vibrational bands show a very small blue shifting on Dy^{3+} substitution. The photoluminescent EM-spectra for all the samples were examined under the excitation wavelength 347 nm gives two high intense emission bands centered around 484 nm (blue) and 581 nm (yellow) due to the $^4F_{9/2} \rightarrow {}^6H_{15/2}$ and $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transitions respectively, and less intense band appeared at wavelength of 456 nm arising due to the $^4I_{13/2} \rightarrow {}^6H_{13/2}$ transition of Dy^{3+} ions. For the optimum Dy^{3+} concentration (2.0 mol%), the Y/B ratio is found to be 0.8753 which is close to unity hence combination of both yellow and blue emission bands emits white light as validated by CIE chromaticity coordinates calculated values are (x = 0.34959, y = 0.35381). Therefore, the results reveal that Dy^{3+} substituted GTO pyrochlore samples could be used as luminescent material for wLEDs application.

Journal Pre-proof

Highlights

1. We have synthesized a series of Dy^{3+} doped $Gd_2Ti_2O_7$ through solid-state reaction methods.
2. Pyrochlore superstructure are being intact in all samples without any doping induced lattice distortion as observed through XRD and Raman spectroscopy.
3. Dopant Dy^{3+} driven PL emission through the host $Gd_2Ti_2O_7$ contain the most intense band of blue (484 nm) and yellow light (581 nm).
4. The value of CCT for 2.0 mol % Dy^{3+} ion substituted GTO is 4844.95 lying in the standard cool white light calorimetric point.

Journal Pre-proof

1. Introduction

Over the past few years, the trivalent rare earth ions doped materials with electronic transition within 4f orbitals were extensively explored due to their potential application in the field of material science, chemistry, physics, and chemical sciences due to their important role in applications like optical fibre [1], planar waveguides [2], compact micro-chip lasers [3], solid state lasers [4], cathode ray tube [5], field emission display [6], white light emitting diodes [7] and florescent tubes [8] *etc.* Researchers has been focused much on the white light emitting diodes (wLEDs) in the solid-state lighting field despite of fluorescent lamps due to their excellent advantages such as small size, energy saving, high energy efficiency, robustness, longer life times, environmental friendliness, and high brightness [9–12]. The designing of a new high efficiency luminescent device for the production of wLEDs with improved performance can be achieved by selecting appropriate host and trivalent rare earth ions. There are two methods to obtain white light from solid state lighting device [13] (SSLD) *via* **(i)** RGB-LEDs (free from phosphor) method in which devices providing red, green and blue colours are efficiently mixed together to illuminate white light, **(ii)** phosphor integrated SSLD uses an ultra violet (UV)/ near ultra-violet (NUV) light. The limitation of first method is the adjustment of power supply for each LED in accordance to balance the efficient mixing and emission intensity of individual colour. While the serious problem with second method of phosphor integrated SSLD is colour dependence, blue-yellow colour separation poor colour rendering index (the illuminated objects colour do not have their true colour) *etc.* [14].

Also, the UV/NUV LEDs with multiple colours emitting phosphors is an alternative way to form white light emission with good colour rendering index due to ease of colour mixing and control but this method also has serious drawback such as low luminescence efficiency due to blending of red, green and blue emitting phosphors [15]. Hence for the fabrication of wLEDs using UV/NUV LEDs, single phase phosphor is necessary to avoid above problems. Various approaches are used to make white light emission from single host lattice by **(a)** doping single rare earth ion, **(b)** doping of more than one rare earth ions, **(c)** co-doping of various ions and controlling emission *via* energy transfer process, and **(d)** controlling the defect formation through ion concentration.

Ternary pyrochlore oxides are considered to be suitable host for better luminescent active ions because of its high melting point, and very good solubility of rare earth ions despite of their chemical and physical properties [16]. Pyrochlore oxide structures with general formula $A_2B_2O_7$ have several potential applications such as ionic/electric conductors [17], catalysts [18], radioactive waste management [19], and have proven to be suitable host material for the enhancement of luminescence properties [20] of florescent centers in various domains of scientific technology and instrumentation due to their excellent optical, magnetic, electrical, and dielectric properties [21].

Cubic Pyrochlore oxides $A_2B_2O_7$ (eight molecules per unit cell), are a derivative of fluorite structure having Fd3m space group with two types of cations at A-site and B- site. The cations at A-site are large sized rare earth ions having 3+ covalency and B-site contains smaller sized cations like titanium, zirconium *etc.* with 4+ covalency. A-site cations are 8-fold coordinated to oxygen anion and located within distorted cubic coordination polyhedron (contains 6 equally spaced anion oxygens at a slightly shorter distance from central cation) and the B-site cations

are 6-fold coordinated which is located within distorted octahedron (contains all the six anions at equal distances from the central cations) due to the removal of oxygen atoms. Two types of oxygen sites were illustrated by Wyckoff positions 48f and 8b for pyrochlore structures. In the Wyckoff notation 48f, oxygen has (x, 0.5, 0.5) position coordinated to two rare earth cations A^{3+} and two B^{4+} cations. For the another 8b Wyckoff notation, oxygen is located at (0.375, 0.375, 0.375) in a tetrahedron coordinated to only rare earth A^{3+} cations [22]. Pyrochlore structures has one positional parameter 'x' and its value ranges within 0.309 - 0.355 and lattice parameter of nearly ~10 Å. The phase stability of pyrochlore structures were explained on the behalf of relative cationic radii ratio of cation A and B, *i.e.*, r_A/r_B. For the stable pyrochlore oxide structures the value if cation radii ratio lies within 1.460 to 1.80 [23].

The addition of trivalent rare earth ions to these hosts can improve the chemical durability and luminescence intensity. In various type of host, pyrochlore ternary oxide $Gd_2Ti_2O_7$ (GTO) paved the significant role and improvement over the other oxides, phosphate and fluoride glass systems [24]. GTO have a very intact pyrochlore structure. The larger cationic radii ratio 1.75 will provide a larger flexibility of accommodating of different trivalent lanthanide ions without undergoing into phase transformation or phase segregation [25]**.** Apart from structural and chemical flexibility, GTO have a suitable localized environment for strong luminescent, optical, magnetic, and electrical properties [26,27]**.** By varying the Dy^{3+} rare earth ion concentration in the host a strong variation of the yellow/blue intensity ratio can be attained [28]. The present work focuses on the investigation of structural and luminescent properties of trivalent rare earth ion Dy^{3+} doped GTO pyrochlore host with different concentration of Dy^{3+} substitution. The cationic radii ratio r_{Gd}/r_{Ti} and r_{Dy}/r_{Ti} are obtained as 1.7405 and 1.6975, respectively which means the maximum concentration of Dy^{3+} (for 2.5%) at the place of gadolinium cation (A-site) does not affect the basic pyrochlore structure. Moreover, the luminescent results of this host for various Dy^{3+} concentration were recorded and related to the structural properties. The rare earth ions were used to generate different colours like Eu^{3+} (at 616 nm having $^5D_0$-$^7F_2$ transition producing red colour [29]), Tb^{3+} (high intensity PL emission at 545nm is due to $^5D_4$-$^7F_5$ producing green light [20]), Er^{3+} (the emission peaks are green at 545 nm due to $^4S_{3/2}$-$^4I_{15/2}$ optical transitions and red at 647nm due to $^4F_{9/2}$-$^4I_{15/2}$ transition [30]), Sm^{3+} (the most intense emission peak at 606 nm is due to $^4G_{5/2} \rightarrow {}^6H_{7/2}$ transition of samarium ion producing orange light [31]) *etc.* but Dy^{3+} ion generate two different colours; blue (486 nm, $^4F_{9/2}$-$^6H_{15/2}$) and yellow (576 nm, $^4F_{9/2}$-$^6H_{13/2}$) [32]. The intensity of $^4F_{9/2}$-$^6H_{13/2}$ is hypersensitive to the host matrix while the transition $^4F_{9/2}$-$^6H_{15/2}$ is insensitive to its local environment. Thus, Dy^{3+} ion is the most suitable dopant to establish a relation between local environment and emitted Y/B colours ratio and therefore, the substitution of Dy^{3+} at the place of Gd^{3+} in the host have a lot of importance over other rare earth ions for the generation of white light [33]. Hence, Dy^{3+} is selected as the suitable dopant for the particular host. In the present manuscript, a series of Dy^{3+} substituted GTO single phasic pyrochlore oxide samples have been successfully prepared through solid state reaction method to explore its use as potential phosphor for white LEDs by studying structural and luminescence properties in details using XRD, SEM, Raman spectroscopy, and PL spectroscopy provided CIE chromaticity coordinates. In the present work, the optical studies coupled with structural analysis conclude that the optimum concentration of Dy^{3+} substituted GTO series is found to be 2.0 mol% which is having an intact

pyrochlore structure and CIE coordinate set for this pyrochlore composition lies in the standard cool white light calorimetric point.

2. Experimental Details

2.1. Synthesis

Polycrystalline pyrochlore oxide series of Dy^{3+} substituted GTO with various doping concentration of Dy^{3+} substitution (where x = 0.5, 1.0, 1.5, 2.0, and 2.5 mol%), have been synthesized by using conventional solid state reaction technique. In this method initially AR grade raw precursors namely, Gadolinium (III) oxide (Gd_2O_3) (≥99.99% purity, Sigma Aldrich), Dysprosium (III) oxide (Dy_2O_3) (≥99.99% purity, Sigma Aldrich), and Titanium (IV) oxide (TiO_2) (≥99.99% purity, Sigma Aldrich), were taken into stoichiometric ratios and then thoroughly mixed and grounded multiple times using agate mortar pestle [34], and a systematic step wise flowchart for the preparation of Dy^{3+} substituted GTO series was shown in **Figure 1**.

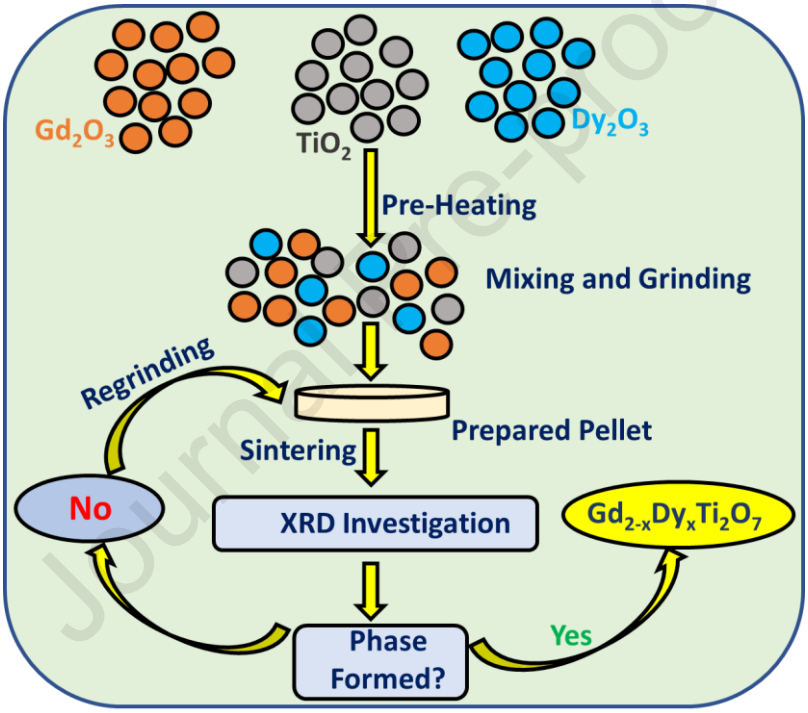


Figure 1. Synthesis process for pure and Dy^{3+} substituted GTO samples.

Rare-earth ion-based oxides get easily oxidized due to their higher reactivity towards CO_2 and water in the exposure of atmospheric environment. Hence, to avoid excess moisture and volatile impurities, all the AR grade powders were placed in alumina crucibles and essentially pre-heated in a programmable muffle furnace at a temperature of 800 ℃ for the time duration of 5 hours. After pre-heating step, stoichiometric ratios of all the AR grade pure reagents Gd_2O_3, Dy_2O_3, and TiO_2 were accurately measured using weighing machine (Shimadzu, Model Number: D307030200) to prepare a batch of 3-gram sample for each composition and was reported in **Table 1**.

Table 1. Raw powders of Gd_2O_3, Dy_2O_3, and TiO_2 weight used for the synthesis of 3-gram batch of $Gd_{2-x}Dy_xTi_2O_7$ series.

Dy^{3+} (%)	**$Gd_{2-x}Dy_xTi_2O_7$**	**Gd_2O_3 (gram)**	**TiO_2 (gram)**	**Dy_2O_3 (gram)**
0	$Gd_2Ti_2O_7$	2.0824	0.9176	0
0.5 %	$Gd_{1.99}Dy_{0.01}Ti_2O_7$	2.0718	0.9175	0.0107
1.0 %	$Gd_{1.98}Dy_{0.02}Ti_2O_7$	2.0612	0.9174	0.0214
1.5 %	$Gd_{1.97}Dy_{0.03}Ti_2O_7$	2.0506	0.9173	0.0321
2.0 %	$Gd_{1.96}Dy_{0.04}Ti_2O_7$	2.0399	0.9172	0.0428
2.5 %	$Gd_{1.95}Dy_{0.05}Ti_2O_7$	2.0293	0.9171	0.0535

After weighing stoichiometric amounts of all the precursors, all were mixed and grounded properly for 9 hours using agate mortar pestle with ethanol as dispersed medium to form uniform and homogeneous final powdered product. The finely grounded products were transformed into the pellets of 10 mm diameter and 2 mm thickness, using hydraulic machine by exerting a pressure of 4 tons using a stainless-steel die-set of 10 mm diameter. Now all these pellets were kept in alumina crucible and then heated in a programmable muffle furnace at three temperatures of 1200 ℃, 1250 ℃ and 1300 ℃ for 24 hours, 36 hours, and 48 hours, respectively with a heating rate of 5 ℃/min and cooling rate of 2 ℃/min, and these steps were represented well by **Figure 1**. Series of final product of $Gd_{2-x}Dy_xTi_2O_7$ (x = 0, 0.01, 0.02, 0.03, 0.04, and 0.05) were formed through multi-step solid state method from the initial raw powders Gd_2O_3, Dy_2O_3, and TiO_2 following the reaction (1);

$$(1 - 0.5 * x).Gd_2O_3 + (0.5 * x).Dy_2O_3 + TiO_2 \xrightarrow{1200\,℃\,(24\,hours),\ 1250\,℃\,(36\,hours),\ 1300\,℃\,(48\,hours)} Gd_{2-x}Dy_xTi_2O_7$$

$$(x = 0, 0.01, 0.02, 0.03, 0.04, and\ 0.05) \ldots\ldots\ldots (1)$$

2.2. Characterization Tools

Structural investigation for the complete series of Dy^{3+} substituted GTO was evaluated *via* X-ray diffraction (XRD). Crystalline phases were investigated by plotting the XRD data within range $10° \leq 2\theta \leq 80°$ (step-size of 0.02°, and scanning rate 0.5°/min) using a CuK_α source of wavelength λ = 1.54056 Å in a D8-Advance Brucker XRD machine at Delhi Technological University (DTU), New Delhi. Raman spectroscopy for all the samples were recorded using a Nd: YAG laser source of wavelength 532 nm using WiTec alpha 300 RA system with a 50× objective lens at Jawaharlal Nehru University (JNU), Delhi. For the complete series, photoluminescence (PL) excitation and emission were examined through a Jasco-8300 fluorescence spectrometer coupled with Xenon lamp, at DTU, New Delhi. The measurements were performed at room temperature.

3. Results and Discussion

3.1. X-ray Diffraction

Room temperature XRD for the as synthesized series of Dy^{3+} substituted GTO with various concentration of Dy^{3+}, has been performed through XRD and were well described by **Figure 2(a)**. All the compositions having the same XRD patterns and were marked by miller indices. Pyrochlores oxides have cubic crystal structure with cell parameter nearly ~10 Å lattice parameter (a = b = c), and belongs to Fd-3m space group. Pyrochlore structure have two types of intensities; The high intensity XRD peaks with even miller indices (222), (400), (440), (622), (444), and (800) at the 2θ value ~ 30.42°, 35.31°, 50.75°, 60.35°, 63.43°, and 74.75 respectively, were represents basic fluorite structure on the other hand low intensity XRD peaks with odd miller indices (111), (311), (331), (511), and (531) at the 2θ value ~ 15.18°, 29.15°, 38.67°, 46.38°, and 53.32° respectively, represents the highly ordered pyrochlore superstructure (marked by a club (♣) in **Figure 2(a)**). The XRD patterns for all the compositions have single phase pyrochlore structure without any impurity peak which confirms the complete dissolution of Dy^{3+} ions into the GTO host and also substitution of Dy^{3+} at the Gd^{3+} doesn't affect the crystal structure and the stability of each composition was confirmed by the following r_A/r_B cationic radii ratio (2);

$$\left[\frac{r_{Gd,Dy}}{r_{Ti}}\right] = \left[\frac{r_{Gd} \times atomic\ \% + r_{Dy} \times atomic\ \%}{r_{Ti}}\right] \ldots\ldots\ldots (2)$$

Here, value of cationic radii for Gd (1.053) is larger as compared to the cationic radii of Dy ion (1.027), therefore the ratio of r_A/r_B goes on decreasing from 1.7405 (0 % Dy^{3+}) to 1.7396 (2.5 % Dy^{3+}) for maximum substitution. As the value of Dy ion is smaller as compared to gadolinium ion therefore there is nearly negligibly small shift of XRD peaks towards the lower angle 2θ side with increasing concentration of dysprosium ion.

Journal Pre-proof

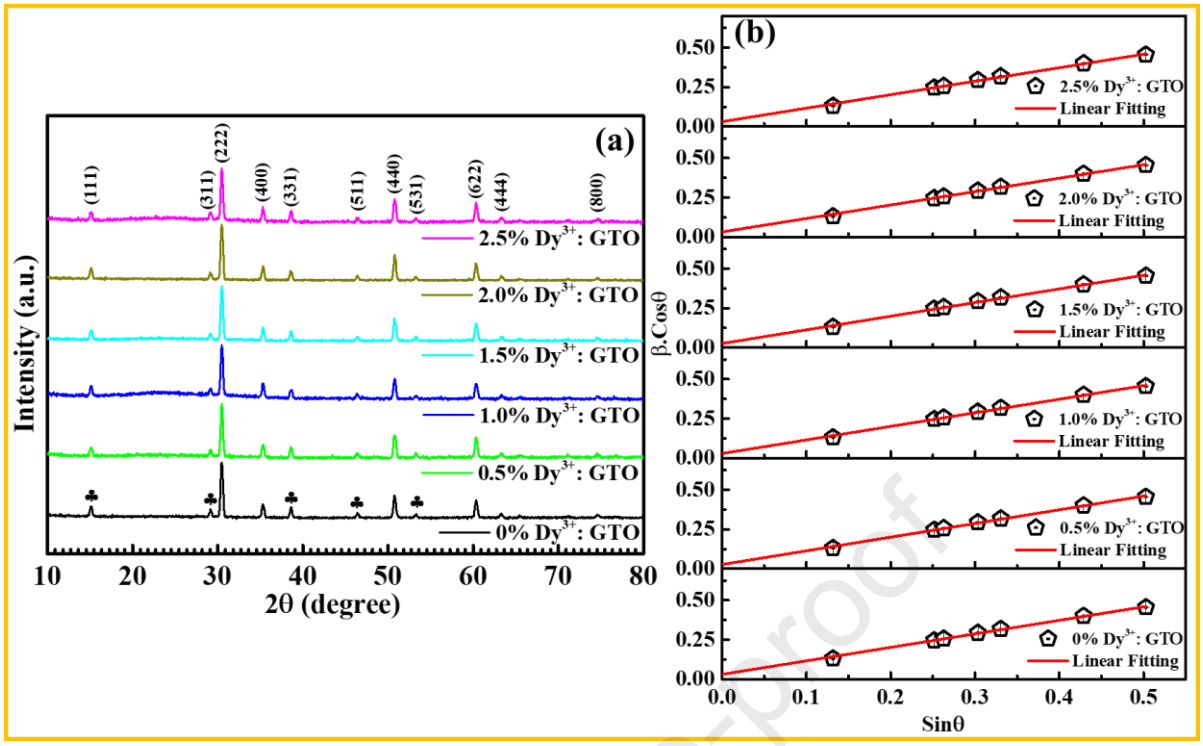


Figure 2. (a) XRD patterns for pure and Dy^{3+} substituted GTO, and **(b)** WH-Plot for all the compositions.

The strain and crystallite for all the compositions of Dy^{3+} substituted GTO were calculated by using Williamson-hall plot (WH-P) formalism [35] as given below *via* equation (3);

$$\beta_{h,k,l} \cos\theta = \left(\frac{K \times \lambda}{D}\right) + 4\varepsilon \sin\theta \ldots\ldots\ldots (3)$$

The value of factor K (structure factor) is 0.94 for the cubic crystal structures and 'D' is the average crystallite size. The parameter '$\beta_{h,k,l}$' represents the full width of diffraction half maxima (FWHM) of first seven dominating high intensity XRD peaks and '2θ' is the diffraction angle corresponding to xrd diffraction peaks. Here 'λ' (1.5406 Å) is the wavelength of CuK_α source and parameter 'ε' is the lattice micro-strain related to the pyrochlore ternary oxide samples. To find the values of parameters 'D' and 'ε' for the complete Dy^{3+} substituted GTO series, a graph (straight-line) is plotted between $\beta_{h,k,l}$ cos θ (along y-axis) and sin θ (along x-axis), as shown by WH-P in **Figure 1(b)**, and results were given in **Table 2**. As on increasing Dy^{3+} concentration, the cationic radii ratio (r_A/r_B) is changed only by less than 0.1 %. It is very clear that this small amount of Dy^{3+} substitution is not going to affect the phase forming sintering temperature and therefore, the structural properties depict insignificant changes as reported in **Table 2**. The calculated crystallite size as well as lattice strain does not significantly vary as a function of Dy^{3+} concentration.

Table 2. Structural parameters measured for undoped and Dy^{3+} substituted GTO for all compositions *via* XRD results.

Dy^{3+} (%)	$Gd_{2-x}Dy_xTi_2O_7$	r_A/r_B	Crystallite Size (D) in Å	Strain
0%	$Gd_2Ti_2O_7$	1.7405	46.744 ± 0.007	0.214±0.004
0.5 %	$Gd_{1.99}Dy_{0.01}Ti_2O_7$	1.7403	53.654 ± 0.008	0.217±0.005
1.0 %	$Gd_{1.98}Dy_{0.02}Ti_2O_7$	1.7401	49.764 ± 0.008	0.215±0.005
1.5 %	$Gd_{1.97}Dy_{0.03}Ti_2O_7$	1.7399	59.590 ± 0.008	0.216±0.005
2.0 %	$Gd_{1.96}Dy_{0.04}Ti_2O_7$	1.7397	44.448 ± 0.007	0.213±0.004
2.5 %	$Gd_{1.95}Dy_{0.05}Ti_2O_7$	1.7395	46.44 ± 0.008	0.214±0.004

For more structural investigations all the compositions were further refined by Rietveld refinement with FullProf Suite software. Linear interpolation between a set background point with refinable heights, was used to refine background with number of ten cycles, and for fitting experimental data pseudo-Voigt function was used for each composition. During refinement instrumental parameter, lattice constant, shape parameter, scale factors, occupancies, and oxygen 48f positions parameters were also fixed for every refinement and the occupancies of Gd and Ti ions didn't change much in the refinement step. The degree of disorder parameter 'x' of 48f oxygen anion and lattice parameter were calculated for complete series of Dy^{3+} substituted GTO. Rietveld refinement plot for all the compositions with increasing content of Dy^{3+} ions at the place of Gd^{3+} were depicted by **Figure 3** and lattice parameter, Braggs factor (R_B), volume (V), weight factors R_{wp} and $R_{exp,}$ and goodness of fitting (χ^2), were tabulated in **Table 3**. In refined plot for each composition, black circle represents experimental intensity data, red line represents the refined intensity/observed intensity, blue line represents the difference experimentally observed and calculated intensities, and the green perpendicular line symbol represents Bragg's peak positions. For all the refined compositions, the value of goodness lying within 1.0 to 2.0 value, that represents best fitting. On increasing the Dy^{3+} cation at A-site in GTO, the value of lattice parameter doesn't show any change up to two decimal places which is the accuracy of XRD detector with 0.02 step size. These results show that the substitution of smaller sized cation at the A-site with very small amount is not going to made any significant change in unit cell of GTO even for 2.5 % doped composition. Since x-parameter of 48f oxygen atoms approaches to 0.3125, the composition will acquire a perfect pyrochlore structure. Therefore, the decreasing value of O48f from 0.3269 to 0.3242 as on increasing Dy^{3+} content shows that the crystalline unit cell is trying to acquire a more ordered pyrochlore structured cell [35].

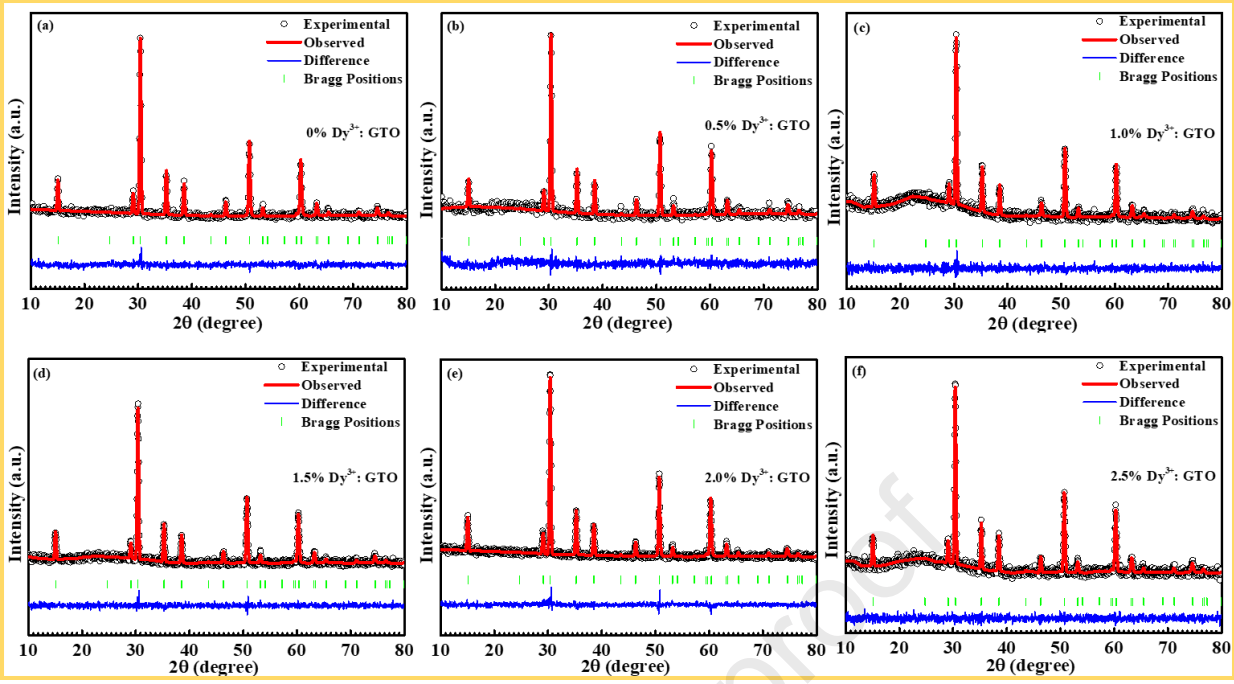


Figure 3. Rietveld refinement plot for pure and Dy^{3+} substituted GTO for all compositions.

Table 3. Rietveld refinement structural parameters for the diffraction patterns of Dy^{3+} substituted GTO for complete series.

Dy^{3+} (%)	Composition ($Gd_{2-x}Dy_xTi_2O_7$)	Lattice Parameter (Å)	48f oxygen 'x' parameter	R_P	R_{wp}	R_{exp}	χ^2	V (Å^3)
0 %	$Gd_2Ti_2O_7$	10.18	0.3269	43.7	26.0	24.0	1.176	1055
0.5 %	$Gd_{0.99}Dy_{0.01}Ti_2O_7$	10.18	0.3268	62.9	33.3	29.7	1.255	1055
1.0 %	$Gd_{0.98}Dy_{0.02}Ti_2O_7$	10.18	0.3240	59.5	31.4	31.2	1.016	1055
1.5 %	$Gd_{0.97}Dy_{0.03}Ti_2O_7$	10.18	0.3259	48.1	26.8	26.0	1.062	1055
2.0 %	$Gd_{0.96}Dy_{0.04}Ti_2O_7$	10.18	0.3311	37.7	22.9	20.7	1.223	1055
2.5 %	$Gd_{0.95}Dy_{0.05}Ti_2O_7$	10.18	0.3242	56.1	30.7	31.0	1.026	1055

3.2. *micro*-Raman Spectroscopic Analysis

The Raman vibrational bands for the Dy^{3+} substituted GTO for all the compositions have been found in the frequency range 150-800 cm^{-1}, and were well represented in **Figure 4**. Detailed

group study of pyrochlore structure $A_2B_2O_7$ suggesting the presence of six vibrational Raman active modes; one A_{1g}, one E_g and four T_{2g} and in all these six vibrational modes only oxygen atom position is displaced [36]. **Figure 4(a-f)** represents the deconvoluted Raman spectra of $Gd_{2-x}Dy_xTi_2O_7$ powders at increasing concentration of Dy^{3+} with x=0.01, 0.02, 0.03, 0.04, and 0.05 containing various vibrational modes.

$$\Gamma = A_{1g} + E_g + 4T_{2g}$$

In the pure GTO sample and Dy^{3+} substituted GTO with various concentrations of Dy^{3+}, all the vibrational bands are in good agreement with the previous results. It was observed that Dy^{3+} ions get completely dissolved in the host GTO without creating structural distortion. Out of all the six active Raman bands, the A_{1g} and E_g vibrational modes are originating due to parallel polarization spectra, on the other hand T_{2g} vibrational bands are formed due to perpendicular polarization [36]. Among these 4 T_{2g} vibrational modes, only one T_{2g} mode is due to O_{8b} oxygen atom while the other three T_{2g} modes are due to O_{48f} oxygen atom vibrations [37]. Raman spectra have five intense Raman bands (~210 cm^{-1}, ~312 cm^{-1}, ~517 cm^{-1}, ~548 cm^{-1}, and ~680 cm^{-1}). Most intense Raman band consists of two vibrational modes E_g and T_{2g} (with very close wave number) and formed due to the O-Gd-O bending mode, centered around ~312 cm^{-1} [37]. Other remaining three T_{2g} vibrational bands were found to be centered around ~210 cm^{-1}, ~548 cm^{-1}, and ~680 cm^{-1} respectively. The peak centered at ~517 cm^{-1} corresponds to the A_{1g} mode, originating due to relative vibrations of internal Gd-O stretching mode. With the introduction of Dy^{3+} ions, it was found that a very less intense band around ~447 cm^{-1} appeared due to the TiO_2 trace amount [20]. The cationic radii ratio r_A/r_B decreases with the Dy^{3+} ion substitution in GTO host from 1.7405 to 1.7396 and the introduction of smaller cation would reduce the size of GTO unit cell or in other words, substitution of Dy ion made an enhancement in force constant within the bands. Therefore, due to increasing force constant or decreasing bond length, all the bands shifted slightly towards higher wavenumber side *i.e.*, blue shifting but the shifting is within error and the Raman results. Although these small substitution concentrations are not induced significant structural changes as they have not been observed through XRD. The observed Raman vibrational bands and their corresponding wave numbers for the Dy^{3+} substituted GTO pyrochlore series were well depicted in **Table 4** with increment of Dy^{3+} ions.

Table 4. Observed Raman Vibrational bands and their respective peak position.

Sample	**Peak Positions**					
	1st Peak	**2nd Peak**	**3rd Peak**	**4th Peak**	**5th Peak**	**6th peak**
0% Dy^{3+}	207.97±0.7	312.19±0.1	448.92±0.5	517.47±0.2	548.47±2.9	681.03±1.8
0.5% Dy^{3+}	208.37±0.8	312.26±0.1	449.32±0.9	517.51±0.2	548.45±3.3	680.87±1.8
1.0% Dy^{3+}	209.62±0.8	312.33±0.1	450.19±4.7	517.38±0.2	550.24±3.5	679.51±1.8
1.5% Dy^{3+}	210.21±0.7	313.51±0.1	448.57±2.3	518.47±0.2	548.79±2.6	682.46±1.7
2.0% Dy^{3+}	210.15±0.6	313.67±0.1	448.32±3.1	518.80±0.2	551.10±3.0	681.78±1.9
2.5% Dy^{3+}	210.51±0.9	313.21±0.1	446.94±4.2	518.17±0.2	549.25±3.0	679.40±2.1

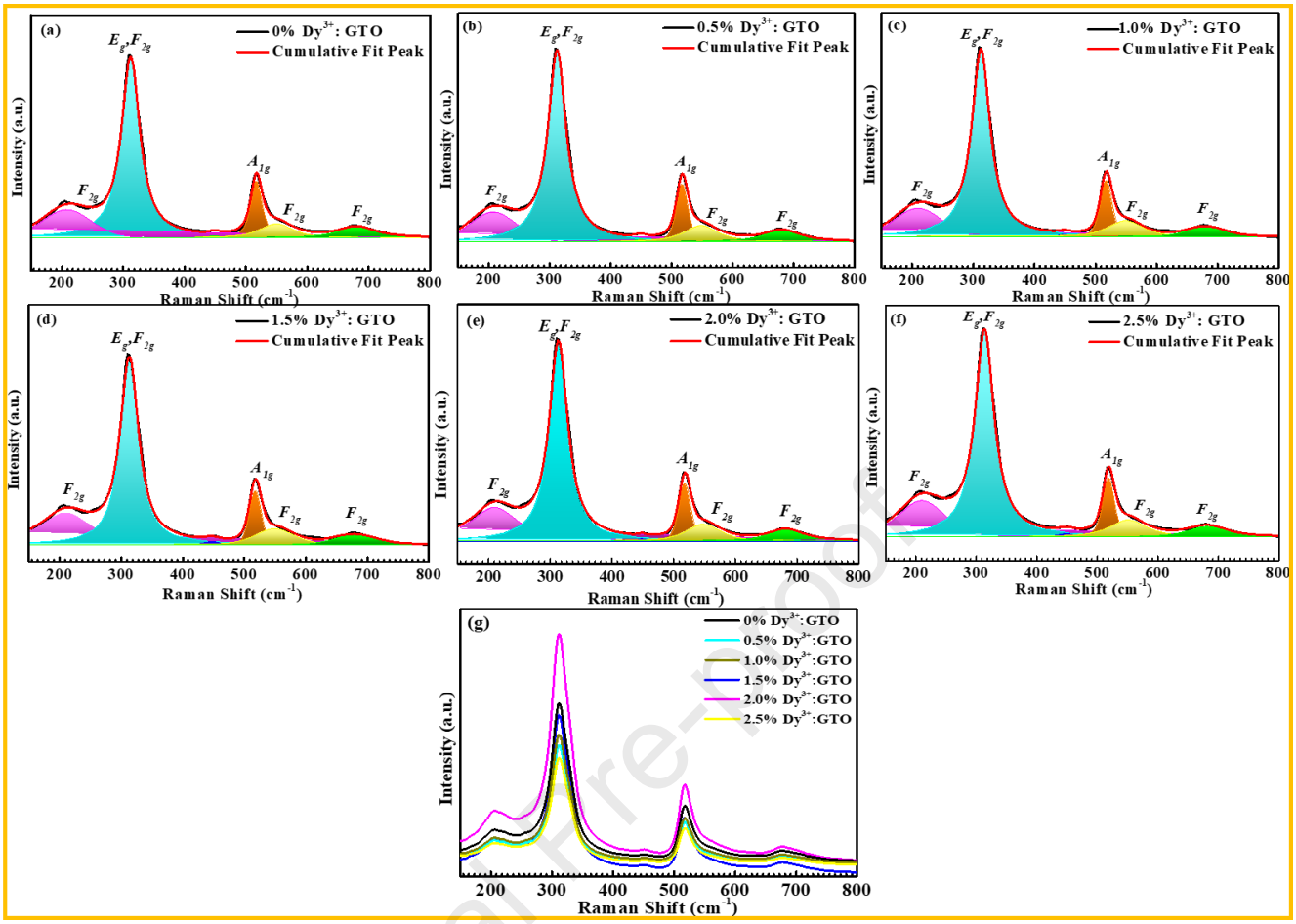


Figure 4(a-f). Deconvoluted Raman spectra for pure GTO and Dy^{3+} substituted GTO series and **(g)** Comparative Raman Spectra for complete series of samples.

3.3. Photoluminescence (PL) Spectroscopy

In consequence of Dy^{3+} substitution through PL analysis, the PL excitation (EX) spectra and the PL emission (EM) spectra of trivalent ion Dy^{3+} substituted GTO with various concentrations (0.5%, 1.0%, 1.5%, 2.0%, and 2.5%) are well depicted by **Figure 5 & 6**. **Figure 5** represents the photoluminescence EX-spectra in the spectral range from 280 to 400 nm for the complete series of Dy^{3+} substituted GTO along with pure GTO pyrochlore oxide by selecting emission wavelength of 484 nm. The EX-spectra contain six sharp bands centered around 322 nm, 347 nm, 352 nm, 361 nm, 380 nm, and 389 nm due to the intra band 4f to 4f transition from ground state $^6H_{15/2}$ to various excited states $^6P_{3/2}$ + $^4M_{17/2}$, $^6P_{7/2}$, $^6P_{5/2}$ + $^4I_{11/2}$, $^4I_{13/2}$ + $^4F_{7/2}$, $^4G_{11/2}$, and $^4I_{15/2}$ respectively [15,29–31].

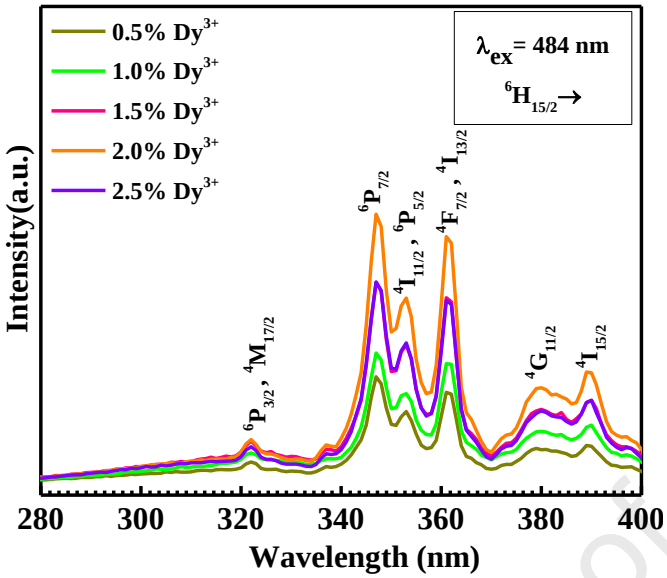


Figure 5. EX-spectra of Dy^{3+} substituted GTO at the place of Gd^{3+}.

The strongest excitation band $^6P_{7/2}$ is centered around ~347 nm and the second strongest excitation band $^4I_{13/2}$ + $^4F_{7/2}$ is appeared at ~360 nm. Hence, to plot PL emission results, 347 nm wavelength has been selected as excitation wavelength among all the excitation peaks.

Figure 6 represents the EM-spectra plotted in the spectral region 370 nm to 650 nm for the complete series of Dy^{3+} substituted GTO pyrochlore oxide with various concentration of Dy^{3+} cations (0%, 0.5%, 1.0%, 1.5%, 2.0%, and 2.5%) by fixing excitation wavelength of 347 nm. The EM-spectra for pure sample (GTO: represented by black line) has emission band around ~ 428 nm due to the Gd^{3+} ions present in the pure host GTO [20]. The complete EM-spectra contains three bands and among these bands two strong intensity bands centered around 484 nm and 581 nm, which can be assigned due to the $^4F_{9/2} \rightarrow {}^6H_{15/2}$ and $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transitions, respectively and the less intense band appeared at the wavelength of 456 nm arising due to the $^4I_{13/2} \rightarrow {}^6H_{13/2}$ transition of Dy^{3+} ions. The emission peaks at 484 nm and 581 nm indicate that the Dy^{3+} cation could behave as luminescent centers and the holes and electrons generated by the host GTO recombined at the localized Dy^{3+} centers which transfers the excitation energy to the Dy^{3+} cations [41]. From many studies it was noticed that the transition peak centered around 581 nm (yellow) wavelength is frequently significant due to localized position of Dy^{3+} ion with low-symmetry without inversion center while the blue emission (484 nm) shows more strength than the yellow one when it occupied the high symmetry local site with an inversion center [16].

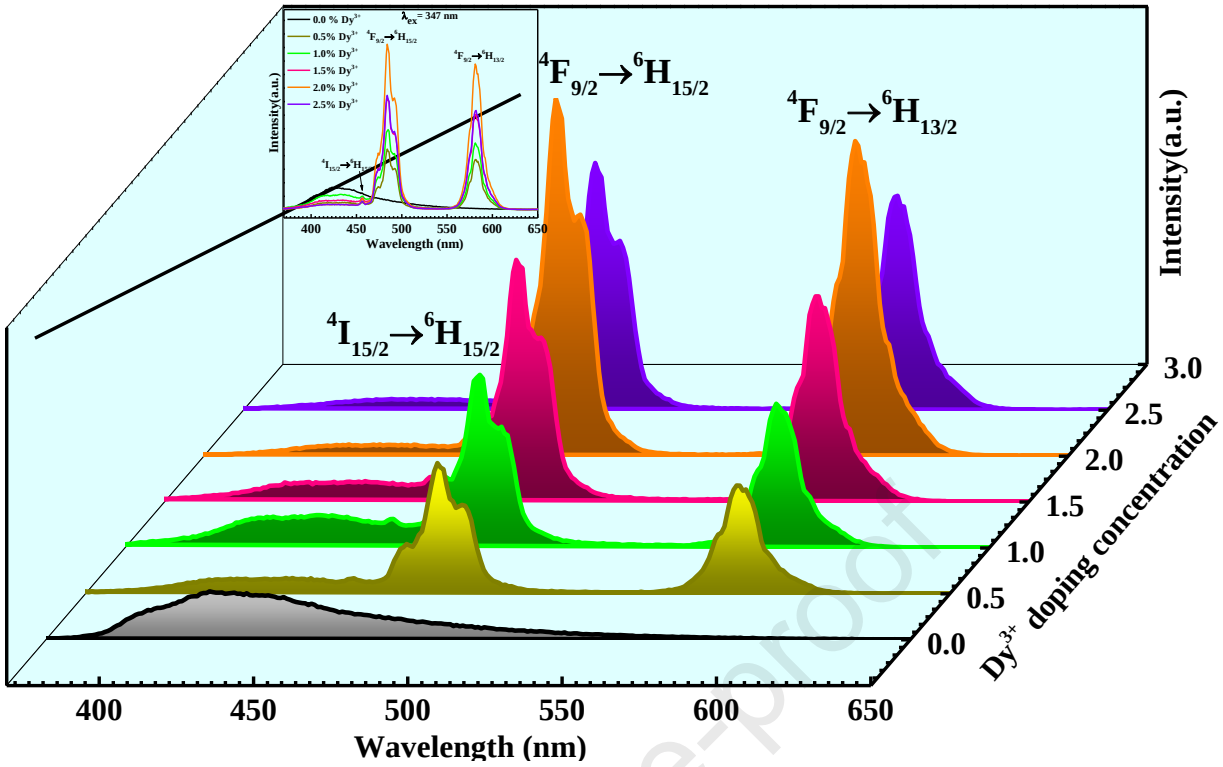


Figure 6. EM-spectra of Dy^{3+} substituted GTO at λ_{ex} = 347 nm.

The peak at 484 nm with $^4F_{9/2} \rightarrow {}^6H_{15/2}$ transition corresponds to magnetic dipole and its emission intensity is independent of crystal field of GTO and on the other side the transition peak $^4F_{9/2} \rightarrow {}^6H_{13/2}$ at 581 nm (yellow) belongs to a forced electric dipole transition [15]. When active Dy^{3+} located at local site with low symmetry without the inversion centre in host GTO, the emission band at 484 nm *i.e.*, blue is more prominent in the complete emission spectra as compared to the other transition peak at 581 nm (yellow) [42,43]. In the host GTO pyrochlore oxide, two types of cations Gd^{3+} and Ti^{4+} are present. The Dy^{3+} ions are supposed to occupy Gd^{3+} sites due to having same valance state. This can be explained as the size of Dy^{3+} (1.027) ions are smaller than the Gd^{3+} (1.053) ions therefore the Dy^{3+} ions can easily occupy the Gd^{3+} sites in GTO with low symmetry and it is in good acceptance with the results found by XRD analysis. From Figure 6 it is clearly noticeable that the emission intensity first increases with an increment of Dy^{3+} cations in GTO at the place of Gd^{3+} site, up to to a maximum content of 2.0% mole of Dy^{3+} ions, but beyond 2.0% of Dy^{3+} in GTO, the emission intensity gradually decreases (for 2.5% of Dy^{3+}) due to the well-known concentration quenching effect (caused by non-radiative energy transfer). This is due to the reason that as the amount of Dy^{3+} ion increases, the distance between luminescent centers decreases which can cause the non-radiative energy transfer. Energy transfer process from the one Dy^{3+} ion to another Dy^{3+} ion, in the present work depends on the critical distance (D_c) between the Dy^{3+}-Dy^{3+} cations. According to Blasse [44,45], the critical distance between adjacent Dy^{3+} ions could be calculated through the following expression (4);

$$D_c = 2 \times \left(\frac{3V}{4\pi x_c n}\right)^{\frac{1}{3}} \ldots\ldots\ldots (4)$$

Here, 'V' represents the unit cell volume calculated through the Rietveld refinement measurements, the term 'x_c' denotes the optimum concentration of the Dy^{3+} ions, and 'n' represents the number of cations per unit cell. From experimental data analysis it was observed that for the optimized concentration x_c = 0.04, V = 1055 $Å^3$, and n = 8. For the current system the value of critical distance is found to be 18.4680 Å. According to Van Uitert the energy transfer is associated with the phenomenon of exchange energy, re-absorption of radiation, and multiple interactions [46]. Out of these all interactions the exchange interaction is responsible for the energy transfer for the forbidden transition and the critical distance for this should be shorter than 3-5 Å [45]. In present study, the value of critical distance between adjacent Dy^{3+} ions are found to be greater than 5 Å, therefore multipolar interactions will play important role in this case rather than exchange interactions.

The luminescence intensity 'I' and the mole fraction of the activator ions 'x' could be corelated to each other through the following expression (5) [47];

$$\log\left(\frac{I}{x}\right) = -\frac{Q}{3}\log(x) + A \ldots\ldots\ldots (5)$$

Here, 'A' is a constant and the quantity 'Q' is the type of interaction between the rare-earth ions. If the value of Q come out to be 6, 8 and 10, then the interactions corresponding to electric dipole-dipole, dipole quadrupole, and quadrupole-quadrupole respectively [48]. The relationship between log(I/x) and log(x) plotted along y-axis and x-axis respectively for the Dy^{3+} substituted GTO with various concentrations under the excitation wavelength of 347 nm wavelength were well represented by **Figure 7**. The value of 'Q' parameter is come out to be 4.05 which is close to 6 indicates that the effect of concentration quenching in the Dy^{3+} substituted GTO is due to the electric dipole-dipole interaction [48].

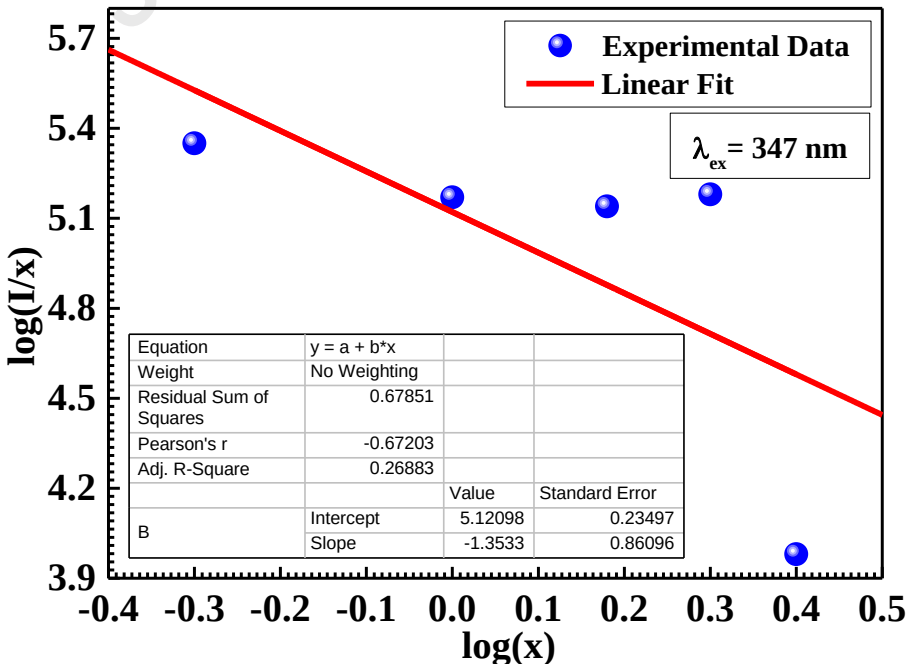


Figure 7. Relationship of log(I/x) with log(x) for complete series of Dy^{3+} substituted GTO under 347 nm excitation.

The Yellow/blue (Y/B) ratio for the Dy^{3+} substituted GTO pyrochlore oxide has been studied with respect to function of Dy_2O_3 concentration and it is important for white light emission. The degree of covalence between the Dy^{3+}-O^{2-} ions can be identified by the Y/B ratio. Greater value of Y/B ratio indicating the good degree of covalency between dysprosium and oxygen ions. For the excitation wavelength 347 nm, the value of Y/B ratio for the complete series of Dy^{3+} substituted GTO for concentrations 0.5 %, 1.0 %, 1.5 %, 2.0 %, and 2.5 %, were calculated and tabulated in **Table 5**, were found to be nearly close to unity, indicates the stability of the colour coordinates for all concentrations. The value of Y/B ratio increases with increasing Dy^{3+} up to 2.0 mol % and then decreases for further increment of Dy^{3+}. The results are in accordance to the known Dy^{3+} ions doped single-phase $BaNb_2O_6$ phosphors [15], Dy^{3+} doped ZnO [32], and Dy^{3+} doped $Mg_2Al_4Si_5O_{18}$ phosphor [43]. We conclude that the Y/B ratio and CCT provide a basis for the design of efficient white phosphor for displays and WLED's.

Table 5. Y/B ratios, x and y CIE chromaticity coordinates, and CCT for Dy^{3+} substituted GTO with various doing concentration of Dy^{3+} ions.

λ_{ex} = 347 nm			
X (Dy^{3+} concentration in mol%)	**Yellow/Blue (Y/B) Ratio**	**(x, y) coordinates**	**CCT (K)**
0.5 %	0.8346	(0.31730, 0.31024)	6385.51
1.0 %	0.8383	(0.32988, 0.31403)	5638.77
1.5 %	0.8507	(0.32780, 0.32651)	5732.52
2.0 %	0.8753	(0.34959, 0.35381)	4844.95
2.5 %	0.8671	(0.34217, 0.34587)	5102.82

Figure 8, depicts the systematic energy level diagram consisting excitation, emission, resonance energy transfer, and cross-relaxation processes involved in Dy^{3+} substituted GTO. The radiation excites the Dy^{3+} ions by using energy of 347 nm (upward black arrow represents excitation line) and then relaxes to the energy state $^4F_{9/2}$ via non-radiative transfer (denoted by downward green arrow) and then emitting three visible light emission lines 456nm ($^4I_{13/2}\rightarrow{}^6H_{13/2}$), 484 nm ($^4F_{9/2}\rightarrow{}^6H_{15/2}$), and 581 nm ($^4F_{9/2}\rightarrow{}^6H_{13/2}$) radiatively from $^4F_{9/2}$ energy levels. The possible non-radiative transitions could be due to resonance energy transfer corresponds to transition from $^4F_{9/2}+{}^6H_{15/2} \rightarrow {}^6H_{15/2}+{}^4F_{9/2}$ and cross relaxation channels such as ($^4F_{9/2}+{}^6H_{15/2}\rightarrow{}^6F_{11/2}$, $^6H_{9/2}+{}^6F_{5/2}$), ($^4F_{9/2}+{}^6H_{15/2}\rightarrow{}^6F_{9/2}$, $^6H_{7/2}+{}^6F_{5/2}$) and ($^4F_{9/2}+{}^6H_{15/2}\rightarrow{}^6F_{1/2}+{}^6F_{11/2}$,$^6H_{9/2}$) are well represented by CRC1, CRC2, and CRC3 respectively.

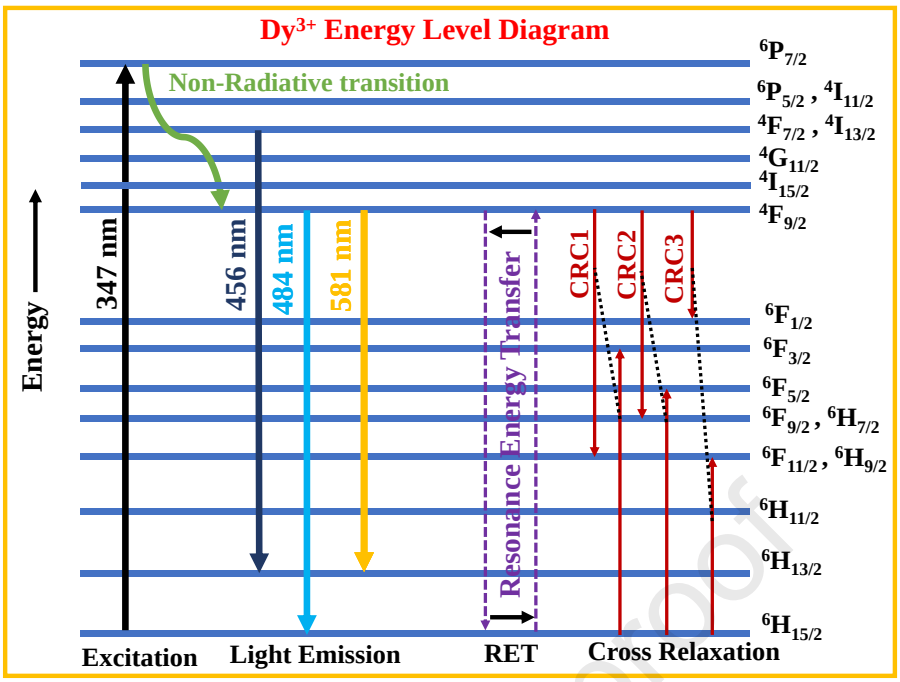


Figure 8. Schematic representation of excitation, emission and energy transfer processes involving in Dy^{3+} substituted GTO.

3.4. CIE Calorimetric Postulate

The Commission International de L' Eclairage (CIE) chromaticity coordinates are helpful to identify the colour of any luminous material [31], specified on a chromaticity diagram as shown in **Figure 9**, for Dy^{3+} substituted GTO under the excitation wavelength of 347 nm for the current series. The PL results were used to calculate the colour coordinates by calculating α(λ), β(λ) and γ(λ) three essential colour matching functions described in CIE 1931 diagram [33]. For a particular spectral density D(λ), amount of inducement necessary to exactly match the colour of power density D(λ) is explained by the following expressions:

$$X = \int_{\lambda} \alpha(\lambda)\mathrm{D}(\lambda)d\lambda \ldots\ldots\ldots (6)$$

$$Y = \int_{\lambda} \beta(\lambda)D(\lambda)d\lambda \ldots\ldots\ldots (7)$$

$$Z = \int_{\lambda} \gamma(\lambda)D(\lambda)d\lambda \ldots\ldots\ldots (8)$$

Here X, Y, and Z are representing the tristimulus values providing power for red, green and blue colours to match the colour of D(λ). Now to identify the actual colour these dimension X, Y, and Z were used to match the colour of D(λ), done by CIE 1931 diagram. From X, Y and Z values x and y can be measured through the following expressions (9 & 10) [33];

$$x = \frac{X}{X+Y+Z} \ldots\ldots\ldots (9)$$

$$y = \frac{Y}{X+Y+Z} \ldots\ldots\ldots (10)$$

The CIE coordinates for the Dy^{3+} substituted GTO for various concentration were calculated from the PL EM-spectra measured under 347 nm excitation wavelength. The value of CIE coordinates is come out to be (0.31730, 0.31024), (0.32988, 0.31403), (0.32780, 0.32651), (0.34959, 0.35381), and (0.34217, 0.34587) for the various doing of Dy^{3+}, 0.5 mol%, 1.0 mol%, 1.5 mol %, 2.0 mol %, and 2.5 mol % respectively, and well depicted in **Table 5**. It was noticed that for all the various concentration the value of CIE coordinates is close to standard equal energy white light coordinates (0.333, 0.333) [15].

The quality of white light can be estimated through chromaticity coordinates using McCamy's approximate expression (11) [39];

$$CCT = -449s^3 + 3525s^2 - 6823.3s + 5520.33 \ldots\ldots\ldots (11)$$

Here, then term 's' represents the inverse slope line where $s = \frac{x-x'}{y-y'}$. And the term '$x' = 0.332$' and '$y' = 0.186$' are the epicenter coordinates. The values of colour correlated temperature (CCT) calculated through equation (11) for Dy^{3+} substituted GTO series with various concentration for excitation at 347 nm were tabulated in Table 4, and were lying in the range 4845-6385 and falls in cool-white light region. From many studies it was found that the lamps with CCT < 3200 K emit warm white light whereas the lamp with CCT > 4000 K are tend to give cool white light [33]. For the optimized concentration (2 mol%) of Dy^{3+} substituted GTO value of CCT is 4845, indicating the cool white light emission and it is close to the ideal white colour region in CIE diagram. Higher value of CCT indicating the large brightness and better visual acuity. The value of CCT lying in cool white light region signifies the use of Dy^{3+} substituted GTO pyrochlore oxide for better w-LED applications.

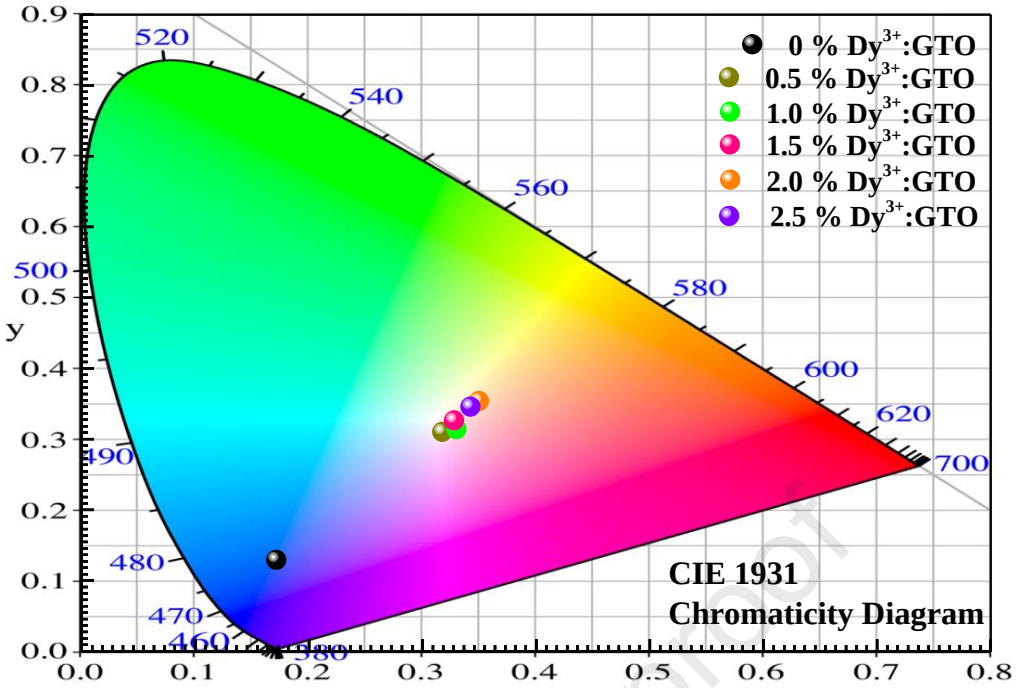


Figure 9. CIE 1931 chromaticity diagram for Dy^{3+} substituted GTO for various concentration.

4. Conclusion

The GTO and series of Dy^{3+} doped GTO were successfully prepared through standard solid state reaction technique. For the complete series, the structural analysis was done through XRD and Raman analysis. XRD results accompanied with Rietveld refinement analysis suggested that all the samples are crystalline in nature and freeze with single pyrochlore phase. This small substitution is not too significant to produce any structural distortion as observed through XRD. which causes lattice contraction of host GTO with Dy^{3+} substitution. Raman profile indicating the presence of functional group corresponding to active vibrational modes which are in good agreement with XRD results. Unlike XRD, Raman spectroscopy exhibits a very small blue shift as the effect of smaller cationic substitution. The EX-spectra for every sample show the appearance of excitation peak at 347 nm wavelength and the EM-spectra have two sharp intensity bands centered around 484 nm and 581 nm, originating due to the $^4F_{9/2} \rightarrow {}^6H_{15/2}$ and $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transition respectively and the less intense band appeared at wavelength of 456 nm formed due to the $^4I_{13/2} \rightarrow {}^6H_{13/2}$ transition of Dy^{3+} ions. The optimum concentration of Dy^{3+} substituted GTO series is found to be 2.0 mol% after this concentration quenching effect occurs. The Y/B ratio for the complete series of samples were found to close to unity and the combination of both yellow and blue emission bands emit white light with CIE coordinates values for the 2.0 mol % of Dy^{3+} ion in GTO host is (0.34959, 0.35381) with CCT value 4844.95 lying in the standard cool white light calorimetric point. Overall, the Dy^{3+} substituted GTO pyrochlore samples could be used as luminescent material for WLEDs application.

Journal Pre-proof

5. Acknowledgement

One of the authors (Umang Berwal) is thankful to the Delhi Technological University (DTU), New Delhi for providing fellowship grant to carry out Ph.D. work. Authors also acknowledge Dr. Rajender Singh Dhaka, Department of Physics Indian Institute of Technology New Delhi, for providing high temperature furnace during sample preparation. AIRF, JNU is acknowledged for Raman measurements and DTU is acknowledged for XRD and PL studies, respectively.

6. Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Journal Pre-proof

References

[1] J. Ballato, E. Snitzer, concentrations for Faraday isolator applications, Appl. Opt. (1995) 1–7.

[2] B. Jacquier, Laser spectroscopy of rare earth-doped glass waveguides, J. Alloys Compd. 225 (1995) 15–19. https://doi.org/10.1016/0925-8388(94)07111-X.

[3] S. Kuck, Laser-related spectroscopy of ion-doped crystals for tunable solid-state lasers, Appl. Phys. B-Lasers Opt. 72 (2001) 515–562. https://doi.org/10.1103/PhysRevB.24.5.

[4] G. Boulon, Fifty years of advances in solid-state laser materials, Opt. Mater. (Amst). 34 (2012) 499–512. https://doi.org/10.1016/j.optmat.2011.04.018.

[5] T. Welker, Recent developments on phosphors for fluorescent lamps and cathode-ray tubes, J. Lumin. 48–49 (1991) 49–56. https://doi.org/10.1016/0022-2313(91)90075-7.

[6] Z. Chen, X. Yang, L. Song, X. Wang, Q. Xiao, H. Xu, Q. Feng, S. Ding, Extraction and complexation of trivalent rare earth elements with tetraalkyl diglycolamides, Inorganica Chim. Acta. 513 (2020) 119928. https://doi.org/10.1016/j.ica.2020.119928.

[7] H. Jiao, Y. Wang, $Ca_2Al_2SiO_7:Ce^{3+}$, Tb^{3+}: A White-Light Phosphor Suitable for White-Light-Emitting Diodes, J. Electrochem. Soc. 156 (2009) J117. https://doi.org/10.1149/1.3097191.

[8] M. Luo, B. Chen, X. Li, J. Zhang, S. Xu, X. Zhang, Y. Cao, J. Sun, Y. Zhang, X. Wang, Y. Zhang, D. Gao, L. Wang, Fluorescence decay route of optical transition calculation for trivalent rare earth ions and its application for Er^{3+}-doped $NaYF_4$phosphor, Phys. Chem. Chem. Phys. 22 (2020) 25177–25183. https://doi.org/10.1039/d0cp04379f.

[9] S. H. Park, K.H. Lee, S. Unithrattil, H.S. Yoon, H.G. Jang, W. Bin Im, Melilite-structure $CaYAl_3O_7:Eu^{3+}$ phosphor: Structural and optical characteristics for near-UV LED-based white light, J. Phys. Chem. C. 116 (2012) 26850–26856. https://doi.org/10.1021/jp307192y.

[10] H.S. Jang, W. Bin Im, D.C. Lee, D.Y. Jeon, S.S. Kim, Enhancement of red spectral emission intensity of $Y_3Al_5O_{12}:Ce^{3+}$ phosphor via Pr co-doping and Tb substitution for the application to white LEDs, J. Lumin. 126 (2007) 371–377. https://doi.org/10.1016/j.jlumin.2006.08.093.

[11] L. Ighting, The promise and challenge of solid-state lighting THE PROMISE AND CHALLENGE OF SOLID -STATE, 54 (2001). https://doi.org/10.1063/1.1445547.

[12] M.H. Crawford, LEDs for Solid-State Lighting : Performance Challenges and Recent Advances, 15 (2009) 1028–1040.

[13] X. Wang, Y. Wang, Synthesis , Structure , and Photoluminescence Properties of Ce^{3+}-Doped $Ca_2YZr_2Al_3O_{12}$:A Novel Garnet Phosphor for White LEDs, (2015). https://doi.org/10.1021/acs.jpcc.5b01552.

[14] H. Kaur, M. Jayasimhadri, M.K. Sahu, P.K. Rao, N.S. Reddy, Synthesis of orange emitting Sm^{3+} doped sodium calcium silicate phosphor by sol-gel method for photonic device applications, Ceram. Int. (2020). https://doi.org/10.1016/j.ceramint.2020.04.224.

[15] A. Manuscript, adjust margins, (2015). https://doi.org/10.1039/C5DT02436F.

[16] N. Vijaya, K. Upendra Kumar, C.K. Jayasankar, Dy^{3+}-doped zinc fluorophosphate

glasses for white luminescence applications, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 113 (2013) 145–153. https://doi.org/10.1016/j.saa.2013.04.036.

[17] J.A. Díaz-Guillén, A.F. Fuentes, M.R. Díaz-Guillén, J.M. Almanza, J. Santamaría, C. León, The effect of homovalent A-site substitutions on the ionic conductivity of pyrochlore-type $Gd_2Zr_2O_7$, J. Power Sources. 186 (2009) 349–352. https://doi.org/10.1016/j.jpowsour.2008.09.106.

[18] S.F. Weng, H.C. Hsieh, C.S. Lee, Hydrogen production from oxidative steam reforming of ethanol on nickel-substituted pyrochlore phase catalysts, Int. J. Hydrogen Energy. 42 (2017) 2849–2860. https://doi.org/10.1016/j.ijhydene.2016.11.181.

[19] C. Meng, X. Ding, W. Li, J. Zhao, H. Yang, Phase structure evolution and chemical durability studies of Ce-doped zirconolite–pyrochlore synroc for radioactive waste storage, J. Mater. Sci. 51 (2016) 5207–5215. https://doi.org/10.1007/s10853-016-9822-x.

[20] U. Berwal, V. Singh, R. Sharma, Key role of Tb^{3+} doping on structural and photoluminescence properties of $Gd_2Ti_2O_7$ pyrochlore oxide, Ceram. Int. (2022). https://doi.org/10.1016/j.ceramint.2022.04.231.

[21] P.H. Crystal, S. Fujihara, K. Tokumo, Cebc3Bdc56E9C3Cade247932C565B7Cc.Pdf, (2005) 5587–5593.

[22] S. K. Sharma, V. Grover, A. K. Tyagi, D. K. Avasthi, U. B. Singh, and P. K. Kulriya, Probing the Temperature Effects in the Radiation Stability of $Nd_2Zr_2O_7$ Pyrochlore under Swift Ion Irradiation, Materialia **6**, 100317 (2019).

[23] M.A. Subramanian, G. Aravamudan, G.V.S. Rao, OXIDE REVIEW, 15 (1983).

[24] J. Nair, P. Nair, G.B.M. Doesburg, J.G. Van Ommen, J.R.H. Ross, A.J. Burggraaf, F. Mizukami, Sintering of lanthanum zirconate, J. Am. Ceram. Soc. 82 (1999) 2066–2072. https://doi.org/10.1111/j.1151-2916.1999.tb02042.x.

[25] A. Kumar, P.K. Kulriya, S.K. Sharma, V. Grover, A.K. Tyagi, V.K. Shukla, Structural and compositional effects on the electronic excitation induced phase transformations in $Gd_2Ti_{2-y}Zr_yO_7$ pyrochlore, J. Nucl. Mater. 539 (2020) 152278. https://doi.org/10.1016/j.jnucmat.2020.152278.

[26] N. Cioatera, E.A. Voinea, E. Panaintescu, A. Rolle, S. Somacescu, C.I. Spinu, R.N. Vannier, Changes in structure and electrical conductivity of rare-earth titanate pyrochlores under highly reducing atmosphere, Ceram. Int. 42 (2016) 1492–1500. https://doi.org/10.1016/j.ceramint.2015.09.095.

[27] J.E. Greedan, Frustrated rare earth magnetism: Spin glasses, spin liquids and spin ices in pyrochlore oxides, J. Alloys Compd. 408–412 (2006) 444–455. https://doi.org/10.1016/j.jallcom.2004.12.084.

[28] M. Zekri, A. Herrmann, R. Turki, C. Rüssel, R. Maâlej, K. Damak, Experimental and theoretical studies of Dy^{3+} doped alkaline earth aluminosilicate glasses, J. Lumin. 212 (2019) 354–360. https://doi.org/10.1016/j.jlumin.2019.04.041.

[29] C. Bouzidi, M. Ferhi, H. Elhouichet, M. Ferid, Structural and luminescence properties of $(Ba_{1-x}Eu_x)MoO_4$ powders, J. Lumin. 179 (2016) 230–235. https://doi.org/10.1016/j.jlumin.2016.07.014.

[30] A. Moadhen, C. Bouzidi, H. Elhouichet, R. Chtourou, M. Oueslati, Concentration and temperature dependence of visible up-conversion luminescence in sol-gel SnO_2 doped with erbium, Opt. Mater. (Amst). 31 (2009) 1224–1227. https://doi.org/10.1016/j.optmat.2009.01.002.

[31] S. K. Gupta, C. Reghukumar, K. Sudarshan, P.S. Ghosh, N. Pathak, R.M. Kadam, Journal of Physics and Chemistry of Solids Orange-red emitting $Gd_2Zr_2O_7$: Sm^{3+}: Structure-property correlation , optical properties and defect spectroscopy, J. Phys. Chem. Solids. 116 (2018) 360–366. https://doi.org/10.1016/j.jpcs.2018.01.031.

[32] G. Amira, B. Chaker, E. Habib, Spectroscopic properties of Dy^{3+} doped ZnO for white luminescence applications, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 177 (2017) 164–169. https://doi.org/10.1016/j.saa.2017.01.039.

[33] P. Suthanthirakumar, K. Marimuthu, Investigations on spectroscopic properties of Dy^{3+} doped zinc telluro-fluoroborate glasses for laser and white LED applications, J. Mol. Struct. 1125 (2016) 443–452. https://doi.org/10.1016/j.molstruc.2016.06.080.

[34] S. K. Sharma, H. S. Mohanty, D. K. Pradhan, A. Kumar, V. K. Shukla, F. Singh, and P. K. Kulriya, Structural, Dielectric and Electrical Properties of Pyrochlore-Type $Gd_2Zr_2O_7$ Ceramic, J. Mater. Sci. Mater. Electron. (2020).

[35] S.K. Sharma, V. Grover, R. Shukla, A. Hussain, A. Mishra, R.C. Meena, P.K. Kulriya, Evidence of improved tolerance to electronic excitation in nanostructured $Nd_2Zr_2O_7$, J. Appl. Phys. 129 (2021). https://doi.org/10.1063/5.0039390.

[36] T.T.A. Lummen, I.P. Handayani, M.C. Donker, D. Fausti, G. Dhalenne, P. Berthet, A. Revcolevschi, P.H.M. Van Loosdrecht, Phonon and crystal field excitations in geometrically frustrated rare earth titanates, Phys. Rev. B - Condens. Matter Mater. Phys. 77 (2008) 1–11. https://doi.org/10.1103/PhysRevB.77.214310.

[37] A. Kumar, S. K. Sharma, V. Grover, Y. Singh, V. Kumar, V. K. Shukla, and P. K. Kulriya, Probing the Short-Range Ordering of Ion Irradiated $Gd_2Ti_{2-y}Zr_yO_7$ ($0.0 \leq y \leq 2.0$) Pyrochlore under Electronic Stopping Regime. Journal of Nuclear Materials, 564 (2022) 153682). https://doi.org/10.1016/j.jnucmat.2022.153682

[38] M. Gökçe, D. Koçyiğit, Spectroscopic investigations of Dy^{3+} doped borogermanate glasses for laser and wLED applications, Opt. Mater. (Amst). 89 (2019) 568–575. https://doi.org/10.1016/j.optmat.2019.02.004.

[39] R. Liu, M. Chen, X. Zhu, Y. Zhou, F. Zeng, Z. Su, Luminescent properties and structure of Dy^{3+} doped germanosilicate glass, J. Lumin. 226 (2020) 117378. https://doi.org/10.1016/j.jlumin.2020.117378.

[40] K. Linganna, C.S. Rao, C.K. Jayasankar, Optical properties and generation of white light in Dy^{3+}-doped lead phosphate glasses, J. Quant. Spectrosc. Radiat. Transf. 118 (2013) 40–48. https://doi.org/10.1016/j.jqsrt.2012.12.002.

[41] S.M. Wang, Z.L. Xiu, M.K. Lü, A.Y. Zhang, Y.Y. Zhou, Z. Sen Yang, Combustion synthesis and luminescent properties of Dy^{3+}-doped $La_2Sn_2O_7$ nanocrystals, Mater. Sci. Eng. B Solid-State Mater. Adv. Technol. 143 (2007) 90–93. https://doi.org/10.1016/j.mseb.2007.07.070.

[42] Q. Su, Z. Pei, L. Chi, H. Zhang, Z. Zhang, F. Zou, The yellow-to-blue intensity ratio (Y/B) of Dy^{3+} emission, J. Alloys Compd. 192 (1993) 25–27.

https://doi.org/10.1016/0925-8388(93)90174-L.

[43] Z. Ci, Q. Sun, S. Qin, M. Sun, X. Jiang, X. Zhang, Y. Wang, Warm white light generation from a single phase Dy^{3+} doped $Mg_2Al_4Si_5O_{18}$ phosphor for white UV-LEDs, Phys. Chem. Chem. Phys. 16 (2014) 11597–11602. https://doi.org/10.1039/c4cp00357h.

[44] G. Blasse, Energy transfer in oxidic phosphors, Philips Res. Reports. 24 (1969) 131.

[45] Q. Xu, J. Sun, D. Cui, Q. Di, J. Zeng, Synthesis and luminescence properties of novel $Sr_3Gd(PO_4)_3:Dy^{3+}$ phosphor, J. Lumin. 158 (2015) 301–305. https://doi.org/10.1016/j.jlumin.2014.10.034.

[46] L.G. Van Uitert, Characterization of Energy Transfer Interactions between Rare Earth Ions, J. Electrochem. Soc. 114 (1967) 1048. https://doi.org/10.1149/1.2424184.

[47] D.L. Dexter, A theory of sensitized luminescence in solids, J. Chem. Phys. 21 (1953) 836–850. https://doi.org/10.1063/1.1699044.

[48] H. Li, R. Zhao, Y. Jia, W. Sun, J. Fu, L. Jiang, S. Zhang, R. Pang, C. Li, Photoluminescence Properties, (2014) 1–7.

Journal Pre-proof

Journal Pre-proof

Declaration of 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA) | 978-1-6654-5834-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCSEA54677.2022.9936576

Supervision Software using Artificial Intelligence

Shantanu Shishodia
Information Technology
Delhi Technological University
shantanushishodia10@gmail.com

Tanmay Garg
Information Technology
Delhi Technological University
tanmaygarg43@gmail.com

Shubhanshu Chaudhary
Information Technology
Delhi Technological University
shubhanshuchaudhary1504@gmail.com

Dr. Kapil Sharma
Information Technology
Delhi Technological University
kapil@ieee.org

***Abstract*—With the dependency on the internet growing immensely everyday, the newer generation is exposed to threats and risks. The aim of this paper is to supervise the activities of a supervisee using Keylogger and Screenshot analyser. Different machine learning algorithms are applied on the logged data and the screenshots which helps in classifying the data in different categories. The supervisor is alerted each time the model raises a red flag. It shows promising results in keeping the track of the user's activities. This helps in evaluating the user's web practices, from an innovative, lawful and moral perspective.**

***Index Terms*—Natural Language Processing, Keyloggers, Spyware, Malware, Deep Learning, ImageNet, MobileNetV2**

1. INTRODUCTION

The development in spyware techniques in this fast-paced world is going on at a rapid rate. One of the most common spyware tools is keylogger because they are quite difficult to detect in a system and provide the information of the user in a hassle-free manner. Violence is growing day by day and the youth is getting affected from it at an alarming rate [17]. As a result, we can observe how 35% of all the internet downloads are related to pornography. Also, according to the world happiness report 2021, the happiness index of almost all the countries has declined compared to 2020 and previous years. Youth is getting exposed to gore materials on the internet at a very young age which is resulting in more irritation and aggression in teenagers, increasing suicide rate and deterioration of mental health. The new generation is heavily dependent on software that uses the internet, like social media apps, so they play an essential role in the daily life of the users [15]. There has been advances in the field of IoT, but there still lies the problem of data security which our proposed software tries to overcome to a certain extent [14].

In this paper, we are proposing how keyloggers and screenshots when combined with some Natural Language Processing techniques and Deep Learning layers can result in an overall surveillance system for a person that can raise an alert whenever they are exposed to some malicious content.

1.1 What is a Keylogger?

A hardware device or program which is used to capture the activities of a person without their consent. It can be utilized by an attacker with ill intentions to get the private data of the victim to the attacker. Keyloggers can be used to get data from any keyboard, laptop or any device that has an input method. Some spyware can be utilized genuinely, if not in every case morally. Guardians and the higher authorities can utilize keystroke data to supervise the Internet conduct of those they are liable for [6], and organizations can utilize it to better focus on the Internet perusing and shopping experience for clients (e.g., Reference [8]). The technique of keylogging can be used in both ethical and unethical methods. This is a choice that depends on the path that the implementer takes while defining the keylogger's application. This project can be considered to be a positive outcome of implementing the keylogger.

1.2 Types of Keyloggers

1.2.1 Hardware Keylogger

These are physical devices that can be plugged into devices like desktops, laptops, smart devices, etc. The shapes and sizes of these are different for each along with different kinds of port that they go into. These are generally small in size. They can be attached to ports behind the computer, the WiFi router or under the keyboard. The latest developments in the field of keylogger include optical keylogger which can be used for capturing the keylogging data using cameras.

1.2.2 Software Keylogger

These are the kind of keyloggers which generally are pieces of code written in different programming languages as per the skills of the user. It can also be defined as a nonphysical method employed for capturing the keystrokes is called a software keylogger. Generally, software keyloggers are more suitable when compared to hardware keyloggers. This is because they don't require physical presence to become operational, they can be shared online through different means. Software Keylogger is a program that can be introduced onto a computer, which monitors all the user activities on the computer [16].

There are multiple subsections in this paper. Similar research conducted in this field is described in the second section of the report. It also contains some knowledge on the basics of keyloggers. As far as the tools used and technologies implemented are considered, they are described in the third section of this report. Section four consists of the proposed methodology of the paper which is the flow of the software describing the acquiring and processing of the data. Fifth section consists of all the experimental analysis which includes the training data, text pre-processing and classifiers for the keylogged data, image analysis and the model for screenshots.

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:34:13 UTC from IEEE Xplore. Restrictions apply.

Fifth section also explains the results obtained through machine learning algorithms and Deep Learning models. Sixth section consists of the conclusion that we derived from proposing this means to use keylogger and screenshots combined with these technologies for a novel purpose.

The objective of this paper is:

- Create a supervision software
- Implement different ML techniques to make the software effective and efficient
- Make it easier for the concerned authority to keep an eye on the user and get notified in a convenient way if an alert is raised

2. Related Work

Numerous papers have been published on spywares, especially highlighting the detection and prevention of keyloggers.

(Reference [1], 2003)	The paper has proposed a strategy for shielding devices from Keylogger assaults. They utilize an identification and prevention server for taking a look at the systems/devices for malevolent assaults. However, this strategy can be utilized simply restricted to the systems/devices in the customer's organization.
(Reference [2], 2020)	This paper talks about the positive as well as the negative aspects of the keyloggers, the positives aspects include parental monitoring, improve employee productivity, Investigate writing, Ethical hacking, Forensic investigations.
(Reference [7], 2015)	This paper has made a survey on the Keyloggers. The paper depicts the different vindictive assaults that are done utilizing Keyloggers and they have given a brief with regards to the different positive uses of the Keyloggers.
(Reference [11], 2018)	Keylogger is a tool that was mostly created on the grounds that this application is once in a while perceived as a malicious program by antivirus, keylogger will record all exercises identified with keystrokes, the recording system is refined by utilizing string matching technique. The utilization of string matching with strategy during the time spent recording the console is to help the administrator in knowing what the client got to on the PC.
(Reference [19], 2013)	This paper has proposed a structure for the distinguishing proof and avoidance of novel Keylogger attacks. They have used a nectar pot in the organization for really taking a look at the gadgets of the customers. Here on the off chance that the aggressor sends the customer key log using email tends to information base, it is weak against attacks.

3. Terminologies

3.1 Malware

It alludes to any product/executable that is intended to hurt any PC or a PC organization. Organized frameworks often face threats from malicious programs in the form of trojan horses, PC viruses and internet worms [12]. At the point when we download any executable from the web then we probably won't know whether it is safe to execute that product/executable record on our framework or not. In this way, to approve the executable, different malware examination strategies are sent. The quantity of new malware samples and toolboxes for automated malware development are developing dramatically, while the examination capacity and information are going down [13].

3.2 Spyware

Spyware is a programmed tool/code which is secretly deployed in a user's system without their consent. When dynamic, it quietly screens the conduct of users, records their online activities, and takes their passwords [4]. There are multiple kinds of spywares including but not limited to adware, trojan, keyloggers, etc. Spyware is circulated in various manners. One of the most well-known is getting clients to tap on a connection that prompts a malignant site. Those connections can be in messages, instant messages, spring up windows in a program, and advertisements on website pages. Adware is a spyware which target advertisements that are dependent on perusing propensities, outperforms the Internet flag promotions [18].

3.3 NLP

Natural Language Processing (NLP) allows the computer to process the data the same way humans process information [3]. With NLP we can get to know the meaning behind sentences or paragraphs, their intention, their structure and the nature behind them. The biggest application of NLP is Sentiment Analysis. Sentiment Analysis is figuring out the highlighted opinions in text or images [5]. This helps to figure out the behaviour, attitude and emotional state of a person. NLP will help us classify textual data into anger, fear and sadness.

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:34:13 UTC from IEEE Xplore. Restrictions apply.

3.4 CNN

CNN is being used to classify images in almost all the places such as Instagram, Facebook, Traffic Surveillance, Self-driving cars etc. They are really efficient and provide really accurate results [10]. Image Classification is a process in which an image is provided to a CNN model and it outputs a number which belongs to a class, or a probability of how sure the network is that it belongs to that class.

4. Proposed Methodology

To monitor the activities of the concerned person, we have proposed the following method using deep learning models and have reached a quality monitoring software.

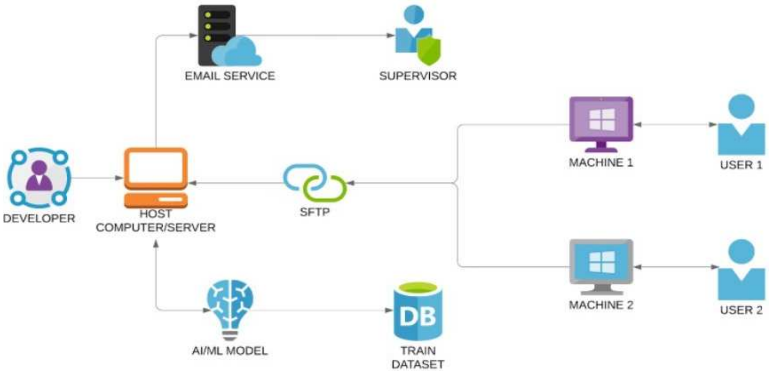


Figure 3 Overview Diagram of the proposed model

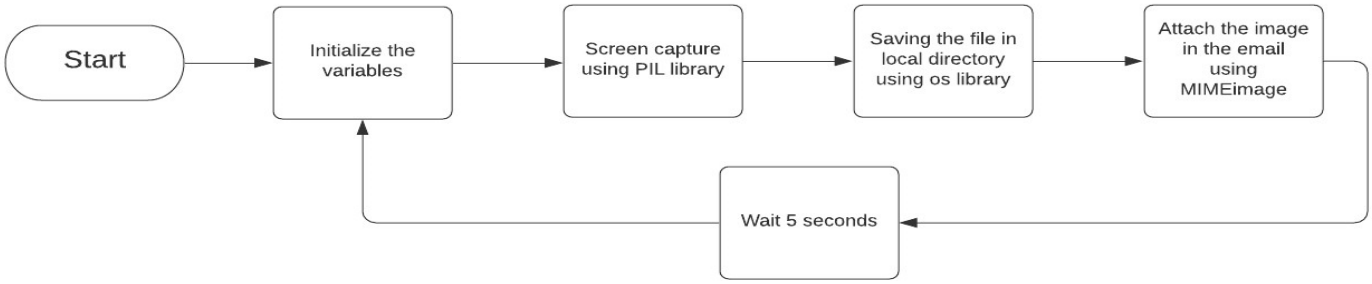


Figure 2 Screenshot Flow

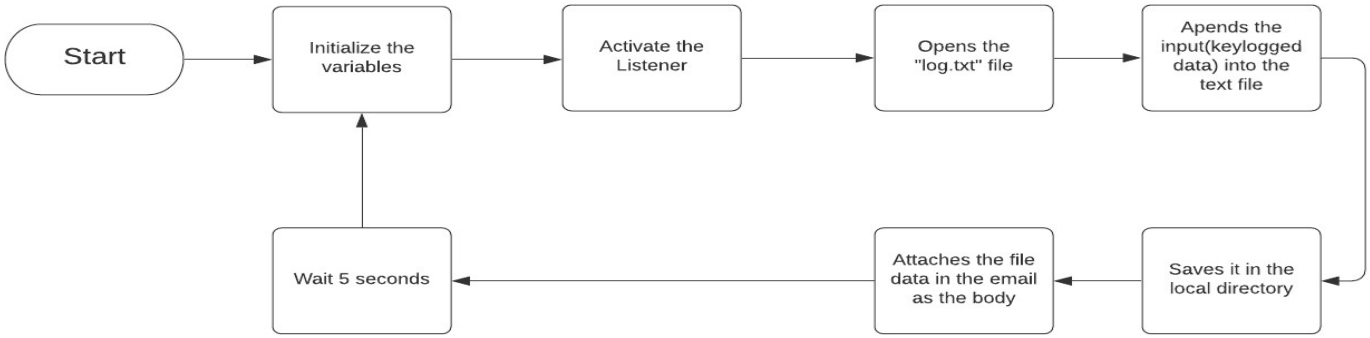


Figure 1 Keylogger Flow

4.1 Flow of the proposed software

The application logs all the keystrokes and captures the screenshot of the user's screen, which are stored temporarily on the computer. The data is then passed through different machine learning algorithms to analyze the log text file and the screenshot. If any kind of red flag is raised, the authority is then informed of the same. This process is repeated after a time interval which is provided by the user (Figure 1). This task is done with the help of different python libraries like PIL, pynput, etc.

4.2 Flow of Screenshot

As the process starts, the variables are initialised and then the screenshot is captured using the Imagegrab.grab() function in the PIL library (Figure 2). The screenshot received will be sent to a transfer learning (Mobilenet V2) model and will be further processed by several flatten and dropout layers to reduce the vector dimensions and size which will further classify the image into: sexy, porn and neutral. If the result of the screenshot is a non-neutral category, then the screenshot would be shared with the concerned authority.

4.3 Flow of Keylogger

It starts by first initialising the variables. Then the listener is activated and then the "log.txt" file is opened. In the text file, the keylogged data is appended (Figure 3). The received keylogging data will be processed through the proposed NLP pipeline and sent to the LSTM model for classification into: fear, anger, sadness and neutral. The output will predict the status of all the keystrokes and the keystrokes would be shared with the concerned authority.

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:34:13 UTC from IEEE Xplore. Restrictions apply.

5. EXPREIMENTAL ANALAYSIS

5.1 Training dataset

Two types of training datasets have been used namely NSFW dataset and Dailydialog dataset for images and text respectively. The NSFW dataset has three categories namely Porn, Sexy and Neutral. The dataset contains 105825 images of the 3 categories. About 5% of the total images are used for testing. The Dailydialog dataset has three categories namely Anger, Sadness and Neutral. The dataset contains 102,000 sentences which are equally divided into the three categories having 34,0000 sentences per category which are further divided into train and test data.

5.2 Analysis of Keylogged Data

5.2.1 Text Pre-processing

- Noise Removal:
  The data is processed and all html markups, URLs, non-ascii symbols, trailing whitespace etc are removed.

- Negation handling:
  All the words such as not, no, cannot, shouldn't, wouldn't that can change the polarity of a sentence sometimes becomes the main sources of low accuracy in sentiment analysis hence, all the negations within the whole scope are handled and data is passed onto the next step in the pipeline.
- Tokenization:
  In this step, the sentences are broken down into smaller parts so that each word can be taken as a discrete unit representing the whole document as a vector which further turns these unstructured strings into numerical data suitable for machine learning tasks.

- Stopword removal:
  In this step, some of the most common words that occur across all the documents such as articles and pronouns are removed as they play no important part in the NLP tasks and hence are not very discriminative.

- Stemming or lemmatization:
  It is the analysis of words that is carried out to return them in the form of a dictionary. It is run on the sentences to highlight the meaning of the keywords.
  Stemming is the analysis of words carried out to remove the suffixes and prefixes from words which turns the word into present tense.

5.2.2 Text Representation:

We have vectorized out logged data using Term-Frequency technique (Term Frequency (TF) — Inverse Dense Frequency (IDF)) (mentioned below):

- First we tokenize the logged data with our preprocess_and_tokenize() method.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

- Then we find it's Term Frequency and Inverse Dense Frequency

5.3 Classifiers

After getting the vectorized text, we pass these vectors into four machine learning models:

5.3.1 Naive Bayes:

It is one of the probabilistic machine learning classifiers used in the paper to help classify vectorized logged data gathered from the keyloggger into the above-mentioned classes. It is based on Bayes theorem through which we predict the probability of a vector belonging in a class based on the given formulae.

$$P(A|B) = P(B|A) \times P(A) \Big/ P(B)$$

where, P(A|B) is the probability of a sample belonging to class A when it already belongs to class B

P(A) is the probability of class belonging to class A

P(B|A) is the probability of a sample belonging to class B when it already belongs to class A

P(B) is the probability of class belonging to class B

5.3.2 Random Forest:

It fits various decision tree classifiers based on sub samples of dataset and uses a parameter such as entropy or Gini index to improve predictive accuracy and prevent overfitting.

5.3.3 Logistic Regression:

It is a type of statistical analysis used for predictive analysis and modelling. In this approach the variable is categorical which helps in predicting the likelihood of an event happening [9]. Mathematically it can be expressed as function where p is the logistic regression function and we try to achieve the optimal value of a and b in the equation stated below.

$$P = e^{a+bX} \Big/ 1 + e^{a+bX}$$

5.3.4 SVM:

It is based upon fitting the data with respect to a line having a decision boundary that can segregate an n dimensional plane into various classes. It creates a hyperplane with the help of extreme case points in the data known as support vectors.

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:34:13 UTC from IEEE Xplore. Restrictions apply.

5.3.5 BERT:

It is a transformer-based ML model developed by Google for NLP applications in 2018. It stands for Bidirectional Encoder Representations from Transformers. It is used in the paper because it outperforms many previous language classifiers tested among various datasets.

5.4 Analysis of Screenshot Image

- The used model was trained in Keras Framework. The total dataset has around 100000 images which were divided into the training and testing dataset as mentioned below.

  Total number of train data is: 31503 (neutral) + 55217 (porn) + 13108 (sexy) = 99828

  Total number of test data is: 1998 (neutral) + 1999 (porn) + 2000 (sexy) = 5997

- All these images were downscaled to 224×224. Then the images were augmented by horizontal flipping, rotating, etc.

- The model was trained over 100 Epochs where the base model that we used was the MobileNetV2 model along with the weights of ImageNet dataset.

- Generally, the MobileNetV2 is used to classify images into 1000 different classes but the top layer of the model was removed and various custom dense Neural Network layers were added as per our requirement to classify the image into the three desired classes: porn, sexy and neutral.

- The Loss Function used in the model: Categorical Cross Entropy. The Optimizer function used in the model: Stochastic Gradient Descent

5.5 Working of proposed model

In our general model, the screenshot and the keylogged data are taken as an input for the image analysis and the text analysis model respectively. Then the image and the log data are classified into their respective classes. If the classified image lies in porn or sexy, then the respective authority will be alerted. Similarly, if the logged data is classified into anger, sadness or fear then the concerned authority would be alerted as well. Until the program is aborted, it will keep analysing the data after a pre-defined time interval (Figure 4).

5.6 Results

The textual data was vectorised using the TF/IDF model. Then, the vectorized text data was initially passed through 4 different classifiers: Random Forest, Naïve Bayes, SVM and Logistic Regression. The accuracy of each classifier is shown in

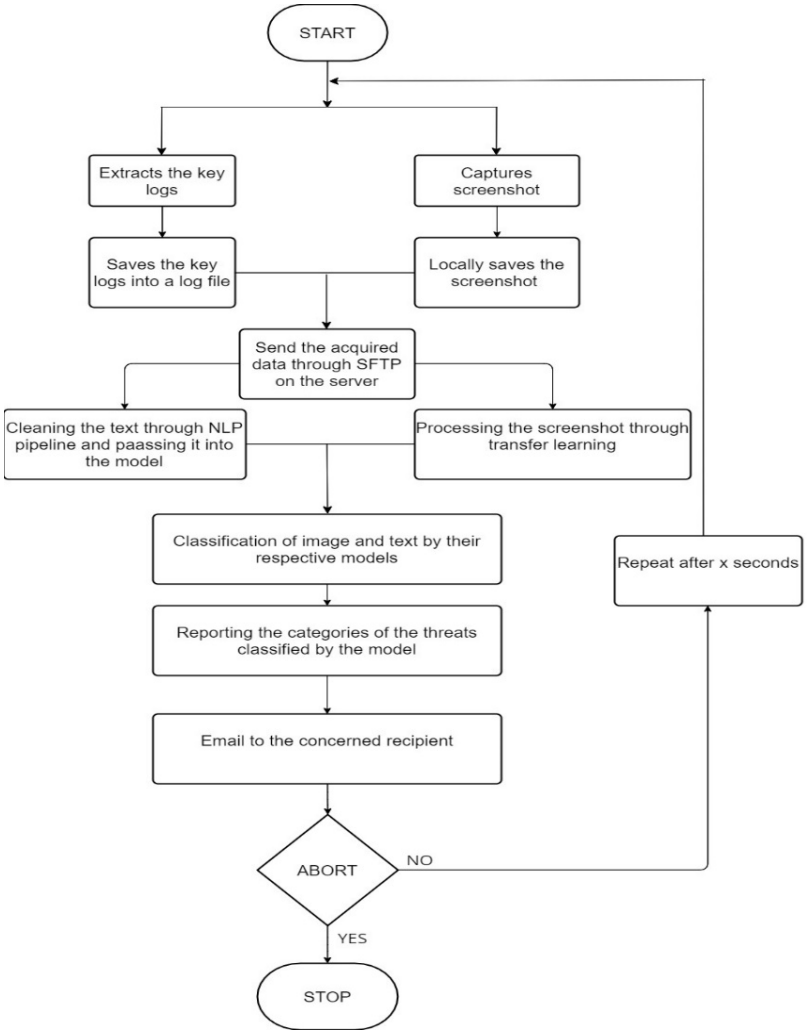


Figure 4 Flowchart of proposed model using different transfer learning techniques

Table 1.

Classifiers	Random Forest	Naïve Bayes	Logistic Regression	Support Vector Machine
Accuracy (%)	63.07	67.02	69.35	72.71

Table 1. Summary of the accuracies achieved by different models

Since these were less than the desired accuracy, we used a transfer learning model, BERT. Table 2 shows the comparison between the best accuracy from the initial four models and BERT.

Classifiers	Support Vector Machines	BERT
Accuracy (%)	72.71	83.06

Table 2. Comparison of BERT and SVM

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:34:13 UTC from IEEE Xplore. Restrictions apply.

Hence, BERT was selected to go ahead as a model of choice for the analysis of the textual data because of its best accuracy as shown in Figure 5.

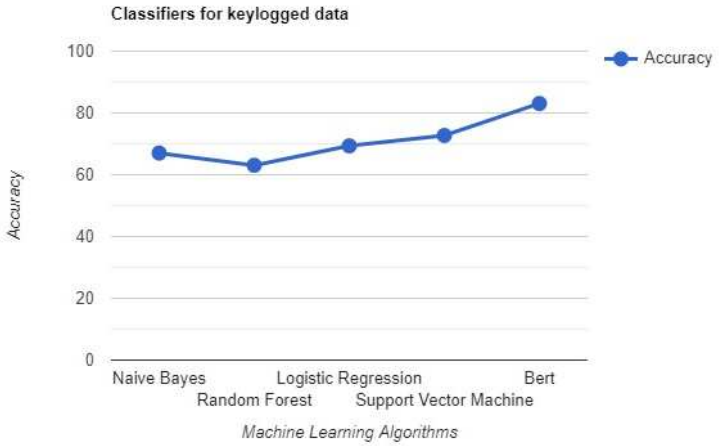


Figure 5 Graph for accuracy of keylogged data classifiers

Using the modified MobileNetV2 model to classify the screenshots, we were able to achieve an accuracy of 89% using the weights of ImageNet. Hence, the model was able to efficiently classify the screenshots into their respective classes which further improved the alert response of the proposed software.

6. CONCLUSION

In this paper, we have presented a novel approach to use spywares like keylogger for monitoring the activities of a user's computer. We have used different python libraries like PIL, pynput, etc. in conjunction with multiple Machine Learning techniques and models to help achieve the goal of tracking the activities of a user efficiently. The screenshots and the keylogged data were successfully classified into the aforementioned categories successfully with accuracies of 89% and 83.06% respectively. This was achieved by using the BERT model for the keylogged data and the MobileNetV2 model for the screenshots. The concerned authority is alerted if the models predict that the data is inappropriate for the user. The models can also be tweaked for different applications like parental protection, employee monitoring system, university labs/cybercafé monitoring systems, etc. as per the requirements of the user.

In the future we can work on implementing more categorical and accurate classifications by increasing the number of epochs to train the model and to use the call-back function to monitor the loss and validation accuracy. We can also augment the dataset by applying more augmentation techniques such as colour augmentation, scaling and padding for an overall increase in the number of images in the training dataset and hence the model accuracy.

REFERENCES

[1] Akhil S, Neeraja M Nair, Asst Prof. Arun R. (2014). DETECTION AND PREVENTION OF KEYLOGGER SPYWARE AT[8] Hagerty, J. and D. Berman (2003), "Caught in the Net: New Battleground over Web Privacy: Ads that Snoop," Wall Street Journal, August 27, p.A1.

[2] Bhardwaj, A., Goundar, S. (2020). Keyloggers: silent cyber security weapons. Network Security, 2020(2), 14–19. https://doi.org/10.1016/s1353-4858(20)30021-0

[3] Hirschberg, Julia, and Christopher D. Manning. "Advances in natural language processing." *Science* 349, no. 6245 (2015): 261-266.

[4] Egele, Manuel Kruegel, Christopher Kirda, Engin Yin, Heng Song, Dawn. (2007). Dynamic Spyware Analysis.

[5] Rajput, Adil. "Natural language processing, sentiment analysis, and clinical analytics." In *Innovation in Health Informatics*, pp. 79-97. Academic Press, 2020.

[6] Ferrer, D. and M. Mead (2003), "Uncovering the Spy Network," Computers in Libraries, (23)5, p. 16.

[7] Hemita Pathak, Apurva Pawar, Balaji Patil. (2015). A Survey on Keylogger: A malicious Attack. International Journal of Advanced Research in Computer Engineering Technology. Published. http://ijarcet.org/wpcontent/uploads/IJARCET-VOL-4-ISSUE-4-1465-1469.pdf

[8] Naider, A.V. (2004), "Testimony of Mr. Avi Z. Naider," U.S. Senate Committee on Commerce, Science and Transportation, http://commerce.senate.gov/ pdf/naider032304.pdf (current July 14, 2004).

[9] Nick, Todd G., and Kathleen M. Campbell. "Logistic regression." *Topics in biostatistics* (2007): 273-301.

[10] Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu et al. "Recent advances in convolutional neural networks." *Pattern Recognition* 77 (2018): 354-377.

[11] Rahim, R., Nurdiyanto, H., Saleh A, A., Abdullah, D., Hartama, D., Napitupulu, D. (2018). Keylogger Application to Monitoring Users Activity with Exact String Matching Algorithm. Journal of Physics: Conference Series, 954, 012008. https://doi.org/10.1088/1742-6596/954/1/012008

[12] Rieck, K., Holz, T., Willems, C., Dussel, P.,¨ Laskov, P. (2008). Learning and Classification of Malware Behavior. Detection of Intrusions and Malware, and Vulnerability Assessment, 108–125. https://doi.org/10.1007/978-3-540-70542-0 6

[13] Rodrigo Rubira Branco, Shamir, U. (2010). Architecture for automation of malware analysis. 2010 5th International Conference on Malicious and Unwanted Software. Published. https://doi.org/10.1109/malware.2010.5665786

[14] Sharma, Kapil, Himanshu Anand, Himanshu Nandanwar, and Anamika Chauhan. "Taxonomy of Routing Protocols." In 2022 International Conference for Advancement in Technology (ICONAT), pp. 1-14. IEEE, 2022

[15] Sharma, Kapil, Rakesh Garg, C. K. Nagpal, and Ramesh Kumar Garg. "Selection of optimal software reliability growth models using a distance based approach." IEEE Transactions on Reliability 59, no. 2 (2010): 266-276.

[16] Solairaj, A., Prabanand, S. C., Mathalairaj, J., Prathap, C., Vignesh, L. S. (2016). Keyloggers software detection techniques. 2016 10th International Conference on Intelligent Systems and Control (ISCO). Published. https://doi.org/10.1109/isco.2016.7726880

[17] Stafford, T. F., Urbaczewski, A. (2004). Spyware: The Ghost in the Machine. Communications of the Association for Information Systems, 14. https://doi.org/10.17705/1cais.01415

[18] Townsend, K. (2003), "Spyware, Adware and Peer-to-eer Networks: The Hidden Threat to Corporate Security," PestPatrol White Paper, http://www.pestpatrol.com/ Whitepapers/CorporateSecurity0403.asp (current July 17, 2004

[19] Wazid, M., Katal, A., Goudar, R., Singh, D., Tyagi, A., Sharma, R., Bhakuni, P. (2013). A framework for detection and prevention of novel keylogger spyware attacks. 2013 7th International Conference on Intelligent Systems and Control (ISCO). Published. https://doi.org/10.1109/isco.2013.6481194

978-1-6654-5834-4/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:34:13 UTC from IEEE Xplore. Restrictions apply.

Journal of Polymers and the Environment
https://doi.org/10.1007/s10924-022-02638-4

ORIGINAL PAPER

Sustainable Xylose-Based Non-Isocyanate Polyurethane Foams with Remarkable Fire-Retardant Properties

Pooja Singh[1] · **Raminder Kaur**[1]

Accepted: 10 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Self-blowing non-isocyanate polyurethane foams have been synthesized by reacting xylose with dimethyl carbonate and hexamethylene diamine using citric acid as a natural crosslinker via the process of transurethanisation. The resulting foams exhibit exceptional flame retardancy in comparison to conventional PU foams. The structure of the foam cells was examined using scanning electron microscopy. Solid-state NMR and infrared spectroscopy were used to analyze the functional groups present in the products formed. The formation of the urethane bridge was confirmed by the emergence of peaks at 158.24 ppm and 1693 cm^{-1} respectively. The thermal properties were evaluated by using TGA and DSC, which further indicated a glass transition temperature of around 112.82 °C. The dependence of viscosity of polymeric solutions on the different shear rates was also studied using a parallel plate rheometer. Limiting Oxygen Index was employed to study the flammability of the prepared foams, and it was observed that they exhibit considerable fire resistance (26.9%),which is quite comparable to the fire resistance offered by the isocyanate-based polyurethane foams with added fire retardants. The diffraction pattern of the prepared NIPU was analyzed by PXRD.

Keywords LOI · NIPU foams · rheology · solid-state NMR · xylose

Introduction

Polyurethanes (PU) are among the most extensively used polymer globally, with a broad array of applications in both academic and industrial settings, mainly attributed to their abrasion resistance, high flexibility, and other desirable characteristics. PU foams are an important class of polymers that make up 50% of the market for polymer foam and account for 67% of global PU utilization. PU foams serve many purposes such as acoustic insulation, automotive seating, mattresses, and thermal insulation [1]. However, there are a few concerns associated with the preparation and use of PU. They are typically synthesized by polyaddition of a polyol to a polyisocyanate, which, along with its precursor phosgene, is exceedingly hazardous. Research groups are now aiming to divest from the usage of hazardous chemicals and severe reaction conditions in PU production with the use of greener intermediates and procedures, considering the current state of sustainable chemistry development. Moreover, the depletion of non-renewable sources and the ramifications of global warming have prompted the innovators to create novel routes of PU production, that have a minimal environmental impact and fulfill all new environmental regulations [2]. Vegetable oil is the most often used renewable resource to generate precursors for PU products, since it is widely accessible [3, 4]. Although these renewable sources have successfully yielded polyols, multiple studies have been carried out to incorporate various fillers and anti-flaming chemicals into these foams to enhance their anti-flammable, thermal, and mechanical abilities [5–12]. In addition, as the toxic polyisocyanates used to make PU could be dangerous to humans, researchers are seeking a green and sustainable route of producing PUs from renewable sources, without the use of polyisocyanates [3]. Consequently, Non-isocyanate polyurethanes (NIPU), a novel category of PU free from the harmful effects of isocyanates, were generated. NIPUs are traditionally synthesized by using cyclic carbonates and amines [13]. The reaction produces two isomeric products, one with a secondary hydroxyl group and the other with a

✉ Raminder Kaur
rkwalia14@gmail.com

[1] Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University (Formerly Delhi College of Engineering), 110042 Delhi, India

Published online: 03 November 2022

methylol group, both of which include urethane linkage and a hydroxyl group. These hydroxyl groups may be able to establish intermolecular hydrogen bonds by combining them with the urethane group [14]. The ring-opening reaction of bis-cyclic carbonates with diamines and polycondensation of carbamates are the two most frequently used approaches for producing NIPU [15]. The most promising strategy is the utilization of cyclic carbonates, which generate minimal chemical intermediates or by-products. The chemical resilience and hydrolytic stability of the polymers are enhanced as additional hydroxyl groups are generated in the NIPU structure in the ring-opening reaction [14]. Transurethane reactions, also called transcarbamoylation, are caused by the condensation of alcohol on carbamate which leads to the formation of urethanes often producing alcohols (with low molar masses) as a byproduct [16, 17]. However there are other methods as well, which are relatively less explored. One such method involves the production of cyclic carbonates from dimethyl carbonate, by reacting with o-hydroxyl compounds followed by reacting with a diamine. Since dimethyl carbonate is an environmentally compatible and non-toxic chemical and the reaction takes place under simple conditions, this route of manufacture is highly valued [18].

Despite the positive potential for the economic viability and environmental stewardship of non-isocyanate polyurethane, commercial production of non-isocyanate polyurethanes is yet to be implemented. Research on NIPUs is currently flourishing, and the advanced characteristics of NIPUs are predominantly being used to develop coatings and cross-linked materials with prospective uses as thermosets, elastomers and space applications also [19–22]. Nevertheless, due to the difficulty in foaming the NIPU matrix, existing examples of NIPU foams are relatively restricted. However, NIPU foams using tannin, glucose, lysine, and lignin have been reportedly prepared [18, 23, 24].

Due to the compelling rise in oil and natural gas prices, polymeric biomaterials; particularly agro-resource-based polymers have been progressing rapidly in recent years. After preprocessing and hydrolysis, the plant biomass hydrolyzates consist of various hexoses and pentoses. Even though the sugar content of hydrolyzates, from different feedstocks, varies, the dominant sugars present in most hydrolyzates from terrestrial plant biomass are glucose and xylose [25]. The motivation behind this study is to achieve an advanced understanding of how the alternative synthesis route influences the structure and properties of the polyurethane foams synthesized without isocyanates. Xylose is an agriculturally workable material that can serve as a feasible replacement for petroleum-based polymers in a life-cycle context. The aim of this research work is to explore the viability of xylose in the synthesis of polyurethane foams and to employ an environmentally acceptable way to generate non-isocyanate polyurethane. Citric acid is employed as a natural crosslinker in this study since it is biodegradable, non-toxic, and completely safe during manufacture. Though many alternative techniques and renewable raw materials have previously been used to produce bio-based NIPU foams, no previous research has mentioned the development of xylose-based polyhydroxyurethanes (PHUs) by the processes of transcarbamoylation. Compared to the already reported procedures for the synthesis of NIPU foams, which are quite complicated, due to the use of sophisticated equipment, expensive techniques, high temperature, or high pressure, the exquisiteness of this investigation lies in the fact that these NIPU foams were synthesized using renewable sources at comparatively lower temperatures (90 ℃) and ambient pressure.

Fourier transform infrared (FTIR) spectroscopy, Differential scanning calorimetry (DSC), Powder X-ray diffraction (PXRD), and Thermal Gravimetric Analysis (TGA), were used to explore the structure and mechanical properties of the synthesized foams. Limiting oxygen index (LOI) was used to determine the fire resistance of all foams.

Materials and Methodology

Materials

Carbamates were prepared by utilizing commercially available hexamethylene diamine (HMDA), which was acquired from Sigma Aldrich. D-Xylose and dimethyl carbonate (DMC) were purchased from Loba Chemie (India). Citric acid was purchased from TCI Chemicals (Belgium). All chemicals and solvents used in the analytical measurement procedures were of analytical grade and utilized exactly as received.

Methodology

In this study, NIPU foams with various combinations of precursors were prepared. The carbamates were prepared by the method already reported in the literature [26]. The reaction of DMC with HMDA was allowed to proceed for 60 min at 70 °C with continuous stirring, and the synthesized intermediate was termed RXCL. Further, aqueous solution of D-xylose was added to a beaker containing RXCL and it was heated at 80 °C, with constant stirring for 20 min. During heating, the viscosity of the reaction mixture increased gradually, and its color changed to dark brown. The beaker was then placed in the oven for 2 h at 90 °C and the resulting foam product was labeled RXCO. A similar procedure was applied to prepare NIPU foams with citric acid. The dark brown reaction mixture was supplied with varying amounts of citric acid to prepare RXC1 and RXC2. The foaming beaker was then kept in the oven for 2 h at 90 °C. The prepared foams were further kept for curing for 24 h at room

temperature. Detailed formulations of the synthesized foams have been mentioned in Table 1. A schematic representation of the synthesis of Xylose-based NIPU Foam is shown in Fig. 1.

Characterization Methods

Rheological Studies

The rheological characteristics of pure NIPU and NIPU/Citric acid solutions were studied, at room temperature, using an Anton Paar Modular Compact Rheometer 302 (MCR). The shear-thinning behavior was explored using a steady shear flow measurement, to determine the viscosity changes in a wide shear rate range (0.1–100 s^{-1}), using a 25 mm parallel plate geometry and 1 mm gap.

Morphological and Structural Studies

The morphology and structure of the developed foams were assessed by powder X-Ray Diffraction (PXRD), utilizing a high-resolution Bruker diffractometer (D8 Discover), coupled with a point detector (scintillation counter), using Cu K α radiation received via a gobel mirror that covers the range of 2θ = 5–80°, with a scan rate of 1.0 s/step and step size 0.02°, at 298 K. Perkin Elmer Frontier ATR-FTIR provided by an ATR Miracle diamond crystal was employed to examine the functional groups of all the prepared foam samples. The powdered samples were placed on the ATR equipment's diamond eye (1.8 mm) and the sample's touch was maintained by securely screwing the clamping equipment. Each sample was scanned with 32 scans at a resolution of 4 cm^{-1} in the wave number range of 600 to 4000 cm^{-1}, to record the spectrum. Thermogravimetric analysis (TGA) was employed to evaluate the thermal stability of the resulting NIPU foams (using TGA 4000, Perkin Elmer). Differential scanning calorimetry (DSC) was used to assess the thermal stability of the resulted foams using DSC 8000, Perkin Elmer. Nitrogen was used as a purge gas. To determine the glass transition temperature, the thermal characteristics were analyzed at 20 °C min^{-1} in the range of −50 to 250 °C. A scanning electron microscope (Jeol JSM-6610LVmodel) was used to examine the morphology and the interior structure of the foams on a lateral cross-section to the rising direction of NIPU foam. Solid-state CP MAS 13 C NMR was used to examine the powdered xylose foam. A Brüker AVANCE 400 MHz (Brüker, Billerica, MA, USA) spectrometer with a 4 mm probe was used to acquire the spectrum at a frequency of 10 kHz. Chemical shifts were evaluated in comparison to tetramethyl silane (TMS). The spectrum was recorded with the spinning sideband suppression.

Table 1 Formulations for the preparation of NIPU foams

Sample	D-Xylose (g)	DMC (g)	HMDA (g)	Citric Acid(g)
RXCO	13	6	7	0
RXC1	13	6	7	1
RXC2	13	6	7	2

Apparent Density and Water Absorption Test

The apparent densities were calculated as the ratio of weight to cubic sample volume by following the standards of ASTMD1622-08. Five samples were taken to determine average density. To explore the water absorption of the synthesized NIPU foams, four samples were evaluated for

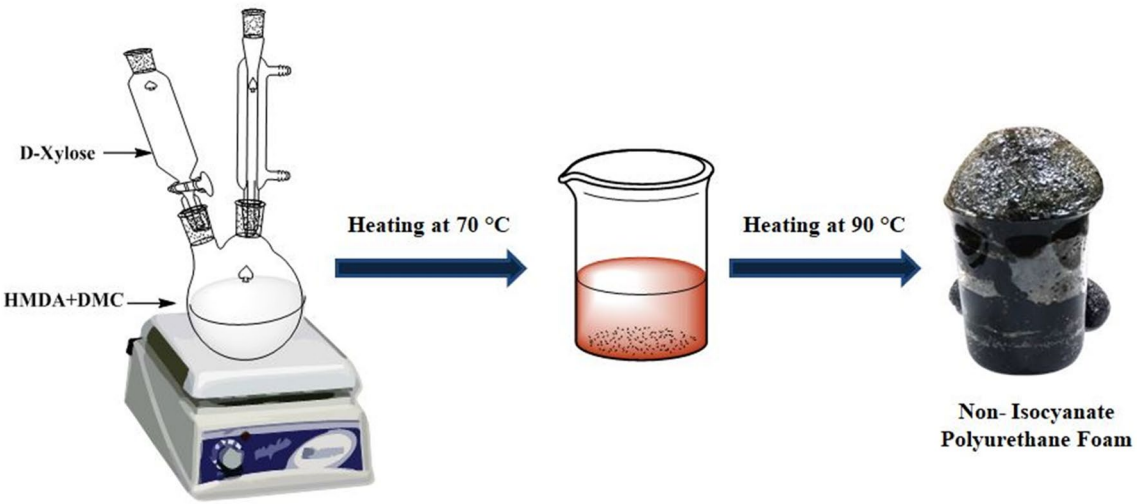


Fig. 1 Schematic representation of the synthesis of Xylose-based NIPU Foam

24-hour water absorption and the effect of densities on water absorption was investigated [27].

Oxygen Index

The limiting oxygen index (LOI) was determined using an Oxygen index meter, according to the method described in the literature [23], (Fire Testing Technologies Ltd, United Kingdom). The specifications of all the samples were taken as 100 mm x 10 mm x 10 mm. The average LOI values for five samples of each formulation product were recorded.

Results and Discussions

Preparation of the Foams

The xylose-based NIPU foams were prepared using a facile method at ambient pressure. The anticipated primary adducts created by various reactions that eventually lead to skeletal networks in the foam are presented in Fig. 2. The addition of tricarboxylic citric acid serves dual purpose as a crosslinking agent [28, 29] and a blowing agent [30] in the synthesis of RXC1 and RXC2 NIPUs. Citric acid-based foams generate their foaming energy through a relatively violent reaction between the citric acid carboxylic group and the NH_2 group from diamine, which leads to self-blowing at

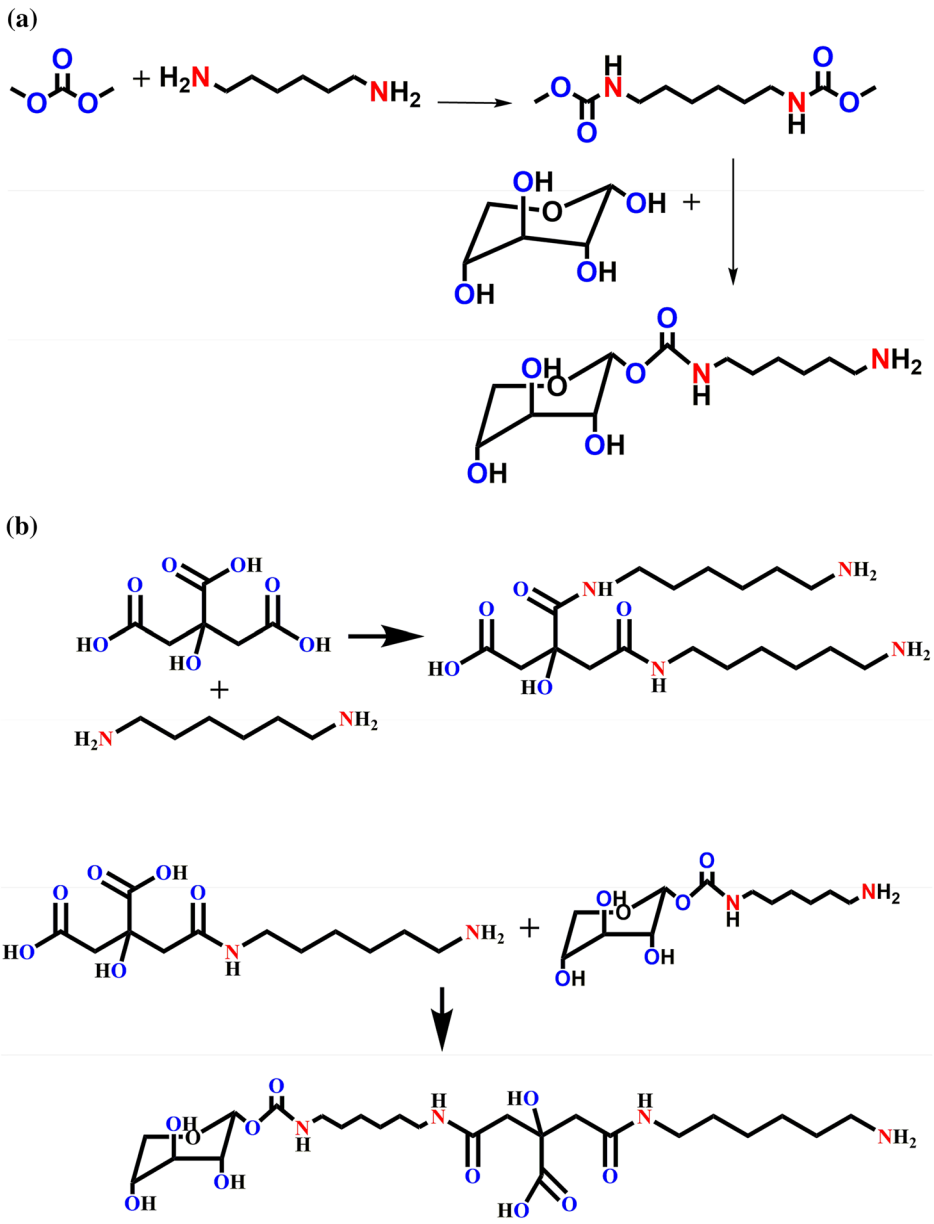


Fig. 2 **a** Schematic Representation of Possible Reaction Mechanism for the Synthesis of Xylose-based NIPU foams **b** Role of Citric Acid

Table 2 Physical Properties of Xylose-based NIPU foams

Sample	Apparent density (g/cm^3)	2 h Water absorption (%)	Pore size Diameter (mm)	LOI Values (%)
RXCO	0.574 ± 0.13	319.42 ± 46	1.11 ± 0.045	22.3% ± 0.11
RXC1	0.4676 ± 0.14	353.82 ± 35.17	1.82 ± 0.33	23.99% ± 0.78
RXC2	0.303 ± 0.16	421.24 ± 31.05	1.95 ± 0.08	26.9% ± 0.24

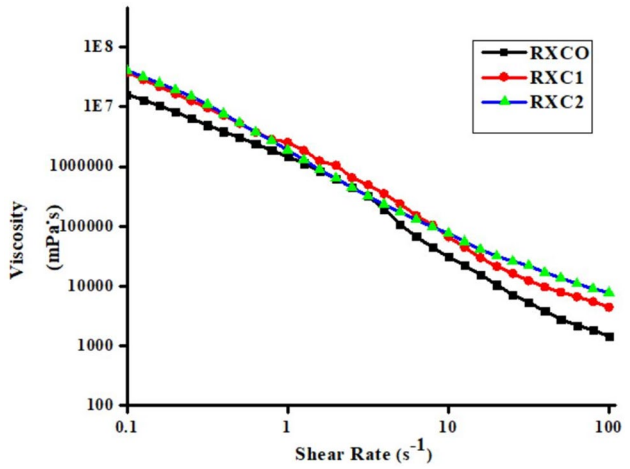


Fig. 3 Plot of viscosity versus shear rate for the polymeric solutions of RXCO, RXC1, RXC2

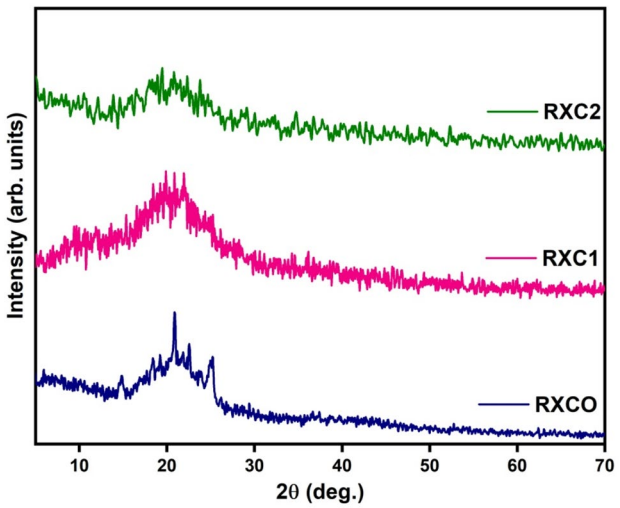


Fig. 4 XRD pattern of the prepared NIPUs

room temperature. A three-dimensional network is formed as a result of this reaction, causing the mixture's viscosity to rise sharply followed by its gelation. The foaming action is caused by heating the reaction mixture at 70 °C in the case of foams prepared without citric acid.

The prepared foams were cured by heating, which is crucial for the accessibility and stability of the synthesized foams and to examine their physical properties [23]. Furthermore, as the amount of citric acid increased, the apparent densities of the prepared foams declined slightly, and water absorption increased consecutively, as reported in Table 2. The foaming energy generated by the interaction of citric acid with the diamine $-NH_2$ groups is responsible for the decrease in foam density [18]. Increasing the amount of citric acid, on the other hand, can yield a smaller apparent density and hence, a larger expansion volume [30].

Rheological Analysis

The shear-thinning characteristics were investigated by employing a steady shear flow measurement to analyze the changes in viscosity of the dark brown reaction mixture of RXCO, RXC1, and RXC2 in a wide range of shear rates (0.1–100 s-1), using a 1 mm gap and a 25 mm parallel plate. The steady shear profiles of samples before curing were studied at 25 °C and are presented in Fig. 3. A significant increase in the viscosity was observed with the addition of citric acid, suggesting a resistance to the flow of the polymeric solution.

The viscosity of polymeric solutions varied with polymer content, owing to the variation in polymer chain entanglement. RXC2, containing the highest concentration of citric acid exhibited the highest viscosity at a low shear rate, and displayed maximum shear thinning behavior when subjected to high shear. The high viscosity may be a result of the existence of a highly entwined polymer structure in the solution due to the weak crosslinks developed by citric acid (before curing), that break down on the application of high shear and result in maximum shear thinning. While in RXCO, the low viscosity at low shear rates, and the lowest shear thinning at high shear rates can be attributed to the lack of crosslinking in absence of citric acid. The shear-thinning behavior of these solutions demonstrates that the application of a high shear rate leads to disentanglement and even realignment of molecular chains in the blend solutions.

Morphological Characterization

In the PXRD pattern presented in Fig. 4, the presence of a dominant peak at 21° implies that the synthesized material is segmented polyurethane (NIPUs in this case) and the broadness of the peak signifies the amorphous nature of the

predominant phase. This is consistent with the previously reported values in the literature for segmented polyurethanes [31–33]. These observations are further confirmed from DSC thermograms, as Tg is observed only in the case of polymers having an amorphous region as in the dominant phase, contrary to crystalline polymers that show sharp transitions related to melting temperature, not glass transition temperature.

Figure 5 displays the Fourier Transform Infrared (FT-IR) spectra of all foams. The occurrence of a band around 1690 cm^{-1} in the FTIR spectrum of RXCL indicates the formation of carbamate linkage and confirms the methoxycarbonylation of HMDA with DMC. In the spectra of NIPU foams RXCO, RXC1, and RXC2, the occurrence of a characteristic band at 1693 cm^{-1} corresponds to urethane C=O bond formation [18, 34]. Furthermore, a band at 3340 cm^{-1} indicates the presence of hydroxyl groups and -NH- moiety of urethane groups. Additionally, the band centered at 1537 cm^{-1} is indicative of the urethane group, implying the formation of carbamate structures [23]. The analysis of FTIR spectra suggests that RXCL, RXCO, RXC1, and RXC2 have chemical structures that are comparable, but not identical to pristine xylose and citric acid, hence indicating that the xylose based-NIPU foams prepared using similar procedures have comparable spectra. This confirms that xylose-based NIPU foams have been successfully fabricated using an isocyanate-free approach at atmospheric pressure.

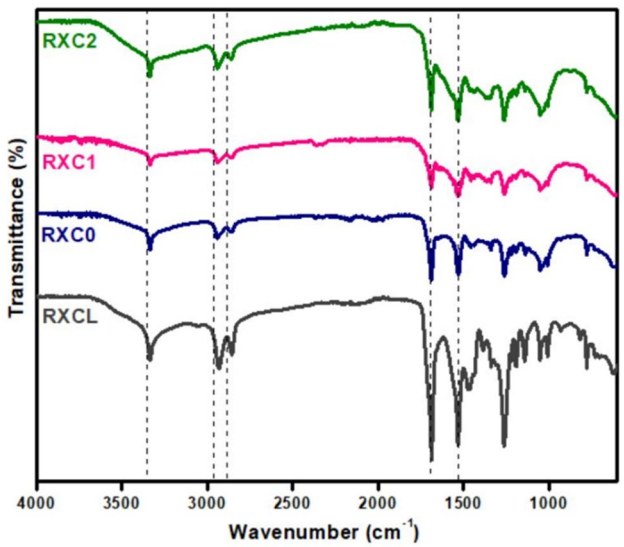


Fig. 5 FTIR spectra of the synthesized products

The morphology of the foams synthesized with varying amounts of citric acid was examined using SEM. Figure 6 displays the SEM micrographs of RXCO, RXC1, and RXC2. This suggests that all foams have a consistent cell morphology (closed cellular structure with few open cells). As evident from Fig. 6 (a), RXCO exhibits considerably round pores, however, the cell shapes got disoriented with the addition of citric acid. Figure 6 (b) and (c) show a considerable degree of open cell morphology in RXC1 and RXC2 compared to RXCO, which may be the result of the addition of citric acid in the resins during the preparation. The citric acid and HMDA reacted relatively vigorously producing greater energy, as a result of which more foamed material or larger foam cells appeared, as reported in other studies also [30]. It was observed while analyzing the apparent density and water absorption of the prepared materials, that they decreased and increased respectively, with an increase in the amount of citric acid owing to the formation of larger pores as depicted in Table 2; Fig. 6. An increase in the swelling index can be attributed to the ease of the solvent penetrating the material more easily due to the formation of larger pores. In addition, some ruptures or wreckage can be seen in all prepared foams that can be attributed to the cutting step in sample preparation. It was observed that, with an increase in the citric acid concentration, the rigidity of synthesized foams increased linearly, due to the formation of stronger cell walls. Thus, the obtained products were considerably hard compared to RXCO, and exhibited resistance to mechanical force. Scattered debris can be seen in Fig. 6 (c), and it was a result of cutting during sample preparation for SEM analysis.

The composition of the prepared foams was examined by solid-state Cross Polarization Magic Angle Spinning (CP

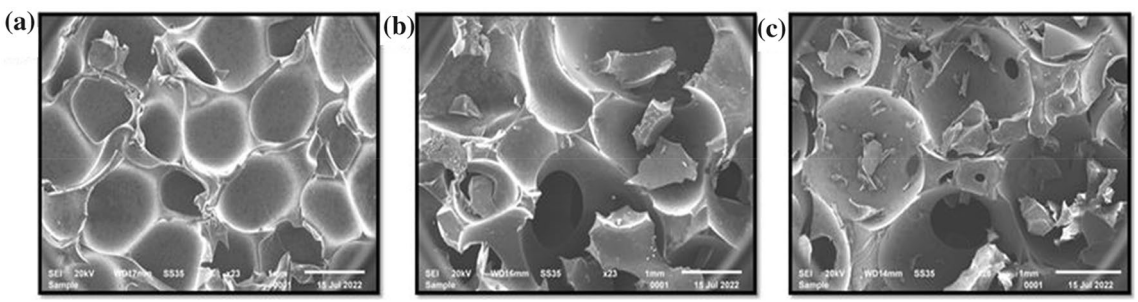


Fig. 6 SEM images of RXCO (**a**), RXC1 (**b**), RXC2 (**c**)

MAS) ^{13}C NMR. Figure 7 displays the ^{13}C NMR spectrum of RXCO. The peaks at 44.71 ppm and around 30.56–28 ppm correspond to the methylene groups (–CH_2–) of the diamine chain, which are directly coupled and not attached to urethane groups, respectively. The presence of a peak at 42 ppm verifies the diamine reaction and the production of urethane groups, which is also confirmed by the FTIR spectrum of RXCO. The peak at 52.53 ppm appears due to the methyl group present in –COO–CH_3 moiety of unreacted carbonated xylose molecules. The notch around 70 ppm can be credited to the carbon atoms of the xylose ring. The peak at 158.24 ppm refers to the –C=O– linkage present in –NH–COO– moiety of the urethane bridge. This peak is extremely prominent, suggesting that the fraction of urethane produced is comparatively high in comparison to that of the 52.53 ppm peak [35].

Thermal Characterization

TGA was employed to evaluate the thermal stability of the synthesized NIPUs in a nitrogen (N_2) atmosphere across a temperature range of 25 to 900 °C. The TGA-DTG curves display two-step mass loss, as presented in Fig. 8 (a), (b), and (c). The initial weight loss step between 30-200 °C involves the decomposition of hexamine and excess acid as well as the discharge of absorbed water to 120 °C and evaporation of monomers/oligomers generated during the course of the reaction below 200 °C [10, 28]. The second weight loss step observed between 200 and 275 °C occurs due to the fragmentation of the polymer backbone as a result of urethane bond cleavage. The final weight loss step occurring above 400 °C (as shown by DTG curves) involves more than 58% mass loss. Possible fragmentation of C–C bonds and the degradation of pyrolysis remanent compounds from the prior two steps might have triggered this process. Similar results have been observed in the literature [36]. Figure 8 (a), (b) and (c) depict the respective degradation temperatures and residual masses of the prepared foams.

DSC measurements were carried out to determine the glass transition temperature (Tg) of the synthesized NIPU foams. The DSC thermogram, presented in Fig. 8 (d), shows that the Tg of the synthesized NIPUs lies in the range of 109–112 °C and is similar to the data previously reported for NIPUs in literature [37]. Tg values of the synthesized NIPUs are higher than room temperature, suggesting a significant level of crosslinking by citric acid. Furthermore, the appearance of a singular Tg peak in the DSC curve of each sample (Fig. 8 (d)) suggests the existence of a homogenous phase in NIPU foam. Thus it can be inferred that the chemical structure of the curing agent has a considerable impact on the Tg values of NIPU foams, and aids in tailoring their thermal properties [38]. The role of citric acid as a blowing agent is demonstrated in the SEM micrographs of the synthesized NIPU foams (Fig. 6) and that as a crosslinker is proven by

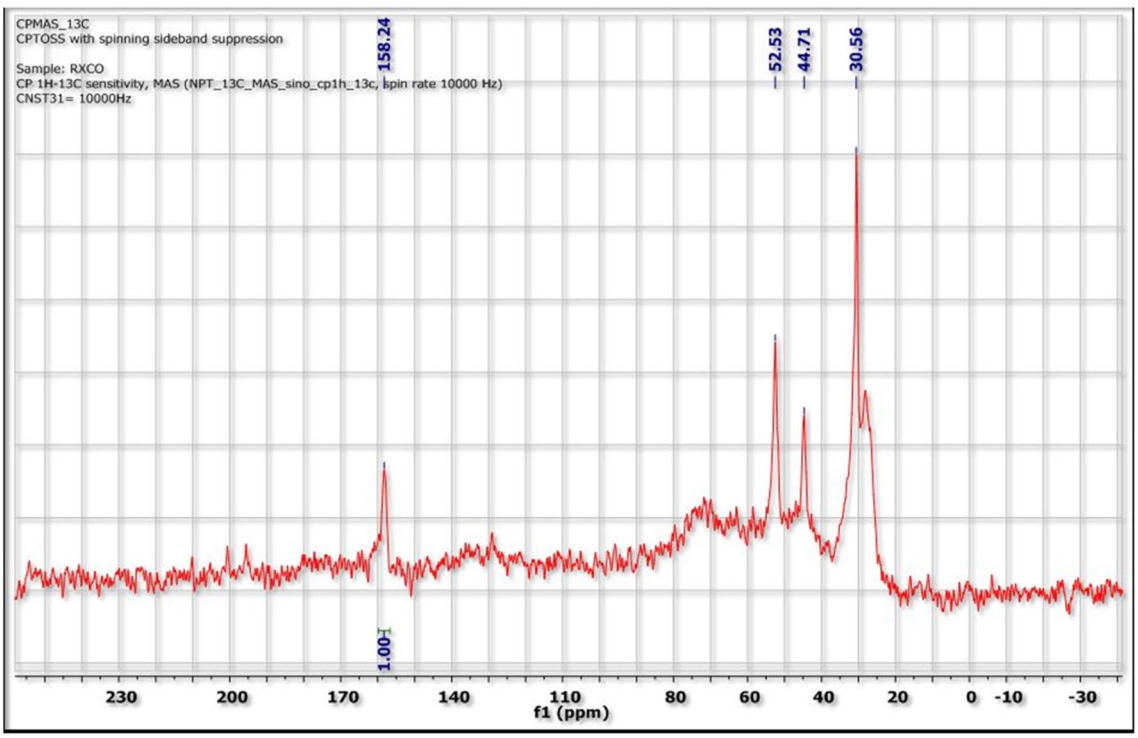


Fig. 7 Solid State ^{13}C NMR of Xylose-based NIPU foam RXCO

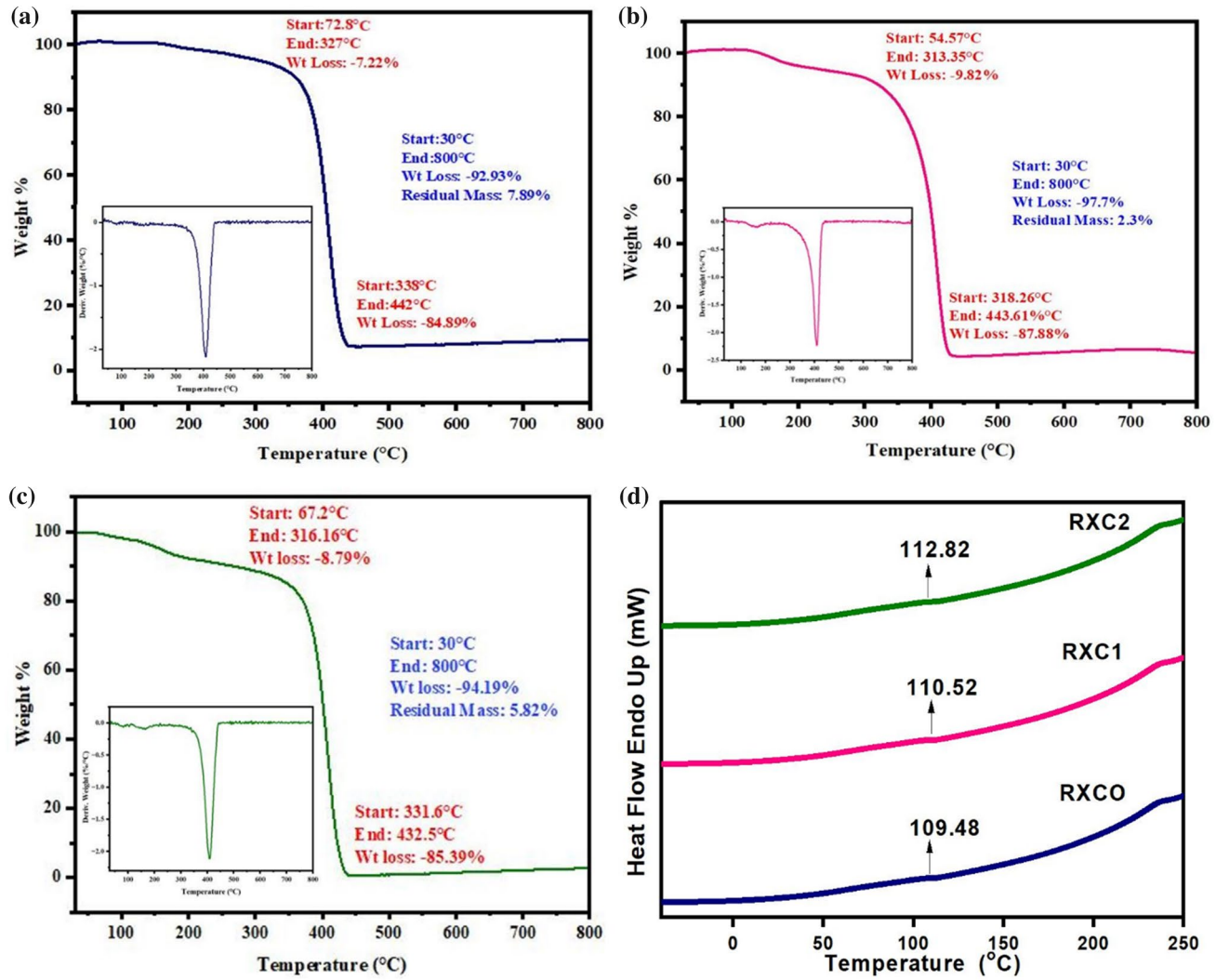


Fig. 8 TGA-DTG curve of **a** RXCO **b** RXC1, and **c** RXC2 under the N_2 atmosphere and **d** DSC Thermogram of the xylose-based NIPU foams (RXCO, RXC1, RXC2)

DSC thermograms (Fig. 8 (d)), rheology data (Fig. 3), and also by LOI values (Table 2). Previous studies also suggest citric acid as a widely reported crosslinker [23].

Flame Retardant Properties

The limiting oxygen index (LOI) is used to measure the flame-retardant behavior of the polymeric materials by assessing the minimal oxygen concentration required for the combustion of polymer in a combined oxygen and nitrogen atmosphere. This technique continues to be one of the most crucial quality control and screening tool to evaluate both the flammability resistance and ignitability in plastics [39]. Materials exhibiting high LOI values show excellent flame retardancy. When the LOI value of a substance exceeds 27%, it is considered flame-retardant. Contrarily, a material is rendered inflammable, when the LOI value is less than 22% [40]. Materials with an LOI of less than 21% are categorized as combustible, whereas those with an LOI of more than 22 are classified as self-extinguishing, since their ignition cannot be perpetuated at ambient temperature, without the addition of external energy [39]. The LOI values of each foam specimen were assessed, and the outcomes are presented in Table 2. The LOI values of pristine xylose-based NIPU foam (RXCO) are around 22.3%, showing that this type of foam is not inflammable, but self-extinguishable. The LOI values of NIPU foams incorporated with citric acid were found to be 23.996% for RXC1 and 26.9% for RXC2. As the amount of citric acid was increased, the LOI values of the samples increased markedly. This can be attributed to the creation and densification of an isolation layer, that grew larger with increasing citric acid amount and generated a more intumescent char [41].

This phenomenon may be explained due to the presence of the carbonaceous char layer that shields the polyurethane surface from the flame. As a result, increasing the amount of citric acid in the foam results in a greater barrier effect, leading to improved fire performance. This can also be established by the fact that citric acid also acts as a crosslinker here, as prior proven by DSC measurements. The higher the density of the crosslinked network, the higher will be the amount of energy required to burn the sample under observation, since the three-dimensional structure will show resistance to burning [42]. As a consequence of which, RXC2 having the maximum amount of citric acid has the highest LOI value, inferring that citric acid contributes in enhancing the flame retardancy of the material. Similar data has already been documented in a few studies [43–44].

Conclusion

Xylose-based non-isocyanate polyurethane foams have been prepared via transurethanisation reaction. The methodology presented in this study allows for the production of sustainable polyurethane foams using xylose as a feedstock and citric acid as a crosslinker and blowing agent, resulting in the materials with high bio-based content. It is observed that the inclusion of citric acid in various proportions in the NIPU foam amplifies the thermal stability and the fire retardancy of the xylose-based NIPU foams. FTIR measurements confirmed the urethane linkages in all the synthesized foam compositions. Solid-state ^{13}C NMR further established the formation of the urethane bridge by displaying a distinctive apex at 158.24 ppm. The three types of foams, displayed comparable regular cell shapes, with lots of open pores visible, but the NIPU foams prepared without citric acid exhibited smaller mean cell sizes and closed cells, compared to the ones prepared using citric acid (RXC1 and RXC2), this might be due to the higher amount of foaming energy produced by citric acid. Homogenous foams can be produced by a compromise between the blowing and crosslinking reaction. The Tg of all the synthesized NIPU foams pertains to the range of 109–112°C, as observed from the DSC thermograms, hence, indicating a considerable degree of crosslinking. The LOI values, determined to study the fire resistance of the xylose-based NIPU foams, reported that the foam with highest citric acid content, i.e. RXC2, exhibited the greatest LOI value i.e. 26.9%, which is comparable to that of being a flame retardant (27%). The non-toxicity and simplicity of the methodology presented here overcome some of the major challenges in NIPU chemistry, including low reactivity, longer reaction times, high temperature, high pressure, and the requirement of expensive equipment. This environment-friendly and cost-effective xylose-based NIPU is well suitable as the replacement of conventional isocyanate-based PU foams, by the incorporation of various reinforcing fillers, for commercial scale manufacturing, as well as its significant potential status as "green material" for futuristic applications. As an outcome of this study, the futuristic approach can be the optimization studies focusing on more specific mechanical and physical properties. This novel robust approach will also open up new possibilities for the construction of more sustainable polyurethane foams and offers an alternative solution to the conventional isocyanate route, due to the outstanding accessibility and diversity of the major components of the formulations.

Acknowledgements The authors are thankful to the Department of Chemistry, Delhi Technological University for its infrastructural facilities and financial aid.

Author Contributions All the authors contributed to the study methodology, design and formal analysis. Material preparation, data collection and validation was performed by PS. The first draft of the manuscript was written by PS. The reviewing and editing of previous versions of the manuscript were carried out by RK and PS. The study conception, funding acquisition, supervision and project administration were performed by RK. All authors read and approved the final manuscript. Concept- Raminder Kaur Experimental Work- Pooja Singh Manuscript Writing-Pooja Singh Manuscript Correction-Raminder Kaur

Funding The authors are thankful to the Department of Chemistry, Delhi Technological University for its infrastructural facilities and financial aid.

Declarations

Conflict of interest The authors state that they have no known contending monetary or personal relationships that could sway the study reported in this paper.

References

1. Kaur R, Singh P, Tanwar S et al (2022) Assessment of bio-based polyurethanes: perspective on applications and bio-degradation. Macromol 2(3):284–314
2. Maisonneuve L, Lamarzelle O, Rix E et al (2019) Isocyanate-free routes to polyurethanes and poly (hydroxy Urethane) s. Chem Rev. https://doi.org/10.1021/acs.chemrev.5b00355
3. Agrawal A, Kaur R, Walia RS (2017) PU foam derived from renewable sources: perspective on properties enhancement: an overview. Eur Polym J 95:255–274. https://doi.org/10.1016/j.eurpolymj.2017.08.022
4. Agrawal A, Kaur R, Walia RS (2019) Development of vegetable oil-based conducting rigid PU foam. e-Polymers 19:411–420. https://doi.org/10.1515/epoly-2019-0042
5. Agrawal A, Kaur R, Walia R (2019) Flame retardancy of ceramic-based rigid polyurethane foam composites. J Appl Polym Sci 136:48250. https://doi.org/10.1002/app.48250
6. Kaur R, Kumar M (2013) Function of silicon oil in the castor oil based rigid polyurethane foams. J Polym Eng 33:875–880. https://doi.org/10.1515/polyeng-2013-0083
7. Kumar M, Kaur R (2017) Glass fiber reinforced rigid polyurethane foam: synthesis and characterization. e-Polymers. https://doi.org/10.1515/epoly-2017-0072

8. Agrawal A, Kaur R (2019) Effect of nano filler on the flammability of bio-based RPUF. Integr Ferroelectr 202:20–28. https://doi.org/10.1080/10584587.2019.1674820
9. Malik M, Kaur R (2018) Influence of aliphatic and aromatic isocyanates on the properties of poly(ether ester) polyol based PU adhesive system. Polym Eng Sci 58:112–117. https://doi.org/10.1002/pen.24537
10. Malik M, Kaur R (2018) Synthesis of NIPU by the carbonation of canola oil using highly efficient 5,10,15-tris(pentafluorophenyl) corrolato-manganese(III) complex as novel catalyst. Polym Adv Technol 29:1078–1085. https://doi.org/10.1002/pat.4219
11. Kathalewar MS, Joshi PB, Sabnis AS, Malshe VC (2013) Non-isocyanate polyurethanes: from chemistry to applications. RSC Adv 3:4110–4129. https://doi.org/10.1039/c2ra21938g
12. Rokicki G, Parzuchowski PG, Mazurek M (2015) Non-isocyanate polyurethanes: synthesis, properties, and applications. Polym Adv Technol 26:707–761. https://doi.org/10.1002/pat.3522
13. Maisonneuve L, Lamarzelle O, Rix E et al (2015) Isocyanate-free routes to polyurethanes and poly(hydroxy Urethane)s. Chem Rev 115:12407–12439. https://doi.org/10.1021/acs.chemrev.5b00355
14. Monie F, Grignard B, Thomassin JM et al (2020) Chemo- and regioselective additions of nucleophiles to cyclic carbonates for the preparation of self-blowing non-isocyanate polyurethane foams. Angew Chemie - Int Ed 59:17033–17041. https://doi.org/10.1002/anie.202006267
15. Xi X, EPizzi A, Gerardin C, et al (2019) Preparation and evaluation of glucose based non-isocyanate polyurethane self-blowing rigid foams. Polym (Basel) 11(11):1802. https://doi.org/10.3390/polym11111802
16. Bähr M, Mülhaupt R (2012) Linseed and soybean oil-based polyurethanes prepared via the non-isocyanate route and catalytic carbon dioxide conversion. Green Chem 14:483–489. https://doi.org/10.1039/c2gc16230j
17. Doley S, Dolui SK (2018) Solvent and catalyst-free synthesis of sunflower oil based polyurethane through non-isocyanate route and its coatings properties. Eur Polym J 102:161–168. https://doi.org/10.1016/j.eurpolymj.2018.03.030
18. Huang Y, Pang L, Wang H et al (2013) Synthesis and properties of UV-curable tung oil based resins via modification of Diels-Alder reaction, nonisocyanate polyurethane and acrylates. Prog Org Coatings 76:654–661. https://doi.org/10.1016/j.porgcoat.2012.12.005
19. Panchireddy S, Grignard B, Thomassin JM et al (2018) Catechol containing polyhydroxyurethanes as high-performance coatings and adhesives. ACS Sustain Chem Eng 6:14936–14944. https://doi.org/10.1021/acssuschemeng.8b03429
20. Chen X, Li J, Xi X et al (2020) Condensed tannin-glucose-based NIPU bio-foams of improved fire retardancy. Polym Degrad Stab 175:109121. https://doi.org/10.1016/j.polymdegradstab.2020.109121
21. Clark JH, Farmer TJ, Ingram IDV et al (2018) Renewable self-blowing non-isocyanate polyurethane foams from lysine and sorbitol. Eur J Org Chem 2018:4265–4271. https://doi.org/10.1002/ejoc.201800665
22. Vlcek T, Cabulis U, Holynska M (2022) Eco-friendlier and non-isocyanate-based polyurethane materials for space applications. CEAS Sp J. https://doi.org/10.1007/s12567-021-00408-x
23. Kim SR, Ha SJ, Wei N et al (2012) Simultaneous co-fermentation of mixed sugars: a promising strategy for producing cellulosic ethanol. Trends Biotechnol 30:274–282. https://doi.org/10.1016/j.tibtech.2012.01.005
24. Li HQ, Cao Y, Li XT et al (2014) Heterogeneous catalytic methoxycarbonylation of 1,6-hexanediamine by dimethyl carbonate to dimethylhexane-1,6-dicarbamate. Ind Eng Chem Res 53:626–634. https://doi.org/10.1021/ie4029945
25. Xi X, Pizzi A, Gerardin C, Du G (2019) Glucose-biobased non-isocyanate polyurethane rigid foams. J Renew Mater 7:301–312. https://doi.org/10.32604/jrm.2019.04174
26. Yao S, Wang BJ, Weng YM (2022) Preparation and characterization of mung bean starch edible films using citric acid as cross-linking agent. Food Packag Shelf Life 32:100845. https://doi.org/10.1016/j.fpsl.2022.100845
27. Garcia PS, Eiras Grossmann MV, Yamashita F et al (2011) Citric acid as multifunctional agent in blowing films of starch/PBAT. Quim Nova 34:1507–1510. https://doi.org/10.1590/S0100-40422011000900005
28. Chen X, Xi X, Pizzi A et al (2020) Preparation and characterization of condensed tannin non-isocyanate polyurethane (NIPU) rigid foams by ambient temperature blowing. Polym (Basel) 12(4):750. https://doi.org/10.3390/POLYM12040750
29. He X, Xu X, Wan Q et al (2019) Solvent- and catalyst-free synthesis, hybridization and characterization of biobased nonisocyanate polyurethane (NIPU). Polymers 11(6):1026. https://doi.org/10.3390/polym11061026
30. Lei W, Fang C, Zhou X et al (2018) Polyurethane elastomer composites reinforced with waste natural cellulosic fibers from office paper in thermal properties. Carbohydr Polym 197:385–394. https://doi.org/10.1016/j.carbpol.2018.06.036
31. Gandara M, Mulinari DR, Monticeli FM, Capri MR (2020) Sugarcane bagasse fibers reinforced in polyurethane for sorption of vegetal oil. J Nat Fibers 00:1–12. https://doi.org/10.1080/15440478.2019.1710653
32. Rokicki G, Piotrowska A (2002) A new route to polyurethanes from ethylene carbonate, diamines and diols. Polym (Guildf) 43:2927–2935. https://doi.org/10.1016/S0032-3861(02)00071-X
33. Xi X, Pizzi A, Delmotte L (2018) Isocyanate-free polyurethane coatings and adhesives from mono- and di-saccharides. Polym (Basel) 10:1–21. https://doi.org/10.3390/polym10040402
34. Błażek K, Kasprzyk P, Datta J (2020) Diamine derivatives of dimerized fatty acids and bio-based polyether polyol as sustainable platforms for the synthesis of non-isocyanate polyurethanes. Polym (Guildf) 205:12276810. https://doi.org/10.1016/j.polymer.2020.122768
35. Zhang L, Luo X, Qin Y, Li Y (2017) A novel 2,5-furandicarboxylic acid-based bis(cyclic carbonate) for the synthesis of biobased non-isocyanate polyurethanes. RSC Adv 7:37–46. https://doi.org/10.1039/c6ra25045a
36. Eltayeb M, Li SX, Okoye PU, Wang S (2021) Carbodiimide-assisted synthesis of high purity bis(cyclic carbonate) under atmospheric conditions for preparation of non-isocyanate polyurethane. J Polym Environ 29:1880–1893. https://doi.org/10.1007/s10924-020-01996-1
37. John MJ (2018) Flammability performance of biocomposites. Green Compos Automot Appl. https://doi.org/10.1016/B978-0-08-102177-4.00002-1
38. Liu J, Chen RQ, Xu YZ et al (2017) Resorcinol in high solid phenol – formaldehyde resins for foams production. J Appl Polym Sci 134:1–11. https://doi.org/10.1002/app.44881
39. Acuña P, Li Z, Santiago-Calvo M et al (2019) Influence of the characteristics of expandable graphite on the morphology, thermal properties, fire behaviour and compression performance of a rigid polyurethane foam. Polym (Basel). https://doi.org/10.3390/polym11010168
40. Liu J, Safronava N, Lyon RE et al (2018) Enhanced thermal property and flame retardancy via intramolecular 5-membered ring hydrogen bond-forming amide functional benzoxazine resins. Macromolecules 51:9982–9991. https://doi.org/10.1021/acs.macromol.8b02033
41. Mengal N, Syed U, Malik SA et al (2016) Citric acid based durable and sustainable flame retardant treatment for lyocell fabric.

Carbohydr Polym 153:78–88. https://doi.org/10.1016/j.carbpol.2016.07.074
42. Kaya M (2017) Super absorbent, light, and highly flame retardant cellulose-based aerogel crosslinked with citric acid. J Appl Polym Sci 134:1–9. https://doi.org/10.1002/app.45315

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.

Computers and Electrical Engineering 104 (2022) 108445

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

SwitchNet: A modular neural network for adaptive relation extraction☆

Hongyin Zhu [a,b,*], Prayag Tiwari [c,*], Yazhou Zhang [d], Deepak Gupta [e], Meshal Alharbi [f], Tri Gia Nguyen [g], Shahram Dehdashti [h,i]

[a] Inspur Electronic Information Industry Co., Ltd., Jinan 250101, China
[b] State Key Laboratory of High-end Server & Storage Technology, Beijing 100085, China
[c] School of Information Technology, Halmstad University, Sweden
[d] Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou 450002, China
[e] Maharaja Agrasen Institute of Technology, Delhi, India
[f] Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, P.O. Box 151, Alkharj 11942, Saudi Arabia
[g] FPT University, Danang 50509, Viet Nam
[h] School of Information Systems, Queensland University of Technology, Brisbane 4000, Australia
[i] Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany

ARTICLE INFO

Keywords:
Relation extraction
Modular neural network
Information flow
Joint optimization
Entity pair

ABSTRACT

This paper presents a portable toolkit, SwitchNet, for extracting relations from textual input. We summarize four data protocols for relation extraction tasks, including relation classification, relation extraction, triple extraction, and distant supervision relation extraction. This neural architecture is modular, so it can take as input data at different stages of the information extraction process (simple text, text and entities or entity pairs as relation candidates) and compute the rest of the process (named entity recognition and relation classification). We systematically design four information flows to integrate the above protocols by sharing network building blocks and switching different information flows. This framework can extract multiple triples (subject, predicate, object) in one pass. This framework enhances the use of relation classification models in end-to-end triple extraction by inferring pairs of entities of interest and using the shared representation mechanism.

1. Introduction

Knowledge Graph (KG) is a technology that can store relational facts in the form of triples. For example, "The Eiffel Tower is located in Paris" can be represented as a machine-readable triple (The Eiffel Tower, locatedIn, Paris) in a knowledge graph. Knowledge graphs are designed to describe various entities or concepts and their relations that exist in the real world, and ultimately constitute a huge semantic network, in which vertices represent entities or concepts, and edges are composed of attributes or relations, thus determining the semantic relations between entities is a critical task. Relation extraction aims to find the semantic relation between entities from the textual input. This task can be used to complement missing triples in the knowledge graph.

☆ This paper is for CAEE special section VSI-hci2. Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. Carlos Enrique Montenegro Marin.

* Corresponding author.
E-mail addresses: zhuhongyin@inspur.com (H. Zhu), prayag.tiwari@ieee.org (P. Tiwari), yzzhang@zzuli.edu.cn (Y. Zhang), deepakgupta@mait.ac.in (D. Gupta), Mg.alharbi@psau.edu.sa (M. Alharbi), tri@ieee.org (T.G. Nguyen), shahram.dehdashti@gmail.com (S. Dehdashti).

https://doi.org/10.1016/j.compeleceng.2022.108445
Received 2 October 2021; Received in revised form 11 October 2022; Accepted 20 October 2022
Available online 5 November 2022
0045-7906/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

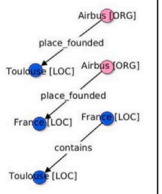
Tasks	Input	Output
Relation classification	[Airbus]ORG is based in [Toulouse]LOC, [France]LOC. (Airbus, relation, Toulouse) (Toulouse, relation, France) (Airbus, relation, France)	
Relation extraction	[Airbus]ORG is based in [Toulouse]LOC, [France]LOC. Relation PAD	
Triple extraction	Airbus is based in Toulouse, France. Entity PAD Relation PAD	

Fig. 1. Examples for text-bound RC, RE, and TE tasks.

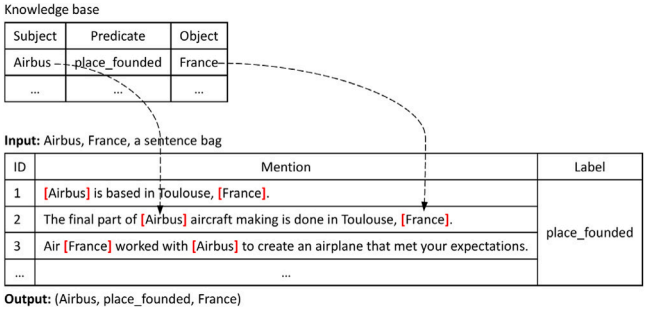


Fig. 2. Example for DS-RE.

Given the input data $\mathcal{X}$ (e.g., simple text and entity mentions), the relation extraction function $\mathcal{F}(\cdot)$ can automatically generate D (a set of triples). In practice, when the data protocol of $\mathcal{X}$ changes to $\hat{\mathcal{X}}$ (e.g., only simple text), existing model architectures cannot be easily transformed to $\hat{\mathcal{F}}(\cdot)$. Minor changes to the data protocol may lead to poor model performance and even make the model unusable. Designing different $\hat{\mathcal{F}}(\cdot)$ independently to adapt to changing data protocols is costly and inefficient. We introduce SwitchNet,[1] a portable toolkit, to resolve the problem of "reinventing the wheel". The main reason for the "wheel" problem lies in the independence and uninheritance of many studies. They implement the same relation extraction setup in different programming ways. Although the basic ideas of different works overlap, there are significant differences in programming implementation that are incompatible with each other. SwitchNet summarizes different protocols for relation extraction, divides the relation extraction process into multiple pluggable modules, and integrates different modules through neural network information flow. By choosing different neural network information flows, the model is adapted to changing data protocols.

This paper investigates the overlap and differences among four relation extraction task forms, including three text-bound tasks, i.e., relation classification (RC), relation extraction (RE) and triple extraction (TE), and the distant supervision relation extraction (DS-RE). Then, we propose a modular neural network that works towards real end-to-end relation extraction. Fig. 1 shows the input–output scheme, where the RC, RE, and TE forms provide the pairs of entity of interest (POEOI), the entity of interest (EOI), and no label respectively. EOI in the text is equivalent to the region of interest (ROI) [1] in a picture. ROI represents any particular portion of the image that seems important for the task. EOI means the important entity spans in the representation matrix of text. POEOI is composed of EOIs, and the holding relation can be inferred from this composition. POEOIs are usually two discontinuous spans, and it is uncertain whether and what kind of relation exists between the two spans.

"PAD" is a special symbol used to fill missing lengths, thus keeping the tensor dimensionality consistent, and allowing program parallelization. In addition to the benefits of PAD for parallelizing batch operations, it guarantees normal processing even if there are no entities or relations in the sample. Fig. 2 demonstrates the DS-RE task, where the input is a bag of sentences that contains the same entity pair. However, there is a noise problem in the data because not all sentences that mention the two entities express the target relation, e.g., the third sentence does not express "*place founded*" relation, so it is important to reduce the impact of wrong labeling. Traditional approaches for relation extraction rely heavily on engineered natural language processing (NLP) pipelines and external resources. Recent approaches use deep learning models to jointly extract entities and relations. These methods might need to design new task schemes or rely on syntactic parsing and separate model training. The new task schemes potentially limit the performance and versatility of their approaches. Ideally, we would like to overcome these two problems, eliminating the reliance on a new task scheme and conducting joint extraction in a modular way.

Existing joint extraction models mainly combine named entity recognition (NER) and RC from the subtask level, rarely considering the overlap and differences between relation extraction tasks. When the data protocol changes, most joint extraction

[1] https://nnntt.github.io/SwitchNet/

models cannot be easily applied to individual subtasks because innovative task schemes may not have an architecture to utilize all the input information or they may choose a simplified relation classification component. Different from them, we use the standard and effective neural components. Although task-specific neural networks have achieved good performance, it is unclear what the relationship between different RE tasks is and how to integrate them adaptively. For humans, different brain building blocks and neural pathways [2] form various cognitive functions. We propose a modular neural network that can share network building blocks by switching different information flows. We map subtasks/subfunctions to modules and use different strategies to selectively activate modules for processing information at different stages. Our main effort lies in systematically designing four information flows that are essential for model integration. In addition, not all triples in the knowledge base are static, e.g., "*place_lived*" and "*company*" relations, and our approach significantly reduces the problem complexity when dealing with multiple data sources for real-time KB updates.

The secondary problem of this framework is how to extract multiple triples in one pass in the TE protocol. This task consists of two subtasks, NER and RE. Prior works might train two subtasks separately, which leads to the drawback that the information between two subtasks cannot be fully exploited. This is because the NER and RE systems are independent of each other during model training, so they cannot be optimized together in the deep learning framework. In contrast, the differentiability of our multi-stage model is achieved through the POEOI inference and the shared representation mechanisms.

Not all context words contribute equally to the relation type, and some less informative words may produce noise. To overcome this issue, we use the attention mechanism [3] to assign different weights to different tokens for relation classification. The RC task can be further divided into single-relation extraction (SRE) and multiple-relations extraction (MRE). Our model can deal with MRE in one pass which resolves the inefficient multi-pass issue. Our model supports 1-of-n and multi-label classification forms corresponding to extracting one or different relations for two entities. Multi-label classification is a generalization of multi-class classification with no limit to the number of classes an instance can be assigned, it maps the input to a binary vector that assigns each element a label of 0 or 1, which naturally resolves the overlapping relation issue [4].

In practice, the advantage of using the proposed framework lies in its modular encapsulation of different functionalities. Task adaptation can be flexibly implemented through the combination of multiple modules and can be applied to a variety of different data protocols to meet various information extraction settings. From the perspective of theoretical research, it supports improving the model architecture in modules to achieve a partial or overall performance improvement, and can also further optimize the collaboration mechanism between different modules. We describe the usage of this toolkit in Section 5. We conduct experiments of TE, RE, RC, and DS-RE tasks on a large but noisy Riedel New York Times (NYT) dataset, a manually annotated SemEval-2018 dataset, a large TACRED dataset, and the NYT Large dataset respectively, as these datasets are well-known and represent different forms of relation extraction. Experimental results show the effectiveness of our approach in different data protocols. This model exhibits a complementary idea for the current model architecture design. In summary, the distinctive properties of this paper can be summarized below:

(i) We systematically design a modular neural network and four information flows to integrate four relation extraction tasks.

(ii) Our model integrates NER and RE subtasks through the POEOI inference and the shared representation mechanisms. We integrate DS-RE with the text-bound relation extraction protocols.

(iii) We conduct experiments on 4 relation extraction data protocols, and the experimental results show that our method can adapt to different data protocols and achieve performance improvements.

The paper is organized as follows: Section 2 presents a literature review on relation classification, joint entity relation extraction, remotely supervised entity extraction, and open information extraction. Section 3 summarizes the task definitions for the 4 relation extraction settings. Section 4 describes the proposed SwitchNet framework with novel neural network information flow and POEOI inference methods in detail. Section 5 describes the toolkit specification and design philosophy. Section 6 presents experimental results and analysis of different relation extraction task settings. Section 7 analyzes the factors that affect model performance. Finally, Section 8 presents the conclusions of this study.

2. Related work

2.1. Relation classification

For the fully-supervised methods, Zeng et al. [5] leverage convolutional deep neural networks to extract lexical and sentence-level features, and to specify entity pairs that should be assigned relation labels, they propose position features to encode the relative distances of words to target entity pairs in the convolutional neural network. Experimental results show that position features are crucial for relation classification. Xu et al. [6] propose a new neural network model, called long short term memory networks along shortest dependency paths (SDP-LSTM), for relation classification. Neural models are necessary to process information in a direction-sensitive manner, and they divide the SDP into two sub-paths, each from an entity to a common ancestor node. The model concatenates the features extracted along the two sub-paths for final classification. A highlight feature of this model is that the shortest dependency path preserves the most relevant information while eliminating irrelevant words in sentences; the multi-channel long short term memory networks (LSTM) network allows for efficient integration of information from heterogeneous sources through dependency paths. They utilize LSTM cells for information dissemination and integration.

2.2. Joint entity-relation extraction

Miwa and Bansal [7] propose an end-to-end model to extract relations between word sequences and entities on dependency structures. Models based on recurrent neural networks capture word sequences and relying on tree substructure information, entities and relations can be represented simultaneously in a single model. The model first detects entities, then extracts relations between detected entities using a single incrementally decoded neural network structure, and jointly updates model parameters with entity and relation labels. This allows models to collectively represent entities and relations within a single model using shared parameters. This model detects entities during training and uses entity information in relation extraction through entity pre-training and scheduled sampling which replaces the predicted labels with gold labels with a certain probability. Zheng et al. [4] propose a new tag scheme that can transform the joint entity-relation extraction task into a sequence labeling problem. Based on this tagging scheme, they investigated different end-to-end models to directly extract entities and their relations without identifying entities and relations separately. They also developed an end-to-end model with a bias loss function to adapt to new labels. This method can enhance the associations between related entities but still falls short in identifying overlapping relations. Zhang et al. [8] propose a graph convolutional network-based neural model for relation extraction, an extension of graph convolutional networks tailored for relation extraction, which can efficiently aggregate information of arbitrary structures in parallel. They propose path-centric pruning to improve the robustness of dependency models by removing irrelevant content without ignoring key information. To incorporate relevant information while maximally removing irrelevant content, they further applied a new pruning strategy to the input tree, allowing words to immediately surround the shortest path between two entities that may be related. Sun et al. [9] propose a graph convolutional network (GCN) based joint model to perform joint type inference for entity-relation extraction tasks. A binary relation classification task is introduced to explore the structure of entity relation bipartite graphs in a more efficient and interpretable manner. To address the joint type inference task, a novel GCN operation on entity-relation bipartite graphs is proposed. Compared with existing joint extraction methods, it provides a new way to explicitly capture interactions of multiple entity types and relation types in a sentence. By introducing the task of binary relation classification, the structure of entity-relation bipartite graphs can be exploited in a more efficient and interpretable way. Wang et al. [10] propose a scheme to simultaneously extract multiple relations and encode the input passages of the MRE task once. They let the self-attention layer know the positions of all entities in the input paragraph, building on a pre-trained self-attention model (Transformer). A structured prediction and entity-aware self-attention layer are proposed on top of BERT. Since all relations are computed at once, it can easily scale to larger datasets.

Bekoulis et al. [11] propose a joint neural model to simultaneously extract entities and relations from textual data. They model the entity recognition task using a conditional random field (CRF) layer, and model the relation extraction task as a multi-head selection problem (i.e., potentially identifying multiple relations for each entity). State-of-the-art performance is achieved in different domains (i.e. news, biomedical, real estate) and languages (i.e. English, Dutch) without relying on any human-crafted features or additional NLP tools. The downside is that this method uses a token to represent an entity. Han et al. [12] propose OpenNRE, an open and extensible toolkit for relation extraction. OpenNRE strikes a balance between system encapsulation, operational efficiency, model scalability, and ease of use. The toolkit prioritizes operational efficiency based on TensorFlow and PyTorch which support fast model training and validation. OpenNRE provides various functional RE modules based on TensorFlow and PyTorch to maintain sufficient modularity and extensibility so that new models can be easily incorporated into the framework. Online systems can also be used to satisfy real-time extraction without training and deployment. For developers whose goal is to train a custom model, they can quickly spin up an OpenNRE-based RE system without knowing too much technical detail and writing tedious glue code. This toolkit did not integrate different components in one model. Our model integrates different modules of the information extraction (IE) process, can jointly optimize the parameters of subtasks, and can extract multiple triples in one pass.

2.3. Distant supervision relation extraction

For the distantly-supervised methods [13], a training set is generated by aligning the relations in an existing knowledge base onto the free text. They address the problem of data annotation. Mintz et al. [13] propose a distant supervision algorithm capable of extracting high-precision patterns for a large number of relations. The combination of grammatical and lexical features provides better performance than separate feature sets. Their algorithms can use large amounts of unlabeled data, and a pair of entities may appear multiple times in the test set. Their experiments use Freebase, a large semantic database of thousands of relations, to provide distant supervision. Their algorithm combines the advantages of supervised IE (combining 400,000 noise pattern features in a probabilistic classifier) and unsupervised IE (extracting large numbers of relations from large corpora in any domain). Existing methods [14] adopt multi-instance learning to alleviate the wrong labeling problem. However, the above methods rely heavily on the quality of featuring engineering. Zeng et al. propose the piece-wise max-pooling strategy over multi-instance learning. This approach can automatically learn features through deep learning models. Lin et al. [15] propose sentence-level attention to alleviate the wrong labeling problem, but the sentence representation needs to be further improved. Reinforcement learning (RL) is also applied to enhance the instance selection process by proposing innovative schemes for the episode, state, action, reward, and optimization, but the stability and computational efficiency of the model are challenges. The DS-RE methods usually compare the precision and recall curves and often put precision before the recall, while the supervised methods obtain determined scores. Our research shows the relationship between text-bound RE and DS-RE.

2.4. Open information extraction

For the Open IE [16], it aims to generate triples from textual input without requiring a pre-specified vocabulary. The subjects and objects extracted by Open IE are fragments in the text, not necessarily entities, such as values, dates, prices, etc. The predicates are also freer than the predicates of our triple extraction task. Open relation extraction methods can be divided into two categories: (1) Relation extraction using explicit rules. Relation extraction is performed by building a combination of methods including dependency parsing, part-of-speech tagging, entity linking, and rule templates. (2) Relation extraction using implicit rules. Supervised training of neural network models on a large human-annotated corpus can learn implicit rules for general triple extraction. However, such methods may extract some meaningless triples and require further filtering and completion of the results.

3. Task definition

Let $\mathcal{X} = [w_1, \ldots, w_n]$ denote a sequence of text, where w_i is the ith token. A subject entity s and an object entity o denote two non-overlapping consecutive spans: $\mathcal{X}_s = [w_{s1}, w_{s1+1} ..., w_{s2}]$ and $\mathcal{X}_o = [w_{o1}, w_{o1+1} ..., w_{o2}]$. R is the set of predefined relations. We formalize the RC, RE, TE, and DS-RE tasks as follows.

- **RC** Given the textual input $\mathcal{X}$ and the positions of all (s, o) pairs, the goal is to predict the corresponding relation $r \in \mathcal{R}$ that holds between each (s, o) pair or no relation otherwise.
- **RE** Given the textual input $\mathcal{X}$ and the positions of all entity spans $\mathcal{E}_e = \{[w_{e11}, w_{e11+1}, \ldots, w_{e12}], \ldots, [w_{ek1}, w_{ek1+1}, \ldots, w_{ek2}]\}$, the goal is to predict the positions of all (s, o) pairs and the relation $r \in \mathcal{R}$ for each (s, o).
- **TE** It only takes as input text $\mathcal{X}$, and the goal is the same as the RE task. This task is equivalent to NER plus RE.
- **DS-RE** It takes as input two entities $(e1, e2)$ and a bag of sentences (C) that mention $e1$ and $e2$ and predict the relation $r \in \mathcal{R}$ (sometimes multiple-relations) that holds between $e1$ and $e2$, or no relation otherwise. This method assumes that any of the relations is supported by at least one sentence in C. Using distant supervision, relations in KBs can be aligned with plain text to produce bags of relation mentions for model training.

4. Approach

Fig. 3 shows the architecture of SwitchNet. This framework mainly contains five modules, i.e., the encoder, the NER module, the RC module, the POEOI inference module, and the DS-RE module. These modules are connected by the designed information flows. Different information flows correspond to different input data protocols. The input mainly consists of three elements, including simple text and two optional elements, the entity annotation (EOI) or the relation annotation (POEOI). In the following, we describe the model architecture, the information flows, and the training methods.

4.1. Model architecture

We divide the model into different modules and then integrate the modules together through neural network information flow to achieve specific functions. In this section, we first introduce the working mechanism of neural network information flow and then introduce the composition of each module.

4.1.1. Neural network information flows

Different tasks can be achieved by switching different information flows. We will describe the mechanism of these information flows to implement adaptive relation extraction, as shown in Fig. 3. We explain the functionality, the location, and the collaboration of the modules in various relation extraction pathways. By controlling the information flow, this model can selectively activate specific operations and features at different stages of information extraction. We use $\mathcal{P}ij$ to indicate the information flow, that is, the functional connection between modules in this neural network. Some information flows may contain specific data operations. We describe the information flows as follows.

(1) When we process the RC task, this model activates the pathway $[\mathcal{P}11, \mathcal{P}12] \rightarrow \mathcal{P}13$, as detailed in Section 4.1.4. The entity relation annotation directly indicates the POEOI through $\mathcal{P}11$. Then, the input to the RC layer comes from the path $\mathcal{P}12 \rightarrow \mathcal{P}13$ which connects the contextual representation and the POEOI modeling directly.

(2) When we process the RE task, this model activates the pathway $[\mathcal{P}21, \mathcal{P}22] \rightarrow \mathcal{P}23$, as detailed in Section 4.1.5. The input provides the pre-annotated EOIs through $\mathcal{P}21$, and then the function of $\Psi(\cdot)$ activates $\mathcal{P}23$ via the generated pair-wise combination of EOIs (POEOIs). Then, the path $\mathcal{P}22 \rightarrow \mathcal{P}23$ connects the contextual representation and the POEOI modeling.

(3) When we process the TE task, this model uses the pathway $\mathcal{P}31 \rightarrow [\mathcal{P}32, \mathcal{P}34] \rightarrow \mathcal{P}33$, as detailed in Section 4.1.5. This model first predicts/infers EOIs via $\mathcal{P}31$ which contains the function of $\Phi(\mathcal{F}_{ner}(\cdot))$. Then, the path $\mathcal{P}32 \rightarrow \mathcal{P}33$ indirectly generates POEOIs by the function of $\Psi(\cdot)$. Then, the path $\mathcal{P}34 \rightarrow \mathcal{P}33$ connects contextual representation and the POEOIs.

(4) When we process the DS-RE task, this model uses the pathway *Distant supervision*$\rightarrow \{\mathcal{P}40 \rightarrow [\mathcal{P}41, \mathcal{P}42] \rightarrow \mathcal{P}43\} \overset{\times k}{\rightarrow} \mathcal{P}44$, as detailed in Section 4.1.6. Suppose there is a knowledge base, the distant supervision technique is used to automatically align the relations in KBs to free text. Each relation will be mapped to a bag of sentences. Then the model takes as input each sentence

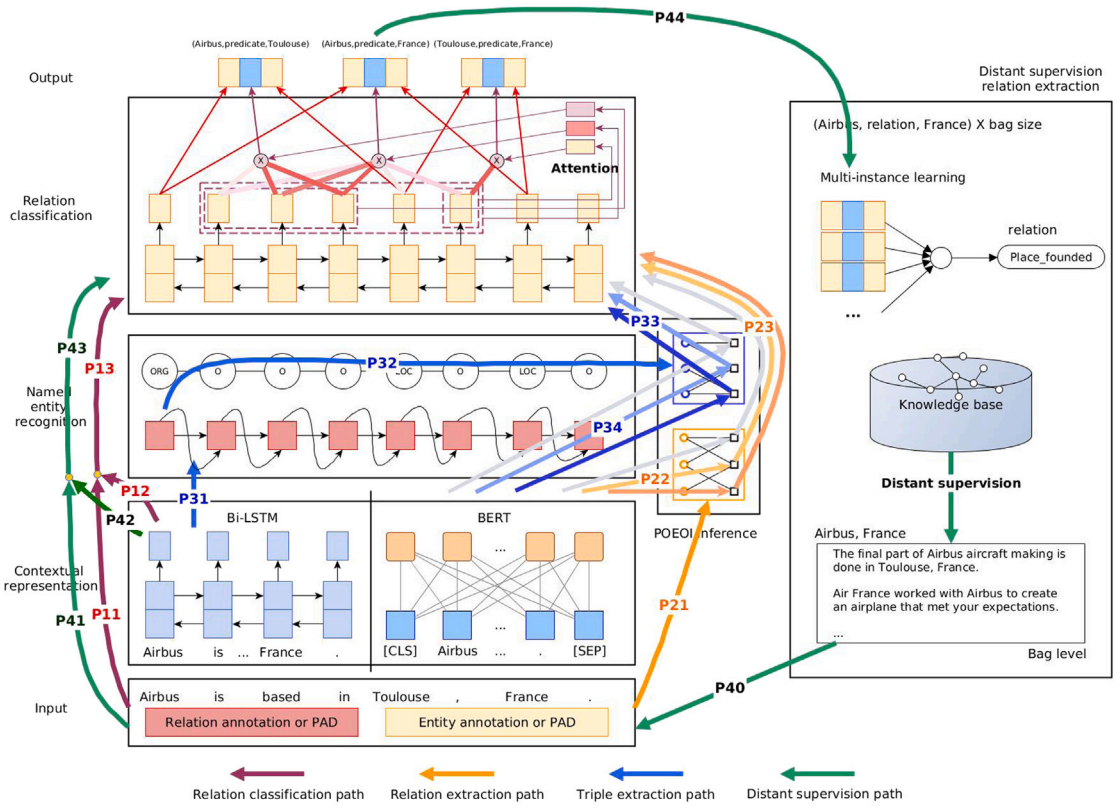


Fig. 3. Neural network architecture overview.

bag via path $\mathcal{P}40$. For each sentence, the information flow that generates the sentence representation is similar to the RC task, i.e., $[\mathcal{P}41, \mathcal{P}42] \rightarrow \mathcal{P}43$. Each sentence only generates one vector. $\overset{\times k}{\rightarrow}$ means this model parallel process k sentences in a bag. Then the model uses MIL to compose a vector representation of the bag.

In the above four pathways, our network adaptively models relation representation based on the POEOIs. The difference among the above RE pathways is whether there are EOI inference and POEOI inference to predict the positions of EOIs and POEOIs as relation candidates during model training.

4.1.2. Encoder

This module encodes the textual input into contextual representation. The neural network can be bidirectional LSTM (Bi-LSTM) or pre-trained language models (e.g., BERT) [17]. Let $[w_1, w_2, \ldots, w_m]$ be the sequence of text with m tokens. The encoder generates the hidden states $[h_1, h_2, \ldots, h_m]$ which can be used as the shared representation for different modules, where $h_i = Encoder(w_i)$. When we use the Bi-LSTM encoder, the $h_i = \overrightarrow{h}_i \oplus \overleftarrow{h}_i$ is generated by the following equation.

$$\overrightarrow{h}_i = \text{LSTM}(\overrightarrow{h}_{i-1}, x_i) \tag{1}$$

$$\overleftarrow{h}_i = \text{LSTM}(\overleftarrow{h}_{i-1}, x_i) \tag{2}$$

where $\overrightarrow{h}_0$ and $\overleftarrow{h}_0$ are initialized to zero vectors. $x_i = v_i \oplus t_i$ is the word representation where $\oplus$ is the concatenation operation. The word representation is composed of the word embedding and character-level representation t_i encoded by a convolutional neural network (CNN). Word embeddings [18] refer to the technique of encoding the meaning of words in the form of low-dimensional dense vectors, so that words that are closer in the vector space are also semantically similar. Character-level embeddings [19] learn vector representations of characters, which are then aggregated through a 1D convolutional neural network or LSTM neural network to generate an overall representation of the word.

Pre-trained language models are first pre-trained on a large-scale corpus based on self-supervised tasks and then fine-tuned for downstream tasks. BERT is pre-trained with the masked language model task and the next sentence prediction task. When we use the BERT as an encoder, [CLS] is a special symbol added in front of each input sample, the hidden state corresponding to this token is used as the aggregated sequence representation for the classification task, [SEP] is a special separator token used to separate discontinuous token sequences (e.g., [CLS] question [SEP] answer [SEP]). An input sentence is denoted as

$[[\text{CLS}], w_1, w_2, \ldots, w_m, [\text{SEP}]]$. BERT consists of multiple layers of transformer encoders that generate contextual representations through the multi-head self-attention mechanism.

$$[h_{cls}, h_1, h_2, \ldots, h_m, h_{sep}] = \mathcal{F}_{bert}([[\text{CLS}], w_1, w_2, \ldots, w_m, [\text{SEP}]]) \tag{3}$$

where $\mathcal{F}_{bert}(\cdot)$ denotes the architecture defined in BERT [17]. We only use the hidden representation of each token $h_i (i = 1, 2, \ldots, m)$. The core idea of the transformer block is a self-attention mechanism through which contextual representations can be generated, as shown in Eq. (4).

$$h_{att} = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{4}$$

The above equation shows the scaled dot-product attention where Q, K, and V are the query, key, and value matrices. The input consists of queries and keys of dimension d_k, and values of dimension d_v.

4.1.3. Named entity recognition

This module aims to discover the EOIs from simple text. BiLSTM-CRF [20] model achieves good performance, and this task can be modeled as finding an optimal label sequence $\mathbf{y}^*$ that maximizes the conditional probability $p(\mathbf{y}|\mathbf{h}; \theta)$ as shown below.

$$\mathbf{y}^* = \arg \max_{\mathbf{y} \in \mathcal{Y}(\mathbf{h})} p(\mathbf{y}|\mathbf{h}; \theta) \tag{5}$$

where θ is the model parameter. $\mathbf{h} = [h_1, \ldots, h_m]$ is the output sequence of the encoder and $\mathbf{y} = [y_1, \ldots, y_m]$ is the label sequence. $\mathcal{Y}(\mathbf{h})$ denotes the set of possible label sequences for $\mathbf{h}$.

$$p(y|\mathbf{h}; W, b) = \frac{\prod_{i=1}^{n} \exp(W^T_{y_{i-1} y_i} h_i + b_{y_{i-1} y_i})}{\sum_{y' \in \mathcal{Y}(h)} \prod_{i=1}^{n} \exp(W^T_{y'_{i-1} y'_i} h_i + b_{y'_{i-1} y'_i})} \tag{6}$$

where $\{[h_i, y_i]\}, i = 1, 2...n$ represents the vector of the ith word h_i and the ith label y_i in the input sequence respectively. $\mathcal{Y}(\mathbf{h})$ denotes all the possible label sequences for the input sequence $\mathbf{h}$. W and b are weight matrix and bias vector, in which W_{y_{i-1}, y_i} and b_{y_{i-1}, y_i} are the weight vector and bias corresponding to the successive labels (y_{i-1}, y_i). $p(y|\mathbf{h}; W, b)$ is the probability of generating this tag sequence over all possible tag sequences.

LSTM-based representations and linear layers can be directly used for class labeling of words, and we introduce the following improvements to this method. In addition to the current input, the LSTM-RNN takes the hidden state of the previous step as input. $\overrightarrow{h}_i^{(ner)} = \text{LSTM}(\overrightarrow{h}_{i-1}^{(ner)}, h_i \oplus \overrightarrow{h}_{i-1}^{(ner)})$ where $\overrightarrow{h}_i^{(ner)}$ is the representation of the ith token in the NER module. The superscript (ner) is used for the notation of the NER module. $\overrightarrow{h}_0^{(ner)}$ is initialized to a zero vector. This representation is used for classification.

$$p(y_i|y_{1:i-1}) = softmax(W \overrightarrow{h}_i^{(ner)} + b) \tag{7}$$

where W and b are weight and bias parameters.

4.1.4. Relation classification

Single-relation extraction For SRE, suppose we have the position of a POEOI (s and o), this module generates the triple vector $v = s^{(rc)} \oplus h_r \oplus o^{(rc)}$, where the superscript (rc) denotes the RC module. The subject $s^{(rc)} \in \mathbb{R}^\mu$ or object $o^{(rc)} \in \mathbb{R}^\mu$ is represented by the average/max pooling of the corresponding hidden states $\mathbf{h}^{(rc)}$. The hidden states are encoded by a Bi-LSTM.

$$\overrightarrow{h}_i^{(rc)} = \text{LSTM}(\overrightarrow{h}_{i-1}^{(rc)}, h_i) \tag{8}$$

$$\overleftarrow{h}_i^{(rc)} = \text{LSTM}(\overleftarrow{h}_{i-1}^{(rc)}, h_i) \tag{9}$$

where $h_i^{(rc)} = \overrightarrow{h}_i^{(rc)} \oplus \overleftarrow{h}_i^{(rc)}$. The predicate h_r is calculated by the attention mechanism. Let $\mathcal{F}_{att} : \mathbb{R}^{n \times \mu} \rightarrow \mathbb{R}^\mu$ denote an attention function that maps n input vectors to a relation vector h_r. μ is the vector dimension. $h_0^{(rc)}$ is initialized to a zero vector.

We introduce the POEOI modeling and the position-aware POEOI modeling. For the POEOI modeling, the most relevant portion of text to determine the relation type is usually the one contained between and including the entities [21]. Not all words contribute equally to the relation type, so we use the weighted aggregation of informative hidden states between two entities to represent the relation. We use a self-attention method [22] to extract informative tokens. This method first captures the main context information of relation by the average-pooling of hidden states as shown in Eq. (10).

$$q = \frac{1}{c_2 - c_1} \sum_{i=c_1}^{c_2} W_Q h_i^{(rc)} \tag{10}$$

where $W_Q \in \mathbb{R}^{\mu \times \mu}$ is model parameter. $h_i^{(rc)} \in \mathbb{R}^\mu$ is the ith hidden state of the relation context. c_1 and $c_2 \in [1, m]$ are the start and end indices of relation context respectively. Then, this method measures the token weights by considering their importance in the context and obtains the normalized importance weights through a softmax function.

$$\alpha_i = \frac{\exp(q^T W_K h_i^{(rc)})}{\sum_{j=c_1}^{c_2} \exp(q^T W_K h_j^{(rc)})} \tag{11}$$

where $h_r \in \mathbb{R}^{\mu}$ is the relation representation. $W_K \in \mathbb{R}^{\mu \times \mu}$ are model parameters.

The relation representation is obtained by calculating the weighted sum of the hidden states. The self-attention mechanism outputs as many vectors as it inputs as Eq. (4). To obtain the representation of text span, inspired by Hierarchical attention networks (HAN) [23], we use a variant of the self-attention mechanism, where the output vector is a weighted aggregation of contextual representations.

$$h_r = \sum_{i=c_1}^{c_2} \alpha_i \cdot W_V h_i^{(rc)} \tag{12}$$

where $W_V \in \mathbb{R}^{\mu \times \mu}$ are model parameters. This is also known as the single-head attention which is highly extensible, because adding more heads will generate multi-head attention.[2]

$$h_r = h_{r,1} \oplus h_{r,2} \ldots \oplus h_{r,n} \tag{13}$$

where n is the number of heads. $h_{r,i}$ represents the ith head.

For the position-aware POEOI modeling, the input to RC module becomes $h_i^{(s,o)} = h_i \oplus p_i^{(s)} \oplus p_i^{(o)}$ where (s, o) is a pair of entities. p_i is the position embedding. Position embedding [5] is a way to encode relative position from current word to s and o. Position embedding is a technique for representing the spatial positional relationship of words in a sentence. Relation classification is a complex task, and structural features in sentences are important, such as the spatial location relationship between context words and entities. Such structural information cannot be captured by word embeddings alone, so position embeddings are needed to obtain spatial positions. For example, as shown in Fig. 1, the relative position of "based" to "Airbus" (s) and "France" (o) are -2 and 4 respectively. Then, the relative position is encoded using a fixed-length vector that is randomly initialized and jointly trained. This feature explicitly encodes the position of POEOI, so it allows the model to consider the entire token sequence. The single-head position-aware POEOI modeling is similar to the position-aware attention LSTM (PA-LSTM) model to some extent. The main difference is that our model can perform MRE in one pass.

This module supports the above two methods. The former is more computationally efficient but may lose some information. The latter can consider more contexts, but add more computation. The two methods achieve comparable results in the TE and RE tasks because the predicted POEOIs is not always correct. Using position embedding in these tasks may increase noise. When the input indicates the POEOI, we prefer to use the latter method because this allows the model to consider more context information. Finally, the triple vector is input into a single layer neural network for classification.

$$p(y^{(rc)}) = softmax(W^{(rc)} v + b^{(rc)}) \tag{14}$$

where $W^{(rc)} \in \mathbb{R}^{\tau \times \rho}$ and $b^{(rc)} \in \mathbb{R}^{\tau}$ are weight and bias parameters. ρ and τ are the input dimension and class number respectively.

Multiple-relations extraction For MRE, suppose we have multiple POEOIs. We use the above process to model each triple and stack them to generate the triple matrix A, i.e., $[v_1, \ldots, v_u]$ where u is the number of triple candidates. Then the model classifies all relation candidates through a softmax function on columns.

$$p(Y^{(rc)}) = softmax(W^{(rc)} A + b^{(rc)}) \tag{15}$$

where $W^{(rc)} \in \mathbb{R}^{\tau \times \rho}$ and $b^{(rc)} \in \mathbb{R}^{\tau}$ are weight and bias parameters. This setting allows the model to update parameters from more perspectives, and each sample is encoded only once.

4.1.5. Inference for pairs of entities of interest

When the POEOI is uncertain, each entity phrase can be a subject or an object, so we consider the relation type and direction or no relation. The output of NER subtask is completely different from the input of RE subtask. We propose the POEOI inference scheme to integrate NER and RE subtasks in the joint optimization process by predicting the POEOIs as below.

$$A = \mathcal{F}_{rc}(\mathbf{h}, \Psi(\Phi(\mathcal{F}_{ner}(\mathbf{h})))) \tag{16}$$

where A is the triple matrix. The inference of this framework means "prediction" instead of traditional pipeline reasoning. Each triple can be predicted as a specific relation type (including relation direction) or NA. $\mathcal{F}_{ner}(\cdot)$ is the NER function that predicts the label sequence. $\Phi(\cdot)$ denotes the function that identifies the EOIs. This can be seen as an action selection process for discovering EOIs. $\Psi(\cdot)$ is the function that infers POEOIs as relation candidates. The POEOIs are generated by using all possible pair-wise combinations of EOIs. POEOI inference is an operation that generates the pair-wise entity combination. For example, for the three entities A, B, and C, after this operation, three combinations of AB, AC, and BC will be formed. For n entities, C_n^2 combinations will be generated.

$$\Psi(\{A, B, C\}) = \{AB, AC, BC\} \tag{17}$$

The different number of relation candidates leads to different learning speed and performance. Extensive experiments will be discussed in Section 7.2. $\mathcal{F}_{rc}(\cdot)$ denotes the function of the relation classification module that encodes triple representation. Eq. (15) can be seen as an action selection process for predicting the relation type. $\mathcal{F}_{ner}(\cdot)$ and $\mathcal{F}_{rc}(\cdot)$ are cascaded subtasks, and we connect

[2] Multi-head attention performs nearly the same as the single-head attention in text-bound RE. For DS-RE, it achieves significant improvement.

them using the POEOI inference and the shared representation mechanisms. The inference function $\Psi(\Phi(\mathcal{F}_{ner}(\cdot)))$ generates a heuristic to add regularization to the network, similar to the effect of dropout operations. This heuristic selects and emphasizes the locally optimal features [24]. However, the function $\Psi(\cdot)$ is a discrete action selection process, so there is a sort of information loss under the gradient-based optimization framework. It is difficult to solve the discrete optimization problems [25] using Stochastic gradient descent (SGD) like optimization algorithms. Other non-SGD optimization mechanisms (e.g., genetic algorithms, ranking mechanisms, etc.) can be explored to augment the model optimization.

We integrate the process of discovering EOIs and modeling POEOIs. We optimize the model at the end of the forward pass. The triple matrix contains the memory of all possible relation candidates. Finally, we predict the relations through Eq. (15). This method works towards minimizing the loss function based on the ground truth, so it is supervised learning with inference. The POEOI inference scheme makes the RC module widely compatible with other SRE models, e.g., PCNN and the PA-LSTM. This expansion enhances the use of existing SRE/MRE methods to make them conduct joint entity and relation extraction.

We introduce the techniques for controlling the ratio of candidate relations in detail in Section 7.2, and this paper does not control the ratio of entities, and related work can be further studied in the future. There are two ways to implement batch processing. (1) Favorable way for program parallelization. It is to first determine the coordinates of each entity, then generate a combination of coordinates, and obtain slices of tensors according to the coordinates (for example, using functions such as torch.index_select() or tensorflow.gather()), which is realized by carefully designed tensor dimension expansion and reshape. (2) Unfavorable way of program parallelization. This method is relatively simple and is implemented through a "for loop". The entity vector set of each sample is processed separately, combined according to the entity pair, and finally all the relation candidate vectors are concatenated, and then the gradient of the batch samples is accumulated by gradient accumulation. To reduce the computational complexity, we abandon the POEOIs that have a distance longer than 50 words.

4.1.6. Distant supervision relation extraction

The distant supervision assumption is that if two entities preserve a relation in a KB, then all sentences that mention the two entities express this relation. This assumption improves the efficiency of automatic data annotation but introduces noise into the training data. In order to alleviate the wrong labeling problem, multi-instance learning (MIL) is a reasonable choice for DS-RE. We further describe how to expand the RC information flow to the DS-RE task.

Suppose there is a bag C containing n sentences, each of which mentions $e1$ and $e2$, i.e. $C = \{s_1, s_2, \ldots, s_n\}$. This model takes as input C and generates the representation $V = \{A_1, A_2, \ldots, A_n\}$. Note that each A_i contains only one POEOI. This task can be seen as n parallel RC processes. Sentence-level attention approach [15] is reasonable and achieves high performance to de-emphasize the noisy sentences, so we use this method as the base. This method aims to calculate the bag representation by weighted aggregation of sentences as Eq. (18). The bag representation is derived from all sentences that mention the two entities.

$$c = \sum_i \beta_i A_i \tag{18}$$

where β_i is the weight of each sentence vector A_i as shown in Eq. (19). This weight is calculated by the softmax function. A_i is calculated by the word-level attention mechanism in the RC module.

$$\beta_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)} \tag{19}$$

where e_i is calculated by Eq. (20) which reflects how well the input sentence s_i matches the relation r. We use the bilinear form as follows.

$$e_i = A_i^T M \mathbf{r} \tag{20}$$

where M is a weighted diagonal matrix, and $\mathbf{r}$ is the query vector associated with a relation r. $\mathbf{o}$ is the output of this module and it represents the scores for all relation types, as shown below.

$$\mathbf{o} = Oc + b^{(ds)} \tag{21}$$

where O is the representation matrix of relations. $b^{(ds)}$ is a bias vector. Finally, we compute the conditional probability of $p(y_i|C;\theta)$ through a softmax function.

$$p(y_i|C;\theta) = \frac{\exp(o_i)}{\sum_{j=1}^{k} \exp(o_j)} \tag{22}$$

where k is the total number of all relations. In the RC module, the single/multi-head attention first extracts informative tokens for instance representation. Then, the MIL selects high-quality instances for bag representation. The entire process constructs hierarchical attention.

4.2. Training methods

Different tasks require different training objectives because the modules have different collaborative methods in various relation extraction pathways. This framework needs to optimize the model to adapt to the corresponding tasks. This framework supports all

the training objectives for different tasks. For the TE and RE tasks, we can optimize the combined objective function at the end of the forward pass.

$$\min_{\Theta} \mathcal{L} = \sum_{i=1}^{|\mathbb{D}|} (\mathcal{L}_i^{(ner)} + \lambda \mathcal{L}_i^{(re)}) \tag{23}$$

where $\mathcal{L}_i^{(ner)}$ and $\mathcal{L}_i^{(re)}$ are the objectives of two subtasks respectively. $|\mathbb{D}|$ is the dataset size. Θ is the model parameter. λ is a hyper-parameter to weight the influence of RE subtask. The training loss is to sum the deviation of the relation prediction and the deviation of entity prediction. We adopt the negative log-likelihood as the objective, as shown below.

$$\mathcal{L}^{(ner)} = -\sum_{j} \log p^{(ner)}(y_j^{(ner)}|\mathbf{x};\Theta) \tag{24}$$

$$\mathcal{L}^{(re)} = -\sum_{k} \log p^{(re)}(y_j^{(re)}|\mathbf{x};\Theta) \tag{25}$$

where $p^{(ner)}(\cdot)$ and $p^{(re)}(\cdot)$ represent the probabilities of true class of NER and RE subtasks respectively. $\mathbf{x}$ denotes the input sequence. j and k denote the indices of entity and relation respectively. We can train the two tasks jointly, and they interact with each other through the encoder. The POEOI inference and the shared representation mechanisms enable the cascaded subtasks to learn together.

By learning the multi-label classification model, we can also predict different relations between two entities. This model will perform binary classification for each relation type separately. The objective of $\mathcal{L}^{(re)}$ becomes the log loss in Eq. (26).

$$\mathcal{L}^{(re)} = -\sum_{k}\sum_{c} y_{k,c}^{(re)} \log p^{(re)}(y_{k,c}^{(re)}|\mathbf{x};\Theta) + (1 - y_{k,c}^{(re)}) \log(1 - p^{(re)}(y_{k,c}^{(re)}|\mathbf{x};\Theta)) \tag{26}$$

where k and c are the triple index and class index respectively. The output gives a probability distribution over all labels, and a threshold is used to determine the final subset of labels.

For RC task, we already know the POEOIs in the sentence, so we can model POEOIs directly. The training objective is shown blew.

$$\min_{\Theta} \mathcal{L} = \sum_{i}^{|\mathbb{D}|} \mathcal{L}_i^{(re)} \tag{27}$$

For DS-RE task, each sentence bag is used to model one POEOI. The training objective is shown blew.

$$\min_{\Theta} \mathcal{L} = -\sum_{i}^{|\mathbb{K}|} \log p(y_i^{(re)}|C_i;\Theta) \tag{28}$$

where $|\mathbb{K}|$ is the number of sentence bags. C_i is a bag of sentences. By selecting different information flows and corresponding training objectives, we can train different functional neural networks for relation extraction tasks.

The relation extraction task can be viewed as solving a mathematical problem. Indicating all entities in advance can reduce the number of independent variables, making the task more deterministic, which is also in line with the laws of solving mathematical problems. The final result is obtained by continuously solving the unknowns. In cases where entities are not indicated, intermediate results need to be inferred from existing information during model training, which can lead to noise. The limitation of this framework is that first of all, the programming of the whole framework is very difficult. Although this version implements basic functions, there is still a large room for optimization in performance. It can be achieved by using better design patterns and programming tricks. When the framework cannot extract entities and contains nested entities, the relation extraction module will be deactivated. A low-quality dataset makes the model learn the wrong rules. Noise in the training set interferes with the model getting a high-quality representation. These noises are harmful, but the model itself is unaware of these mistakes. For relation extraction tasks, it relies on hidden state representations of deep learning models, so high-quality representations are critical. Training a model on low-quality data can cause the model to fail to track the cause of the error, thereby exacerbating the black-box nature of the model.

Integrating existing models is an important function, and in fact, we provide a new perspective, rethinking the contributions of previous research. By reorganizing and improving previous methods, making them suitable for new relation extraction tasks. The data format is a very important issue, especially when agents want to achieve big data intelligence, they need to face various changes and uncertainties in the network environment. When faced with special data that needs to be processed, a single approach will fail, allowing the agent to adapt to a variety of complex data. The work of this paper is an effort to solve this important problem. Joint extraction methods often change the form of the task, which leads developers to lose sight of a standardized entity recognition and relation extraction pipeline. For the two-stage relation extraction pipeline, although the standardized entity recognition and relation extraction pipelines are retained, it may be difficult to jointly optimize the subtasks. Few methods can clearly conform to standardized named entity recognition and relation extraction pipelines while enabling joint optimization of subtasks, and our method addresses this problem. This model is potential to be expanded to consider the longer context in other tasks, e.g., coreference resolution and document-level relation extraction.

5. Toolkit instructions

The working mode of this toolkit is determined by the data protocol. This toolkit can automatically switch to the corresponding task state according to the format of the input data without additional configuration. The input data is stored in JSON format, and

Table 1
Description of input data fields.

Fields	Description
token	Token sequence of text.
pos	Part-of-speech tagging sequence of text.
dep	Dependency parsing tree of text, including the edge types.
relation	Ground truth triple labels, expressed in the form of $[(relation_k, (e_i^{start}, e_i^{end}), (e_j^{start}, e_j^{end})), \dots]$
ner	Ground truth NER label sequences of text.
ner-feature	NER tags extracted by using other toolkits such as Stanford CoreNLP, Spacy, etc.
text	The original input text.
relation-set	List of all predefined relation types.
entity-set	List of all predefined entity types.

the main data fields include "`token, pos, dep, relation, ner, ner-feature, text, relation-set, entity-set`". The description of these fields is shown in Table 1. We chose the JSON format because it is a common data transfer format for Internet data, including the response data returned after an agent requests a service.

The working mode is mainly determined by `ner` and `relation` fields. For model prediction, when the `ner` field is `None`, it means that the data protocol does not provide entity tags, so it is a TE task. When the `ner` field contains entity labels, the RE or RC tasks will be executed. When the `ner` labels contain only two entities, the model works in the RC mode. When the `ner` labels contain three or more entities, the model works in the RE mode. When the input data fields contain `head` and `tail`, the model works in the DS-RE mode. For model training, when the `ner` labels contain three or more entities, the model is trained in the TE mode.

In the model configuration, the main parameters related to word vectors are `word_representation` and `fine_tune`. The `word_representation` parameter can be set to the path of the context-independent word vectors, e.g. GloVe, FastText, etc. or the path of the BERT model. Setting the `fine_tune` parameter to True means that the word vectors or the BERT will be fine-tuned during the training process, and vice versa. In addition, we can choose whether to use the CRF model to decode NER results. The `head_number` parameter can be used to set the number of heads used for relation representation. For the DS-RE model, we mainly use the OpenNRE [12] framework as the baseline and integrate our sentence-level multi-head relation representation.

6. Experiments

6.1. Experimental settings

6.1.1. Dataset

NYT dataset[3] is developed by [14] by aligning Freebase relations with New York Times news articles. Provided for public access from March 2007, Freebase[4] is an open, shared world knowledge sharing database, which consists of a large number of collaboratively edited cross-linked data (currently 1.9 billion triples). Freebase is constructed by extracting entities and relations from knowledge bases such as Wikipedia, WordNet, etc. to form a structured Wikipedia. The training data[5] contains 1.18M sentences with 47 entity types, e.g., *person, location, organization, etc.* and 24 relation types, e.g., *nationality*, *place_founded*, *employee_of*. For more detailed dataset structures, please refer to Appendix A and Figure A.1. We exclude the *None* label (*NA*) relation, as [4], since the relation positions are uncertain. The test set contains 395 samples manually annotated by Hoffmann et al. This dataset is created for DS-RE, but the data format satisfies the settings of the text-bound TE. We use this dataset to evaluate the performance of sentence-level TE. This enhances the use of automatically annotated data in supervised relation extraction. **NYT Large** dataset is further developed by Zeng et al. and Lin et al.[6] based on Riedel NYT. This dataset has 53 relation labels including the *NA* labels. The training set contains 522,611 sentences, 279,786 pairs of entity, and 18,252 facts which cover all sentences in Riedel NYT. We use this dataset to evaluate the DS-RE task.

SemEval-2018 Task 7 dataset is provided for the task of semantic relation extraction and classification in scientific papers. This dataset defines 6 relation types, e.g., *PART_WHOLE*, *USAGE*. We adopt the dataset for subtask 2. The training/test data is composed of 350/150 abstracts of scientific publications from the Association for Computational Linguistics (ACL) Anthology with manually annotated entities and relations. For more detailed dataset structures, please refer to Appendix A and Figure A.2. One of the main challenges is the limited size of the training data. To overcome it, we also select a part of the data from the noisy data of subtask 1.2 to extend the training set which finally contains 3419 sentences.

TACRED dataset is introduced in Zhang et al.. It contains 106k sentences with entity mention pairs drawn from the yearly Text Analysis Conference (TAC) Knowledge Base Population (KBP)[7] challenge. They organized crowdsourced annotations on the Mechanical Turk platform, enabling the annotation of subject and object entity spans and relation types. Sentences are annotated

[3] http://iesl.cs.umass.edu/riedel/ecml/
[4] https://en.wikipedia.org/wiki/Freebase_(database)
[5] https://github.com/shanzhenren/CoType
[6] https://github.com/thunlp/NRE
[7] https://tac.nist.gov/2017/KBP/index.html

with 41 person- and organization-oriented relation types, e.g., *per:employee_of*, *org:founded*, and no relation for negative examples. The advantage of TACRED is that it contains a large number of high-quality relation instances, which enables adequate training of model parameters, and that these entity and relation types are more generic to downstream applications. They annotated all negative instances that emerged during data collection, making the data fit the contextual complexity of relational expressions in the real-world text. For more detailed dataset structures, please refer to Appendix A and Figure A.3. Entity mentions are typed, with subjects classified into person and organization, and objects classified into 16 fine-grained types (e.g., date and location).

6.1.2. Hyper-parameters

We conduct experiments based on the 200-D pre-trained GloVe, the 300-D pre-trained FastText, the 300-D randomly initialized word vectors, and the bert-base-uncased[8] representation respectively. For the SemEval-2018 dataset, we train the domain-specific word embeddings, like (Rotsztejn et al. 2018), using GloVe and Fasttext respectively. Our corpus is composed of all the abstracts since 2001 (5.4 million tokens) collected using the application programming interface (API)[9] on arXiv.org and the ACL Anthology Reference Corpus (ARC)[10] (90 million tokens). We trained word embeddings for 500 epochs with 60 threads. We use 50-D randomly initialized character embeddings. For the TACRED and NYT Large datasets, we use the 300-D GloVe embeddings, like Zhang et al..

For regularization we apply dropout with $p = 0.5$. The output dimension of character CNN is 100-D. We set the hidden size of the LSTM unit to 300D. We set $\lambda = 1.0$. To select better models, we divide each training epoch of NYT into 100 subdivisions and the training epoch of SemEval-2018 into 4 subdivisions randomly and evenly, and each subdivision retains a model for model selection, as this method can generate better models within the epoch. For the NYT and SemEval-2018 dataset, we use the Adam optimization algorithm to update the model parameters with an initial learning rate = 0.001, decay rate = 0.9, batch size = 32 and epoch = 30. For the TACRED dataset, we use Stochastic Gradient Descent (SGD) with the initial learning rate = 0.3, decay rate = 0.9, batch size = 64, cutoff = 5.0 for gradient clipping and epoch = 50. For NYT Large, we use SGD with the learning rate = 0.5, batch size = 64 and epoch = 30. When we fine-tune the model using BERT representation, we set learning rate = 3e−5, batch size = 8 and epoch = 5. We conduct experiments on an Intel(R) Xeon(R) CPU E7-4830 v3 @ 2.10 GHz (Mem: 976G) and the GPU Tesla K40c and TITAN RTX.

6.1.3. Evaluation

For the NYT dataset, we adopt the standard micro F1 score, recall (R), and precision (P) as metrics for NER and RC subtasks. A correct prediction is that the extracted triple matches the ground truth including two entities, relation direction, and type. For the SemEval-2018 dataset, we use the official script. The evaluation is carried out in two steps. First, the relation label and directionality are ignored, so it only evaluates the quality of entity pairs by the F1 score. Second, the evaluation of relation classification is the macro-average F1 score. For the TACRED dataset, we report the micro F1 score as Zhang et al.. For the DS-RE task, we adopt the held-out evaluation as [15], which is an effective evaluation method for a large dataset without costly human intervention. We compare the precision and recall curve. The curve is drawn by ranking all predicted instances according to their confidence scores and traversing the ranking list from the high score to the low score to measure the precision and recall at each position.

6.2. Results of triple extraction

The first part of Table 2 lists some baseline systems for joint entity and relation extraction. The other part is our methods based on context-independent and context-dependent representation. To eliminate the influence of random factors, we did 3 separate runs and took the average. When there is massive training data, our model can achieve good results even without using pre-trained word embeddings. This is because our model enhances the relation classification module. The joint training scheme enables two subtasks to interact with each other through shared representation. Pre-trained word embeddings generally improve model performance. Using GloVe embedding in this model can increase the F1 by 3.42%. Feature-based (BERT-FB) and fine-tuning (BERT-FT) based representation further improve results. BERT representation helps the model consider context information of different granularities. Table 3 lists the results of two subtasks. Pre-trained word embeddings also help to improve the NER subtask. BERT-based representation brings more improvements to the performance of the RE subtask. This indicates that the RE subtask needs to consider more context information.

To observe the training process, we divide the first training epoch into 25 intervals and evaluate the model on the test set. Fig. 4 shows the model performance. We observe that using pre-trained word embeddings achieves a faster learning speed than using the randomly initialized word vectors. The result of the RE subtask changes more significantly because the context of RE is more diverse. Our model learns the two subtasks jointly, making the learning process effective. The training set of the NYT dataset contains a lot of noise introduced by automatic data annotation, and denoising methods can be further studied in the future. Our method could be a reasonable choice when the training data is noisy. The benefits of deep learning lie in its powerful representation capabilities and end-to-end learning schemes. Deep learning can combine pre-trained word embeddings with the capacity to model contextual representations. When using word embeddings, it can achieve significant improvement. Even with randomly initialized word vectors, the deep learning model can update the parameters of the model through downstream tasks to obtain acceptable results. Without pre-annotated entities, the relation extraction results may be unstable, and the problem of enhancing model stability will be further studied in the future.

[8] https://huggingface.co/bert-base-uncased
[9] https://arxiv.org/help/api/index
[10] http://acl-arc.comp.nus.edu.sg/

Table 2
Results on the NYT dataset.

Methods	P	R	F1
MultiR	33.8	32.7	33.3
DS-Joint	57.4	25.6	35.4
Line	33.5	32.9	33.2
FCM	55.3	15.4	24.0
SPTree[a]	37.3	15.4	23.4
CoType	42.3	51.1	46.3
LSTM-LSTM-Bias	61.5	41.4	49.5
Transition	64.3	42.1	50.9
MRT	**67.4**	42.0	51.7
SwitchNet+Random	55.09	49.66	**52.23**
SwitchNet+GloVe	60.28	**52.27**	**55.95**
SwitchNet+FastText	58.61	**53.11**	**55.66**
BERT+TE	50.68	**56.20**	**53.30**
SwitchNet+BERT-FB	64.27	**52.35**	**57.69**
SwitchNet+BERT-FT	62.82	**57.49**	**60.02**

[a]This experiment is conducted by Wu et al. in the ReQuest System.

Table 3
Results of subtasks on the NYT dataset.

Tasks	NER			RE		
Metrics	P	R	F1	P	R	F1
SwitchNet+Random	91.00	87.22	89.06	55.09	49.66	52.23
SwitchNet+GloVe	92.44	91.77	92.10	60.28	52.27	55.95
SwitchNet+FastText	92.01	91.55	91.76	58.61	53.11	55.66
SwitchNet+BERT-FB	89.73	91.03	90.37	64.27	52.35	57.69
SwitchNet+BERT-FT	91.13	93.43	92.26	62.82	57.49	60.02

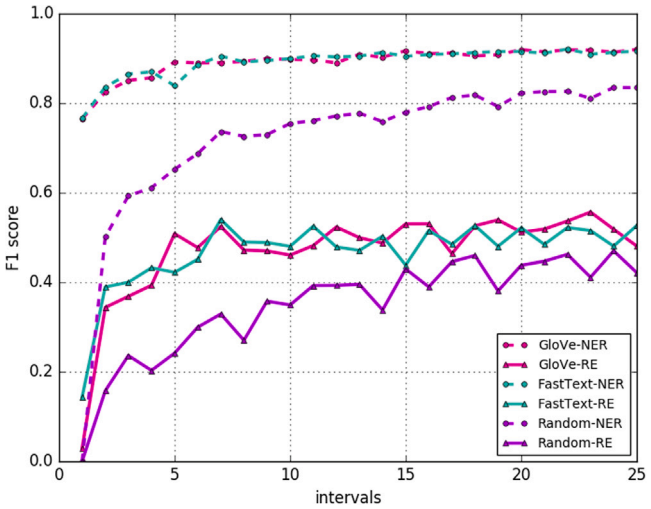


Fig. 4. Learning process in the first epoch (average sampling of 25 intervals).

6.3. Results of relation extraction

To evaluate the performance of the RE task, we conduct experiments on the SemEval-2018 dataset. Entity mentions in each sample are available, so our model does not need to predict the EOIs actively. Formulation (16) becomes $A = \mathcal{F}_{rc}(\mathbf{h}, \Psi(\mathbf{y}))$ where $\mathbf{y}$ is the label sequence. Our model achieves good performance as shown in Table 4. The first part of Table 4 lists baseline models and the second part is our approach. For the RC subtask, our model score is higher than the second (UWNLP) on the official leaderboard. For the RE subtask, our model score is higher than the third (SIRIUS-LTG-UiO). However, the official results were obtained within a limited time and the number of submissions. Their systems used NLP pipelines to extract some features, while the CNN-based system that did not use much feature engineering achieved 18.46% Macro-F1 on the RC subtask. Feature engineering can provide much prior knowledge directly when the model did not fully learn the features. ETH-DS3Lab system achieves the highest score, but this system relies heavily on feature engineering and data augmentation, while our model is jointly trained.

Table 4
Results on the SemEval-2018 dataset.

Systems	Relation instance			Relation classification		
Metrics	P	R	F1	P	R	Marco-F1
UC3M-NII (5th)	–	–	35.4	–	–	18.5
Bf3R (4th)	–	–	33.4	–	–	20.3
SIRIUS-LTG-UiO (3rd)	–	–	37.4	–	–	33.6
UWNLP (2nd)	–	–	**50.0**	–	–	39.1
ETH-DS3Lab (1st)	–	–	48.8	–	–	**49.3**
BERT+RE	–	–	40.86	–	–	42.10
SwitchNet+Random	47.19 ± 2.46	19.52 ± 3.22	27.39 ± 2.80	30.45 ± 1.09	23.57 ± 0.97	26.57 ± 0.95
SwitchNet+GloVe	35.25 ± 0.88	49.77 ± 3.49	41.18 ± 0.66	43.50 ± 2.12	41.05 ± 3.86	42.04 ± 1.30
SwitchNet+FastText	34.63 ± 0.70	50.59 ± 6.81	40.87 ± 1.94	44.99 ± 8.63	36.39 ± 4.13	39.34 ± 0.94
SwitchNet+BERT-FB	34.62 ± 2.60	59.76 ± 2.18	43.84 ± 2.61	52.93 ± 3.19	49.83 ± 5.81	45.58 ± 0.71

Table 5
Results of relation classification.

Settings	P	R	Marco-F1
SwitchNet+Random	66.44 ± 1.40	72.85 ± 1.81	69.46 ± 0.41
SwitchNet+GloVe	74.53 ± 1.89	79.16 ± 2.82	77.35 ± 1.80
SwitchNet+FastText	75.96 ± 0.67	79.98 ± 1.04	77.91 ± 0.62
SwitchNet+BERT-FB	71.20 ± 0.34	73.83 ± 2.37	72.01 ± 0.94

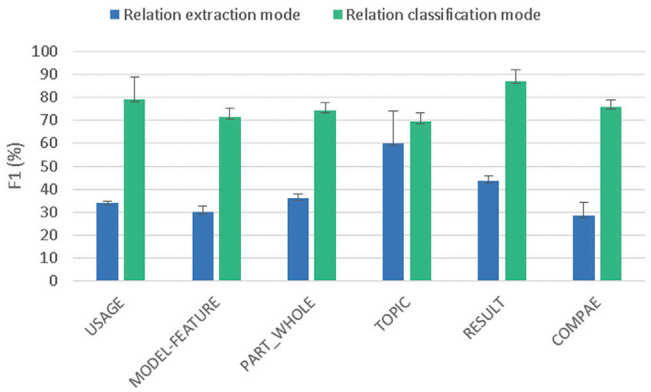


Fig. 5. Class-level comparison of RC and RE.

We only input tokens of simple text and entities. Compared with the models not using much feature engineering, our model achieves significant improvements (8.44% F1 and 27.12% Macro-F1) on RE and RC results respectively. This is because our model generates more relation candidates by using POEOI inference and joint training. These candidates provide rich supervision signals that can update model parameters from more perspectives. The SemEval-2018 dataset is manually annotated but still noisy, so mechanisms to improve data annotation can be investigated in the future. The experimental results are affected by hyper-parameter tuning. In addition, programming skills also directly affects the process of data processing and model parameter learning. In the future, we will further optimize the design patterns and programming tricks. Other systems do a lot of optimization in hyper-parameter tuning, feature engineering, model selection, model integration, etc. To win the competition, these systems require a lot of data processing and model optimization work. Unlike their motivations, this paper mainly focuses on evaluating the adaptability of the proposed framework to different data protocols. We simplify many processes, and the model sometimes achieves relatively good but not optimal results. We do not delve into data optimization strategies and techniques beyond the model, and such work can be further investigated in the future.

6.4. Results of multiple relation classification

To evaluate the influence of pre-annotated relation instances for the RC task, we conduct experiments based on Section 6.3. Note that this is a MRE task because each sentence may preserve multiple relations. Formulation (16) becomes $A = \mathcal{F}_{rc}(\mathbf{h}, \text{POEOIs})$ because POEOIs are provided in the input sample. Compared with Table 4, the pre-annotated POEOIs significantly improve the results of relation extraction.

For the BERT-FB, domain adaption of context-dependent representation requires more supervision signals than other pre-trained word embeddings, while the supervision signals are more sparse in this task. We visualize the prediction results of each class in RE

Table 6
Results on the TACRED.

Systems	P	R	F1
LR	**73.5**	49.4	59.4
SDP-LSTM	66.3	52.7	58.7
Tree-LSTM	66.0	59.2	62.4
PA-LSTM	65.7	64.5	65.1
C-GCN	69.9	63.3	66.4
C-GCN+PA-LSTM	71.3	65.4	**68.2**
TRE	70.1	65.0	67.4
BERT	67.23	64.81	66.00
ERNIE	69.97	**66.08**	67.97
SwitchNet	71.6	61.1	65.9
SwitchNet+C-GCN	73.3	62.9	67.6

and RC tasks in Fig. 5 where the x- and y- axes are the relation type and the F1 score respectively. The RE and RC results are from Table 4 (blue bars) and Table 5 (green bars) respectively. This figure shows that the RC task achieves a more balanced result than the RE task. This is because, during model training, using the pre-annotated POEOIs enhances the correlation of the entity pairs and improves the quality of supervision signals. During testing, the input explicitly indicates the relation positions so the prediction uncertainty is reduced and the false positive prediction also decreases.

6.5. Results of single relation classification

To analyze the generality of the RC component, we also conduct experiments on the TACRED. Note that this is the SRE task because each sentence only focuses on one POEOI. Following Zhang et al. we concatenate word embedding, position embedding, POS tag, and entity label embedding and consider the entire sentence. Our model achieves a higher result than the PA-LSTM as shown in Table 6. This is because the query vector q of our attention captures the global context of a sentence and the hidden representation contains position information.

The RC result is slightly lower than GCN, and this indicates that the POEOI inference scheme has more contribution. GCN has proven very effective in relation classification because its representation captures long-range syntactic relations by encoding the dependency structure. We added GCN to encode syntactic features, that is, $y = \mathcal{F}_{rc}(\mathbf{h} \oplus \mathcal{F}_{gcn}(\mathbf{x}), \text{POEOIs})$ where $\mathbf{h}$ and $\mathbf{x}$ are the hidden states and the dependency tree respectively. $\mathcal{F}_{gcn}(\cdot)$ denotes the GCN model. In this setting, the F1 score is increased by 1.7%, because syntactic information can provide complementary benefits when we extract informative tokens by a sequence model. ERNIE achieves a higher score because it incorporates entities in KGs to enhance the pre-trained language model, but the downside is the computation costs.

We observe that "*per:age*", "*per:title*"and "*org:founded*" relations are easier to extract. "*org:member_of*" and "*per:country_of_death*" are more problematic because model performance is concerned with the class imbalance problem and the context length ($s - o$ distance) of the input text. There are few instances of these missing relations, e.g., "*per:country_of_death*" only has 63 samples, so these relations are not fully learned. The context length of each relation type also influences the result. We observe that relations typically expressed in a longer context tend to be more difficult to extract. We visualize this dataset in Appendix A in supplementary material.[11] It is promising to introduce entity embedding of knowledge graphs or explore the few-shot learning to alleviate this problem. Our proposed method does not achieve the best results because the prototype proposed in this paper may not be good enough in program implementation. C-GCN and PA-LSTM have many optimizations in data preprocessing, dropout and gradient clipping, and these projects are also better organized in terms of design patterns and programming techniques. In the future, we can further optimize the design patterns and programming implementation of the framework.

6.6. Results of distant supervision relation extraction

For DS-RE, prior works show that sentence-level attention effectively reduces the effects of noisy instances. Following [15], we concatenate the position embeddings. For all the methods, we use the common GloVe word embeddings. Table 7 lists the precision of the top-ranked relations and the area under the curve (AUC) of the model. AUC can measure model performance, as shown in Fig. 6. The higher the AUC, the better the model. We observe that the single-head SwitchNet (S-SwitchNet) and multi(12)-head SwitchNet (M-SwitchNet) both achieve significant improvements in terms of P@N and AUC. As the recall increases, our methods have higher precision than other settings. This means that our methods can extract more high-quality relations.

We observe that SwitchNet+ONE achieves a higher AUC than CNN/PCNN+ONE because the word-level attention enhances the sentence representation. Sentence-level attention (SwitchNet+ATT) further improves model performance because sentence-level attention can alleviate the effects of noisy instances, making the model more robust. After we add multi-head attention, this model improves 1.8% AUC value. This is because the multi-head attention can generate relation representation by attending to different tokens from multiple perspectives. Ablation studies can be found in Section 7.1.2. In summary, our model uses hierarchical attention to better encode bag representation.

[11] https://github.com/nnntt/SwitchNet/blob/master/Appendix/appendix.pdf

Table 7
P@N for relation extraction in the entity pairs with different number of top-ranked predictions.

P@N(%)	100	200	300	Mean	AUC(%)
CNN+ONE	77.2	72.1	69.1	72.8	33.6
CNN+ATT	74.3	72.1	66.1	70.8	33.5
PCNN+ONE	79.2	79.1	71.1	76.5	34.0
PCNN+ATT	73.3	71.1	66.8	70.4	34.1
SwitchNet+ONE	81.1	75.1	68.4	74.9	35.0
SwitchNet+ATT	**82.2**	**77.1**	71.4	**76.9**	36.1
+Multi-head	81.2	74.1	**72.1**	75.8	**37.9**

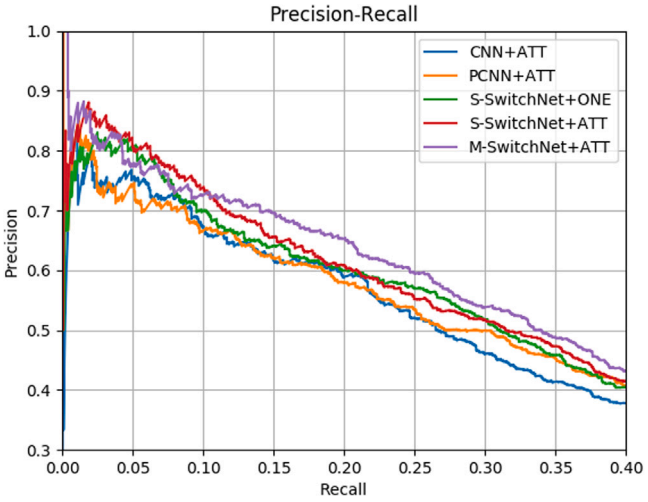


Fig. 6. Precision and recall curves of different settings.

7. Analysis

In this section, we performed ablation studies on the text-bound and distant supervision relation extraction respectively. Then, we analyze the influence of relation candidate selection on model training. Finally, we use a case study to visualize the model's attention distribution.

7.1. Ablation study

7.1.1. Text-bound relation extraction

Table 8 shows an ablation study of multi-task training and pipeline training on the NYT dataset. Two systems denote a fine-tuned NER system and a SwitchNet system. We use a BERT model as the NER system. We first compare the multi-task training in different settings. A first observation is that relying on the fine-tuned NER system slightly reduces the final results by 1.72% F1. Although the NER system is enhanced, some entities still hurt the final precision. Because extracting more entities does not always help the relation extraction subtask and some entities may increase the risk of predicting false positives. The above training process is a multi-stage pipeline and relation candidates are determined by the NER system. We first fine-tune the NER system and then predict the label sequence $\hat{Y}_{ner}$ which are also written to disk. Then we train the joint entity and relation extraction model that can use the $\hat{Y}_{ner}$ information. When we replace $\hat{Y}_{ner}$ with the ground truth NER labels, the result is significantly improved (5.03% F1). This means that a high-quality label sequence can help improve the final results.

Removing the multi-task training degrades the performance by 0.94% F1. This means that multi-task training benefits the RE subtask from the NER subtask. After removing the NER system, the joint optimization process does not rely on the NER system, and the performance improves by 1.72% F1. Then, we remove the NER system and multi-task training. We first train the NER subtask, then freeze the parameters of the shared layer, and then train the RE subtask. The F1 score drops by 2.51% because the RE subtask cannot be encoded at the lower layer to interact with the NER subtask. This means that multi-task joint training is critical for subtask interaction. When we do not freeze the lower layer, the result drops significantly. Because when we train the RE subtask, the memory of the NER subtask is weakening. This phenomenon is also known as catastrophic forgetting.

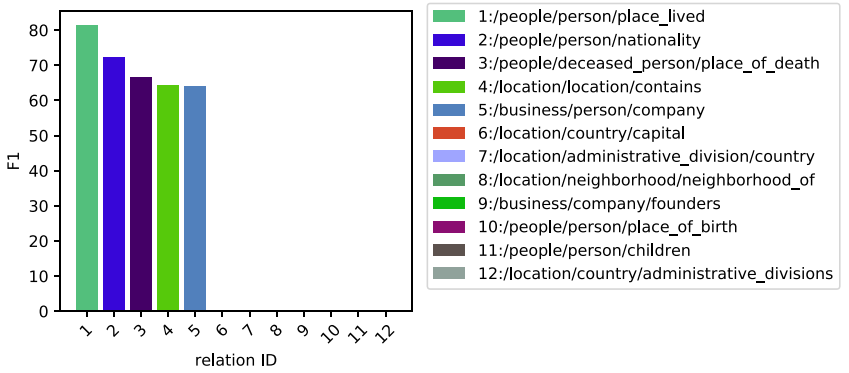
We also apply our training pipeline to BERT. We fine-tune a simple BERT model for TE. Table 9 shows the results. When only using multi-task training, BERT achieves slightly lower results than our SwitchNet. When we remove the multi-task, NER system, and Frozen NER, this model almost forgets all the NER memory, so the RE subtask fails. Fine-tuning models can achieve different functions, which also means that BERT is sensitive to parameter changes.

Table 8
SwitchNet setting ablation.

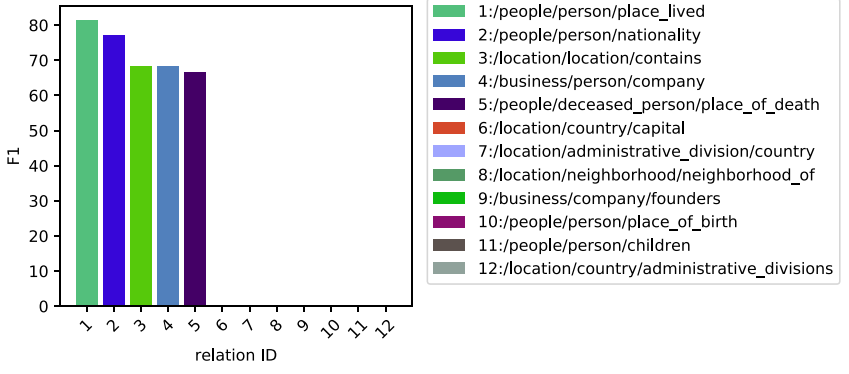
Model	P	R	F1
Two systems + multi-task	56.78	51.89	54.23
+ NER label	61.18	57.46	59.26
– multi-task	55.64	51.13	53.29
– NER system	60.28	52.27	55.95
– multi-task, NER system	50.33	56.96	53.44
– multi-task, NER system, Frozen NER	55.88	43.29	48.78

Table 9
BERT model setting ablation.

Model	P	R	F1
Two systems + multi-task	51.30	54.93	53.05
+ NER label	54.89	61.01	57.79
– multi-task	50.46	54.43	52.37
– NER system	50.68	56.20	53.30
– multi-task, NER system	46.46	54.93	50.34
– multi-task, NER system, Frozen NER	–	–	–



(a) Multi-task training



(b) Multi-task training + NER label

Fig. 7. Class-aware result.

Fig. 7 shows the class-aware prediction result. Fig. 7(a) and 7 (b) denote the model with or without the ground truth NER labels when extracting triples. We observe "*place_lived*", "*nationality*", "*place_of_death*", "*contains*" and "*company*" relations are easier to extract. Other relations in the test set sometimes are not extracted. This is because some sparse relations are not fully learned and there is a class imbalance problem in the training data. We visualize and analyze this dataset in Appendix A. We observe that providing NER labels enhances the extraction of well-trained relations, while other relations are not obviously improved.

Table 10
SwitchNet setting ablation.

Model	100	200	300	Mean	AUC(%)
M-SwitchNet+ATT	81.2	74.1	72.1	75.8	37.9
+ subject, object	79.2	72.1	69.8	73.7	35.0
– Multi-head	82.2	77.1	71.4	76.9	36.1
– Multi-head, ATT	81.1	75.1	68.4	74.9	35.0
– Multi-head, Single-head	73.3	73.6	69.8	72.2	35.1
– Multi-head, Single-head, ATT	75.2	71.6	70.8	72.6	34.7

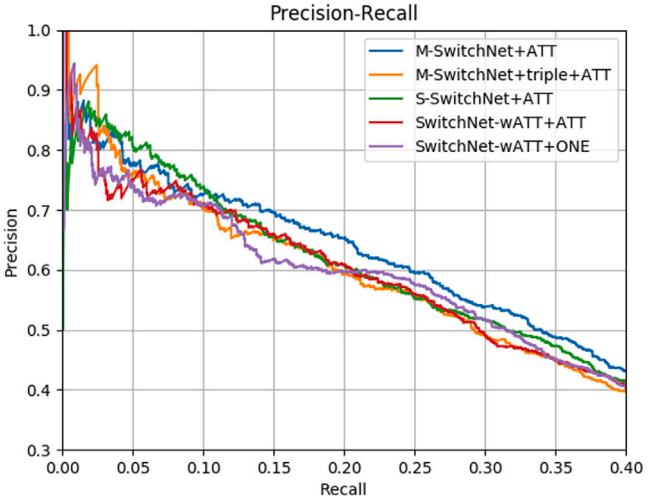


Fig. 8. Precision and recall curves of different settings.

7.1.2. Distant supervision relation extraction

We performed ablation studies on DS-RE. Table 10 shows the results of different settings. A first observation is that adding representations of subject and object to the multi-head attention reduces the results. This is because all positive and noisy samples in the bag contain the same subject and object, so using this information may hurt model training. When we remove the multi-head attention and keep only one head, the AUC decreases, but the P@N of the top-ranked relations improves. When we also remove the sentence-level attention, the result drops. This means that sentence-level attention is important in this framework. When we remove only the word-level attention, the result also drops. This means the word-level and sentence-level attention mechanisms complement each other to form the hierarchical attention mechanism. When we remove the two-level attention, the results will drop further.

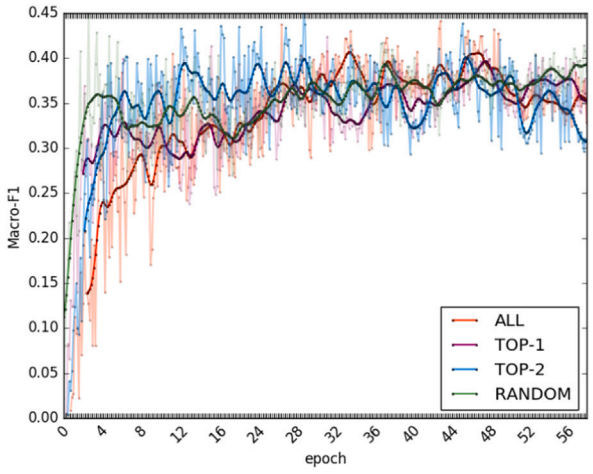
Fig. 8 shows the precision and recall curves of different settings. We observe that when we do not use the **w**ord-level **att**ention (wATT), the curves generally move down. Precision drops faster as recall increases. This means that word-level attention can help achieve higher performance. When we remove the sentence-level attention (ATT), the curve fluctuations increase. This means that the sentence-level selection enhances the stability of the model.

7.2. Analysis of relation candidates selection

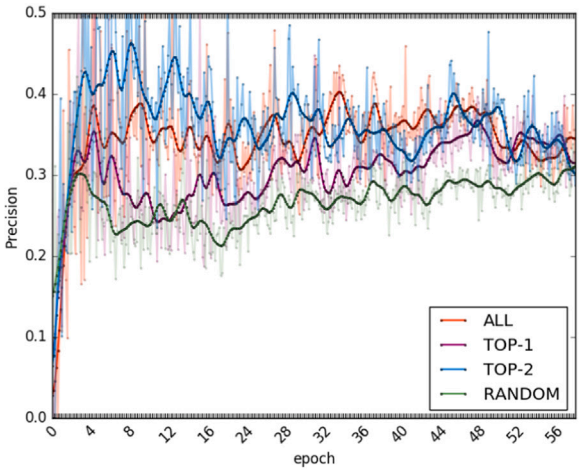
Candidates' feedback can be positive and negative and both types of feedback have great potential to boost recall and precision. However, the number of possible relations is $\mathcal{O}(n^2)$, where n is the number of entities, which potentially increases computational complexity and may lead to the class imbalance problem. Existing approaches for this question typically perform random sampling, which might include some inefficient relation candidates.

The relation extraction task is also a mathematical model. When there are too many candidate relationships in a sample, the generalization of the model will be affected during the learning process, and it will face the problems mentioned in the support vector machine (SVM). SVM is based on the VC dimension theory of statistical learning and the principle of minimum structural risk to obtain the best generalization ability. Inspired by the idea of the optimal hyperplane, we can think of these indistinguishable relation candidates as an approximation of the support vectors. This setting not only reduces the computational complexity but also helps to enhance the generalization performance of the model, especially for noisy datasets.

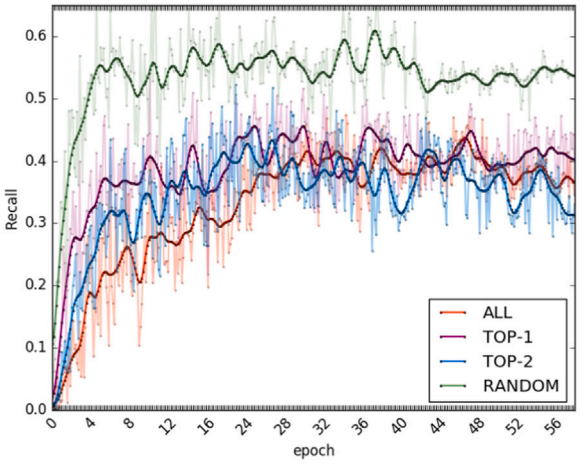
Inspired by the idea of SVM, a few support vectors are effective to decide the classification hyperplane. We assume that difficult candidates that are closer to the classification hyperplane can improve the classifier more effectively. This can reduce computational complexity. During model training, we rank the negative candidates according to their prediction probability of being a None relation and only keep top-k negative ones with the least probabilities. These relations are more likely not to be predicted as NA. k is a



(a) Macro-F1



(b) Precision



(c) Recall

Fig. 9. Prediction results on SemEval-2018 based on different selection strategies for relation candidates.

hyper-parameter, and we experimented with $k \in \{1, 2, \infty\}$. ∞ denotes all candidates. This ranking mechanism filters a vast number of negative candidates leaving the classifier with a small set.

We compared several selection strategies, i.e. random selection (RANDOM), top-k selection (TOP-K), and all candidates (ALL), as shown in Fig. 9. We find that using top-k selection can achieve comparable F1 scores to using all candidates, but the precision and recall might be not balanced enough. When more negative candidates are retained, the recall will decrease and the precision will increase. This method potentially enhances the use of the ranking mechanism to reduce the computational complexity of RE models.

7.3. Case study

It is instructive to analyze which words the model attends to when classifying relations. We hand-picked some examples in the SemEval-2018 dataset and visualize the attention patterns of these samples. Fig. 10 shows how our model extracts informative words for relation representation. The first column is the sample id. The second column contains the extracted triples, and the third column shows the textual input. The first sentence shows that this model is capable of focusing on informative words to identify the "*PART_WHOLE*" relation type for "*English-Chinese Bitexts*" and "*Web*". The second sentence shows this model resolves the comparative relation by attending to "*narrower than*". The third sentence shows that "*are described in*" means the "*TOPIC*" relation.

The fourth sentence shows that "*produced by means*" for "*Translations*" and "*beam search decoder*" denotes the "*MODEL-FEATURE*" relation, while the fifth sentence shows this model extracts the relation type "*RESULT*" by attending to "*produce best*". This suggests

Id	Triple	Sentences
1	(English-Chinese bitexts, PART_WHOLE, Web)	This piece of work has also laid a foundation for exploring and harvesting English-Chinese bitexts in a larger volume from the Web .
		This piece of work has also laid a foundation for exploring and harvesting English-Chinese bitexts in a larger volume from the Web .
2	(domains, COMPARE, MUC-4 terrorism domain)	These previous domains were much narrower than the MUC-4 terrorism domain .
		These previous domains were much narrower than the MUC-4 terrorism domain .
3	(paper, TOPIC, overview)	An overview of HowNet and information structure are described in this paper .
		An overview of HowNet and information structure are described in this paper .
4	(beam-search decoder, MODEL-FEATURE, Translations)	Translations are produced by means of a beam-search decoder .
		Translations are produced by means of a beam-search decoder .
5	(Bayesian classifiers, RESULT, recall performance)	In our evaluation , Bayesian classifiers produce the best recall performance of 80 % but the precision is low (60%) .
		In our evaluation , Bayesian classifiers produce the best recall performance of 80 % but the precision is low (60%) .
6	(WordNet, USAGE, Word Sense Disambiguation (WSD) task)	WordNet has been used extensively as a resource for the Word Sense Disambiguation (WSD) task , both as a sense inventory and a repository of semantic relationships .
		WordNet has been used extensively as a resource for the Word Sense Disambiguation (WSD) task , both as a sense inventory and a repository of semantic relationships .

Fig. 10. Visualization of some cases where the underlined phrase represents an entity in the extracted triple and the red degree denotes the word weight for the relation representation.

that this model considers entity semantics and sentence context. The sixth sentence shows that this model can extract informative tokens "*resource for*" instead of the important verb in a long context. These results indicate that the model considers the entity semantics and the relational context while attending to the informative words for relation representation.

This model also extracts some wrong results. For the sentence *"In essence this can start to rewrite the history of photography", said Grant Romer, director of the advanced residency program in photograph conservation at the George Eastman House in Rochester.*, *George Eastman House in Rochester* implies that (*Rochester, contains, George Eastman House*). However, the result is (*Rochester, place_lived, George Eastman House*). *"Grant Romer ... at George Eastman House in Rochester"* indicates a state but does not mean the relation type *"place_lived"*. The understanding of the temporal state of the relation needs to be further improved. There are still some errors when extracting the relation of the spatial position, such as `neighbor_of`. The understanding of the spatial relations needs to be further improved.

8. Conclusion

Our SwitchNet is implemented for different information extraction configurations, namely NER, RC, RE, TE, and DS-RE. This framework designs a modular neural architecture and uses different information flows to tackle different relation extraction tasks, which significantly reduces the problem complexity for relation extraction tasks. Our model predicts EOIs using a standard NER setup and uses an attention-based classifier to classify relations of POEOIs. Reducing the unknowns of model prediction by indicating the POEOIs can help improve the performance. We also address the secondary problem of jointly learning NER and RE through the POEOI inference and the shared representation mechanisms. However, the challenge is that there is a sort of information loss in the discrete entity detection process. Extensive experiments are conducted on multiple datasets, and our framework achieves good performance.

The relation classifier of this framework is highly extensible, and the joint optimization method is beneficial to upgrade the pipeline method to an end-to-end information extraction model. The systematic design of information flow and modular neural networks is a promising way to reduce the problem complexity when designing an artificial intelligence (AI) system with diverse functions. In the future, we will combine this framework with entity linking technology so that the extracted knowledge can be directly used in the knowledge base. Furthermore, we plan to extend this framework to open information extraction.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.compeleceng.2022.108445.

Data availability

Github link is provided: https://nnntt.github.io/SwitchNet/

References

[1] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, p. 580–7.
[2] Poo M-m, Du J-l, Ip NY, Xiong Z-Q, Xu B, Tan T. China brain project: Basic neuroscience, brain diseases, and brain-inspired computing. Neuron 2016;92(3):591–6.
[3] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of NIPS. 2017, p. 5998–6008.
[4] Zheng S, Wang F, Bao H, Hao Y, Zhou P, Xu B. Joint extraction of entities and relations based on a novel tagging scheme. In: Proceedings of ACL. Association for Computational Linguistics; 2017, p. 1227–36.
[5] Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation classification via convolutional deep neural network. In: Proceedings of COLING 2014. 2014, p. 2335–44.
[6] Xu Y, Mou L, Li G, Chen Y, Peng H, Jin Z. Classifying relations via long short term memory networks along shortest dependency paths. In: Proceedings of EMNLP. The Association for Computational Linguistics; 2015, p. 1785–94.
[7] Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures. In: Proceedings of ACL. The Association for Computer Linguistics; 2016, p. 1105–16.
[8] Zhang Y, Qi P, Manning CD. Graph convolution over pruned dependency trees improves relation extraction. In: Proceedings of EMNLP. Association for Computational Linguistics; 2018, p. 2205–15.
[9] Sun C, Gong Y, Wu Y, Gong M, Jiang D, Lan M, et al. Joint type inference on entities and relations via graph convolutional networks. In: Proceedings of ACL. Association for Computational Linguistics; 2019, p. 1361–70.
[10] Wang H, Tan M, Yu M, Chang S, Wang D, Xu K, et al. Extracting multiple-relations in one-pass with pre-trained transformers. In: Proceedings of ACL. Association for Computational Linguistics; 2019, p. 1371–7.
[11] Bekoulis G, Deleu J, Demeester T, Develder C. Joint entity recognition and relation extraction as a multi-head selection problem. Expert Syst Appl 2018;114:34–45.
[12] Han X, Gao T, Yao Y, Ye D, Liu Z, Sun M. OpenNRE: An open and extensible toolkit for neural relation extraction. In: Proceedings of EMNLP-IJCNLP - system demonstrations. Association for Computational Linguistics; 2019, p. 169–74.
[13] Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. In: Proceedings of ACL. The Association for Computer Linguistics; 2009, p. 1003–11.
[14] Riedel S, Yao L, McCallum A. Modeling relations and their mentions without labeled text. In: Joint European conference on machine learning and knowledge discovery in databases, vol. 6323. Springer; 2010, p. 148–63.
[15] Lin Y, Shen S, Liu Z, Luan H, Sun M. Neural relation extraction with selective attention over instances. In: Proceedings of ACL. 2016, p. 2124–33.
[16] Banko M, Cafarella MJ, Soderland S, Broadhead M, Etzioni O. Open information extraction from the web. In: Proceedings of IJCAI 2007. p. 2670–6.
[17] Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT. Association for Computational Linguistics; 2019, p. 4171–86.
[18] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of ICLR workshop track. 2013.
[19] Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification. In: Proceedings of NIPS, vol. 28. 2015, p. 649–57.
[20] Ma X, Hovy EH. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In: Proceedings of ACL. The Association for Computer Linguistics; 2016, p. 1064–74.
[21] Rotsztejn J, Hollenstein N, Zhang C. ETH-DS3lab at SemEval-2018 task 7: Effectively combining recurrent and convolutional neural networks for relation classification and extraction. In: Proceedings of the international workshop on semantic evaluation. 2018.
[22] He R, Lee WS, Ng HT, Dahlmeier D. An unsupervised neural attention model for aspect extraction. In: Proceedings of the ACL. 2017, p. 388–97.
[23] Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. In: Proceedings of NAACL. 2016, p. 1480–9.
[24] Vafaie H, Imam IF. Feature selection methods: Genetic algorithms vs. greedy-like search. In: Proceedings of the international conference on fuzzy and intelligent control systems, vol. 51. 1994, p. 28.
[25] Dorigo M, Caro GD, Gambardella LM. Ant algorithms for discrete optimization. Artif Life 1999;5(2):137–72.

Hongyin Zhu received the B.A. degree in Automation from Shandong University, Jinan, China, in 2014, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2020. He worked as a Postdoctoral Researcher with the Department of Computer Science and Technology, Tsinghua University, Beijing. He is currently working at Inspur Electronic Information Industry Co., Ltd., Beijing.

Prayag Tiwari received his Ph.D. degree from the University of Padova, Italy. He is currently working as an Assistant Professor at Halmstad University, Sweden. Previously, he worked as a Postdoctoral Researcher at the Aalto University, Finland, and Marie Curie Researcher at the University of Padova, Italy. His research interests include Machine Learning, Deep Learning, Quantum Machine Learning, NLP, Healthcare, and IoT.

Yazhou Zhang received his Ph.D. degree in the College of Intelligence and Computing from Tianjin University (Tianjin, China) in 2020. He is currently a Lecturer in Software Engineering College at Zhengzhou University of Light Industry (Zhengzhou, China). He is also a Postdoctoral Fellow in Tianjin University-China Mobile Communications (Tianjin) Joint Laboratory. He was a visiting researcher at Harbin Institute of Technology in 2013, Beijing Institute of Technology in 2019, and University of Macau in 2019.

Deepak Gupta received a B.Tech. Degree in 2006 from the Guru Gobind Singh Indraprastha University, Delhi, India. He received an M.E. degree in 2010 from Delhi Technological University, India, and Ph.D. degree in 2017 from Dr. APJ Abdul Kalam Technical University (AKTU), Lucknow, India. He completed his Post-Doc from the National Institute of Telecommunications (Inatel), Brazil, in 2018. He has co-authored more than 207 journal articles, including 168 SCI papers and 45 conference articles.

Meshal Alharbi is an Assistant Professor of Artificial Intelligence in the Department of Computer Science at Prince Sattam Bin Abdulaziz University in the Kingdom of Saudi Arabia. He received Ph.D. degree in Computer Science from Durham University, UK, in 2020, and M.Sc. degree in Computer Science from Wayne State University, USA, in 2014. He has 10 years of Experience in Teaching/Research,/Industry. His research interests lie in the Artificial Intelligence

Applications and Algorithms, Agent-Based Modelling and Simulation Applications, Disaster/Emergency Management and Resilience, Optimization Applications, and Machine Learning.

Tri Gia Nguyen received the B.Ed. degree in computer science from the Hue University of Education, Vietnam, in 2011, the M.Sc. degree in computer science from Duy Tan University, Vietnam, in 2013, and the Ph.D. degree in computer science from Khon Kaen University, Thailand, in 2017. He is currently the Head of Department of Information Security, FPT University, Danang, Vietnam. His research interests include the Internet of Things, sensor networks, wireless communications, wireless energy harvesting networks, mobile computing, edge computing, software-defined networking, network functions virtualization, and network security.

Shahram Dehdashti received his Ph.D. degree from the University of Isfahan, Iran. He worked as an assistant researcher at Zhejiang university, China, from 2014 to 2017. He then moved to Madison, USA, to work at the University of Wisconsin 2017–2018. He has spent time at Queensland University of Technology, employed as an assistant researcher from 2018 to 2021. Currently he is employed as senior researcher at Technical University of Munich, Germany. His research interests include Quantum Machine Learning, Quantum Networks, Quantum Cognition, etc.

Supporting Information

Tailoring the output performance of PVDF based piezo-tribo hybridized nanogenerator via B, N co-doped reduced graphene oxide

Shilpa Rana[a], Vishal Singh[a], Bharti Singh[a]*

[a]Department of Applied Physics, Delhi Technological University, Main Bawana Road, Delhi 110042, India

Corresponding author E-mail: bhartisingh@dtu.ac.in

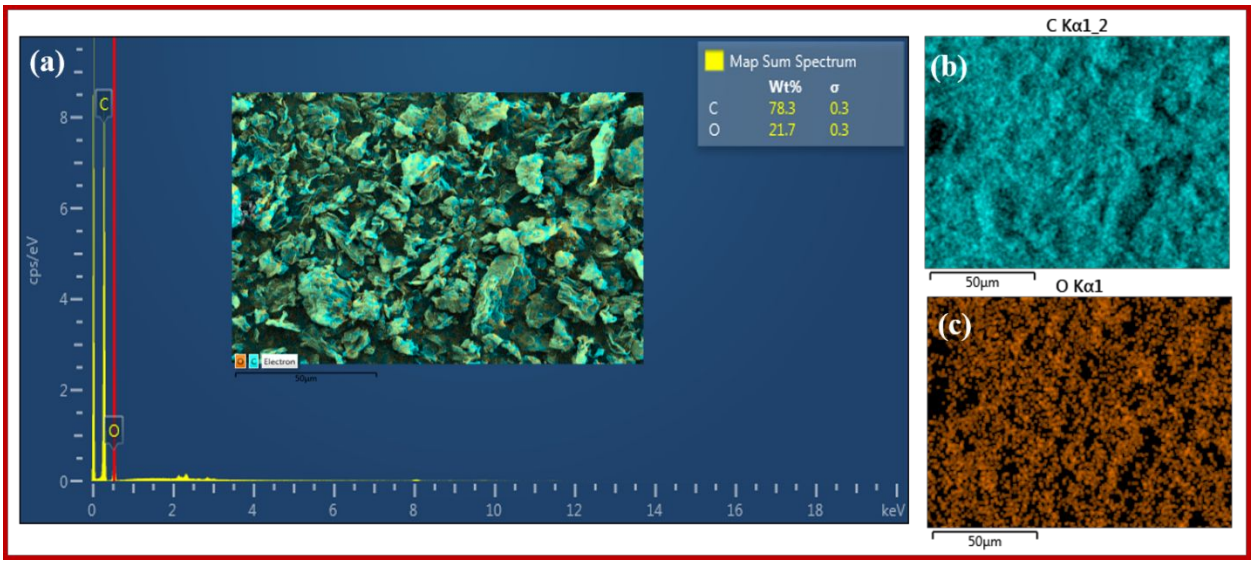


Figure S1.(a-c) EDX spectrum with the elemental mapping image of all elements present in rGO.

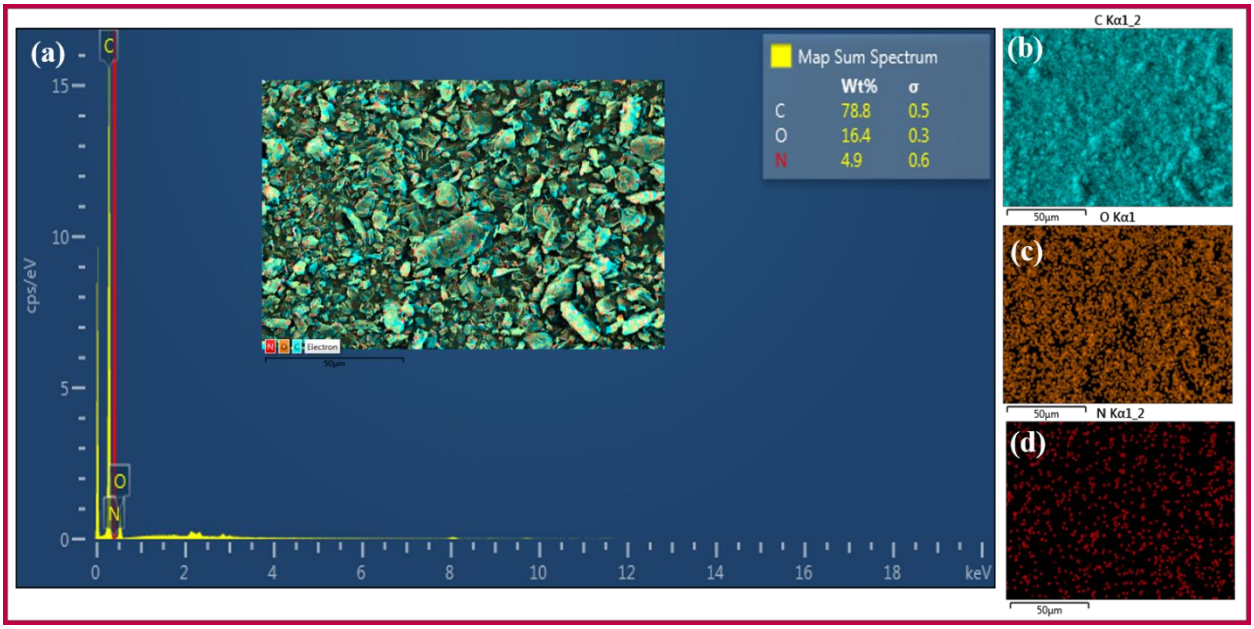


Figure S2. (a-d) EDX spectrum with the elemental mapping image of all elements present in N-rGO.

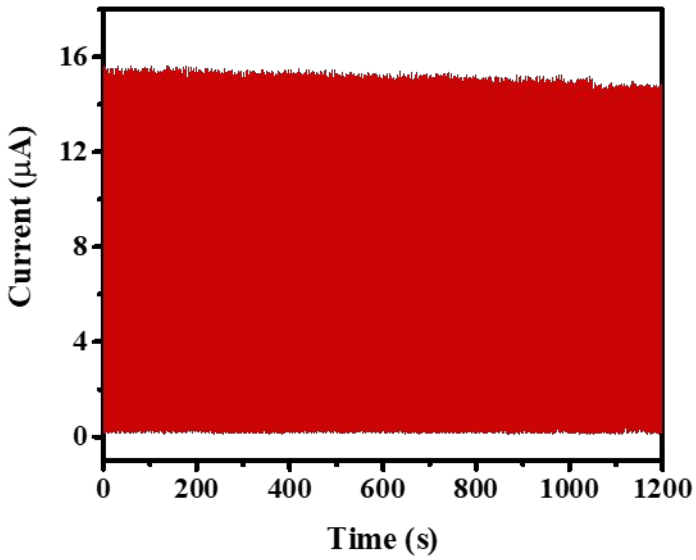


Figure S3. Durability test of PVDF/BN-rGO based PENG device for 1200 s.

Figure S4 shows the working mechanism of an HNG working in vertical contact separation mode. In the initial state S4 (a), no charges are accumulated on the surface of PDMS film and Al electrode because of the insufficient frictional force. When an external force is applied to the HNG, the two layers come in contact with each other and acquire opposite charges on the surface of two materials due to contact electrification. The PDMS surface acquires negative charges, while the Al surface loses electrons and becomes positive charges Figure S4(b). In addition to the triboelectric charge, the piezoelectric composite film ie. PVDF/BN-rGO also undergoes compression, resulting in the generation of piezoelectric potential arising due to the resultant dipole moment. At this stage, there is no flow of charge between the electrodes of the two components because of electrostatic equilibrium. Therefore, when we start releasing the external force the piezoelectric PVDF composite film experienced opposite strain which results in the generation of opposite piezoelectric potential in PVDF composite film as shown in Figure S4 (b). At the same time, the negative charge on PVDF/BN-rGO induces positive electrostatic charges on the bottom Al electrode because of the electrostatic induction effect. As a result, the electrons start flowing from the bottom electrode to the upper electrode of PENG component as shown in Figure S4 (c). These charges flow until triboelectric charge distribution reached to electrical equilibrium state fig d at this stage external force is completely removed and both the Al and PDMS films are reverted back to their original position. When the external force is applied again (Figure S4 (e)), the PENG component returns to its initial charge distribution as shown in Figure S4 (b), and produces the piezoelectric output opposite to that released by an external force, whereas the triboelectric charges generated on the two surfaces will facilitate the flow of electrons from the top Al electrode of PVDF composite film to the electrode attached to PDMS film, thus producing an opposite output

signal. As a result, HNG under periodic contact and separation produces an electrical output due to the combination of the synergetic effect of the piezoelectric and triboelectric effects.

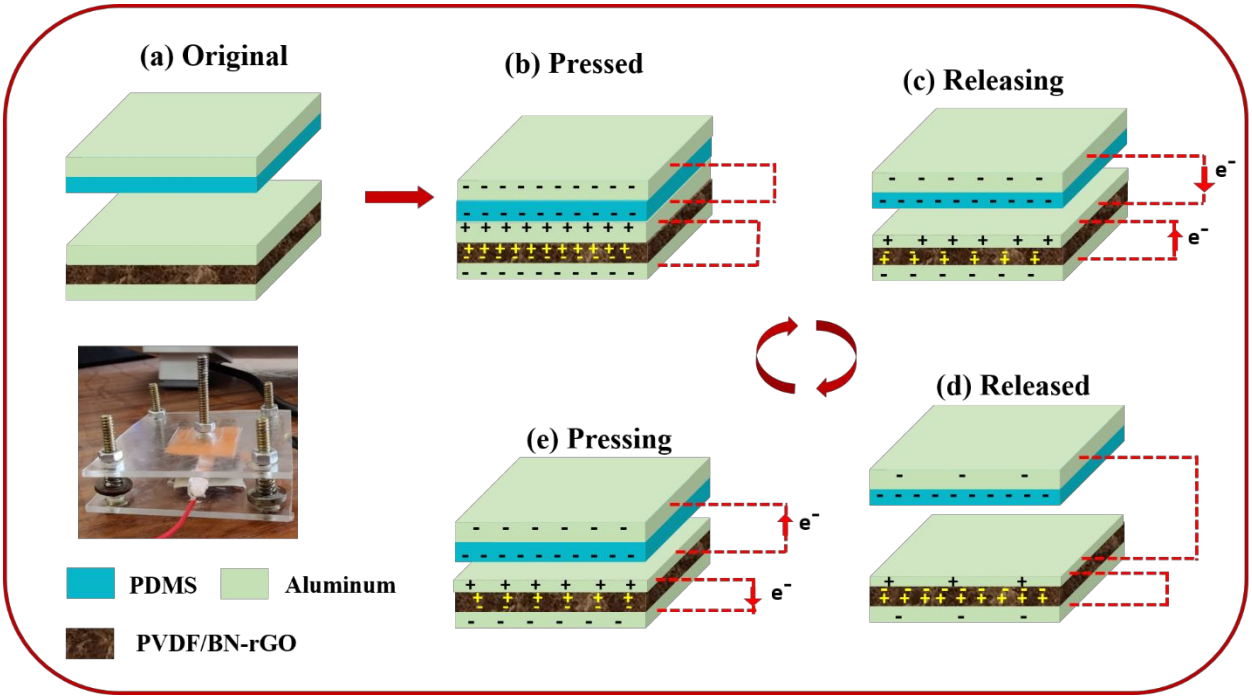


Figure S4. (a-e) A schematic showing the working mechanism of the HNG device including the real photographic image of the HNG

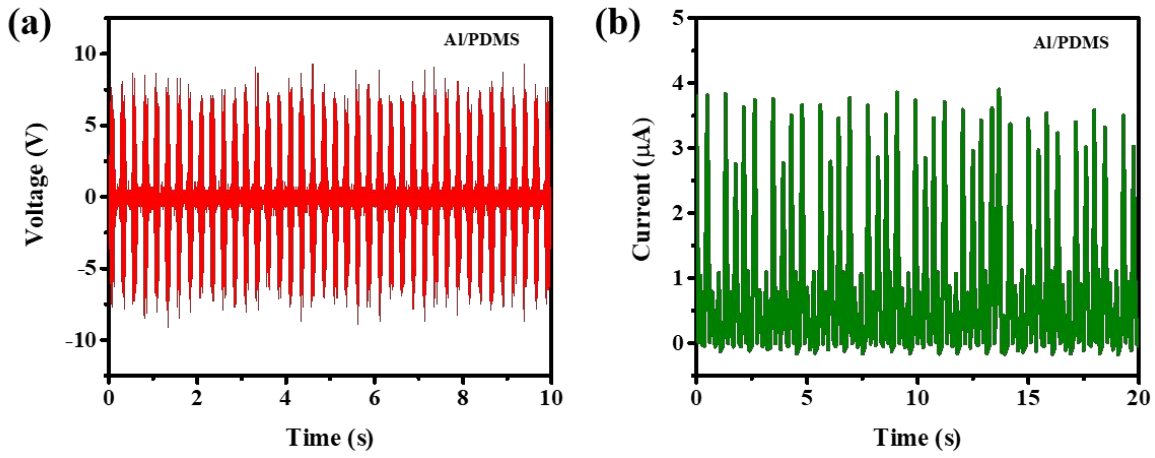


Figure S5. (a) The open-circuit voltage and (b) short-circuit current measurement of TENG with aluminum and PDMS as active triboelectric layer.

DOI: 10.5958/0976-2469.2022.00033.1

LIBRARY HERALD
Vol 60 No 3
September 2022

Textbooks and Their Free Availability on WWW: A Case Study of University of Delhi

LALITA*
NARENDER KUMAR**

1 INTRODUCTION

New normal: online teaching, online/open-book examination, and lockdownlike frequent lock down, work from home etc.,after COVID-19 pandemic has forced educational institutions to revisit their policies for imparting education to their students. Physical classroom teaching is being replaced with online instruction. Paper pen examination mode is being replaced by online examination. Libraries have been giving more impetus on online services rather than traditional services.The pandemic forced millions of students and faculty members to pursue their studies from their respective homes. This new normal has also demanded that the libraries prioritize providing digital access to textbooks and academic materials. The new normal have compelled the libraries to shift their focus from traditional print to digital resources. Within the digital resources, the libraries are focusing more on e-textbookslisted in syllabito meet the demands of studentsforcontinuing their studies.

2 E-TEXTBOOKS – ISSUES

The following are significant issues related to syllabus textbooks in digital format:

2.1 NON-AVAILABILITY

Mosttextbooks are not available in e-book format as publishers prefer to sell these titles in print format.

* Lalita, Assistant Librarian, Delhi Technological University. lalitakumar69@gmail.com

** Narender Kumar, Former Librarian, University of Delhi, narenderkumar1959@gmail.com

2.2 EXORBITANT PRICE

In case of availability, the institutional license price is too high for a library to buy/subscribe to them. The primary reason for this hyperbolized price is a lack of competition. The extremely high cost of e-textbooks makes them unaffordable and unsustainable for academic libraries. Angered by the unexpected hike in price during the pandemic, academicians, students, and librarians have been raising their voices against the unjustifiable pricing of e-textbooks.

In an open letter to the UK Government, thousands of academicians, including librarians, demanded an investigation into the academic e-textbooks market during the pandemic. The letter stated that

- "Copyright law disallows educational establishment from whole scanning books they own in print."
- "Price rises are common, sudden and appear arbitrary. Recently two well-known academic publishers raised the cost of a single-user e-book by 200% or more with no warning earlier this year"[1]

A subject librarian, Johanna Anderson, lambasted the publishing industry while saying, "Publishers are manipulating the market and price gouging from COVID. We are trying to support students during an unprecedented public health crisis, and they are making it so much harder."[2]Supporting the argument, Shankar, a prominent Indian scientist, says, "The commercial publishers, who run scientific publishing like a fiefdom, charging exorbitant subscriptions or publishing fees making exponential profits, and treating the intellectual output of scientists and institutions as if it was all their personal property."[3]

3 FREE ACCESS TO E-TEXTBOOKS

During the pandemic, some of the leading publishers offered a free trial of their e-textbooks, including referenced textbooks but then withdrew after a short period. The new normal "library-access-from-home" and anunprecedented price increaseof e-textbooks compelled users to visit the World Wide Web (WWW)to access freely available full-text of these e-textbooks.

4 OBJECTIVES

The objectives of the study are to find out availability of free PDF version of:

- Textbooks listed in the syllabi of all courses offered at undergraduate and post graduate level by the University of Delhi.
- Faculty wise analysis of available e-textbooks listed in the syllabi of undergraduate and postgraduates' courses offered by different faculties at the University of Delhi.

- Unique titles of textbooks listed in the syllabi of all courses offered by the University of Delhi.

5 LITERATURE SURVEY

Open/free digital sources have become a thing of routine. Scholars/ academicians have entering into an agreement for Creative Common CC) Licensed. "CC copyright licenses and tools forge a balance inside the traditional all rights reserved setting that copyright law. The combination of our tools and our users a vast and growing digital commons, a pool of content that can be copied, distributed, edited, remixed, and build upon, all within the boundaries of copyright law".[4] Libraries have been providing links to the websites for obtaining free and low-cost e-textbook.

Das and Hazara conducted a study on four online databases providing free access to electronic resources i.e., Education Resource Information Centre (ERIC); COnnectingREpository (CORE); PDF Drive; and Project Gutenberg. One of the major conclusions of the study is that "in search type category PDF Drive achieved highest score whereas CORE got the lowest score".[5] In another study an overview has been given about the open access resources in science and technology. In the study they have categorized PDF Drive as a source for open access books.[6]Anyira and Idubor also listed PDF Drive as an "open Access Books and Journals"[7]

The academic community does not demand textbooks to be available free of cost. Still, they expect a reasonable price that makes it profitable for all parties - students, libraries, and publishers. Efforts are being made to make textbooks affordable.[8]To support the open e-textbook movement, the US Department of Education has distributed $7 million in grants in 2020 to help open textbook creation.[9] Another notable effort is from Open Educational Resources (OER) Commons. OER Commons is a platform for educational resources. The main objective of OER Common is to "explore, create, and collaborate with educators around the world to improve curriculum".[10]

Ramaiah mentioned that "in the US several state governments have started free digital textbooks to save money and improve twenty-first century learning. A few non-profitable organizations have started offering free opensources as an initiative".[11]In 2008 Arunachalam mentioned a brief list of sources of open access books.[12]

The authors have not been able to find any study keeping in view the open e-textbooks sources specific for titles listed in syllabus of an academic institute.

6 METHODOLOGY

The University of Delhi offers 76 postgraduate and 75 undergraduate courses through 10 faculties and 19 Academic Centers. A database of references

listed in 173 syllabi/papers available on www.du.ac.in[13]has been compiled in Microsoft Excel. There wereapproximately 49,691 references, including various types of documents in different languages. These references were segregated according to the kind of documents like textbooks, serials, and other resources. The textbooks were further segregated as per their languages.

After said segregation,the textbooks in the English language were searched on WWW for the availability of their free PDF versions. To verify the PDF availability, retrieved URLs of the references were noted and analyzed. While analyzing, it was observed that a major part of the textbooks was found through a search engine called PDF DRIVE. After considering it, all the references found on WWW were categorized into two categories, i.e., Open Source and PDF DRIVE. Open sources include websites like www.archive.org, www.gutunburg.org, www.ignca.org, etc.

PDF Drive established in March 2017 with the sponsorship of Asaha Inc., San Diego, United Statesis a free search engine which allows to search, preview, and download millions of PDF files. Its crawlers are constantly scanning the WWW to add pdf files to its database. PDF Drive respects the intellectual property rights of others and expects its users to do the same. In the case that PDF files are withdrawn from the web, then they are also immediately withdrawn from PDF Drive search results.[14]

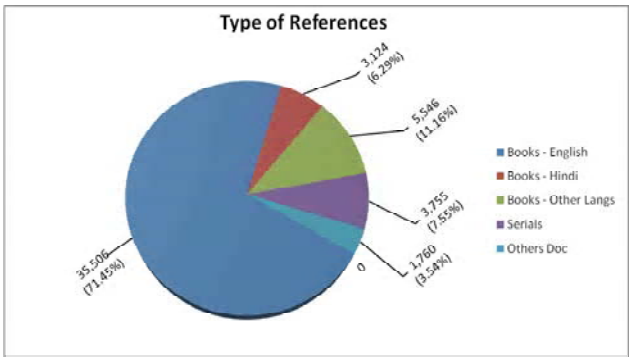
The results have been presented in the form of tables and graphs with their analysis.

7 ANALYSIS

7.1 TOTAL REFERENCES

The total references have been categorized on various parameters like language, type of documents,and presented through Tables and Graphs. References in English language books have been analyzed in detail.

Graph-1



Graph-1 depicts the various type of references listed in 173syllabi of courses/papers offered by the University of Delhi at undergraduate and postgraduate levels. The graph reveals that there are 49,691 total references of various types of documents. The graph also indicates that out of these 49,691 references, 35,506 (71.45%) belong to books in the English language, 3,124 (6.29%) to books in the Hindi language, and 5,546 (11.16%) to books in other languages. The references of other documents include websites andYouTube etc., account for 1,760 (3.54%). Various courses/papers also have references from serial publications accounting for 3,755references (7.55%).

Inference: Graph-1 reveals that majority of references belong to books in the English language (71.45%), followed by books in other languages (11.16%) Hindi language (6.29%). References from serials constitute 7.55%, followed by references from numerous websites (3.54%).

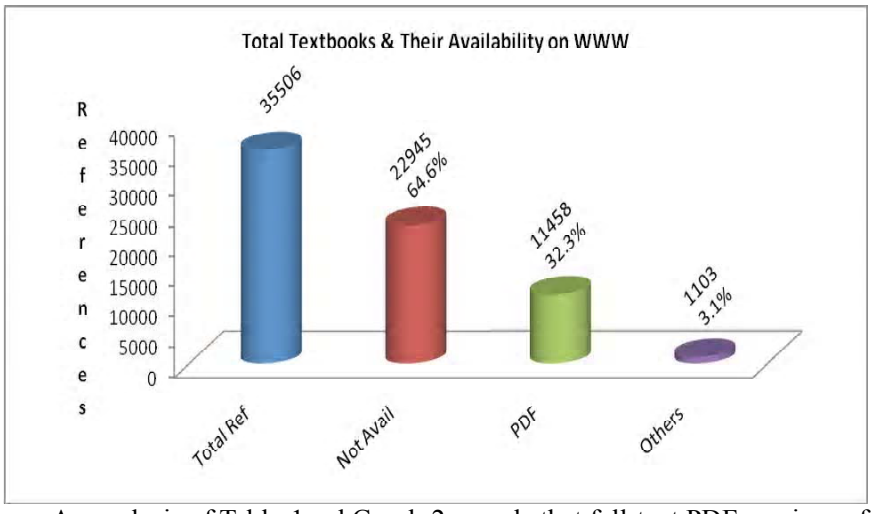
7.2 TOTAL TEXTBOOKS AND THEIR FULL-TEXT PDF AVAILABILITY ON WWW

Tables and Graphs from data related to textbooks in English language have been created to depict the availability of full-text PDF versionsof UG and PG courses offered by the University of Delhi. There are some courses/ papers where no English language books have been listed in the references. An analysis of the available full-text PDF versionsof textbooks has also been carried out based on the faculties offering these UG and PG courses.

Table-1: Total Textbooks and Their Full-Text PDF Availability

Faculties	Undergraduate (UG)					Postgraduate (PG)					Total Ref (UG+PG)
	No of Courses	Open Sources	PDF Drive	Not Found	Grand Total	No of Courses	Open	PDF Drive	Not Found	Total	
Applied Social Sciences and Humanities	12	19	472	1,197	1,688						1,688
Arts	34	153	859	1,521	2,533	26	131	1,270	2,406	3,807	6,340
Cluster Innovation Centre	2	24	61	115	200	1	1	46	60	107	307
Commerce and Business Studies	5	11	271	987	1,269	4	33	328	684	1045	2,314
Education	1	15	18	74	107						107
Inter-Disciplinary and Applied Sciences	5	4	501	1,076	1,581	7	3	451	277	731	2,312
Law						2	150	153	804	1,107	1,107
Management Studies						3	27	565	988	1,580	1,580
Mathematical Sciences	15	8	530	411	949	3		250	226	476	1,425
Music and Fine Arts	5	4	18	320	342	3	12	2	222	236	578
Science	20	129	1,020	2,663	3,812	10	11	982	1,132	2,125	5,937
Social Science	13	86	778	2,432	3,296	7	282	2,883	5,350	8,515	11,811
Grand Total	**112**	**453**	**4,528**	**10,796**	**15,777**	**66**	**650**	**6,930**	**12,149**	**19,729**	**35,506**

Graph-2



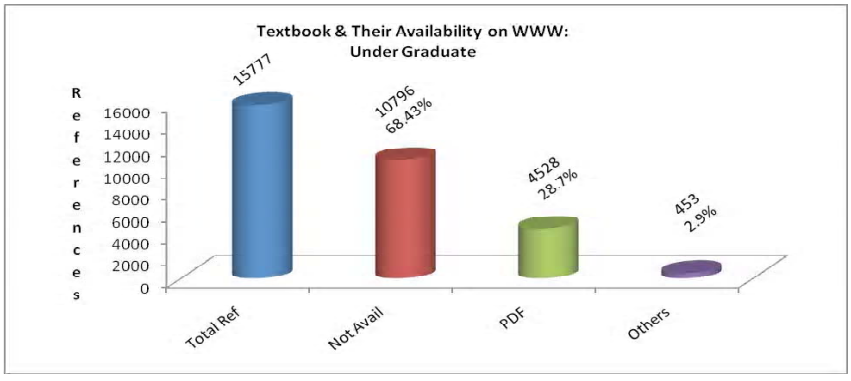
An analysis of Table-1and Graph-2 reveals that full-text PDF versions of 35.4% (12,561) of total textbooks (35,506) are available on WWW. Out of the available full-text textbooks, 32.3% (11,458) are available from PDF DRIVE and 3.1% (1,103) from other open sources.

Inference: Full-text PDF of nearly one-third of the textbooks listed in UG and PG courses in the English language are available on WWW.

7.3 UG TEXTBOOKS AND THEIR FULL-TEXT PDF AVAILABILITY ON WWW

Data has been further analyzed based on courses offered by various faculties at UG and PG levels. Graph-3 given below provides a glimpse of the availability of full-text PDF versions oftextbooks for UG courses offered by the University of Delhi.

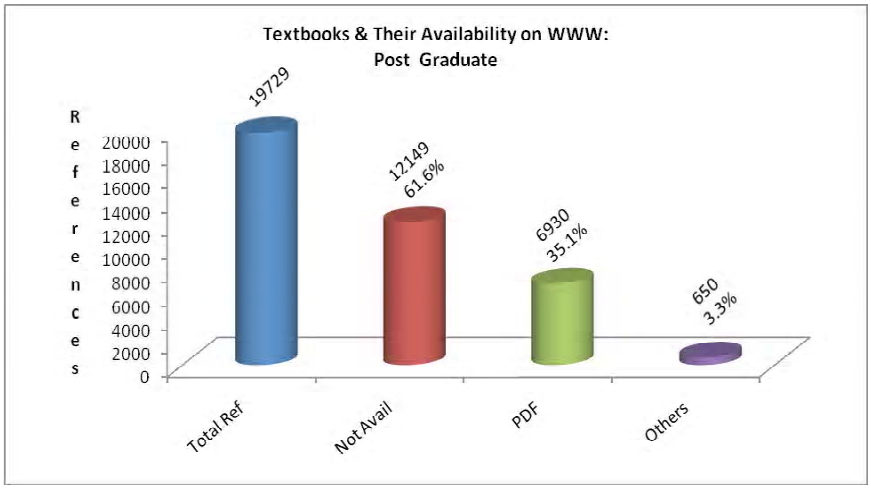
Graph-3



An analysis of Table-1 and Graph-3 reveals that there are 15,777 textbooks in the English language listed in various courses/papers offered by the different faculties at the UG level. While analyzing their availability on WWW, it has been found that full text PDF versions of 31.6% (5,035) of total UG textbooks are freely available. However, the most accessible UG textbooks are available on PDF DRIVE (28.7%), and the remaining (2.9%) are available on other open sources.

Inference: Full-text PDF of 31.6% textbooks listed under UG courses/ papers offered by the University of Delhi are available on WWW.

Graph-4



In-depth analysis of Table-1 Graph-4 further reveals that out of 35,506 textbooks, 19,729 are listed in PG courses offered by the various faculties of the University of Delhi. While analyzing the availability of full text PDF versions of these PG textbooks, 38.4% (7,560) of the total are available on WWW. Further analysis of available textbooks shows that 35.1% (6,930) PG textbooks are available on PDF DRIVE and 3.3% (650) on other open sources.

Inference: Full-text PDF of 38.4% textbooks are available on WWW. Further generalization of the analysis is concluded that there are more chances of full text PDF availability of PG textbooks (38.4%) than UG (31.6%).

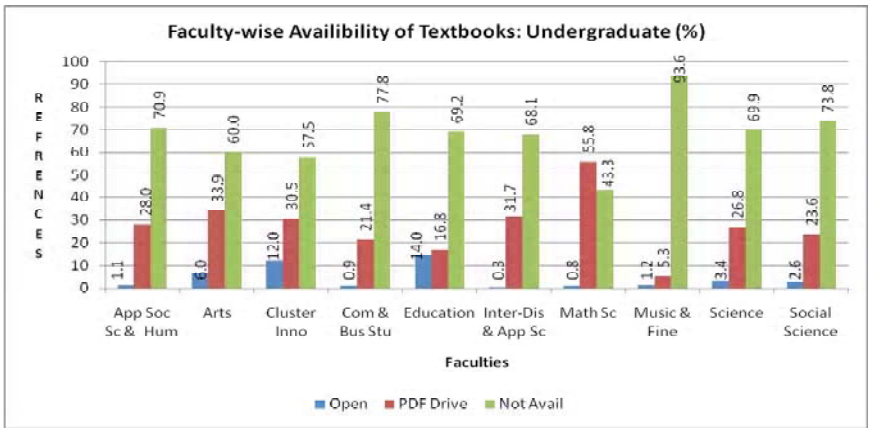
7.4 FACULTY-WISE AVAILABILITY OF TEXTBOOKS AT UNDERGRADUATE LEVEL

The study on the availability of full text PDF versions of textbooks has also been done based on courses offered by different faculties of the University of Delhi. This analysis has been based on courses offered at UG and PG levels.

Graph-5, given below, analyses the availability of textbooksfrom UG courses. The Graph shows an exciting picture in the case of the Faculty of Mathematical Sciences. It reveals that full text PDF versions of more than half of the textbooks (56.7%) given in various UG courses offered by the Faculty of Mathematical Sciences are available on WWW.The Faculty of Mathematical Sciences is followed by Cluster Innovation Centre (42.5%) and Faculty of Arts (40%) for the availability of full text PDF versions of textbooks for UG courses available on WWW. The textbooks of UG courses/papers offered by the Faculty of Music and Fine Arts has the lowest (6.4%) availability of full text PDF versions on WWW.

The faculties where availability of full text PDF versions of textbooks ranges between 26% to 32% are the Faculty of Inter-Disciplinary and Applied Sciences (32%), the Faculty of Sciences (31.1%), the Faculty of Education (30.8%), the Faculty of Social Sciences and Humanities (29.1%), andthe Faculty of Social Sciences (26.2%).

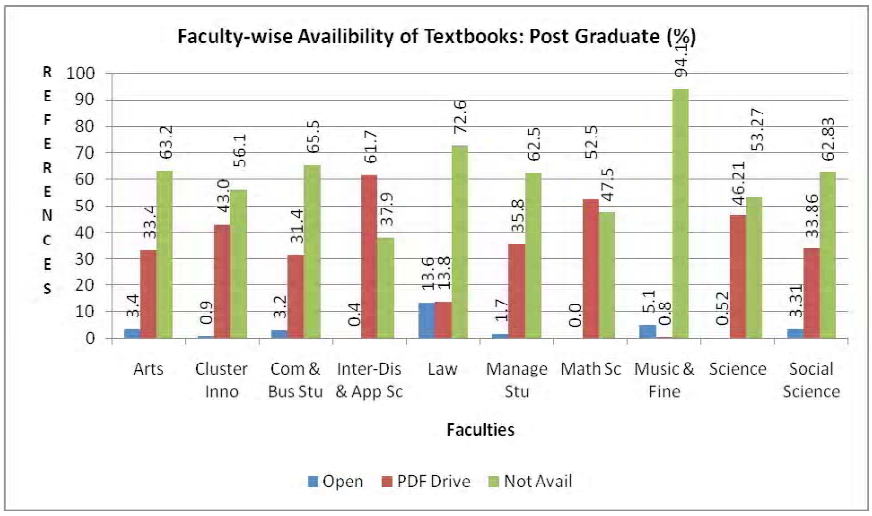
Graph



7.5 FACULTY-WISE AVAILABILITY OF TEXTBOOKS AT POSTGRADUATE LEVEL

Graph-6 below describes the faculty wise availability of full text PDF versions oftextbooks listed in postgraduate courses/papers offered by various faculties of the University of Delhi.

Graph-6



While analyzing the availability of full text PDF versions of textbooks listed in PG courses/papers from various faculties of the University of Delhi, an almost similar trend as that with the references of UG courses has been observed as depicted in Graph-6. However, compared to the UG courses, a more significant percentage of full-text PDF versions were found to be available on WWW of textbooks listed for PG courses/papers.

The analysis revealed that 62.1% of full-text PDF versions of textbooks of PG courses/papers of theFaculty of Inter-Disciplinary and Applied Sciences are available on WWW. It has been followed by the Faculty of Mathematical Sciences (52.5%), and the lowest available number of textbooks belonged to the Faculty of Music and Fine Arts. In other words, two-third of total textbooks of PG courses offered by the Faculty of Inter-Disciplinary and Applied Sciences are available either on PDF Drive (61.7%) or on Open Sources (0.4%). It is also worth noticing that more than half the textbooks of the PG course offered by the Faculty of Mathematical Sciences are available on WWW. The lowest availability of full text PDF versions belongs to the Faculty of Music and Fine Arts in UG and PG.

The availability of full text PDF versions of textbooks listed in PG courses/ papers offered by different faculties are as follows: the Faculty of Sciences (46.7%), the Cluster Innovation Centre (43.9%), the Faculty of Management Studies (37.5%), the Faculty of Social Sciences (37.2%), the Faculty of Arts (36.8%), the Faculty of Commerce and Business Studies (34.6%), the Faculty of Law (27.4%).

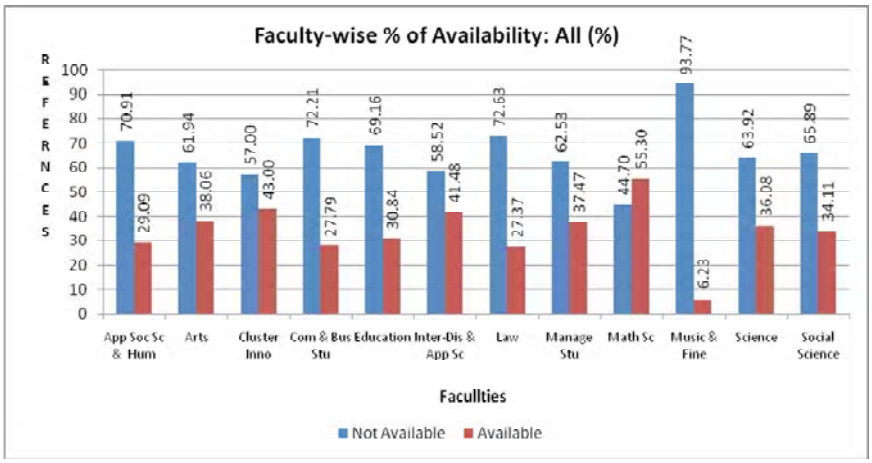
Inference: There is more possibility of finding full-text PDF versions of references listed in PG courses than UG courses offered by the University of Delhi. The full text PDF versions of textbooks of the courses/papers offered

by the Faculty of Inter-Disciplinary and Applied Sciences are most likely to be found on WWW. In contrast, in the case of the Faculty of Music and Fine Arts it is least likely.

7.6 TOTAL AVAILABILITY OF FULL-TEXT PDF VERSIONS ON WWW

A comprehensive analysis, faculty-wise, of all the full text PDF versions of textbooks available on WWW (PDF Drive+Open Sources) irrespective of UG level or PG level, has been carried out. This result for the same has been depicted in the form of given below Graph-7.

Graph-7



An analysis of Graph-7 shows that full text PDF versions of textbooks on WWW from all the sources are highest in the Faculty of Mathematical Sciences, where available resources (55.3%) have surpassed the non-available (44.7%). The Faculty of Mathematical Sciences is followed by the Cluster of Innovation Centre (43%) and the Faculty of Inter-Disciplinary and Applied Sciences (41.48%). Subsequently, these are followed by different faculties, including the Faculty of Arts (38.06%), the Faculty of Management Studies (37.47%), the Faculty of Sciences (36.08%), the Faculty of Social Sciences (34.11%), the Faculty of Education (30.84%), and the Faculty of Law (27.37%). The lowest availability of full text PDF versions of textbooks has been accounted to the Faculty of Music and Fine Arts, where only 6.23% of textbooks are available on WWW.

Inference: Full-text PDF versions of textbooks listed by the Faculty of Mathematical Sciences and the Faculty of Inter-Disciplinary Applied Sciences in their different courses/papers offered are more like to be found on WWW than Social Sciences or Arts faculties.

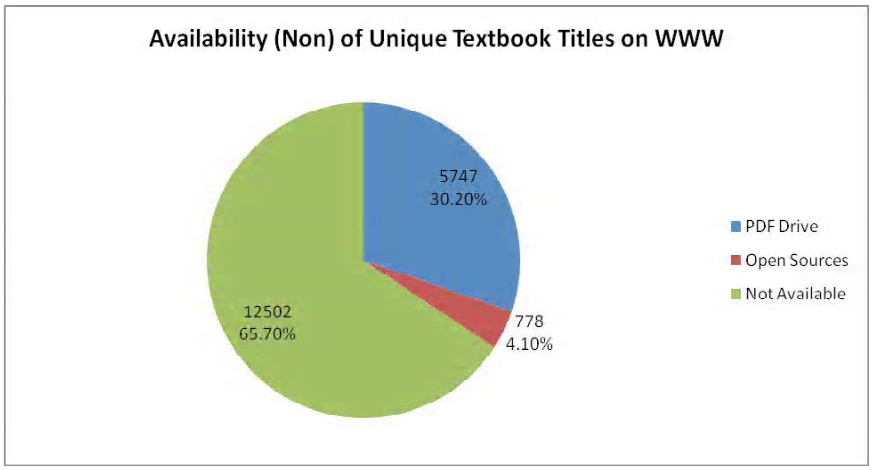
7.7 TOTAL UNIQUE TEXTBOOKS TITLES AND THEIR AVAILABILITY ON WWW

Table-2 below depicts those 19,027 unique textbooks titles cover 35,506 English language book references listed in all the courses/papers and their availability on WWW. This means that a single textbook title may have been listed as a reference in more than one paper/course.

Table-2: Availability (Non) of Unique Titles

Source	**Availability(Non) of**	**% Of Availability (Non) of**
	Unique Titles	
PDF Drive	5747	30.2
Open Sources	778	4.1
Not Available	12502	65.7
Total	**19027**	**100**

Graph-8



Analysis of Table-2 and Graph-8 reveals that out of 19,027 unique textbooks titles, 5,747 (30.2%) are available as full text PDF versions on PDF Drive, whereas 778 (4.1%) are available on other open sources. In total, the full text of 6,525 (34.4%) unique textbooks titles are available on WWW. These 6,525 titles cover 12,561 references in English language books, accounting for 35.38% of the English language references.

Inference: Almost one-third of total references of textbooks in the English language listed in various courses of UG and PG are available on WWW.

8 CONCLUSION

Full-text PDF versionsformore than one-third of textbooks (12,561, 35.4%)in the English language listed in UG and PG courses/papers of the University of Delhi are available on WWW. The primary source of free pdf full text is PDF Drive, accounting for 32.3% (11,458) oftotal textbook. Other open sources contribute 3.1% (1,103). Full-text PDF of 31.6% of textbooks listed under UG courses/papers and 38.4% of textbooks of PG references offered by the University of Delhi are available on WWW. It is further concluded that there are more possibilities of finding Full-text PDF of PG courses than UG courses offered by the University of Delhi. Full-text PDF of textbooks of the courses/papers offered by the Faculty of Mathematical Sciences (55.3%) and Faculty of Inter-Disciplinary and Applied Sciences(41.5%) are more likely to be found on WWW.In contrast, in the Faculty of Music and Fine Arts case, it is least (6.3%)likely. Out of a total (19,025) unique textbook titles, full text PDF versions of 6,525 (34.4%) accounting for 12,561 references are available on WWW. PDF Drive accounts for 30.2% (5,747) of total unique textbooks titles, whereas other open sources contribute 4.1% (778).

9 SUGGESTIONS

The non-availability of listed textbooks in digital format and skyrocketing prices, if available have made it very difficult for academic institutions to acquire these e-textbooks. Keeping these in view, the following suggestions are being put forward:

- The faculties should recommend alternative titles available on open sources.
- The faculties should concentrate on creating their academic material fulfilling the demands of the students for syllabus textbooks and make them available on open sources.

REFERENCES

1. COMPAIGN TO investigage the academic ebook market, 2021. https://academicebookinvestigation.org/ (Accessed on May 7, 2021)
2. FAZACKERLEY (A). Price gouging from Covid: student ebooks costing up to 500% more than in print, 29 January 2021. https://www.theguardian.com/education/2021/jan/29/price-gouging-from-covid-student-ebooks-costing-up-to-500-more-than-in-print (Accessed on May 7, 2021)
3. SHANKAR RAMAN (T R). Why I won't review or write for Elsevier and other commercial scientific journals., 4 April 2021. https://

science.thewire.in/the-sciences/why-i-wont-review-or-write-for-elsevier-and-other-commercial-scientific-journals/ (Accessed on May 7, 2021)

4. CREATIVE COMMONS. About the licenses, [Online]. Available: https://creativecommons.org/licenses/ [Accessed 19 Aug 20121].
5. DAS (R) and HAZRA (U K). As assessment of educational resources through ERIC, CORE, PDF drive and Project Gutenberg, *Library Philosophy and Practice (e-journal),* vol. 2181, 2019.
6. AHMED (H F A) and SOLUCHANA (A). Open access resources in science and technology: an overview, *Library Philosophy and Practice,* pp. 1-12, 2019.
7. ANYIRA (I E) and LBUDOR (I).Open access resources: Option for Libraries in a Challenged Economy, *Library Philosophy and Practice (e-journal),* no. 4202, 2020.
8. MAKE TEXTBOOKS affordable, 2021. https://studentpirgs.org/campaigns/make-textbooks-affordable/ (Accessed on May 7, 2021)
9. NAGLE (C). The 2020 open textbook pilot: Why it matters and how to apply., 17 September 2020. https://studentpirgs.org/2020/09/17/the-2020-open-textbook-pilot-why-it-matters-and-how-to-apply/ (Accessed on May 7, 2021)
10. O. E. R. Commons. Explore, create, collaborate, 2021 https://www.oercommons.org/ (Accessed on May 7, 2021)
11. RAMAIAH (CK). E-books: past, present and future, *DESIDOC J. Libr. Inf. Technol.*, March, 2012,**32**(2), 79-82.
12. ARUNACHALAM (S). Open access to scientific knowledge. DESIDOC J. Libr. Inf. Technol., Jan, 2008, **28** (1), 7-14.
13. DELHI (University of -). Revised CBCS syllabi, 2018. http://du.ac.in/index.php?page=revised-cbcs-syllabi (Accessed on May 7, 2021)
14. PDF Drive," [Online]. Available: https://www.pdfdrive.com/home/about [Accessed 19 Aug 2021]

arXiv:2210.13467v1 [math.CV] 24 Oct 2022

Toeplitz Determinants for a Class of Holomorphic Mappings in Higher Dimensions

Surya Giri and S. Sivaprasad Kumar

Abstract

In this paper, we establish the sharp bounds of certain Toeplitz determinants formed over the coefficients of mappings from a class defined on the unit ball of complex Banach space and on the unit polydisc in $\mathbb{C}^n$. Derived bounds provide certain new results for the subclasses of normalized univalent functions and extend some known results in higher dimensions.

Keywords: Quasi-convex mappings; Toeplitz determinants; Coefficient inequalities.

AMS Subject Classification: 32H02, 30C45.

1 Introduction

Let $\mathcal{S}$ be the class of analytic univalent functions in the unit disk $\mathbb{U} = \{z \in \mathbb{C} : |z| < 1\}$ having the form $g(z) = z + \sum_{n=2}^{\infty} b_n z^n$ and $\mathcal{K}(\alpha) \subset \mathcal{S}$ denote the class of convex functions of order α, $0 \leq \alpha < 1$. A function $g \in \mathcal{K}(\alpha)$ if and only if

$$\operatorname{Re}\left(1 + \frac{zg''(z)}{g'(z)}\right) > \alpha, \quad z \in \mathbb{U}.$$

For $\alpha = 0$, the class $\mathcal{K}(\alpha)$ reduces to the class of convex functions $\mathcal{K} := \mathcal{K}(0)$. A function $g \in \mathcal{S}$ is said to be starlike of order α if and only if

$$\operatorname{Re}\left(\frac{zg'(z)}{g(z)}\right) > \alpha, \quad z \in \mathbb{U}.$$

The class of all starlike functions of order α is denoted by $\mathcal{S}^*(\alpha)$ and let $\mathcal{S}^* := \mathcal{S}^*(0)$. Recently, Ali et al. [2] obtained the bounds of certain Toeplitz determinants whose entries are the Taylor series coefficients of functions in $\mathcal{S}$ and some of its subclasses. For $g(z) = z + \sum_{n=2}^{\infty} b_n z^n$, the Toeplitz matrix is given by

$$T_{m,n}(g) = \begin{bmatrix} b_n & b_{n+1} & \cdots & b_{n+m-1} \\ b_{n+1} & b_n & \cdots & b_{n+m-2} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n+m-1} & b_{n+m-2} & \cdots & b_n \end{bmatrix}.$$

In particular, the second order Toeplitz determinant is

$$\det T_{2,2}(g) = b_2^2 - b_3^2 \tag{1.1}$$

and the third order Toeplitz determinant is given by

$$\det T_{3,1}(g) = \begin{vmatrix} 1 & b_2 & b_3 \\ b_2 & 1 & b_2 \\ b_3 & b_2 & 1 \end{vmatrix} = 2b_2^2 b_3 - 2b_2^2 - b_3^2 + 1. \tag{1.2}$$

Toeplitz matrices and Toeplitz determinants have various applications in pure as well as in applied mathematics. They occur in a variety of fields including partial differential equations, image processing and differential geometry. For more details and applications, we refer [18].

Numerous articles have recently focused on finding sharp estimates for the Toeplitz and Hermitian-Toeplitz determinants for various classes, but in one dimensional complex plane. Ahuja et al. [1] established the sharp bounds of $|\det T_{2,2}(g)|$ and $|\det T_{3,1}(g)|$ for the class $\mathcal{K}$ and its subclasses. For more work in this direction, we refer [12, 11, 4] and the references cited therein. For the class $\mathcal{K}$ and $\mathcal{K}(\alpha)$, the following bounds are proved in [1].

Theorem A. *[1] If $f \in \mathcal{K}$, then $|\det T_{2,2}(f)| \leq 2$. The bound is sharp.*

Theorem B. *[1] If $f \in \mathcal{K}$, then $|\det T_{3,1}(f)| \leq 4$. The bound is sharp.*

Theorem C. *[1] If $f \in \mathcal{K}(\alpha)$, then the following sharp inequality hold:*

$$|\det T_{2,2}(f)| \leq \frac{2(1-\alpha)^2(2\alpha^2-6\alpha+9)}{9}.$$

Theorem D. *[1] If $f \in \mathcal{K}(\alpha)$ and $\alpha \in [0, 1/2]$, then the following sharp inequality hold:*

$$|\det T_{3,1}(f)| \leq \frac{8\alpha^4-34\alpha^3+71\alpha^2-72\alpha+36}{9}.$$

In this paper, we generalize the above results in higher dimensions for a class of holomorphic mappings defined on the unit ball in complex Banach space and on the unit polydisc in $\mathbb{C}^n$. Let X be a complex Banach space with respect to norm $\|\cdot\|$ and $\mathbb{B} = \{z \in X : \|z\| < 1\}$ be the unit ball. When $X = \mathbb{C}$, $\mathbb{B}$ is denoted by $\mathbb{U}$. Let $\mathbb{C}^n$ denote the space of $n-$complex variables $z = (z_1, z_2, \cdots, z_n)'$ and $\mathbb{U}^n$ be the Euclidean unit ball in $\mathbb{C}^n$. The boundary and distinguished boundary of $\mathbb{U}^n$ are denoted by $\partial\mathbb{U}^n$ and $\partial_0\mathbb{U}^n$, respectively.

Let $L(X, Y)$ denote the set of all continuous linear operators from X into a complex Banach space Y. For each $z \in X \setminus \{0\}$, let

$$T_z = \{l_z \in L(X, \mathbb{C}) : l_z(z) = \|z\|, \|l_z\| = 1\}.$$

By the Hahn-Banach theorem, this set is non-empty.

By $\mathcal{H}(\Omega, \Omega')$, we denote the set of holomorphic mappings from a domain $\Omega \subseteq X$ into a domain $\Omega' \subseteq Y$ and let $\mathcal{H}(\Omega) = \mathcal{H}(\Omega, X)$. If $g \in \mathcal{H}(\mathbb{B})$ and $z \in \mathbb{B}$, then for each $k = 1, 2, \cdots$, there is a bounded symmetric $k-$linear mapping

$$D^k g(z) : \prod_{j=1}^{k} X \to X,$$

called the k^{th} order Fréchet derivative of g at z such that

$$g(w) = \sum_{k=0}^{\infty} \frac{1}{k!} D^k g(z)((w-z)^k)$$

for all w in some neighborhood of z.

A mapping $g \in \mathcal{H}(\Omega)$ is said to be biholomorphic if $g(\Omega)$ is a domain in X and the inverse g^{-1} exists and is holomorphic on $g(\Omega)$. If the Fréchet derivative $Dg(z)$ has a bounded inverse for each $z \in \Omega$, then $g \in \mathcal{H}(\Omega)$ is called locally biholomorphic mapping on Ω. Analog of the class $\mathcal{S}$, let $\mathcal{S}(\mathbb{B})$ denote the class of biholomorphic mappings g from $\mathbb{B}$ into X, satisfying $g(0) = 0$ and $Dg(0) = I$, where I represents the linear identity operator from X into X. It is easily seen that $\mathcal{S}(\mathbb{B})$ is not a normal family when the dimension is greater than one [3]. A mapping $g \in \mathcal{S}(\mathbb{B})$ is said to be starlike if $g(\mathbb{B})$ is starlike with respect to the origin. Also, a mapping $g \in \mathcal{S}(\mathbb{B})$ is said to be convex if $g(\mathbb{B})$ is convex.

On a bounded circular domain $\Omega \subset \mathbb{C}^n$, the first and the m^{th} Fréchet derivative of a holomorphic mapping $g : \Omega \to X$ are written by $Dg(z)$ and $D^m g(z)(a^{m-1}, \cdot)$ respectively. The matrix representations are

$$Dg(z) = \left(\frac{\partial g_j}{\partial z_k}\right)_{1 \leq j,k \leq n},$$

$$D^m g(z)(a^{m-1},\cdot)=\left(\sum_{p_1,p_2,\cdots,p_{m-1}=1}^{n}\frac{\partial^m g_j(z)}{\partial z_k\partial z_{p_1}\cdots\partial z_{p_{m-1}}}a_{p_1}\cdots a_{p_{m-1}}\right)_{1\le j,k\le n},$$

where $g(z)=(g_1(z),g_2(z),\cdots g_n(z))'$ and $a=(a_1,a_2,\cdots a_n)'\in\mathbb{C}^n$.

Liu and Liu [13] defined the following class:

Definition 1.1. [13] Suppose $\alpha\in[0,1)$ and $g:\mathbb{B}\to X$ is a normalized locally biholomorphic mapping. If

$$\operatorname{Re}\left\{l_z[(Dg(z))^{-1}(D^2g(z)(z^2)+Dg(z)(z))]\right\}\ge\alpha\|z\|,\quad l_z\in T_z,\ z\in\mathbb{B}\setminus\{0\},$$

then f is called a quasi convex mapping of type B and order α on $\mathbb{B}$.

If $\mathbb{B}=\mathbb{U}^n$ and $X=\mathbb{C}^n$, then the above condition reduces to

$$\left|\frac{q_k(z)}{z_k}-\frac{1}{2\alpha}\right|<\frac{1}{2\alpha},\quad\forall z\in\mathbb{U}^n\setminus\{0\},$$

where

$$q(z)=(q_1(z),q_2(z),\cdots,q_n(z))'=(Dg(z))^{-1}(D^2g(z)(z^2)+Dg(z)(z))$$

is a column vector in $\mathbb{C}^n$ and k satisfies

$$|z_k|=\|z\|=\max_{1\le j\le n}\{|z_j|\}.$$

For $\mathbb{B}=\mathbb{U}$ and $X=\mathbb{C}$, the relation is equivalent to

$$\operatorname{Re}\left(1+\frac{zg''(z)}{g'(z)}\right)>\alpha,\quad z\in\mathbb{U}.$$

Let $\mathcal{K}_\alpha(\mathbb{B})$ denote the class of quasi convex mappings of type B and order α.

When $\alpha=0$, Definition 1.1 is the definition of quasi convex mapping of type B introduced by Roper and Suffridge [15].

Definition 1.2. Let $\Phi:\mathbb{U}\to\mathbb{C}$ be a biholomorphic function such that $\operatorname{Re}\Phi(z)>0$, $\Phi(0)=1$, $\Phi'(0)>0$ and $\Phi''(0)\in\mathbb{R}$. Let $\mathcal{M}_\Phi$ be the class of mappings given by

$$\mathcal{M}_\Phi=\left\{p\in\mathcal{H}(\mathbb{B}):p(0)=Dp(0)=I,\ \frac{l_z(p(z))}{\|z\|}\in\Phi(\mathbb{U}),\ z\in\mathbb{B}\setminus\{0\},l_z\in T_z\right\}.$$

In case of $\mathbb{B}=\mathbb{U}^n$ and $X=\mathbb{C}^n$, we have

$$\mathcal{M}_\Phi=\left\{p\in\mathcal{H}(\mathbb{B}):p(0)=Dp(0)=I,\ \frac{p_j(z)}{z_j}\in\Phi(\mathbb{U}),\ z\in\mathbb{U}^n\setminus\{0\}\right\},$$

where $p(z)=(p_1(z),p_2(z),\cdots,p_n(z))'$ is a column vector in $\mathbb{C}^n$ and j satisfies $|z_j|=\|z\|=\max_{1\le k\le n}\{|z_k|\}$.

In 1999, Roper and Suffridge [15] gave a sufficient condition for a normalized biholomorphic convex mapping on the Euclidean unit ball in $\mathbb{C}^n$. Later, Zhu [20] provided a brief proof of this theorem. Xu et al. [17] obtained the sharp bounds of Fekete-Szegö inequality for the class of quasi-convex mappings of type B and order α defined on the unit ball $\mathbb{B}$ and on the unit polydisc in $\mathbb{C}^n$. Liu and Liu [14] derived the sharp estimates of all homogenous expansions for a subclass of holomorphic mappings of quasi-convex mappings of type B and order α in higher dimensions. Contrary to the coefficient inequalities for many subclasses of $\mathcal{S}$, only few are known for homogeneous expansions for subclasses of biholomorphic mappings in the case of several complex variables [5, 9, 8, 10, 6, 19].

In case of one complex variable, many coefficient problems are studied for the class $\mathcal{K}$ such as Theorems A-D. A natural question arises that how to retain these results in higher dimensions. Providing an answer to this question is the aim of this study.

2 Main Results

The main results of the paper are stated and proved in this section.

Theorem 2.1. *Let $g \in \mathcal{H}(\mathbb{B}, \mathbb{C})$ with $g(0) = 1$, $g(z) \neq 0$, $z \in \mathbb{B}$ and suppose that $G(z) = zg(z)$. If $(DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z)) \in \mathcal{M}_\Phi$ such that Φ satisfies*

$$|\Phi''(0) + 2(\Phi'(0))^2| \geq 2\Phi'(0) > 0,$$

then

$$\left| \left(\frac{l_z(D^2G(0)(z^2))}{2!||z||^2} \right)^2 - \left(\frac{l_z(D^3G(0)(z^3))}{3!||z||^3} \right)^2 \right| \leq \frac{(\Phi'(0))^2}{4} + \frac{(\Phi'(0))^2}{36} \left(\frac{1}{2} \frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0) \right)^2.$$

The bound is sharp.

Proof. Fix $z \in X \setminus \{0\}$ and let $h : \mathbb{U} \to \mathbb{C}$ be defined by

$$h(\zeta) = \begin{cases} \dfrac{l_z((DG(\zeta z_0))^{-1}(D^2G(\zeta z_0)((\zeta z_0)^2) + DG(\zeta z_0)\zeta z_0))}{\zeta}, & \zeta \neq 0, \\ 1, & \zeta = 0, \end{cases}$$

where $z_0 = \frac{z}{\|z\|}$. Then $h \in \mathcal{H}(\mathbb{U})$ and $h(0) = \Phi(0) = 1$. Since $(DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z)) \in \mathcal{M}_\Phi$, therefore, we have

$$\begin{aligned} h(\zeta) &= \frac{l_z((DG(\zeta z_0))^{-1}(D^2G(\zeta z_0)((\zeta z_0)^2) + DG(\zeta z_0)\zeta z_0))}{\zeta} \\ &= \frac{l_{z_0}((DG(\zeta z_0))^{-1}(D^2G(\zeta z_0)((\zeta z_0)^2) + DG(\zeta z_0)\zeta z_0))}{\zeta} \\ &= \frac{l_{\zeta z_0}((DG(\zeta z_0))^{-1}(D^2G(\zeta z_0)((\zeta z_0)^2) + DG(\zeta z_0)\zeta z_0))}{\|\zeta z_0\|} \in \Phi(\mathbb{U}), \quad \zeta \in \mathbb{U}. \end{aligned}$$

Applying a similar method used in [7, Theorem 7.1.14], we get

$$(DG(z))^{-1} = \frac{1}{g(z)} \left(I - \frac{\frac{zDg(z)}{g(z)}}{1 + \frac{Dg(z)z}{g(z)}} \right). \tag{2.1}$$

A simple computation using the fact $G(z) = zg(z)$ yields

$$D^2G(z)(z^2) + DG(z)(z) = (D^2g(z)(z^2) + 3Dg(z)(z) + g(z))z. \tag{2.2}$$

By using (2.1) and (2.2), it follows

$$(DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z)) = \frac{D^2g(z)(z^2) + 3Dg(z)(z) + g(z)}{g(z) + Dg(z)(z)} z. \tag{2.3}$$

Consequently

$$l_z((DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z))) = \frac{D^2g(z)(z^2) + 3Dg(z)(z) + g(z)}{g(z) + Dg(z)(z)} \|z\|. \tag{2.4}$$

Using (2.4), we obtain

$$\begin{aligned} h(\zeta) &= \frac{l_{\zeta z_0}((DG(\zeta z_0))^{-1}(D^2G(\zeta z_0)((\zeta z_0)^2) + DG(\zeta z_0)\zeta z_0))}{\|\zeta z_0\|} \\ &= \frac{D^2g(\zeta z_0)((\zeta z_0)^2) + 3Dg(\zeta z_0)(\zeta z_0) + g(\zeta z_0)}{g(\zeta z_0) + Dg(\zeta z_0)(\zeta z_0)}. \end{aligned}$$

Equivalently, we can write

$$h(\zeta)(g(\zeta z_0)+Dg(\zeta z_0)(\zeta z_0))=D^2g(\zeta z_0)((\zeta z_0)^2)+3Dg(\zeta z_0)(\zeta z_0)+g(\zeta z_0).$$

The series expansion in terms of ζ gives

$$\left(1+h'(0)\zeta+\frac{h''(0)}{2}\zeta^2+\cdots\right)\left(1+2Dg(0)(z_0)\zeta+\frac{3Dg(0)(z_0^2)}{2}\zeta^2+\cdots\right)$$
$$=1+4Dg(0)(z_0)\zeta+\frac{9Dg(0)(z_0^2)}{2}\zeta^2+\cdots.$$

Comparison of the homogenous expansions of either sides of the above equality provide $h'(0)=2Dg(0)(z_0)$. That is

$$h'(0)\|z\|=2Dg(0)(z). \tag{2.5}$$

Also, we have

$$\frac{D^2G(0)(z^2)}{2!}=Dg(0)(z)z,$$

which gives

$$\frac{l_z(D^2G(0)(z^2))}{2!}=Dg(0)(z)\|z\|.$$

Now, using $|h'(0)|\leq\Phi'(0)$ with (2.5), we obtain

$$\left|\frac{l_z(D^2G(0)(z^2))}{2!\|z\|^2}\right|\leq\frac{\Phi'(0)}{2}. \tag{2.6}$$

Moreover, for $\lambda\in\mathbb{C}$, Xu et al. [17, Theorem 3.1] proved that

$$\left.\begin{array}{l}\left|\dfrac{l_z(D^3G(0)(z^3))}{3!||z||^3}-\lambda\left(\dfrac{l_z(D^2G(0)(z^2))}{2!||z||^2}\right)^2\right|\\ \leq\dfrac{|\Phi'(0)|}{6}\max\left\{1,\left|\dfrac{1}{2}\dfrac{\Phi''(0)}{\Phi'(0)}+\left(1-\dfrac{3}{2}\lambda\right)\Phi'(0)\right|\right\},\quad z\in\mathbb{B}\setminus\{0\}.\end{array}\right\} \tag{2.7}$$

Since $|\Phi''(0)+2(\Phi'(0))^2|\geq2\Phi'(0)$, therefore the above inequality gives

$$\left|\frac{l_z(D^3G(0)(z^3))}{3!||z||^3}\right|\leq\frac{\Phi'(0)}{6}\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)}+\Phi'(0)\right). \tag{2.8}$$

Also, note that

$$\left|\left(\frac{l_z(D^3G(0)(z^3))}{3!||z||^3}\right)^2-\left(\frac{l_z(D^2G(0)(z^2))}{2!\|z\|^2}\right)^2\right|$$
$$\leq\left|\frac{l_z(D^3G(0)(z^3))}{3!||z||^3}\right|^2+\left|\frac{l_z(D^2G(0)(z^2))}{2!\|z\|^2}\right|^2.$$

The required bound follows from the above inequality together with the bounds given in (2.6) and (2.8).

The result is sharp for the function G given by

$$DG(z)=I\exp\int_0^{T_u(z)}\frac{\Phi(it)-1}{t}dt,\quad z\in\mathbb{B},\ \ \|u\|=1. \tag{2.9}$$

Clearly, $(DG(z))^{-1}(D^2G(z)(z^2)+DG(z)(z))\in\mathcal{M}_\Phi$ and

$$\frac{D^3G(0)(z^3)}{3!}=-\frac{1}{6}\left(\frac{\Phi''(0)}{2}+(\Phi'(0))^2\right)(l_u(z))^2z\ \text{ and }\ \frac{D^2G(0)(z^2)}{2!}=\frac{i\Phi'(0)}{2}l_u(z)z,$$

which immediately gives

$$\frac{l_z(D^2G(0)(z^2))}{2!} = \frac{i\Phi'(0)}{2} l_u(z)\|z\|$$

and

$$\frac{l_z(D^3G(0)(z^3))\|z\|}{3!} = -\frac{1}{6}\left(\frac{\Phi''(0)}{2} + (\Phi'(0))^2\right)(l_u(z))^2\|z\|^2.$$

Taking $z = ru$ $(0 < r < 1)$, we get

$$\frac{l_z(D^3G(0)(z^3))}{3!\|z\|^3} = -\frac{1}{6}\left(\frac{\Phi''(0)}{2} + (\Phi'(0))^2\right) \text{ and } \frac{l_z(D^2G(0)(z^2))}{2!\|z\|^2} = \frac{i\Phi'(0)}{2}. \tag{2.10}$$

According to the above equations, we have

$$\left|\left(\frac{l_z(D^3G(0)(z^3))}{3!\|z\|^3}\right)^2 - \left(\frac{l_z(D^2G(0)(z^2))}{2!\|z\|^2}\right)^2\right| = \frac{(\Phi'(0))^2}{4} + \frac{1}{36}\left(\frac{\Phi''(0)}{2} + (\Phi'(0))^2\right)^2,$$

which establishes the sharpness of the bound and completes the proof.

Theorem 2.2. *Let $g \in \mathcal{H}(\mathbb{B}, \mathbb{C})$ with $g(0) = 1$, $g(z) \neq 0$, $z \in \mathbb{B}$ and suppose that $G(z) = zg(z)$. If $DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z) \in \mathcal{M}_\Phi$ such that Φ satisfy*

$$2\Phi'(0) - 2(\Phi'(0))^2 \leq \Phi''(0) \leq 4(\Phi'(0))^2 - 2\Phi'(0),$$

then

$$|2a_2^2a_3 - a_3^2 - 2a_2^2 + 1| \leq 1 + 2(\Phi'(0))^2 + \frac{(\Phi'(0))^2}{4}\left(\frac{\Phi''(0)}{2\Phi'(0)} - 3\Phi'(0)\right)\left(\frac{\Phi''(0)}{2\Phi'(0)} + \Phi'(0)\right),$$

where

$$a_3 = \frac{l_z(D^3G(0)(z^3))}{3!||z||^3} \text{ and } a_2 = \frac{l_z(D^2G(0)(z^2))}{2!||z||^2}.$$

The bound is sharp.

Proof. Since $2\Phi'(0) \leq 4(\Phi'(0))^2 - \Phi''(0)$, inequality (2.7) gives

$$\left|\frac{l_z(D^3G(0)(z^3))}{3!||z||^3} - 2\left(\frac{l_z(D^2G(0)(z^2))}{2!||z||^2}\right)^2\right| \leq \frac{\Phi'(0)}{6}\left(2\Phi'(0) - \frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)}\right) \tag{2.11}$$

for $z \in \mathbb{B} \setminus \{0\}$. Also, since Φ satisfy $\Phi''(0) + 2(\Phi'(0))^2 \geq 2\Phi'(0)$, therefore by (2.7), we have

$$\left|\frac{l_z(D^3G(0)(z^3))}{3!||z||^3}\right| \leq \frac{\Phi'(0)}{6}\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0)\right), \quad z \in \mathbb{B} \setminus \{0\}. \tag{2.12}$$

Also, we have

$$|2a_2^2a_3 - 2a_2^2 - a_3^2 + 1| \leq 1 + 2|a_2|^2 + |a_3||a_3 - 2a_2^2|.$$

The required bound follows directly from the above inequality along with the bounds given in (2.6) and (2.12), and the bound of $|a_3 - 2a_2^2|$ given by (2.11).

Equality case holds for the function $G(z)$ defined in (2.9) as for this function, we have $a_2 = \frac{i\Phi'(0)}{2}$, $a_3 = -\frac{1}{6}\left(\frac{\Phi''(0)}{2} + (\Phi'(0))^2\right)$ and hence

$$2a_2^2a_3 - 2a_2^2 - a_3^2 + 1 = 2(\Phi'(0))^2 + \frac{(\Phi'(0))^2}{12}\left(\frac{\Phi''(0)}{2\Phi'(0)} - 3\Phi'(0)\right)\left(\frac{\Phi''(0)}{2\Phi'(0)} + \Phi'(0)\right) + 1,$$

which establish the sharpness of the result.

Theorem 2.3. *Let $g \in \mathcal{H}(\mathbb{U}^n, \mathbb{C})$ with $g(0) = 1$, $g(z) \neq 0$, $z \in \mathbb{U}^n$ and suppose that $G(z) = zg(z)$. If $(DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z) \in \mathcal{M}_\Phi$ such that Φ satisfies*

$$|\Phi''(0) + 2(\Phi'(0))^2| \geq 2\Phi'(0),$$

then

$$\left.\begin{aligned} &\left\| \left(\frac{D^3G(0)(z^3)}{3!} \right)^2 - \left(\frac{D^2G(0)(z^2)}{2!} \right)^2 \right\| \\ &\leq \frac{(\Phi'(0))^2\|z\|^6}{36}\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0) \right)^2 + \frac{(\Phi'(0))^2\|z\|^4}{4}, \quad z \in \mathbb{U}^n. \end{aligned} \right\} \tag{2.13}$$

The bound is sharp.

Proof. For $z \in \mathbb{U}^n \setminus \{0\}$ and $z_0 = \frac{z}{\|z\|}$, define $h_k : \mathbb{U} \to \mathbb{C}$ such that

$$h_k(\zeta) = \begin{cases} \dfrac{\zeta z_k}{p_k(\zeta z_0)\|z_0\|}, & \zeta \neq 0, \\ 1, & \zeta = 0, \end{cases} \tag{2.14}$$

where $p(z) = (DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z))$ and k satisfies $|z_k| = \|z\| = \max_{1\leq j\leq n}\{z_j\}$. By (2.3), we have

$$h_k(\zeta) = \frac{D^2g(\zeta z_0)((\zeta z_0)^2) + 3Dg(\zeta z_0)(\zeta z_0) + g(\zeta z_0)}{g(\zeta z_0) + Dg(\zeta z_0)(\zeta z_0)},$$

or, equivalently

$$h_k(\zeta)(g(\zeta z_0) + Dg(\zeta z_0)(\zeta z_0)) = D^2g(\zeta z_0)((\zeta z_0)^2) + 3Dg(\zeta z_0)(\zeta z_0) + g(\zeta z_0).$$

Comparison of same homogeneous expansions in the Taylor series expansions in terms of ζ yield

$$h_k'(0) = 2Dg(0)(z_0). \tag{2.15}$$

Furthermore, from $G(z_0) = z_0g(z_0)$, we have

$$\frac{D^2G_k(0)(z_0^2)}{2!} = Dg(0)(z_0)\frac{z_j}{\|z\|}. \tag{2.16}$$

Combining (2.15) and (2.16) with the fact $|h_k'(0)| \leq \Phi'(0)$ gives

$$\left| \frac{D^2G_k(0)(z_0^2)}{2!}\frac{\|z\|}{z_j} \right| \leq \frac{\Phi'(0)}{2}.$$

If $z_0 \in \partial\mathbb{U}^n$, then we get

$$\left| \frac{D^2G_k(0)(z_0^2)}{2!} \right| \leq \frac{\Phi'(0)}{2}.$$

Since

$$\frac{D^2G_k(0)(z_0^2)}{2!}, \quad k = 1, 2, \cdots n$$

are holomorphic functions on $\overline{\mathbb{U}}^n$, therefore by the maximum modulus theorem of holomorphic functions on the unit polydisc, we have

$$\left| \frac{D^2G_k(0)(z_0^2)}{2!} \right| \leq \frac{\Phi'(0)}{2}, \quad z_0 \in \mathbb{U}^n, k = 1, 2, \cdots n.$$

That is

$$\left| \frac{D^2G_k(0)(z^2)}{2!} \right| \leq \frac{\Phi'(0)\|z\|^2}{2}, \quad z \in \partial\mathbb{U}^n, k = 1, 2, \cdots n. \tag{2.17}$$

For $\lambda \in \mathbb{C}$, Xu et al. [17, Theorem 3.2] established that

$$\left.\begin{aligned}\left|\frac{D^3G_k(0)(z^3)}{3!} - \lambda\frac{1}{2}D^2G_k(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\right| \\ \leq \frac{|\Phi'(0)|\|z\|^3}{6}\max\left\{1, \left|\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \left(1 - \frac{3}{2}\lambda\right)\Phi'(0)\right|\right\}.\end{aligned}\right\} \tag{2.18}$$

Since $|\Phi''(0) + 2(\Phi'(0))^2| \geq 2\Phi'(0)$, therefore, from (2.18), we get

$$\left|\frac{D^3G_k(0)(z^3)}{3!}\right| \leq \frac{\Phi'(0)\|z\|^3}{6}\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0)\right) \tag{2.19}$$

for $z \in \mathbb{U}^n$ and $k = 1, 2, \cdots n$. Using the bounds from (2.17) and (2.19), we have

$$\begin{aligned}&\left|\left(\frac{D^3G_k(0)(z^3)}{3!}\right)^2 - \left(\frac{D^2G_k(0)(z^2)}{2!}\right)^2\right| \\ &\qquad \leq \frac{(\Phi'(0))^2\|z\|^6}{36}\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0)\right)^2 + \frac{(\Phi'(0))^2\|z\|^4}{4}\end{aligned}$$

for $z \in \mathbb{U}^n$ and $k = 1, 2, \cdots n$. Therefore,

$$\begin{aligned}&\left\|\left(\frac{D^3G(0)(z^3)}{3!}\right)^2 - \left(\frac{D^2G(0)(z^2)}{2!}\right)^2\right\| \\ &\qquad \leq \frac{(\Phi'(0))^2\|z\|^6}{36}\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0)\right)^2 + \frac{(\Phi'(0))^2\|z\|^4}{4}, \quad z \in \mathbb{U}^n,\end{aligned}$$

which is the required bound.

To prove the sharpness of the bound, consider the function G given by

$$DG(z) = I\exp\int_0^{z_1}\frac{\Phi(it) - 1}{t}dt. \tag{2.20}$$

It can be showed that $(DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z)) \in \mathcal{M}_\Phi$ and for $z = (r, 0, \cdots, 0)'$ in (2.20), the equality case holds in (2.13).

Theorem 2.4. *Let $g \in \mathcal{H}(\mathbb{U}^n, \mathbb{C})$, $g(0) = 1$, $g(z) \neq 0$, $z \in \mathbb{U}^n$ and suppose that $G(z) = zg(z)$. If $(DG(z))^{-1}(D^2G(z)(z^2) + DG(z)(z)) \in \mathcal{M}_\Phi$ and Φ satisfy*

$$2\Phi'(0) - 2(\Phi'(0))^2 \leq \Phi''(0) \leq 4(\Phi'(0))^2 - 2\Phi'(0),$$

then

$$\begin{aligned}&\|2a_2^2a_3 - a_3^2 - 2a_2^2 + 1\| \\ &\qquad \leq 1 + \frac{(\Phi'(0))^2\|z\|^6}{36}\left(2\Phi'(0) - \frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)}\right)\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0)\right) + \frac{(\Phi'(0))^2\|z\|^4}{2},\end{aligned}$$

where

$$a_3 = \frac{D^3G(0)(z^3)}{3!} \quad \text{and} \quad a_2^2 = \frac{1}{2}D^2G(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right).$$

The bound is sharp.

Proof. Since $G(z) = zg(z)$, we have

$$\frac{1}{2}D^2G_k(0)\left(z_0, \frac{D^2G(0)(z_0^2)}{2!}\right)\frac{z_k}{\|z\|} = \left(\frac{D^2G_k(0)(z_0^2)}{2!}\right)^2, \quad k = 1, 2, \cdots n,$$

where $z_0 = \frac{z_k}{\|z\|}$ and k satisfies $|z_k| = \|z\| = \max_{1\leq j\leq n}\{|z_j|\}$ (see [16]). If $z_0 \in \partial\mathbb{U}^n$, then

$$\left|\frac{1}{2}D^2G_k(0)\left(z_0, \frac{D^2G(0)(z_0^2)}{2!}\right)\right| = \left|\frac{D^2G_k(0)(z_0^2)}{2!}\right|^2.$$

Considering the function $h_k(\zeta)$ given in (2.14) and following the same methodology as in the proof of Theorem 2.3, we obtain (2.17), which together with the above relation yields

$$\left|\frac{1}{2}D^2G_k(0)\left(z_0, \frac{D^2G(0)(z_0^2)}{2!}\right)\right| \leq \frac{(\Phi'(0))^2\|z\|^4}{4}. \tag{2.21}$$

Also, since Φ satisfy $2\Phi'(0) \leq 4(\Phi'(0))^2 - \Phi''(0)$, therefore from (2.18), we obtain

$$\left.\begin{aligned}\left|\frac{D^3G_k(0)(z^3)}{3!} - \quad & D^2G_k(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\right| \\ & \leq \frac{|\Phi'(0)|\|z\|^3}{6}\left(2\Phi'(0) - \frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)}\right), \quad k = 1, 2, \cdots n.\end{aligned}\right\} \tag{2.22}$$

Thus, from (2.19), (2.21) and (2.22), we have

$$\begin{aligned}&\left|1 + D^2G_k(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\left(\frac{D^3G_k(0)(z^3)}{3!}\right) - D^2G_k(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\right. \\ &\left. - \left(\frac{D^3G_k(0)(z^3)}{3!}\right)^2\right| \\ &\leq 1 + \left|\frac{D^3G_k(0)(z^3)}{3!}\right|\left|\frac{D^3G_k(0)(z^3)}{3!} - D^2G_k(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\right| \\ &\quad + \left|D^2G_k(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\right| \\ &\leq 1 + \frac{(\Phi'(0))^2\|z\|^6}{36}\left(2\Phi'(0) - \frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)}\right)\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0)\right) + \frac{(\Phi'(0))^2\|z\|^4}{2}\end{aligned}$$

for $z \in \mathbb{U}^n$ and $k = 1, 2, \cdots n$. Therefore

$$\begin{aligned}&\left\|1 + D^2G(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\left(\frac{D^3G(0)(z^3)}{3!}\right) - D^2G(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right)\right. \\ &\left. - \left(\frac{D^3G(0)(z^3)}{3!}\right)^2\right\| \\ &\leq 1 + \frac{(\Phi'(0))^2\|z\|^6}{36}\left(2\Phi'(0) - \frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)}\right)\left(\frac{1}{2}\frac{\Phi''(0)}{\Phi'(0)} + \Phi'(0)\right) + \frac{(\Phi'(0))^2\|z\|^4}{2},\end{aligned}$$

which is the required bound.

Sharpness of the bound can be seen from the function $G(z)$ defined in (2.20) by taking $z = (r, 0, \cdots, 0)$, which completes the proof.

Note that if $g \in \mathcal{H}(\mathbb{B})$ and $(Dg(z))^{-1}(D^2g(z)(z^2) + Dg(z)(z)) \in \mathcal{M}_\Phi$, then various choices of Φ give different subclasses of holomorphic mappings. For instance, when $\Phi(z) = (1 + (1 - 2\alpha)z)/(1 - z)$ and $\Phi(z) = (1 + z)/(1 - z)$, we easily obtain $g \in \mathcal{K}_\alpha(\mathbb{B})$ and $g \in \mathcal{K}(\mathbb{B})$, respectively. For these classes, Theorem 2.1 to Theorem 2.4 yield the following results.

Corollary 2.5. *Let $g \in \mathcal{H}(\mathbb{B}, \mathbb{C})$ and $G(z) = zg(z) \in \mathcal{C}_\alpha(\mathbb{B})$. Then the following inequality holds:*

$$\begin{aligned}&\left|\left(\frac{l_z(D^2G(0)(z^2))}{2!||z||^2}\right)^2 - \left(\frac{l_z(D^3G(0)(z^3))}{3!||z||^3}\right)^2\right| \\ &\qquad \leq \frac{2(1-\alpha)^2(2\alpha^2 - 6\alpha + 9)}{9}, \quad l_z \in T_z,\ z \in \mathbb{B}\setminus\{0\}.\end{aligned}$$

If $\mathbb{B} = \mathbb{U}^n$ *and* $X = \mathbb{C}^n$, *then*

$$\left.\begin{aligned}\left\|\left(\frac{D^3G(0)(z^3)}{3!}\right)^2 - \left(\frac{D^2G(0)(z^2)}{2!}\right)^2\right\| \\ \leq (1-\alpha)^2\|z\|^4 + \frac{(2\alpha^2-5\alpha+3)^2\|z\|^6}{9},\ z \in \mathbb{U}^n.\end{aligned}\right\} \tag{2.23}$$

All these bounds are sharp.

Remark 2.1. In case of $n = 1$, $\mathbb{B} = \mathbb{U}$ and (2.23) reduces to the following:

$$\left|\left(\frac{G^{(3)}(0)}{3!}\right)^2 - \left(\frac{G''(0)}{2!}\right)^2\right| \leq \frac{2(1-\alpha)^2(2\alpha^2-6\alpha+9)}{9},$$

which is equivalent to Theorem C.

Corollary 2.6. *Let* $g \in \mathcal{H}(\mathbb{B}, \mathbb{C})$ *and* $G(z) = zg(z) \in \mathcal{C}_\alpha(\mathbb{B})$. *Then for* $\alpha \in [0, 1/2]$, *the following sharp bound holds:*

$$|2a_2^2a_3 - a_3^2 - 2a_2^2 + 1| \leq \frac{8\alpha^4 - 34\alpha^3 + 71\alpha^2 - 72\alpha + 36}{9}, \quad z \in \mathbb{B} \setminus \{0\},$$

where

$$a_3 = \frac{l_z(D^3G(0)(z^3))}{3!||z||^3} \quad \text{and} \quad a_2 = \frac{l_z(D^2G(0)(z^2))}{2!||z||^2}.$$

Corollary 2.7. *Let* $g \in \mathcal{H}(\mathbb{U}^n, \mathbb{C})$ *and* $G(z) = zg(z) \in \mathcal{C}_\alpha(\mathbb{U}^n)$. *Then for* $\alpha \in [0, 1/2]$, *the following sharp inequality holds:*

$$\|2a_2^2a_3 - a_3^2 - 2a_2^2 + 1\| \leq 1 + 2\|z\|^4(1-\alpha)^2 + \frac{(1-\alpha)^2(9 - 18\alpha + 8\alpha^2)\|z\|^6}{9} \tag{2.24}$$

for $z \in \mathbb{U}^n$, *where*

$$a_3 = \frac{D^3G(0)(z^3)}{3!} \quad \text{and} \quad a_2^2 = \frac{1}{2}D^2G(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right).$$

Remark 2.2. When $n = 1$, Corollary 2.7 is equivalent to Theorem D.

In particular, for $\alpha = 0$, we obtain the following results for the class $\mathcal{C}$ in higher dimensions.

Corollary 2.8. *Let* $g \in \mathcal{H}(\mathbb{B}, \mathbb{C})$ *and* $G(z) = zg(z) \in \mathcal{C}(\mathbb{B})$. *Then the following holds:*

$$\left|\left(\frac{l_z(D^2G(0)(z^2))}{2!||z||^2}\right)^2 - \left(\frac{l_z(D^3G(0)(z^3))}{3!||z||^3}\right)^2\right| \leq 2, \quad l_z \in T_z,\ z \in \mathbb{B} \setminus \{0\}.$$

If $\mathbb{B} = \mathbb{U}^n$ *and* $X = \mathbb{C}^n$, *then*

$$\left\|\left(\frac{D^3G(0)(z^3)}{3!}\right)^2 - \left(\frac{D^2G(0)(z^2)}{2!}\right)^2\right\| \leq \|z\|^4 + \|z\|^6, \quad z \in \mathbb{U}^n. \tag{2.25}$$

All these bounds are sharp.

Remark 2.3. For $n = 1$, (2.25) is equivalent to Theorem A.

Corollary 2.9. *Let* $g \in \mathcal{H}(\mathbb{B}, \mathbb{C})$ *and* $G(z) = zg(z) \in \mathcal{C}(\mathbb{B})$. *Then the following sharp bound holds:*

$$|2a_2^2a_3 - a_3^2 - 2a_2^2 + 1| \leq 4, \quad z \in \mathbb{B} \setminus \{0\},$$

where

$$a_3 = \frac{l_z(D^3G(0)(z^3))}{3!||z||^3} \quad \text{and} \quad a_2 = \frac{l_z(D^2G(0)(z^2))}{2!||z||^2}.$$

Corollary 2.10. *Let $g \in \mathcal{H}(\mathbb{U}^n, \mathbb{C})$ and $G(z) = zg(z) \in \mathcal{C}(\mathbb{U}^n)$. Then for $\alpha \in [0, 1/2]$, the following sharp estimation holds:*

$$|2a_2^2 a_3 - a_3^2 - 2a_2^2 + 1| \leq 1 + 2\|z\|^4 + \|z\|^6$$

for $z \in \mathbb{U}^n$, where

$$a_3 = \frac{D^3G(0)(z^3)}{3!} \quad \text{and} \quad a_2^2 = \frac{1}{2}D^2G(0)\left(z, \frac{D^2G(0)(z^2)}{2!}\right).$$

Remark 2.4. when $n = 1$, Corollary 2.10 is equivalent to Theorem B.

Declarations

Funding

The work of the Surya Giri is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No. 1112/(CSIR-UGC NET JUNE 2019).

Conflict of interest

The authors declare that they have no conflict of interest.

Author Contribution

Each author contributed equally to the research and preparation of manuscript.

Data Availability

Not Applicable.

References

[1] O. P. Ahuja, K. Khatter and V. Ravichandran, Toeplitz determinants associated with Ma-Minda classes of starlike and convex functions, Iran. J. Sci. Technol. Trans. A Sci. **45** (2021), no. 6, 2021–2027.

[2] M. F. Ali, D. K. Thomas and A. Vasudevarao, Toeplitz determinants whose elements are the coefficients of analytic and univalent functions, Bull. Aust. Math. Soc. **97** (2018), no. 2, 253–264.

[3] H. Cartan, Sur la possibilitéd'étendre aux fonctions de plusieurs variables complexes la théorie des fonctions univalentes. In: Lecons sur les Fonctions Univalentes ou Multivalentes. Paris: Gauthier-Villars, 1933.

[4] S. Giri, and S. S. Kumar, Hermitian-Toeplitz determinants for certain univalent functions, arXiv preprint arXiv:2203.10429 (2022).

[5] I. Graham, H. Hamada, T. Honda, G. Kohr and K. H. Shon, distortion and coefficient bounds for Carathéodory families in $\mathbb{C}^n$ and complex Banach spaces, J. Math. Anal. Appl. **416** (2014), no. 1, 449–469.

[6] I. Graham, H. Hamada and G. Kohr, Parametric representation of univalent mappings in several complex variables, Canad. J. Math. **54** (2002), no. 2, 324–351.

[7] I. Graham and G. Kohr, Geometric function theory in one and higher dimensions, Monographs and Textbooks in Pure and Applied Mathematics, 255, Marcel Dekker, Inc., New York, 2003.

[8] H. Hamada and T. Honda, Sharp growth theorems and coefficient bounds for starlike mappings in several complex variables, Chinese Ann. Math. Ser. B **29** (2008), no. 4, 353–368.

[9] H. Hamada, G. Kohr and M. Kohr, The Fekete-Szegö problem for starlike mappings and nonlinear resolvents of the Carathéodory family on the unit balls of complex Banach spaces, Anal. Math. Phys. **11** (2021), no. 3, Paper No. 115, 22 pp.

[10] G. Kohr, On some best bounds for coefficients of several subclasses of biholomorphic mappings in $\mathbf{C}^n$, Complex Variables Theory Appl. **36** (1998), no. 3, 261–284.

[11] B. Kowalczyk, O. S. Kwon, A Lecko, Y. J. Sim and B. Śmiarowska, The third-order Hermitian Toeplitz determinant for classes of functions convex in one direction, Bull. Malays. Math. Sci. Soc. **43** (2020), no. 4, 3143–3158.

[12] A. Lecko, Y. J. Sim and B. Śmiarowska, The fourth-order Hermitian Toeplitz determinant for convex functions, Anal. Math. Phys. **10** (2020), no. 3, Paper No. 39, 11 pp.

[13] X.-S. Liu and M.-S. Liu, Quasi-convex mappings of order α on the unit polydisk in $\mathbf{C}^n$, Rocky Mountain J. Math. **40** (2010), no. 5, 1619–1644.

[14] X. Liu and T. Liu, Sharp estimates of all homogeneous expansions for a subclass of quasi-convex mappings of type B and order α in several complex variables, Acta Math. Sci. Ser. B (Engl. Ed.) **36** (2016), no. 6, 1808–1818.

[15] K. A. Roper and T. J. Suffridge, Convexity properties of holomorphic mappings in $\mathbf{C}^n$, Trans. Amer. Math. Soc. **351** (1999), no. 5, 1803–1833.

[16] Q. H. Xu and T. S. Liu, On the coefficient inequalities for a class of holomorphic mappings, Complex Var. Elliptic Equ. **65** (2020), no. 9, 1474–1487.

[17] Q. Xu, T. Liu and X. Liu, The coefficient inequalities for a class of holomorphic mappings in several complex variables, Chinese Ann. Math. Ser. B **41** (2020), no. 1, 37–48.

[18] K. Ye and L.-H. Lim, Every matrix is a product of Toeplitz matrices, Found. Comput. Math. **16** (2016), no. 3, 577–598.

[19] Q. Xu, T. Yang, T. Liu and H. Xu, Fekete and Szegö problem for a subclass of quasi-convex mappings in several complex variables, Front. Math. China **10** (2015), no. 6, 1461–1472.

[20] Y.-C. Zhu, Biholomorphic convex mappings on $\mathbb{B}_p^n$, Chinese Annals Math **24** (2003), 269-278.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: spkumar@dtu.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: suryagiri456@gmail.com

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-14008-2

Underwater image enhancement using multiscale decomposition and gamma correction

Amarendra Kumar Mishra [1] · Mahipal Singh Choudhry [1] · Manjeet Kumar [1]

Received: 21 April 2022 / Revised: 3 August 2022 / Accepted: 12 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Underwater image enhancement has been attracting much attention due to its significance in marine engineering and aquatic robotics. Captured underwater images usually suffer from contrast degradation, low illumination, color cast, and noise. Many underwater image enhancement and restoration algorithms have been developed but are not able to solve all these problems. In this paper, a new single image retinex algorithm using gamma correction is proposed. Here the input image is decomposed into illumination and Reflectance. Illumination contains brightness variation, and Reflectance preserves the details information. Then Reflectance decomposed into multiple layers, which carried out gamma correction and contrast enhancement. Whereas illumination carried out brightness adjustment. Finally, these layers are combined to obtain an enhanced image. The proposed method produces high-quality enhanced images compared to the existing state-of-art method such as the Hue-preserving-based approach for underwater color image enhancement, Underwater image processing using a hybrid technique, and Underwater dark channel before using a guided image filter. The proposed method is tested for the underwater image enhancement benchmark data set and compared with the existing state-of-art method. Qualitative and quantitative results demonstrate the effectiveness of the proposed method in terms of seven parameters such as measure of enhancement (EME), discrete entropy (DE), peak signal to noise ratio (PSNR), Structure similarity index measure (SSIM), underwater color image quality evaluation (UCIQE), underwater image quality measure (UIQM), and patch-based contrast quality index (PCQI) for

✉ Manjeet Kumar
manjeetchhillar@gmail.com

Amarendra Kumar Mishra
Amarendra_2k18dce502@dtu.ac.in

Mahipal Singh Choudhry
msc_1976@yahoo.com

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

Published online: 26 October 2022

underwater images. Six parameters of the proposed method performed better compared to an existing method. The visual appearance of the output image of the proposed method has a very high quality.

Keywords Underwater-image enhancement · Real-world underwater images · Retinex · Brightness adjustment · Gamma correction · Gradient-domain

1 Introduction

In recent years, underwater research has increased exponentially because of the shortage of resources on the land, such as manganese, nickel, copper, and cobalt. Acquiring clean underwater images and videos plays a crucial role in finding resources in the ocean. The underwater image processing has also been an essential source of interest in different branches of technology and scientific research, such as the detection of underwater objects, V Pandey et al. [39] enhancing object detection in arial images, A Hooda et al. [25] estimate the surface roughness for super hydrophobic coating by image processing and machine learning, marine biology research, an inspection of aquatic infrastructure and cable, control of underwater vehicles, and archelogy [5]. The leading cause of quality degradation underwater is the absorption of light energy during propagation. The light exponentially decays with the distance and depth. The red channel of the RGB image is absorbed faster than the green and blue channel since light propagates underwater is wavelength-dependent; that's why images appear bluish and green [11]. Pictures and underwater videos are not good because of weak contrast, color distortion, and low illumination. The absorption is caused by reducing light energy and scattering due to the deviation of light from its original path. Figure 1 shows the optical image formation model (IFM) [33] for the underwater image.

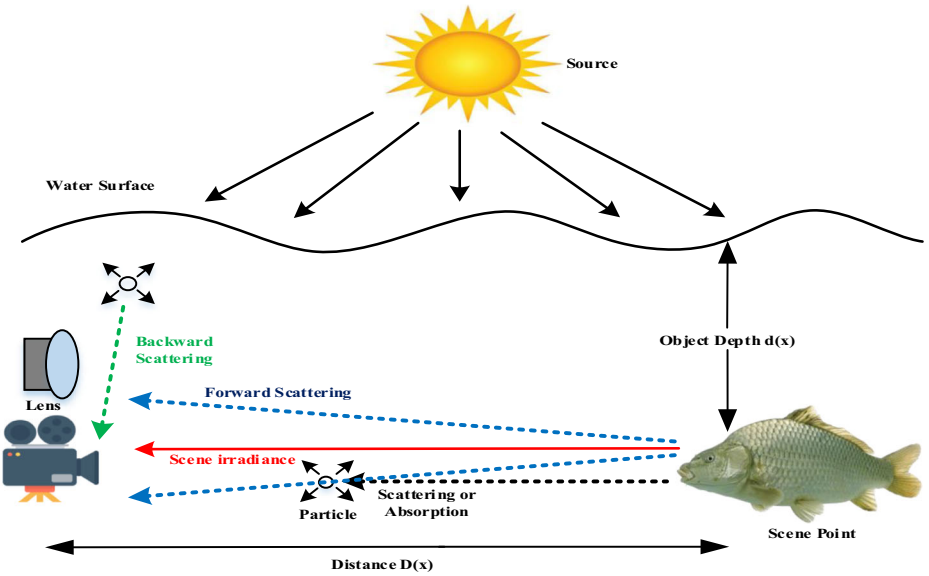


Fig. 1 Underwater optical image formation model (IFM) with forward and backward scattering [33]

The light received by the camera has three components: a scene irradiance, it is also called a direct component that comes directly from an object, a forward scattering component that deviates its path, i.e., it reaches the camera in another direction; and a back-scattering component reflected by the suspended particle present in the water, the light ray come to the camera before reaches to the object. Finally, the underwater image captured by the camera is a linear combination of these three components. Forward Scattering components are responsible for the blurring of the image, and back-scattering components degrade the details of the picture. Due to these two components, underwater photos have lost their visual quality and detailed information. To solve this problem, the researcher gives many techniques like underwater image enhancement, restoration, saliency detection, noise reduction, color constancy and correction, image defogging, and underwater image recovery.

There are three primary directions of underwater image enhancement:

1) Based on Modifying Image Pixel Values: The modifying image pixel value is also called the IFM-free based model, in which contrast and color of the degraded underwater image are mainly based on pixel intensity re-distribution, without considering the particular underwater imaging principles. The underwater image enhancement based on modifying image pixels is further categorized into single and multiple color models. Some work has already been done, such as attenuation map guided color correction and detail preserved dehazing [37]; Underwater target detection with an attention mechanism and improved scale [46]; etc.
2) Based on Underwater Imaging Models: The IFM-based underwater image enhancement method, such as dark channel prior (DCP) [22]; underwater dark channel prior (UDCP) [14]; maximum intensity prior (MIP) [10]; red channel prior (RCP) [20]; underwater light attenuation prior (ULAP) [41]; and other.
3) Based on Deep Learning Methods: Finally, Texture synthesis guided deep hashing for texture image retrieval [8]; deep learning-based underwater image enhancement methods are J Hu et al. [27] use the two-branch deep neural network for underwater image enhancement in 2021; In 2022, B Xu et al. [47] use GAN neural network, etc. for example, Cao et al. [9] used a global CNN network and a refined network to estimate the background light and predict the scene depth map. A Kumar et al. [31] use a machine learning approach for brain tumor detection and segmentation.

The proposed method is based on the retinex algorithm in which the input image is split into the illumination and reflectance parts. Further reflectance decomposed into the base layer, first detail layer, and second detail layer, respectively on the basis of parameters. The reflectance part preserves the edge, whereas the illumination part preserves the brightness. Gamma correction is carried out on the base layer and contrast enhancement in gradient domain on the first detail layer and second detail layer corresponding. Finally, we multiply these layers with a brightness adjustment component to get the enhanced image. The output image has good visual quality compared to the existing state-of-the-art method. Also, the parameters for enhancing the image are better than the existing method.

The remaining part of the paper is arranged in the following way. An overview of existing state-of-art methods for underwater image enhancement and restoration has been reviewed in Section 2. In Section 3, the retinex-based algorithm using gamma correction for underwater image enhancement and restoration is explained. The qualitative and quantitative result analysis have been discussed in Section 4. Finally, Section 5 concludes the paper.

2 Related work

Many underwater image enhancement, restoration, saliency detection, noise reduction, color constancy and correction, image defogging, and underwater image recovery method have emerged in the last few years. The few old image enhancement method, such as contrast limited adaptive histogram equalization [24], which reduce the noise from the underwater image, Deng [13] used a generalized unsharp masking algorithm to enhance contrast and sharpness, reduce the hallo effect, and to solve the out of range problem; A probability method simultaneously calculates illumination and Reflectance [17] All these methods are effective only for ordinary images. Hence, these image enhancement methods are not able to compensate for the contrast degradation of underwater images.

The objective of this research was to make color cast corrections and boost the contrast via the manipulation of an image pixel. The same objective was to realize in a different manner than Iqbal et al. [29] did, the expansion of RGB color space's saturation. In contrast, Fu et al. [19] adopted a two-step enhancement, where a color correction process led to a contrast enhancement process. Ancuti et al. [1] made use of complimentary information associated with multiple images to get an even better result. They incorporated a multi-scale fusion technique by fusing color-corrected images with contrast to enhance one belonging to the same single input image. Ancuti et al. [3] recent time introduced a fundamental color channel compensation (3C) pre-processing to improve image enhancement by color correction. The color channel compensation(3C) algorithm considerably improves the conventional restoration method by reconstructing the lost channel based on the opponent color channel.

In line with this, the objective of the proposed Retinex method too is in accordance with the ones proposed by us. Zhang et al. [49] took the aid of multi-scale Retinex for the enhancement of the underwater image. Fu et al. [16] first correct the color cast, followed by enhancement in contrast by employing a Retinex-based model. Dai et al. [12] address issues related to both lowly illuminated images and underwater images by first utilizing different techniques on the incident light and subsequently improving the contrast of reflectance.

To avoid the presence of artifacts, over-enhancement, or distortion in color, enhancement-based methods refrain from the physical properties of underwater mechanisms. The same mechanism would have led to the amplification of noise caused by suspended particles or marine snow.

2.1 Image formation optical model

The optical image formation model is given by McGlamery [23]; according to this model, the total radiation O_r Received the underwater optical image system has three components. These are the following components: (a) Direct component, which is reflected by an object without scattering and is represented by D_c. (b) The back-scatter component, B_s It is an ambient light scattered by suspended particles. (c) The Forward-scatter component is represented by F_s This component deviates from its line-of-sight propagation and reaches the camera. The total optical radiation O_r is a linear combination of the above three components. Mathematically, represented by

$$O_r(x,y) = D_c(x,y) + F_s(x,y) + B_s(x,y) \tag{1}$$

Where (x, y) is the coordinate of the pixel. We assume that the distance between object and camera is very close so that we can neglect the forward-scattering component [45].

The modified optical radiation

$$\begin{aligned} O_r(x,y) &= D_c(x,y) + B_s(x,y) \\ D_c(x,y) &= Q^c(x,y)T^c(x,y) \\ B_s(x,y) &= \boldsymbol{B}^c(1-T^c(x,y)) \end{aligned} \tag{2}$$

Where $Q^c(x,y)$ is scene radiance, $T^c(x,y)$ as a transmission whose value lies between zero and one, its value decreases exponentially. $\boldsymbol{B}^c$ is the background light. Where c belongs to red, green, and blue channels. The image intensity captured by the camera is represented by

$$I^c(x,y) = Q^c(x,y)T^c(x,y) + \boldsymbol{B}^c(1-T^c(x,y)) \quad c\in\{r,G,b\} \tag{3}$$

The transmission is commonly written as an exponentially decaying function [22] [4, 50] based on the Beer-Lambert law [26].

$$T^c\left(\left(x,y\right)\right) = e^{-\delta^c D(x)} \tag{4}$$

Where δ^c Is the attenuation coefficient for channel C, and $D(x)$ is the distance between the imagined device and the scene point.

The proposed article is a novel retinex-based algorithm using gamma correction to enhance underwater images. Firstly, an input image has been decomposed into illumination and Reflectance. Subsequently, brightness adjustment is applied to the illumination part. It shows the display quality of the picture. A weighted list square (WLS) based filter [15] carries on the reflectance part. The reflectance output is the following layers: a base layer, first detail layer, and second detail layer. Then gamma correction is applied on the base layer and contrast enhancement on both detail layers and combined with all three layers. Finally, multiply with brightness adjustment to get an enhanced image. Based on the quantitative comparison, the value of six parameters out of seven which is presented in Table 1, are better in the proposed method than the existing start-of-art method.

The contribution of the authors to the paper is as follows:

- The authors analyzed the loss in quality of underwater images and its causes. Gamma correction and bright modification by multi-level image decomposition have been employed for the enhancement of degraded underwater images.

Table 1 Comparison of the proposed method with existing methods on different parameters for 890 images

Methods	EME	DE	PSNR	SSIM	UIQM	UCIQE	PCQI
MIL [33]	9.4561	5.2121	32.7311	0.7676	33.3598	6.7876	0.9561
UDCPGIF [23]	10.6999	4.3813	59.9829	0.5188	33.9441	4.8822	0.9887
UDCPSM [28, 32]	10.5716	4.4015	60.5113	0.5461	32.4418	3.1095	0.9896
HUIE [35]	12.2903	5.1761	66.1795	0.7285	32.5351	6.4946	0.7124
CBFU [2]	8.2488	4.8614	**68.1791**	0.8469	35.1462	7.1722	0.9906
HPU [26]	8.3625	5.0032	31.6036	0.8526	32.1707	0.6136	0.8015
HTU [30]	12.2903	5.1761	66.1795	0.7285	32.5351	6.4946	0.7124
MSRCR [40]	11.1108	4.8217	61.8387	0.5873	34.1452	6.2737	0.9911
MPORU [38]	5.4428	4.7332	53.1908	0.4624	35.7365	6.9073	0.1488
Proposed Method	**14.0725**	**5.4564**	67.9461	**0.8544**	**36.7576**	**7.3202**	**0.9997**

- An underwater image quality enhancement model based on retinex theory has been implemented. UIEB datasets (available in the public domain), stellar performance is observed.
- Using the proposed method, high-quality visual images have been derived out of raw underwater images under different light intensities.

3 Proposed method

The author proposed a novel method for single image real-world underwater image enhancement. In this section, we can also restore the brightness, color, and contrast. In this proposed method, we have split the input underwater image into the illumination and reflectance components. Further Reflectance decomposed into multi-scale levels in that way base layer, first detail layer, and second detail layer. The gamma correction is carried out on the base layer, and contrast enhancement in the gradient domain is applied on both detail layers. These processes are done for the restoration of weak Illumination, color cast, and noise. The flow diagram of the proposed method shown in Fig. 2.

3.1 Evaluation of the illumination and the reflectance

To obtain the Illumination and the Reflectance, we are using the Retinex model on the underwater image. According to this model, the underwater capture image has two components, i.e., Illumination and Reflectance. Mathematically it is represented as:

$$I^c(x,y) = L^c(x,y) \cdot R^c(x,y), \quad c \in \{r, G, b\} \tag{5}$$

Where $I^c(x,y)$ is capture image, $L^c(x,y)$ is Illumination, and $R^c(x,y)$ is Reflectance. (x,y) is the coordinate of pixel. Dot '·' Represent the multiplication sign. C belongs to the color space of the image, i.e., Red, Green, and Blue.

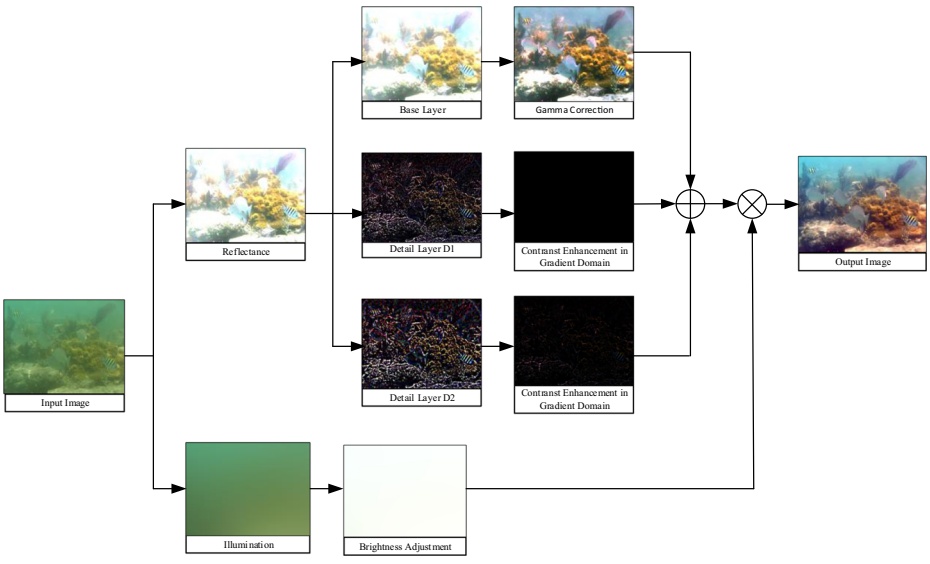


Fig. 2 Proposed retinex-based algorithm using gamma correction for underwater image enhancement

We separate the Illumination from the input underwater image in-ordered to improve brightness and contrast. The illumination part is piece-wise smooth and contains illumination variance according to retinex theory. The reflectance preserved color information and details of the underwater input image while the global incident light of the scenario is captured in the Illumination. Further brightness adjustment, gamma correction, and contrast enhancement in gradient domain strategies are applied to the different image layers to compensate for color, contrast, and Illumination separately.

Now we take logarithmic both sides of eq. (1). The logarithmic function converts multiplication into addition which reduces the complexity of the Algorithm and makes it easy to sense brightness.

After taking logarithmic, the eq. (1) is represented as:

$$\mathsf{i}(x,y) = \mathsf{l}(x,y) + \mathsf{r}(x,y) \tag{6}$$

Where, $\mathsf{i}(x,y) = \log(I^c(x,y))$, $\mathsf{l}(x,y) = \log(L^c(x,y))$, $\mathsf{r}(x,y) = \log(R^c(x,y))$.

Single scale retinex is defined as:

$$\mathsf{r}(x,y) = \mathsf{i}(x,y) - \mathsf{l}(x,y) \tag{7}$$

$$R^c(x,y) = 10^{\mathsf{r}(x,y)} \tag{8}$$

$$\mathsf{l}(x,y) = \log(T(x,y) * I^c(x,y)) \tag{9}$$

Where, $\mathsf{l}(x,y)$ is Illumination, which is the convolution of Gaussian function and the underwater input image, $T(x,y)$ is a gaussian function, $I^c(x,y)$ is the underwater input image, and $R^c(x,y)$ is retinex. The Gaussian function is mathematically defined as:

$$T(x,y:\sigma) = \alpha e^{-\frac{x^2+y^2}{2\sigma_n}} \tag{10}$$

Where α is a constant, (x,y) is the coordinate of the pixel, and σ_n is the variance which maintains the shape of the Gaussian function. Equation (10) is a probability density function so that according to properties of pdf, the integration of Gaussian function over the entire range becomes unity. Mathematically,

$$\iint T(x,y)dxdy = 1 \tag{11}$$

The Reflectance and Illumination of raw images are shown in Fig. 3.

The variational methods used to evaluate Illumination and Reflectance to transfer into the logarithmic domain to simplify the complex problem. Hence, the dissimilarity of gradient magnitude is typically suppressed in the logarithmic environment, which leads to over-smooth Reflectance with edge and texture detail being lost. To overcome the above problem, the weighted variational model proposed by [18] is used to estimate the Illumination and Reflectance simultaneously. The objective function defines according to the weighted variational model:

$$\begin{gathered} F(\mathsf{r},\mathsf{l}) = \arg\min_{\mathsf{r},\mathsf{l}} \|\mathsf{r}+\mathsf{l}-\mathsf{i}\|_2^2 + \mu_1 \|e^{\mathsf{r}} \cdot \nabla \mathsf{r}\|_1 + \mu_2 \left\|e^{\mathsf{l}} \cdot \nabla \mathsf{l}\right\|_2^2 \\ \text{s.t} \quad \mathsf{r} \le 0 \text{ and } \mathsf{i} \le \mathsf{l} \end{gathered} \tag{12}$$

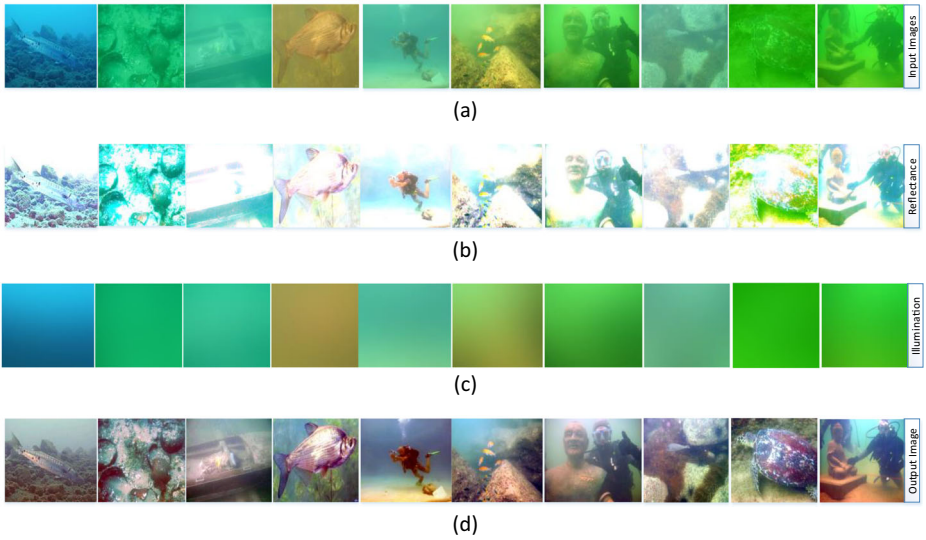


Fig. 3 Intermediate Process of Underwater Image Enhancement (**a**) Input Image (UIEB Dataset) (**b**) Reflectance of the Input Image (**c**) Illumination of the Input Image (**d**) Output Enhanced Image

Where, $\mathsf{i}(x,y) = \log(I^c(x,y))$, $\mathsf{l}(x,y) = \log(L^c(x,y))$, $\mathsf{r}(x,y) = \log(R^c(x,y))$. The μ_1 and μ_2 are two positive parameters. ∇as the mathematical operator gradient and $\|\cdot\|_{\mathrm{P}}$ Denote the p-norm operator.

The first term $\|\mathsf{r}+\mathsf{l}-\mathsf{i}\|_2^2$ is used to maintain reliability and minimize the distance between $(\mathsf{r}+\mathsf{l})$ and i. This term $\|e^{\mathsf{r}} \cdot \nabla \mathsf{r}\|_1$ carry out piece-wise constant on r, and finally, the term $\|e^{\mathsf{r}} \cdot \nabla \mathsf{r}\|_2^2$ enforces special smoothness on l.

To eradicate the impact of the weights e^{r} and e^{l} For the simplification of objective function $F(\mathsf{r},\mathsf{l})$. Now the function can be rewritten as:

$$F\left(\mathsf{r}^k,\mathsf{l}^k\right) = \arg\min_{\mathsf{r}^k,\mathsf{l}^k} \|\mathsf{r}+\mathsf{l}-\mathsf{i}\|_2^2 + \mu_1\left\|R^{k-1}\cdot\nabla\mathsf{r}^k\right\|_1 + \mu_2\left\|L^{k-1}\cdot\nabla\mathsf{l}^k\right\|_2^2 \quad \text{s.t } \mathsf{r}^k \le 0 \text{ and } \mathsf{i} \le \mathsf{l} \tag{13}$$

where k is the k^{th} iteration. To solve the above eq. (9) an alternating direction method of multipliers (ADMM) [21] is used.

3.2 Multi-scale decomposition of the reflectance

The derived parameter estimated by the conventional optical image formation-based method is firmly reliant on precision. After solving the visual image formation model with accuracy still, it may present undesired color artifacts. Even though we have considered the physical characteristics of water, the explicit possibilities and prior are still limited in miscellaneous and composite underwater scenes. However, the goal of this article is to develop a new background for the image enhancement of composite real-world underwater images. We have taken advantage of both model-based and enhancement-based categories to develop a multi-scale gradient-domain contrast enrichment approach stretched into underwater image restoration. In the proposed method, transmission has been calculated by the [30] rather than background light using improved DCP and solving the image formation optical model to indicate the superiority of the observed image. Further providing guidance for the how-to

managed enhancement coefficients which are used in this article. The conventional method has the complex of restoring the underwater appearance; to overcome this complexity, the author has proposed a plan to restore the underwater image properly. The first step is to decompose the Reflectance into a multi-scale layer. To complete this process, we have used weighted least squares (WLS) based filter [15], which splits the Reflectance into multiple layers. With the different values of λ, we obtain corresponding layers. Mathematically it is accepted as:

$$V = F_{\lambda}(R) = (I + \lambda L_R)^{-1} R \tag{14}$$

Where V is defined as the smooth version of the R. I is the matrix, which has a unity determinant, $L_R = D_x^T A_x D_x + D_y^T A_y D_y$ Where (x, y) is the spatial coordinate of a pixel, D_x and D_y are discrete differentiation operators, A_x and A_y Represent smoothness weights. As we increased the value of λ, we obtained the different, smoother image sequence $V^1, \cdots\cdots, V^{k-1}$.

In the second step, the progressive image is subtracted by the previous image, and the output after the subtraction is called multi-scale layers' representation. The V^{k-1} is called the coarsest image, and it is equivalent to base layer $\mathcal{B}$, when we put the value of k is one. This base layer consists of color variation of the Reflectance; on the other hand, the detail layer d_i Mathematically computed as:

$$d_i = V^{i-1} - V^i, \textit{Where } i = 1, \cdots\cdots\cdots k-1 \textit{and } V^0 = R \tag{15}$$

In this experiment, the value of K is set to three.

If we increase the value of parameter λ much high, it will produce a smoother coarsest image. Now by subtracting this roughest image from Reflectance, in this case, the maximum number of the minor textures, like noise and marine snow, can be captured in the output detail layer d_1. Hence, it will lead to noise amplification, so we need to manage the value of parameter λ to compensate for noise amplification. By default, in the weighted least squares (WLS) base filter, the value of parameter λ is set to unity. But in our experiment, the lambda value is set to below unity.

To obtain the information of color, brightness, and contrast of different layers corresponding to different lambda values, then apply the process to restore blurred images and low-light and color distortion problems simultaneously. The flow diagram of the proposed method is shown in Fig. 2, in which input underwater images are split into Illumination and Reflectance. Further Reflectance is divided into multi-scale layers, i.e., base and two detail layers. Gamma correction is carried out on the base $\mathcal{B}$ layer of Reflectance, and contrast enhancement in the gradient domain is applied on both detail layers $\boldsymbol{d_i}$. Contrast enhancement is used in the gradient domain to remove artifacts. The second part of the splitter input image, i.e., Illumination, applies brightness adjustment.

3.3 Brightness adjustment

In the Illumination (£) part, the variation of light happened. The brightness adjustment is used for low-light problems. In this layer, enhancement of Illumination is carried out. We have used the sigmoid function to improve the brightness. The magnitude of the sigmoid function lies between −1 to +1. Mathematically it is represented as:

$$£'(x, y) = Sigmoid(\beta £(x, y)) \tag{16}$$

Where *Sigmoid*(·) is the sigmoid function and β is the controlling parameter. In this experiment, β is set to:

$$\beta = 10 + \frac{(1-£_{mean})}{£_{mean}} \tag{17}$$

Where $£_{mean}$ is the mean of illumination part. If the value of $£_{mean}$ is low, then controlling parameter β will increase, this lead to enhancing the illumination part. The brightness Adjustment example is shown in Fig. 4 given below.

3.4 Gamma correction

Most of the underwater images appear blue or green due to the red component of the color channel disappearing first compared to the green and blue channels. The red channel has less energy in terms of frequency, or it has a greater wavelength than the green and blue channels so that it disappears first. According to retinex model, the input underwater image is split into Illumination and Reflectance. After decomposing the Reflectance into a subsequent layer, the first layer is the base with a colored variance. The base layer has low color brightness, so we need to correct it. So that we have used gamma correction, mathematically, it is represented as:

$$F(x,y) = \alpha(I^c)^{\gamma} \qquad c \in \{r,g,b\} \tag{18}$$

Where α is constant, in this experiment, its value is set to 0.5, and the value of gamma (γ) is set to 1.6. *Gamma* γ is related to the mean value of the image.

3.5 Contrast enhancement

In this section, a contrast enhancement in the gradient domain is used to enhance the underwater image by estimating the gradient domain of Reflectance instead of solving the image formation model. In this process, the undesired artifacts and noise are suppressed at the efficient level so that texture details enhancement performance is good, as shown in eq. (19).

The contrast of underwater input image can be estimated by the optical image formation model [11], which describe in eq. (19)

$$\sum_x \|\nabla P(x)\| = T\sum_x \|\nabla S(x)\| = \sum_x \|\nabla S(x)\| \tag{19}$$

Fig. 4 **a** Raw underwater image, **b** Brightness adjustment, and **c** Output of the proposed method

Where P denotes the underwater input image, and S is the scene radiance. x is the coordinate of the pixel ∇ as the gradient operator, and T is characterized as a transmission that has a value less than or equal to unity, i.e., $T \leq 1$. Its value is estimated by the fast method explained in [42]. The transmission is used as a reciprocal; it works as an enhancement coefficient. If the value of T tends to be zero artifact and noise is dominant in the enhanced image, to compensate for this effect, the average value of transmission T is used in this experiment. Hence, the mathematical expression for multi-scale detail layer in gradient domain is expressed as:

$$\nabla D_i^{'} = \frac{\omega_i}{\overline{T}} \nabla D_i \tag{20}$$

Where $\overline{T}$ is the average value of T and ω_i As the non-negative controlling parameters, it shows the enhancement strength. The $D_1^{'}$ component corresponding to ω_1 contains most of the small details and noise. Which desire to enhance less. While $D_2^{'}$, The part contains a firm edge and significant structure. The $D_2^{'}$ component corresponds to ω_2 need to enhance more strength. So that in this experiment, the value of $\omega_1 = 3$ and $\omega_2 = 4$ fixed. By solving the poison equation on modified gradients, two detail enhance layers $D_1^{'} and\ D_2^{'}$ are obtained. Accordingly, better-quality Reflectance is given by the summation of both the enhanced details layer and gamma-corrected base layer:

$$\begin{aligned} R^{'} = {} & Gamma\ Correction(B) + Enhance\ Detail\ Layer\left(D_1^{'}\right) \\ & + Enhance\ Detail\ Layer\left(D_2^{'}\right) \end{aligned} \tag{21}$$

The different stage output on Reflectance is shown in Fig. 5 below.

Finally, restored output images are a multiplication of improved Reflectance and improved Illumination. Mathematically it is given as:

$$I^{'} = R^{'} \cdot \pounds^{'} \tag{22}$$

3.6 The proposed algorithm for underwater image enhancement is summarized as follows

Step 1: As a first step, an input raw-underwater image has been taken for processing.
Step 2: Decomposed the input image into two parts, i.e., Illumination eq. (9) and Reflectance eq. (8).

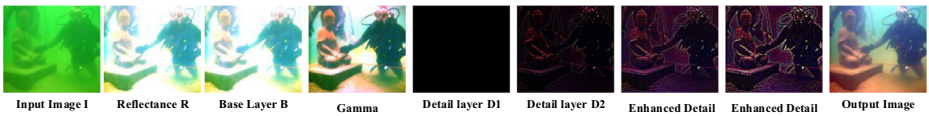


Fig. 5 Different stages of underwater image Reflectance

Step 3: Apply brightness adjustment on the illumination part as shown in eqs. (16) and (17).
Step 4: We perform a multi-scale decomposition of Reflectance in this step by eq. (14).
Step 5: Output of Reflectance are base layer, detail layer1, and detail layer2 as shown in eq. (15).
Step 6: Apply gamma correction on the base layer eq. (18).
Step 7: Contrast enhancement is carried out on the detail layer1 and detail layer2 eq. (20).
Step 8: Linear combination of the base layer, detail layer1, and detail layer2 eq. (21).
Step 9: Finally, multiply the linear combination of these three layers with the brightness adjustment of the illumination part. It will produce an enhanced output image in eq. (22).

4 Experiment and discussion

The experiment was performed on MATLAB2019b software. The experimental results of the proposed method have been compared to the existing underwater image enhancement method, which is also implemented on the same software. These methods are MIL [33], UDCPGIF [23], UDCPSM [28, 32], HUIE [35], CBFU [2], HPU [26], HTU [30], MSRCR [40] and MPORU [38] used to analyse both qualitative and quantitative measure. In this experiment, the Underwater Image Enhancement Benchmark (UIEB) [34] data set have been used. This is a publicly available data set; we have downloaded it from https://li-chongyi.github.io/proj_benchmark.html here only. The data set contain 950 real-world underwater images. Out of 950 images, 890 images consist of reference images. These images are considered to give the best enhancement result. The remaining 60 images have been considered to be challenging images because it is not given satisfactory results. In this experiment, we have implemented source code and fairly evaluated the parameters of different methods and their visual quality. For quantitative evaluation, we have used seven metrics which are the measure of enhancement (EME) [6], Desecrate Entropy (DE) [6], Peak Signal to Noise Ratio (PSNR) [7], Structural Similarity Index (SSIM) [43], Underwater Image Quality Measure (UIQM) [6], Underwater Color Image Quality Evaluation (UCIQE) [48] , and Patch-Based Contrast Quality Index (PCQI) [44]. The value of these metrics used for comparative analysis of the proposed method with other existing state-of-art methods are discussed below:

1. **Measurement of enhancement (EME):** It evaluates the max to min intensity ratio of an image and applies average logarithmic to it. The input image is divided into sub-image of size $k_1 \times k_2$. The parameters are mathematically described as [6]:

$$EME = \frac{1}{k_1 \times k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k2} \left(20 \times ln \left(\frac{I^w_{max,i,j}}{I^w_{min,i,j} + \varepsilon} \right) \right), \tag{23}$$

Where, $I^w_{max,i,j}$ and $I^w_{min,i,j}$ are the maximum and minimum intensities in a particular block $B(i, j)$, and ε is a small positive constant to avoid the indeterminant term.

2. **Discrete entropy (DE)** [6]: Discrete entropy reveals how much information is contained by the image, the discrete entropy of image I is mathematically expressed as

$$DE(I) = -\sum_{\forall k} p(k) log_2 p(k) \tag{24}$$

Where $p(k)$ is the probability density function (PDF), which is defined as

$$p(k) = \frac{h_i(k)}{N} \qquad for\ k = 0, 1, 2 \ldots\ldots K{-}1 \tag{25}$$

Where $h_i(k)$ is the number of pixels with gray-level k, and N is the total number of pixels in all images.

3. **Peak Signal to Noise Ratio (PSNR)** [7]**:** PSNR is used to calculate the peak signal-to-noise ratio. It is measured in terms of decibels. Basically, it is measuring the quality of the enhanced image with respect to the original image. In other words, we can say that PSNR gives the similarity of an underwater enhanced image against a degraded image based on the mean square error of each pixel. Performance of the proposed method as well as compared method is measured by parameters given below:

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left[I(i,j) - \widetilde{I}\,(i,j) \right]^2} \tag{26}$$

$$PSNR = 20\log_{10}\left(\frac{255}{RMSE}\right) \tag{27}$$

where M N is the total number of pixels in both images, i.e., enhanced and original, I as the original image and $\widetilde{I}$ As the enhanced underwater image at a particular level, MSE represents the collective root mean squared error between the underwater enhanced image and reference image. In contrast, PSNR represents a measure of the maximum error in decibels.

4. **Structural Similarity Index (SSIM)** [43]**:** It compares the reference image's structure with an enhanced underwater image. Mathematically structure similarity index is evaluated as:

$$SSIM\left(I, \widetilde{I}\right) = \frac{\left(2\mu_I \mu_{\underset{I}{\sim}} + C_1\right)\left(2\sigma_{\underset{II}{\sim}} + C_2\right)}{\left(\mu_I^2 + \mu_{\underset{I}{\sim}}^2 + C_1\right)\left(\sigma_I^2 + \sigma_{\underset{I}{\sim}}^2 + C_2\right)} \tag{28}$$

Where μ_I is the mean intensity of the image defined as:

$$\mu_I = \frac{1}{N} \sum_{i=1}^{N} I_i \tag{29}$$

and $\sigma_{\underset{II}{\sim}}$ is the standard deviation of the image defined as:

$$\sigma_{\widetilde{II}} = \frac{1}{(N-1)} \sum_{i=1}^{N} (I_i - \mu_i)\left(\widetilde{I}_i - \mu_{\widetilde{I}}\right) \tag{30}$$

C_1 and C_2 are the positive constants and it is used for the indeterminant situation when $\mu_I^2 + \mu_{\widetilde{I}}^2$ and $\sigma_I^2 + \sigma_{\widetilde{I}}^2$ are very close to zero.

5. **Underwater color Image quality evaluation (UCIQE)** [48]: Mathematically, it is defined as

$$UCIQE = c_1 \times \sigma_c + c_2 \times con_1 + c_3 \times \mu_s \tag{31}$$

Where, c_1, c_2 and c_3 are weighted coefficients. σ_c is the standard deviation of chroma, con_1 is the contrast of illuminance and μ_s It shows an average saturation.

6. **Patch-based contrast quality index (PCQI)** [44]**:** The patch-based contrast image quality index is evaluated as

$$PCQI = q_c(x,y) \times q_s(x,y) \times q_i(x,y) \tag{32}$$

Where $q_c(x,y)$ is the contrast change, $q_s(x,y)$ is the structural distortion and $q_i(x,y)$ is the mean intensity, mathematically it is defined as:

$$q_c(x,y) = \frac{4}{\pi} \cdot \arctan\left(\left|\frac{c_2^y}{c_2^x}\right|\right) \tag{33}$$

$$q_s(x,y) = \frac{c_2^y + r^T v_2}{\left\|c_2^y \cdot v_2 + r\right\|}, \tag{34}$$

$$\text{and } q_i(x,y) = e^{\left(-\frac{\left|c_1^x - c_2^y\right|}{\sqrt{NL}}\right)}, \tag{35}$$

4.1 Qualitative analysis

In this section, ten different images are taken, checking their enhanced visual quality using different methods. Usually, a specific state-of-art process works well for a special kind of image only, and it may not be suitable in an extensive logic for another purpose. Figure 6 shows the enhanced visual quality of images of different methods taken from the UIEB [34] database.

Image capture in low-light conditions, as shown in Fig. 6, because of severe light distortion. It can be seen from the results that MPORU [38], HUIE [35], MIL [33], CBFU [2], and

proposed methods show good visual quality. On the other hand, HPU [26], UDCPGIF [23], and UDCPSM [28, 32] are inclined to give an abnormal appearance since the result is a little blurred to show the natural response. The methods MSRCR [40] and HTU [30] offer a bright visual quality that does not look natural. Hence, our method illustrates outstanding performance in the detail enhancement layer in the hazy background region. According to the optical image formation model, light rays get attenuated while traveling underwater because they are wavelength-dependent. Since the red component has less energy than the green and blue parts, i.e., it has the highest wavelength compared to green and blue light. So that red light disappears firstly. However, underwater images appear greenish or bluish, as shown in Fig. 6(1,3,4,8,10). UDCPGIF and UDCPSM methods do not positively affect contrast enhancement for underwater images. The proposed method results are comparable to the HUIE method, but it is superior to preserving more details.

4.2 Quantitative analysis

In this section, we calculate the parameters of enhanced images. These are the following parameters, measure of enhancement (EME) [6], Desecrate Entropy (DE) [6], Peak Signal to Noise Ratio (PSNR) [7], structural similarity index (SSIM) [43], underwater image quality measure (UIQM) [6], underwater color image quality evaluation (UCIQE) [48], and patch-based contrast quality index (PCQI) [44]. In this experiment, the average value of these parameters is calculated over an underwater image enhancement benchmark (UIEB) [36] data

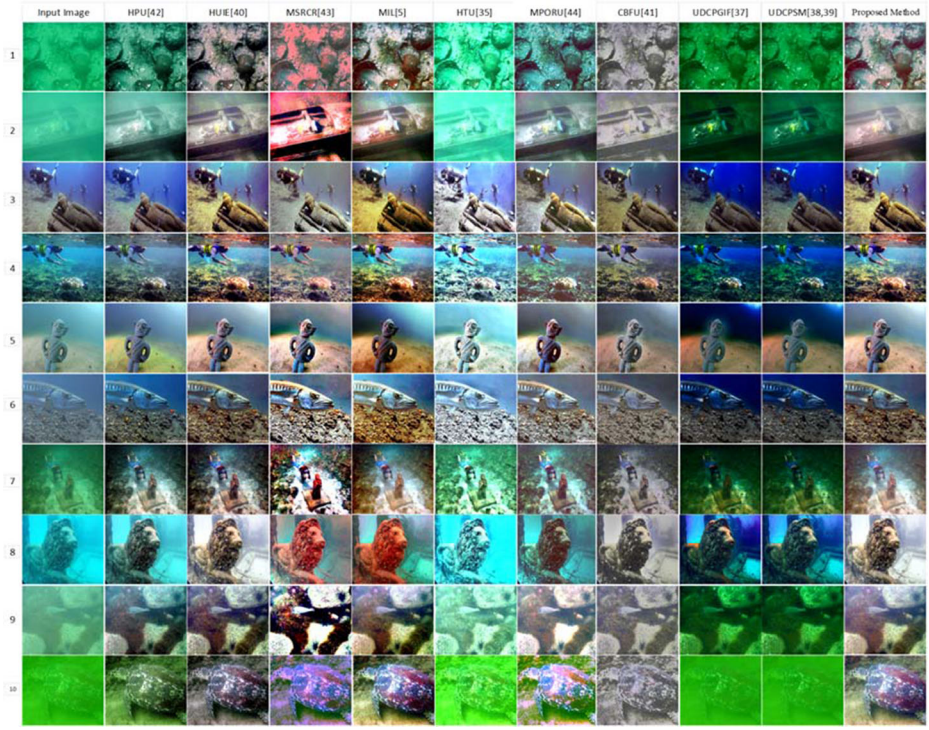


Fig. 6 Qualitative comparison of the proposed method with existing methods on ten images

set which contains 890 raw images and 60 reference images. The evaluated results are shown in Table 1.

Furthermore, in order to conduct a quantitative analysis of the proposed method, we categorize the parameters in two ways, the first one is full-reference evaluation metrics, and the second is non-reference evaluation metrics. We treat the reference images supported by UIEB as the ground truths for full-reference evaluation and compute the EME, PSNR, SSIM, and PCQI among the restored results and corresponding references. A higher value of the PCQI metric better the contrast variation for human visual perception. The MSRCR and proposed method have approximately the same PCQI; the proposed method is slightly greater than MSRCR. Now comes the PSNR metric. A higher value of the peak signal-to-noise ratio means the result is closer to the reference image. The CBFU has a more significant peak signal-to-noise ratio than another method. The proposed method has comparable PSNR results to CBFU, as shown in Table 1. EME parameters measure the enhancement; the higher the value shows the better result. The proposed method has a higher value compared to another method. Another full-reference metric is SSIM; its value lies between zero to one, the value closer to unity; it represents the ideal result in context to the original image. The non-reference criterion is discrete entropy (DE), UIQM, and UCIQE. Discrete entropy shows the information contained in an enhanced image; a higher value is better for the image. In our experiment proposed method has a higher value compared to another method. MIL, HUIE, HPU, and HTU methods have comparable values but are lower than the proposed method. A higher value of UIQM and UCIQE is favorable. The proposed method gives a higher value as compared to the other nine methods and can be observed in Table 1. A higher value of UIQM has increased colorfulness, sharpness, and contrast, while the higher value of UCIQE has increased the color-cast, blurring, and low-contrast.

The experiment shows that the overall performance of the proposed method is better than most of the other existing state-of-art methods. The most important thing is that our method works on any underwater images, and it will accept any contrast level.

5 Conclusion and future work

In this article, the author gives a novel method that is effectively applicable to almost all real-world underwater images with different contrast-degradation levels where most of the existing state-of-art methods do not provide satisfactory results. We capture additional information into the corresponding layer to compensate for the brightness, color, and contrast attenuation. Firstly, we applied a retinex model to the input image and decomposed it into Illumination and Reflectance; Illumination represents brightness variation. Then, Reflectance is decomposed into multi-scale layers. Gamma correction is applied to the smoothest base layer, while contrast enhancement is applied to the gradient domain of detail layers. Adding the output of all reflectance layers, then multiplying it with brightness adjustment, and finally, we get visually satisfying results with improved brightness, contrast, and color. Nine existing state-of-art methods were performed over the proposed method to find out their strength, and it shows that the proposed method gives a very high quality of an enhanced underwater image. The proposed way is not complex due to straightforward operations like addition, subtraction, multiplication, and convolution. Experimental result shows that the proposed scheme produces the highest full-reference quality metric EME, SSIM, PCQI, PSNR, and non-reference quality metric DE, UCIQM, and UIQM. All these metrics show the highest value in the case of the

proposed method, except the PSNR values of CBFU are slightly more significant than our method. All these parameters are evaluated on the same data set, which is publicly available.

Despite the proposed method producing a high-quality image for various types of underwater images. Nevertheless, the proposed method cannot produce a good result when the underwater images contain significant noise, but it is still better than existing state-of-art methods. When noise interference is much more in the underwater pictures, the proposed method does not work efficiently, so we need to work on that in the future.

Data availability No data associated with this manuscript.

Declarations

Ethics approval This article does not contain any studies with animals or human participants performed by any authors.

Conflict of interest All Authors of this work declare no conflict of interest.

References

1. Ancuti C, Ancuti CO, Haber T, Bekaert P (2012) Enhancing underwater images and videos by fusion. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE. pp. 81–88
2. Ancuti CO, Ancuti C, De Vleeschouwer C, Bekaert P (2018) Color balance and fusion for underwater Image enhancement. IEEE Trans Image Process 27:379–393. https://doi.org/10.1109/TIP.2017.2759252
3. Ancuti CO, Ancuti C, De Vleeschouwer C, Sbert M (2019) Color channel compensation (3C): a fundamental pre-processing step for image enhancement. IEEE Trans Image Process 29:2653–2665
4. AUTHOR (n.d.) Automatic Red-Channel underwater image restoration _ Elsevier Enhanced Reader.pdf
5. Bailey GN, Flemming NC (2008) Archaeology of the continental shelf: marine resources, submerged landscapes and underwater archaeology. Quat Sci Rev 27:2153–2165. https://doi.org/10.1016/j.quascirev.2008.08.012
6. Bhandari AK (2020) A logarithmic law based histogram modification scheme for naturalness image contrast enhancement. J Ambient Intell Humaniz Comput 11:1605–1627. https://doi.org/10.1007/s12652-019-01258-6
7. Bhandari AK, Singh VK, Kumar A, Singh GK (2014) Cuckoo search algorithm and wind driven optimization based study of satellite image segmentation for multilevel thresholding using Kapur's entropy
8. Bhunia AK, Perla SRK, Mukherjee P, Das A, Roy PP (2019) Texture synthesis guided deep hashing for texture image retrieval. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE. pp. 609–618
9. Cao K, Peng Y-T, Cosman PC (2018) Underwater image restoration using deep networks to estimate background light and scene depth. In: 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI). IEEE. pp. 1–4
10. Carlevaris-Bianco N, Mohan A, Eustice RM (2010) Initial results in underwater single image dehazing. In: Oceans 2010 Mts/IEEE Seattle. IEEE. pp. 1–8
11. Chiang JY, Chen YC (2012) Underwater image enhancement by wavelength compensation and dehazing. IEEE Trans Image Process 21:1756–1769. https://doi.org/10.1109/TIP.2011.2179666
12. Dai C, Lin M, Wang J, Hu X (2019) Dual-purpose method for underwater and low-light image enhancement via image layer separation. IEEE Access 7:178685–178698
13. Deng G (2011) A generalized unsharp masking algorithm. IEEE Trans Image Process 20:1249–1261. https://doi.org/10.1109/TIP.2010.2092441
14. Drews P, Nascimento E, Moraes F, Botelho S, Campos M (2013) Transmission estimation in underwater single images. In: Proceedings of the IEEE international conference on computer vision workshops. pp. 825–830
15. Farbman Z, Fattal R, Lischinski D, Szeliski R (2008) Edge-preserving decompositions for multi-scale tone and detail manipulation. SIGGRAPH'08 Int. Conf. Comput. Graph. Interact. Tech. ACM SIGGRAPH 2008 Pap. 2008 https://doi.org/10.1145/1399504.1360666

16. Fu X, Huang Y, Zeng D, Zhang X-P, Ding X (2014) A fusion-based enhancing approach for single sandstorm image. In: 2014 IEEE 16th international workshop on multimedia signal processing (MMSP). IEEE. pp. 1–5
17. Fu X, Liao Y, Zeng D, Huang Y, Zhang XP, Ding X (2015) A probabilistic method for Image enhancement with simultaneous illumination and reflectance estimation. IEEE Trans Image Process 24:4965–4977. https://doi.org/10.1109/TIP.2015.2474701
18. Fu X, Zeng D, Huang Y, Zhang XP, Ding X (2016) A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 2782–2790. https://doi.org/10.1109/CVPR.2016.304
19. Fu X, Fan Z, Ling M, Huang Y, Ding X (2017) Two-step approach for single underwater image enhancement. In: 2017 international symposium on intelligent signal processing and communication systems (ISPACS). IEEE. pp. 789–794
20. Galdran A, Pardo D, Picón A, Alvarez-Gila A (2015) Automatic red-channel underwater image restoration. J Vis Commun Image Represent 26:132–145
21. Goldstein T, Osher S (2009) The split Bregman method for L1-regularized problems. SIAM J Imaging Sci 2:323–343. https://doi.org/10.1137/080725891
22. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33:2341–2353
23. He K, Sun J, Tang X (2013) Guided image filtering. IEEE Trans Pattern Anal Mach Intell 35:1397–1409. https://doi.org/10.1109/TPAMI.2012.213
24. Hitam MS, Awalludin EA, Jawahir Hj Wan Yussof WN, Bachok Z (2013) Mixture contrast limited adaptive histogram equalization for underwater image enhancement. Int. Conf. Comput. Appl. Technol. ICCAT 2013. 0–4. https://doi.org/10.1109/ICCAT.2013.6522017
25. Hooda A, Kumar A, Goyat MS, Gupta R (2022) Estimation of surface roughness for transparent superhydrophobic coating through image processing and machine learning. Mol Cryst Liq Cryst 726:90–104
26. Hou G, Pan Z, Huang B, Wang G, Luan X (2018) Hue preserving-based approach for underwater colour image enhancement
27. Hu J, Jiang Q, Cong R, Gao W, Shao F (2021) Two-branch deep neural network for underwater image enhancement in hsv color space. IEEE Signal Process Lett 28:2152–2156
28. Image, S., Removalusing, H.: (12) United States Patent. 2, (2012)
29. Iqbal K, Odetayo M, James A, Salam RA, Talib AZH (2010) Enhancing the low quality images using unsupervised colour correction method. In: 2010 IEEE International Conference on Systems, Man and Cybernetics. IEEE. pp. 1703–1709
30. Krishnapriya TS, Kunju N (2019) Underwater Image Processing using Hybrid Techniques. Proc. 1st Int. Conf. Innov. Inf. Commun. Technol. ICIICT 2019. 13–16. https://doi.org/10.1109/ICIICT1.2019.8741468
31. Kumar A, Chauda P, Devrari A (2021) Machine learning approach for brain tumor detection and segmentation. Int J Organ Collect Intell 11(3):17. https://doi.org/10.4018/IJOCI.2021070105
32. Levin A, Lischinski D, Weiss Y (2008) A closed-form solution to natural image matting. IEEE Trans Pattern Anal Mach Intell 30:228–242. https://doi.org/10.1109/TPAMI.2007.1177
33. Li CY, Guo JC, Cong RM, Pang YW, Wang B (2016) Underwater image enhancement by Dehazing with minimum information loss and histogram distribution prior. IEEE Trans Image Process 25:5664–5677. https://doi.org/10.1109/TIP.2016.2612882
34. Li C, Guo C, Ren W, Cong R, Hou J, Kwong S, Tao D (2019) An underwater image enhancement benchmark dataset and beyond. IEEE Trans Image Process 29:4376–4389
35. Li X, Hou G, Tan L, Liu W (2020) A hybrid framework for underwater image enhancement. IEEE Access 8:197448–197462. https://doi.org/10.1109/ACCESS.2020.3034275
36. Li C, Guo C, Ren W, Cong R, Hou J, Kwong S, Tao D (2020) An underwater Image enhancement benchmark dataset and beyond. IEEE Trans Image Process 29:4376–4389. https://doi.org/10.1109/TIP.2019.2955241
37. Liang Z, Wang Y, Ding X, Mi Z, Fu X (2021) Single underwater image enhancement by attenuation map guided color correction and detail preserved dehazing. Neurocomputing. 425:160–172
38. Mi Z, Li Y, Wang Y, Fu X (2020) Multi-purpose oriented real-world underwater Image enhancement
39. Pandey V, Anand K, Kalra A, Gupta A, Roy PP, Kim B-G (2022) Enhancing object detection in aerial images. Math Biosci Eng 19:7920–7932
40. Parthasarathy S, Sankaran P (2012) An automated multi scale Retinex with color restoration for image enhancement. 2012 Natl. Conf. Commun. NCC 2012. https://doi.org/10.1109/NCC.2012.6176791
41. Song W, Wang Y, Huang D, Tjondronegoro D (2018) A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration. In: Pacific Rim Conference on Multimedia. Springer. pp. 678–688

42. Tarel JP, Hautière N (2009) Fast visibility restoration from a single color or gray level image. Proc. IEEE Int. Conf. Comput. Vis. 2009-Janua, 2201–2208. https://doi.org/10.1109/ICCV.2009.5459251
43. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13:600–612. https://doi.org/10.1109/TIP.2003.819861
44. Wang S, Ma K, Yeganeh H, Wang Z, Lin W (2015) A patch-structure representation method for quality assessment of contrast changed images. IEEE Signal Process Lett 22:2387–2390. https://doi.org/10.1109/LSP.2015.2487369
45. Wang Y, Song W, Fortino G, Qi LZ, Zhang W, Liotta A (2019) An experimental-based review of Image enhancement and Image restoration methods for underwater imaging. IEEE Access 7:140233–140251. https://doi.org/10.1109/ACCESS.2019.2932130
46. Wei X, Yu L, Tian S, Feng P, Ning X (2021) Underwater target detection with an attention mechanism and improved scale. Multimed Tools Appl 80:33747–33761
47. Xu B, Zhou D, Li W (2022) Image enhancement algorithm based on GAN neural network. IEEE Access 10:36766–36777
48. Yang M, Sowmya A (2015) An underwater color Image quality evaluation metric. IEEE Trans Image Process 24:6062–6071. https://doi.org/10.1109/TIP.2015.2491020
49. Zhang S, Wang T, Dong J, Yu H (2017) Underwater image enhancement via extended multi-scale Retinex. Neurocomputing. 245:1–9
50. Zhao X, Jin T, Qu S (2015) Deriving inherent optical properties from background color and underwater image enhancement

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

Springer Nature or its licensor 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.

UNIVERSITY NEWS
Vol. 60 September 19-25
No. 38 2022
Price Rs. 30.00

A Weekly Journal of Higher Education Published by the Association of Indian Universities

In This Issue

ITEMS PAGE

Articles

University of Delhi Library: Celebrating the Century 3

Changing the Higher Education Landscape: Implementing the Mandate of National Education Policy—2020 9

Vision 2030 in Indian Higher Education 14

Attitude of Students of Old Styled Universities towards E-learning: A Study 19

Convocation Address

NALSAR University of Law, Hyderabad 23

Campus News 26

Theses of the Month (Social Sciences) 32

Advertisement 36

New Subscription Tariff
(Effective April 01, 2020)

	Inland		Foreign	
Institutions	Academics/ Students *(at residential address only)*		Airmail	Surface Mail
	Rs.	Rs.	US$	US$
1 year	1250.00	500.00	210.00	170.00
2 years	2200.00	900.00	400.00	300.00

Subscription is payable in advance by Bank Draft/MO only in favour of Association of Indian Universities, New Delhi.

Opinions expressed in the articles are those of the contributors and do not necessarily reflect the views and policies of the Association.

Patron:
Prof. Suranjan Das

Editorial Committee Chairperson:
Dr (Ms) Pankaj Mittal

Editorial Committee:
Dr Baljit Singh Sekhon
Dr Amarendra Pani
Dr Youd Vir Singh

Editor:
Dr Sistla Rama Devi Pani

#Let'sBeatCoronaTogether

University of Delhi Library: Celebrating the Century

Narender Kumar* and Lalita**

Higher Education in Delhi started with the establishment of St. Stephen College (1881). A few more colleges, namely Hindu College (1899), Lady Harding Medical College for Women (1916), and Ramjas College (1917), were also established before the first University – the University of Delhi – came into being in 1922. The University of Delhi was established by an Act of the Central Legislature, except for the Lady Harding Medical College for Women and three other colleges affiliated with the newly established University of Delhi. The University also started two faculties, i.e., Arts and Science, in the same year.

It started with a modest two faculties, three colleges, and about 750 students. Presently, the University of Delhi comprises 16 faculties, 86 departments, 20 centres, three institutes, and 91 colleges. Currently, it offers 540 undergraduate, postgraduate, M.Phil., PhD, certificate, and diploma programmes. The 2020-2021 annual report says that the University has enrolled 2,16,823 students as undergraduates and 30,366 as postgraduates in regular mode. 475 students have been registered for M. Phil and 4011 for PhD, and 5948 in certificate/ diploma/PG diploma. In informal mode, the University has enrolled 5,72,261 students for undergraduate and 33,967 for postgraduate programme (University of Delhi, Annual Report, 2020-21).

Journey of University of Delhi Library

The First Quarter (1922 – 1946)

The University of Delhi is celebrating the year 2022 as its centenary year, so as the Delhi University Library. Delhi University has been at the forefront of library matters in the country since 1922. The University library benefited from the wisdom of the most outstanding Indian Library Scientist and the Father of The Library Movement in India - Dr S.R. Ranganathan and Dr K.L Kaul. These great luminaries of the academic World of their times harboured an exciting vision of the importance of books and libraries, and all of them patronized the libraries in their respective ways. The present article attempts to pen down the historical journey of a hundred years of the Delhi University Library, highlighting the critical milestones.

The story of placing their patronage at the service of the libraries starts right from the first Vice Chancellor, Dr Hari Singh Gour (1922-26), whose devotion to libraries was an extreme passion to the extent that he chaired the meetings of the University Library Committee and bestowed care and consideration in laying the foundation of the office which is

* *Deputy Librarian (Consultant), South Campus, Benito Juarez Road, New Delhi-110021. E-mail: narenderkumar.1959@gmail.com*

***Assistant Librarian, Delhi Technological University, Bawana Road, Shahbad Daulatpur Village, Rohini, New Delhi-110042, Delhi. E-mail: lalitakumar69@gmail.com*

today known by the name of Delhi University Library System. The University Library was established on his initiative under the Chairmanship of Father Reverend F.J. Western, a mathematics missionary teacher from Cambridge University who taught at St. Stephen's College. He stayed for the initial one and a half years and was succeeded by the first Honorary Librarian, Khan Bahadur Pirzada Mohammad Hussain (1923-27), a teacher of oriental languages at St. Stephen's College.

The Vice Chancellor, Rai Bahadur Sir Moti Sagar (1926-1930), was an ardent lover of libraries not only of the University but also of the country. The second honorary Librarian Khan Saheb Maulvi Mohammad Fazludin (1927-31), again from St. Stephen's College, took over the charge of the University Library. The Vice Chancellor, a Fellow of the Lahore Library Association and a Trustee of the Dayal Singh Public Library at Lahore, Chaired the All India Library Conference held at Lahore in December, 1929 and spelt out his vision of the library as an expendable wealth of the country.

> *"Of all forms of wealth, a well-equipped library is the most valuable possession of any nation. Every other emblem of wealth passes empires, kingdoms, great possessions, and hoarded treasure; magnificent surroundings disappear like the dreams of a night. Of all past civilizations and their pruned achievements, the highest and the best heritage left is the thought enclosed in the caskets of books. Wisdom and wit, creations of fancy and imagination, things of beauty, profound thoughts on the great problems that have faced man from the beginning of creation, elaborate doctrines and tenets of faith, an exploration into the secret of nature, the steady progress and achievements of science, all these and more are to be found in a few shelves stocked with books. That nation is poor, however great; its material wealth does not possess well-stocked libraries." (Sagar, 1930*)

The Vice Chancellor Khan Bahadur Dr Abdur Rehman (1930-34), associated with the developmental work of Delhi University Library since its very beginnings, appointed Dr Ishtiq Hussain Qureshi (1931-37) a teacher of history at the St. Stephen's College as the Honorary Librarian in 1931. So far, the Delhi University Library was leading a vagrant life shifting from one building to the other in the University. The Vice Chancellor recognized the need for a permanent house for the Library and, in October 1933, permitted the Library to be shifted to the 'Green Room' of the Vice Regal Lodge - a new and respectable accommodation built in 1922 to entertain the visiting Prince of Wales.

During the tenure of Vice Chancellor Rai Bahadur, Dr Ram Kishore (1934-38) formed a historical relationship between the Delhi University Library and the Indian Library Association (1933). Dr A. Bhattacharya (1937-47), Teacher of English at the Hindu College and later the College's Principal took over as the fourth and the last Honorary Librarian in 1937 and stayed at the helm of Delhi University Library for long ten years till 1947. Delhi University Library hosted the Third All India Library Conference in December 1937. The conference focused on three questions (i) Furtherance of library movement in India; (ii) Promotion of training of librarians; (iii) Improvement of the status of librarians, which have evoked unending and perennial debate and also continue to remain elusive even in our times. Speaking on this occasion, Vice Chancellor highlighted the role of libraries in the changing Indian society:

> *"Knowledge and culture are no longer the monopolies of the few, nor is it considered necessary or fair that they should be so. We live in an age of mass awakening characterized by a determination to abolish illiteracy and to allow everybody to share the blessings of sweetness and light, which have been so long the monopoly of a privileged class living in cultured ease. In these changed circumstances, libraries should naturally function as the necessary instruments of mass awakening and general diffusion of culture. This movement in every advanced country has determined the planning and organization of libraries in most democratic countries. In India, we are on the threshold of constructive democracy, and nothing perhaps is more important today than to initiate a library movement on a national scale as a part of a national campaign against illiteracy, ignorance, and superstition." (*Rehman, 1931)

Then came the titanic age of the Vice Chancellor, Sir Maurice Gwyer, from 1938 to 1950(12 Years). Sir Maurice Gwyer invoked bold and path-breaking initiatives for the reform and upliftment of the University on the whole. Delhi University Library also received due attention of this visionary Vice Chancellor, who was motivated by his conviction of a library which he said:

*"There is no place more fascinating than a library and no place where it is possible to spend happier hours. Those who have benefited from the opportunities a good library alone can give will always be ready to cooperate in extending those opportunities to others." (*Gwyer, *1937-38)*

When he took over the reins of the University, he found Delhi University Library in shambles but he transformed it into an efficient organisation (Shrimali, 1922). In June 1941, meeting the request of the Vice Chancellor Sir Gwyer, Shri GD Birla donated Rs. 25,000 for the reorganization and re-planning of the Delhi University Library. Sir Gwyer supplemented University Library with one more library—the Law Library in 1942). The opportunity to improve library resources came in 1946 when it received a book grant of Rupees 2.5 lakh. Sir Gwyer was the first to realize that the solution to the space problem lay in a separate building and allotted the present site for the future library then. He invited Dr S.R. Ranganathan of Madras University librarian to look into the state of affairs of the Delhi University Library and make recommendations for its improvement and reorganization. Ranganathan accepted the invitation on the condition that he should be allowed to teach and conduct research and not be burdened with administrative work. Sir Maurice acceded to the request and gave him every facility. On the occasion of the Silver Jubilee year of the University, the Vice Chancellor conferred upon Ranganathan the honour of an honoris causa doctorate. Ranganathan concentrated on training and research and started a Diploma of Library Science course (changed to Bachelor of Library and Information Science) in 1946.

2nd Quarter 1947-1971

This quarter witnessed the establishment of libraries at the Central Institute of Education now became the Faculty of Education and Department of Modern European Languages in 1948. The Department later bifurcated into, i.e., the Department of Germanic and Romance Studies and the Department of Slavonic and Finno Ugrian Studies, and both have their libraries. In 1948, the Ratan Trust donated Rs. One lakh to Delhi School of Economics Society for its library building. It is thus, started the Ratan Tata Library began in 1948.

The Master of Library Science course, probably the first master's degree course in librarianship not only in the country but also in the whole of the Commonwealth countries, started in 1948. At Delhi University, Ranganathan drafted his famous Thirty-Year Plan for the Development of Library System in India covering 1950-80, published by Delhi University Press. Ranganathan had suggested expanding the library staff and expressed that to get the best out of the Library in the long run, the University should spend as much on library services as on books. As a first step, the post of Sub-Librarian was created and eventually made into that of the University Librarian. The Vice-Chancellor superbly fast on one of Ranganathan's recommendations - that a young graduate with research experience, a good personality, and a desire for social service be selected (i.e., as paid Librarian) and sent up for training - selected Professor S. Dasgupta for the post of the first professionally trained University Librarian. The University sent him for library training at Madras University under Prof. Ranganathan. This step ended the 25-year-old tradition of honorary librarianship in the University and set an example before other universities. Professor S. Dasgupta served six Vice Chancellors as University Librarians- Sir Maurice Gwyer, Dr S.N. Sen, Dr G.S. Mahajani, Professor N.K: Sidhanta, and Dr C.D. Deshmukh- till his death in 1966. Under Dasgupta's leadership and guidance, the University library flourished, knowing only the sky as its limit.

Vice Chancellor Dr S.N. Sen (1950-53) vigorously pursued Sir Maurice's dream of a new and suitable home for the Library. Vice Chancellor Professor G.S. Mahajani (1953-57 laid down the foundation stone of the new library building in 1955. The new library building took nearly four years (1955-58) to get constructed. Chancellor of the University Dr S. Radhakrishnan inaugurated the present Central Library building in December 1958, while Professor VKRV Rao (1957-60) was the Vice Chancellor. The decade of the 1960s witnessed the development of the Delhi University Library System and started many departmental libraries in the Faculty of Arts, Mathematics, Social Science, and Management. A library was born in the Faculty of Music and Arts also. All the major departments of the Faculty of Science developed their separate collection and libraries between 1962—1965. A new library was established in the Department of Chinese and Japanese Studies, now named East Asian Studies, in 1967. At the end of the decade, another library was established in the Faculty of Management Studies in 1969.

The Vice Chancellor Dr CD. Deshmukh (1962-67), after the demise of Professor Dasgupta (1966), invited Professor S. Bashiruddin (1966-67), who had retired as the University Librarian from Aligarh Muslim University, its second University Librarian. Dr Deshmukh's single monumental contribution was to invite Ford Foundation to study and identify the highs and the lows in the library system. Ford Foundation sent its emissary Dean Carl M White as the library expert in 1965-66. Carl White burnt the midnight oil in the company of Professor S. Dasgupta, and both of them honestly and very candidly spotted the weaknesses of the library system to be corrected with the aid of the Ford Foundation. They also identified areas that could be improved through managerial and administrative changes. Carl White put forth his findings in a book entitled Survey of Delhi University Library, published by the University of Delhi in 1965 (White, 1965). Based on the Carl White survey, Ford Foundation made over a grant of one million dollars to the Library System to update book collections, professional development of the library staff, and induction of reprographics technology in library work. Based on another recommendation, Carl White brought about a change in the Ordinance XVI of the University statute book under which the Library's human resources and financial administration at the University's expense were centralized in the hands of the University Librarian.

Vice Chancellor Professor B.N. Ganguly (1967-69) opened his innings by initiating a large-scale developmental work based on Carl White's survey. The Librarian, Mr Bashuriddin, demitted his office in August 1967, and the reins of the Library System were taken over by A.L. Kapoor (1967-68) as an interim arrangement. University Library System hosted the Seventh Annual Conference of IASLIC in December 1967. Professor Ganguly appointed Dr B.V.R. Rao, who retired from the Indian Institute of Science, Bangalore, as the third University Librarian on a tenure basis from 1968-72.

The Vice Chancellor, Professor K.N. Raj (1969-70), appointed a Working Group on the University Library System to chalk out a program for future development of the libraries of the University.

3rd Quarter 1972 – 1996

The tenure of Vice Chancellor Prof. Sarup Singh (1970-74) stands taller than that of his predecessors because of his bold and outstanding initiatives. He set the expansion of the University as his first goal and made Delhi University a multi-campus University by setting up the South Campus. The South Campus had to have its university-level parallel library to meet the academic and research needs in the new upcoming academic disciplines. The four years tenure of University Librarian Dr B.V.R. Rao ended in December 1972. The post was offered to Professor D.R. Kalia, Director, Central Secretariat, but the Ministry of Education did not spare him. After that Vice Chancellor invited his time-tested colleague Mr K.L. Kaul, a Deputy University Librarian. Dr Kaul joined Delhi University Library as the fourth University Librarian on January 1973 for a year only. With the help of Professor Sarup Singh, Kaul initiated the processes of rationalizing library services and strengthening the libraries according to their status and importance. Dr Kaul concretized the concept of carrying library service right to the doorsteps of undergraduate students by setting up two undergraduate textbook libraries in the East and the South of Delhi City. This experiment proved beneficial to the undergraduate students of the University, leaving the Central Library to cater to the needs of postgraduates and research scholars.

Prof. Sarup Singh also placed the University Librarian, the Deputy Librarians, the Assistant Librarians, and the Documentation Officers at par with the teachers of the University in matters of study leave and duty leave for attending professional conferences. Dr Kaul also took pains to straighten the working and utilization of the Ford Foundation grant by taking prompt action on the pending problems. Dr Kaul, with the support of Professor Sarup Singh, also initiated measures to string together all the libraries into an interactive system called Delhi University Library System (DULS) while retaining their functional autonomy. In December 1973, after Dr Kaul left the University after a year, the officiating charge of the Library System reverted to Mr A.L. Kapoor.

Vice Chancellor Professor R.C. Mehrotra (1974-79) took up the appointment of a University Librarian and succeeded in bringing Professor A.P. Srivastava as the fifth University Librarian. Professor Srivastava joined the University in August 1975 and stayed as the University Librarian for 21 years, demitting his office voluntarily in March 1996. An earlier record of the sort was only of Professor Dasgupta, the University Librarian, for 24 years (1942-66). Of the myriad activities of Professor A.P. Srivastava, the one that stands out as scintillating

is his contribution to increasing the Library's human resources almost threefold. Two more Zonal Libraries for undergraduates were added in the West and North of Delhi. Professor Srivastava played the central role of a midwife in the birth of the Central Science Library, of which Vice Chancellor Professor Mehrotra had a vision. The process of utilization of the Ford Foundation grants successful close by Professor Srivastava. Professor Srivastava organized the 22nd All India Library Conference of the ILA in January 1976.

Vice Chancellor Professor Gurbux Singh (1980-85) will be remembered for his tireless efforts in securing academic status for the university and college librarians from the University Grants Commission. Professor Gurbux Singh also granted more administrative and financial autonomy to the various constituent libraries of DULS. Between 1981 and 1985, different degrees of financial and administrative decentralization resulted in the emergence of the Central Reference Library (CRL), South Campus Library, Central Science Library (CSL), Arts Library, the Ratan Tata Library, and Law Library. However, the Delhi University Library System (DULS), of which these libraries are integral, continues to be headed by the University Librarian.

The university and college librarians will remain grateful to Vice-Chancellor Professor Moonis Raza (1986-90) for extending Merit Promotion Scheme's privileges, re-employment after 60 years of age, and franchise rights to the librarians in the elections of the Academic and Executive Councils. He got significant improvements carried out in the library infrastructure of South Campus, Central Science, and the Ratan Tata. He inducted special provisions in 1987 to enable the teachers and the research students to consult and borrow books from any unit of the Library System with their membership originating from any library. Professor Raza made a modest beginning of computer-based services in the areas of new books added to the Library, back runs of periodicals, and library membership. Professor Moonis Raza, perturbed over the financial crisis afflicting the libraries, made a multi-pronged approach to deal with this disturbing feature by reducing duplication of periodicals, developing networking, seeking enhancement in grants from the UGC, and introducing an internal economy. The Government of India bestowed endowments of Rs. seven crores to the Ratan Tata Library between 1994-2004 for maintaining the journals collection and modernization of the library.

Vice Chancellor Professor Upendra Baxi (1991-95) harboured an absolute faith in Professor A.P. Srivastava. The duo drastically revised the then-existing legal framework and drafted new ordinances for the two-tier governance of the libraries of the University. Professor Baxi brought a high-powered Governing Body with the Vice Chancellor as the Chairman into existence. This body was directly accountable to the Vice Chancellor in library matters of human resources and finances. The second part of the ordinance mandated satisfactory library services to readers as their primary responsibility. The twin ordinances were mutually exclusive in the areas of their duties, but at the same time, they complemented each other, the one being subservient to the other.

4th Quarter 1997 - 2022

Professor V.R. Mehta (1995-2000) took over the reins of the University in 1995, and Professor Srivastava voluntarily demitted his office in March 1996. Before selecting as a regular Librarian, the charge of the DULS moved from one Deputy Librarian to another Deputy Librarian, including Mr A. L. Kapoor, Mr H. C. Jain, and Mr M.L. Saini as officiating Librarian. Professor Mehta selected Mr M.L. Saini as the sixth University Librarian in 1998. Professor Mehta set unto himself the arduous task of inducting ICT in the library and setting up the library computerization mission under the care of South Campus Director Professor Abhai Man Singh, who was able to give consultancy in selecting appropriate computers for DULS. University Librarian and all his senior professional colleagues worked with matching diligence in harnessing the computer hardware and software (TROODON) to library services in the shortest possible period. DULS libraries created databases of library documents.

Professor Deepak Nayyar (2000-2005) became the Vice Chancellor in 2000. Mr M L Saini got retirement voluntarily in 2004. The Vice Chancellor appointed Professor K M Srimali as Officer on Special Duty (OSD) for DULS. During the OSD period, Ordinance XVI got revised, and certain powers vested with University Librarian were curtailed. Professor Nayyar appointed Dr S Majumdar as the seventh University Librarian in 2005. Dr Majumdar initiated several new initiatives to keep pace with the latest ICT environment. These new initiatives include preparing Websites of DULS libraries, subscription of electronic databases, digitization of books and

theses, information literacy and competency program, development of IT infrastructure for the physically challenged, and organization of the International Conference on Academic Libraries (ICAL). The University constituted the Kapthpalia Committee in 2007 to revisit the existing DULS staff promotion policy, find flaws, and recommend remedies. The Committee submitted its recommendations to be considered a breakthrough in human resources deployment a breakthrough. The Executive Council approved its recommendations.

Professor Dinesh Singh (2010-2015) became the Vice Chancellor of the University, and Dr Majumdar superannuated in December 2010. In January 2011, the senior Deputy Librarian, Mr R.N. Vashista, was given charge of the DULS until February 2013. Mr Vashista carried forward the work started by Dr Majumdar. The Government of India granted an additional Rs. Seven crores corpus to the Ratan Tata Library in March 2012. In February 2013, the University gave the charge to another Deputy Librarian, Dr S.C. Jindal. Dr Jindal strengthened the security system of the documents and introduced bar code technology in many libraries and security gates at the South Campus library. The University introduced four years UG program and seven foundation courses for which the library collected and compiled reading material for these foundation courses. Dr Jindal ensured that the gathered reading material was available in braille format for visually challenged students. The University introduced Gyanaodaya – education on the wheel – an innovative and noble program by endowing students with first-hand knowledge of the World beyond the classroom ((*http://www.du.ac.in/uploads/Academics/International%20Relations/04052013gyanodaya.pdf*.). In this project, an educational train with the students and teachers visited places in India. A library was attached to each train to complete the project during their journey. The Library also participated in *Anterdhvani*, a cultural program to inspire students to march to the drumbeat of their inner calling ((*http://www.du.ac.in/uploads/Academics/International%20Relations/Antardhvanihandbook.pdf*.).

Prof Dinesh Singh appointed Dr D.V. Singh as the eighth University Librarian in May 2015. Dr Singh created a computerized catalogue of vernacular languages. Professor Yogesh Tyagi (2016-2021) took over as the Vice Chancellor of the University of Delhi in 2016. Dr D.V. Singh retired in June 2019. In October 2019, Dr Narender Kumar, Deputy Librarian, was given the charge of DULS. Dr Kumar introduced remote access, framed and got the approval of the e-book purchase policy, and purchased e-books listed in various UG and PG syllabi. During Dr Kumar's tenure, the University also agreed to replace the existing library software TROODON with internationally acclaimed software KOHA for which University formed a committee. Prof PC Joshi, Officiating Vice-Chancellor, appointed Dr Rajesh Singh as the ninth University Librarian in December 2020. Dr Singh has been framing guidelines for migrating from TROODON to KOHA; purchasing eBooks. Dr Singh has also launched Delhi University E-Library, providing 24x7 million e-resources everywhere.

Conclusion

Since the inception of the University, the Library has been the resource centre for all the information needs of students, teachers, and researchers and a leader in the academic library system in India. DULS has always been at the forefront in accepting the new challenges and technologies needed for the development of the library. In a hundred years of journey, it has become South Asia's most extensive library system.

References

1. The University of Delhi, 'Annual Report', 2020-21. Available: *http://www.du.ac.in/uploads/new-web/Annual-Report/05042022_98th%20Annual%20Report_Part-II%20(English).pdf.* [Accessed 2022 August 2022].
2. Sagar, M (1930). Message by Vice Chancellor, University of Delhi, *Modern Librarian,* Vol. 1, No. 2, pp. 28-29, December.
3. Rehman A (1931). Message by Vice Chancellor, University of Delhi, *Modern Librarian,* Vol. 1, No. 3-4, pp. 109-10, March-April.
4. Gwyer, M.(1937-38). Message by Vice Chancellor, University of Delhi, *Modern Librarian,* Vol. 8, No. 2, p. 26.
5. Shrimali, K. (1922-2025). History of Delhi University Library System, Available: *http://crl.du.ac.in/aboutus.pdf.* [Accessed 10 August 2022].
6. White, C. M. (1965). Survey of Delhi University Library, University of Delhi, Delhi.
7. *Gyanaodaya*, University of Delhi, Available: *http://www.du.ac.in/uploads/Academics/International%20Relations/04052013gyanodaya.pdf.* [Accessed 10 August 2022].
8. *Anterdhavni,* University of Delhi, Available: *http://www.du.ac.in/uploads/Academics/International%20Relations/Antardhvanihandbook.pdf.* [Accessed 10 August 2022]. ❑

Validation of an Information Diffusion based Health Social Media on Yoga and Ayurveda

RK Yadav
Dept. of Computer Science & Engg.
Delhi Technological University
Delhi, India
rkyadav@dtu.ac.in

Girish Kumar
Dept. of Mechanical Engineering
Delhi Technological University
Delhi,
girish.kumar154@gmail.com

Naveen Kumar
Dept. of Computer Science & Engineering
Delhi Technological University
Delhi, India
thisismenaveenkumar@gmail.com

Pratyaksh Varshney
Dept. of Computer Science & Engineering
Delhi Technological University
Delhi, India
varshneypratyaksh2017@gmail.com

Natwar Jha
Dept. of Computer Science & Engineering
Delhi Technological University
Delhi,
thisismenatwarjha@gmail.com

Amit
Dept. of Mechanical Engineering
Delhi Technological University
Delhi,
prajapatiamitapn@gmail.com

2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA) | 978-1-6654-5834-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCSEA54677.2022.9936537

***Abstract*—The deadly COVID-19 coupled with other diseases has proven to be the biggest challenge humans have seen in ages. Healthcare systems, even in the most developed countries, have completely shattered during the peak of waves. Thus, leaving millions rendered helpless and alone around the globe. This proves the importance of self-care and immunity to being the best possible way for being healthy. Solutions specific to strengthening immunity are available in the Indian sciences of Yoga and Ayurveda and has been scientifically proven. These have been around in India for ages but there is an immense lack of awareness in India and around the globe regarding that. This review paper aims at filling the lack of awareness by proposing a model and validating it further by collecting and comparing the pre and post-implementation data.**

I. Introduction

A. Importance and significance of the topic

The approach of Ayurveda is based on the basic principle of maintaining a holistic relationship within the body, mind, and soul. There are many age-old remedies in ancient Ayurvedic sciences for diseases like diabetic wounds, Asthma, Arthritis, memory loss, etc. for which no effective medicine is available in Allopathy [1].

Ayurveda majorly revolves around the relationship of body, mind, and soul. Any disturbances in this relationship result in different physical ailments. These ancient sciences aim to re-establish the symbiosis between the body, mind, and soul by creating the optimum health environment.

It is realized that it turns out to be quite difficult to maintain a healthy lifestyle, especially in this post-COVID-19 era. There are several uncertainties and worries which revolve around and are related to finances, childcare, elderly parents, job security, and mental health. The constant uncertainty about the future, the never-ending news coverage of all the negatives happening around, and the constant social media platforms are driven flood of messages can increase our sense of anxiety and stress levels.

II. Background

According to the government figures, the health sector has been allocated Rs 86,200.65 crore in the Union Budget, a hike of 16 percent over Rs 73,931 Crore in 2021-22. This was just Rs. 40,000 crores in 2017-2018. This amounts to a growth rate of 16 to 17% The Central Government is further planning to increase it to 2.5% of the country's GDP by the year 2025 [2].

If seen collectively, Our country's per capita expenditure on health services amounts to roughly around 1.6k INR. However, there is a disparity between the several parts of the nation with some spending more, some spend much less than the average one.

This high difference in the spending on basic health services between different classes of people has led to the flourishing of the private health care sector in the nation. About half of the total healthcare infrastructure of the country is in private hands. Adding to the fears, private health insurance is generally the trend for middle and upper-class people [3].

The easy path of following the Yoga, Ayurveda, and Naturopathy by the people of each and every age group can lead to a large deduction in budgetary requirements for Healthcare Sectors and can be directed to other important sectors of Infrastructure, Innovation, Poverty, and Starvation reduction, etc.

III. Explain why the surveyed area has been studied

There are two main reasons why this research area was chosen:

A. Health-Related Misinformation

This is the digital age, and we look to social media to find solutions and updates regarding health-related issues. It is the habit that people blindly believe whatever they are fed on social media. Health care facilities have become so expensive that it is almost impossible for many to afford them. Thus, People try to look for inexpensive

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:32:08 UTC from IEEE Xplore. Restrictions apply.

so-called homemade treatments, which claim no side effects.

Health misinformation is a big deal as health is a sensitive issue. After facing huge criticism from lawmakers and doctors, a majority of the social media platforms made drastic changes in their approach to check the false health information from spreading but still, a large chunk of it is still out there. Some of these false stories garnered more than millions of shares, comments, and likes. These numbers show the grave problems they pose [4].

B. Ayurveda and Yoga – Lost with time:

Ayurveda is one of the best ways to treat diseases and by following the principle of prevention to lead a healthy lifestyle. Yoga originally is a combination of several physicals, mental, and spiritual body practices, exercises, or disciplines.

The entire comprehensive method of treatment in the Ayurveda is quite holistic and individualized. The presence of different unique types of recuperative, preventive, curative, mitigative, and rehabilitative aspects makes Ayurveda holistic. The basic philosophy and aim of Ayurvedic Medicines are promoting good health, prevention of disease, and cure of sickness. Similarly, Yoga works on the principle of prevention [5].

But at present, though these medicines and practices are beneficial and safe to use, they are being neglected [6]. The Google Search results hardly incorporate the cures related to Ayurveda or yoga in treatments to some health-related questions.

At the time of covid, the best method for the survey we found was online mode. Online surveys are one of the most effective methods especially when digital adoption is at peak [7]. According to Statista [8] Internet penetration increased by 12% from 2018-2020. We circulated Google Form in various groups so that we can get diverse responses.

IV. Pre-Implementation Survey results and Inference:

We conducted an elaborate survey to know the Health Awareness scenario and assess the requirements of the general public in the health sector. We were successful in making this survey an effective one as a total of 212 people from diverse backgrounds took part in the survey.

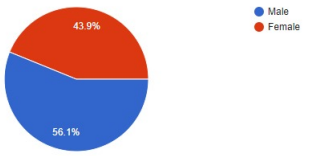
Specific insights from the details of people who participated in the Survey are mentioned below:

A. Age:

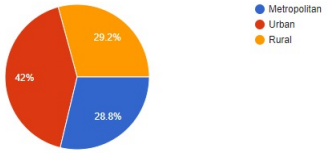
It is very important to understand the thinking and requirements of people of different age categories. People from all the age groups took part in the survey which make the results more inclusive than ever.

Age Group (in years)	Frequency
<= 20	11
21 – 30	44
31 – 40	47
41 – 50	42
51 – 60	36
61 – 70	21
>70	11

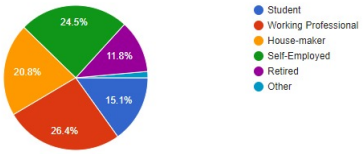
B. Gender



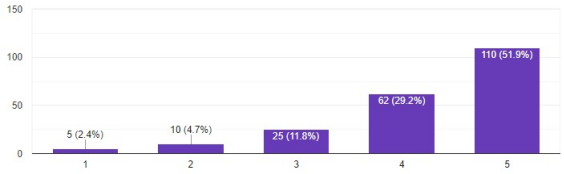
C. Place of Residence



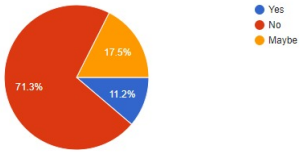
D. Profession



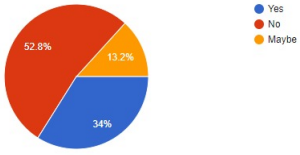
E. Is your Health and Lifestyle affected by COVID-19



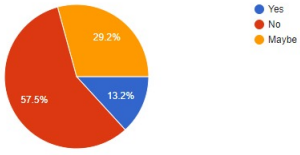
F. How often do you take Doctor Appointment



G. Are you aware of Yoga and Ayurveda

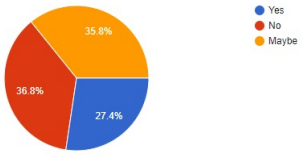


H. Do you practice and follow them in your daily routine



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:32:08 UTC from IEEE Xplore. Restrictions apply.

I. Are you aware that Yoga and Ayurveda can be learned online



V. Proposed Model Outline

1) Data collected from around 200 users through g- form and manual surveying
2) Out of the sample of data, few targeted user is passed through the model for different diseases he mentioned
3) The solution to diseases is provided through yoga, Ayurveda or naturopathy, natural remedies or self/ awareness
4) After successful completion the user experience is taken as node..and passed in graph/algorithm and is diffused in the network....so that it reaches the goal user
5) Then the study is again performed on user and their fresh responses are recorded. Survey -2
6) Data is analyzed and many inferences can be made, its been observed that around health infrastructure load has been reduced.

VI. Proposed Model:

We are proposing a Health Social Network based on Yoga, Ayurveda, and Naturopathy, and here we are validating its requirement as per the current scenario in the healthcare sector and how this model is effective and can reduce health infrastructure load by 43% [9].

In the current scenario, health trends indicate that there is no social network platform where people can share their health experiences and learn from others' experiences regarding how they got a disease, what mistakes they made that led to the occurrence of the disease, and how they cured/ prevented it. .

Also, Community motivation can help follow a good lifestyle, overcome various diseases, and most importantly can help prevent many incurable/severe diseases to occur. Ex- Diabetic person will share the poor lifestyle habits he followed which led to the occurrence of that disease. So, this will act as a warning for the people following the same poor lifestyle and may improve their lifestyle to avoid getting Diabetic and hence decreasing the load on health infrastructure.

This model is based On Information diffusion in social networks. In this model, a person sharing his experiences or a health expert giving some good health tips acts as a seed node i.e. the node which brings the information to the Network Graph.

Information is brought to the network Graph by Seed nodes and then is shared with their neighbors i.e. followers. It appears on the feed page of social networks where non-followers can also see that information. They, based on their interest, can leverage the health tips or can share them with others. Ex- meditation-based good posts can be helpful to everyone in the social network followers/ non-followers.

Also, people can share and forward it on other social networks to their friends, family members, and followers. Although there is a probability(α) associated with the sharing as in the SIR model:

Number of people who will share the info forward in other social networks communities

$$x = \alpha * n \tag{1}$$

Where x=new information spreaders and n= the infected people

Hence the Information is diffused to the other social platforms like WhatsApp Facebook Instagram. It increases the number of followers of the seed nodes and the people to which information would spread in future posts. In this way, the people in the base health social network and people who recovered will increase with the increase in Degree Centrality, as the spreading ability of nodes is directly proportional to the number of nodes to which it is connected.

This model involves multiple social network information diffusions and the degree of the node is the sum of followers in health social network(d) and in other social networks(o):

$$degree = d + o \tag{2}$$

Similarly, a similar kind of experience can be shared by multiple nodes in a similar health social network. So the probability of a node following that health tip will increase on getting its information from multiple nodes. This further increases the probability to share forward the posts(p) using the number of nodes who shared that health tip(y), no of good comments(c), the sum of likes on the post(s) and post quality(q)

$$p = \beta * y + \gamma * s + \delta * c + \alpha * q \tag{3}$$

Where beta, gamma, delta, and alpha are constants.

Hence, as per the above equation at a level where u is the number of people benefitted by the health social network/ seed node, n is the number of people who got the information:

$$U = n * p \tag{4}$$

Since like/ good comments come on good quality posts hence the information absorption or sharing is also dependent on the quality of posts shared by seed nodes. Followers will be more for the nodes sharing good quality posts, thus in the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:32:08 UTC from IEEE Xplore. Restrictions apply.

same post category we can Rank the nodes' post quality(q) by the degree centrality of that node(d):

$$q = \Theta * d \tag{5}$$

where theta is a constant
Using above equations 2,3, 4 and 5

$$u = n * (\beta * y + \gamma * s + \delta * c + \alpha * q) \tag{6}$$

And since u depends on the category and assumes each person is associated by one category only, Total no of people who benefitted from Health Social network at the time t:

$$T = \sum(u) \tag{7}$$

where $\sum$ is the Union of results of all information spreads in the network

With the increase in the number of posts made by a seed node and increase in time, the number of followers of the seed node, degree increases. Hence it becomes a good information spreader and with time the effectiveness of social networks increases:

$$Effectiveness = \int Tdt \tag{8}$$

The information diffusion can be understood using multi-graph SIR model usage. The diagrammatic representation has been shown after SIR Explanation.

A. *Explaining SIR below*

Epidemic prevention models or information diffusion models are like each other. Therefore the Popular Suspected-Infected-Recovered (SIR) model [16] is used to study information diffusion in Social Networks.

In relation to SIR where Susceptible(S(t)) means people who are prone to infection and have the possibility of infection from an infected person, here Susceptible will be the possible post readers who might read the posts shared by health influencers on the platform.

Infective(I(t)) in SIR refers to people who have diseases and can transmit them to others in contact. Here Infective will be the people who will share the posts on health topics of Yoga, Ayurveda, and Naturopathy, Infective may not be the seed nodes.

Recovered(R(t)) in the SIR model refers to people who recover from the disease and cannot transmit the disease to others in contact with people. Here in our model recovered will be the people whose posts have become outdated/ do not want to share the post with anyone else over time t.

Infection Rate α in SIR is the probability of transmission of disease when a susceptible and an infective come in contact, here it is the probability of sharing/forwarding a post after reading.

Recovery Rate β in SIR is the probability that the infective will lose infection and become a recovered one. Here it means that the post will lose its virality.

B. *Why we chose SIR over other models*

A single paragraph containing four models names and associating SIR with our project and rejecting other models including references for all models also

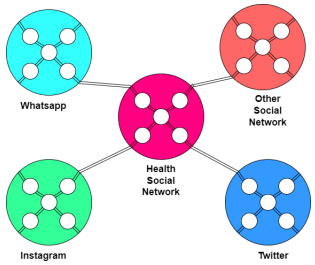
Basic Epidemic models:
Information diffusion processes are the same as epidemic spread. In epidemic transmission, a virus can be transmitted from infected to susceptible, similarly in information diffusion information can be diffused from communicators to recipients. Information diffusion can be understood by learning epidemic models. The basic models are SI (Susceptible Infected) model, SIS (Susceptible Infected Susceptible) model, SIR (Susceptible Infected Removed) model, and SIRS (Susceptible Infected Removed Susceptible) model

The SI [10] model does not allow infected people to get cured after being infected Hence this is impractical according to our information diffusion model as after sharing a post with its followers a user can stop sharing the post. In this way, the user becomes free from infection and is cured. When an individual is cured, he/she can be immune i.e., if the post is reshared with the user, the probability of him/her sharing it again is nearly zero. Hence [11] [12] [13] model also cannot be used to explain our Information diffusion model. The SIRS [14] model is different from all the above, it believes that a cure user becomes susceptible with a probability non zero, but since we are taking the probability of the cured user becoming susceptible to be zero, hence we can explain information diffusion in our case using [15] model.

C. *What are Seed Nodes in our Model:*

In this Social network of Health Experts and Health Conscious people, these are the following kinds of seed nodes—

1) Health experts sharing their knowledge/ health tips.
2) Seed nodes will also be created when a person. connects with a health expert or takes its services like Yoga or consultations and shares his experiences after recovery.
3) Some person recovered from a disease and share his experience of how he got infected.
4) A user following some effective prevention techniques
5) Somebody who read any good health tip and want to share with his followers.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:32:08 UTC from IEEE Xplore. Restrictions apply.

After this diffusion/ post sharing of my seed node its followers increase and hence its spreading ability increases

VII. Final Inference of the Pre–Implementation Survey:

We find out among the people of different age categories, gender, and backgrounds, that there is a lack of awareness of Yoga and Ayurveda amongst them. There are adverse impacts of COVID-19 on the health and lifestyle of people. Many a little people actually follow Yoga and Ayurveda which is the only preventive measure for good health. Moreover, people are actually willing to adopt some online solution, if available, based on Yoga and Ayurveda in order to maintain a healthy lifestyle.

VIII. – Inferences from Post-Implementation Survey:

We conducted an elaborate survey to check how the platform has brought about a change in health awareness in the lives of the general public. We were successful in making this survey an effective one as a total of 268 people from diverse backgrounds and categories took part in the survey.

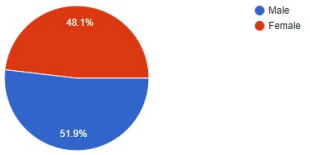
Specific insights from the details of people who participated in the Survey are mentioned below:

A. Age:

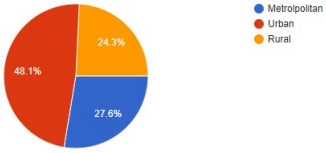
It is very important to understand the thinking and experiences of people of different age categories. People from all the age groups took part in the survey which make the results more inclusive than ever.

Age Group (in years)	Frequency
<= 20	16
21 – 30	67
31 – 40	74
41 – 50	68
51 – 60	24
61 – 70	12
>70	7

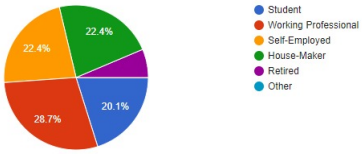
B. Gender



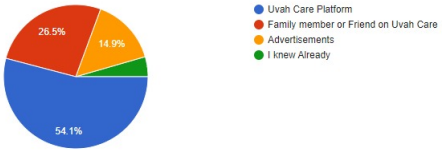
C. Place of Residence



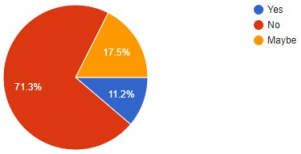
D. Profession



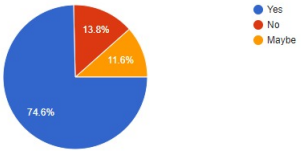
E. From where you got the motivation and information about Yoga and Ayurveda



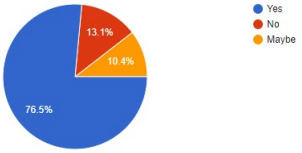
F. Were you practicing Yoga and Ayurveda before joining our platform



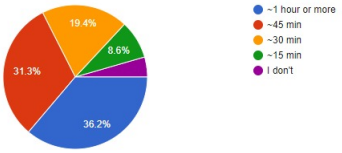
G. Do you find betterment in your body mind and soul



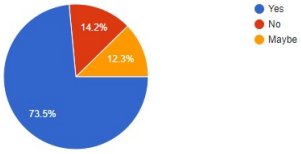
H. Is our platform better than random sources for providing verified health information



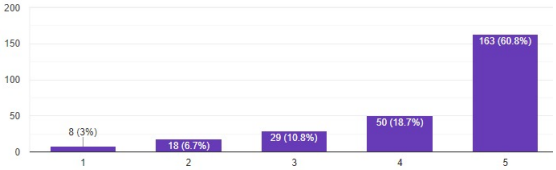
I. How much time you devote to your body for exercises



J. Does your number of appointments to the doctor decrease after following yoga and ayurveda



K. On what scale would you like to rate the platform information dissemination



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:32:08 UTC from IEEE Xplore. Restrictions apply.

IX. How our approach is better than the existing ones

- These above-mentioned platforms are basically for entertainment and connection making. There is a reluctance among people to share health-related information on these platforms.
- Most of the content over these is personal or funny. Thus, for people looking specifically for health-related solutions, it becomes a very easy diversion towards other content.
- The biggest issue is the lack of verification of the available information. The majority of the content served here has proved to mislead and several health experts have expressed their concern about it.

The health social media platform which we have developed addressed all these above-mentioned issues. Following are the ways they are resolved:

- Out platform is a dedicated health platform where only health-related information of several degrees is available. Here people can easily share their health experiences with others.
- Our platform is a community-based platform where people can ask their doubts and they are addressed by the experts or their fellow users.

X. Final Inference of the Post–Implementation Survey

We find out that among all the users of the platform, there are different age categories, gender, and backgrounds. We come to the conclusion that people got motivation and awareness about Yoga and Ayurveda from our platform only. They started to spend enough time on their health. It is also seen that they trusted the platform and are even willing to take a subscription for further services. People found app-based learning useful and even benefitted from it. The main motive of bringing awareness towards Yoga and Ayurveda has been successfully achieved. We aspire to spread the advantages of this platform to as many users as possible.

XI. Conclusion

In this paper, we presented the results of the Pre-Model Survey which we found about the lack of awareness of Yoga and Ayurveda in the general public and their requirements. We proposed a novel model based on Social Information Diffusion which meets all these requirements. We devised this method so as to overcome most of the challenges being faced by the people. We then carried out a Post-Implementation Survey for making out comparisons based on various parameters between the Post and Pre-implementation surveys. We found out that it is well received by the people and this model has successfully fulfilled all the requirements and information gaps in this sensitive area.

References

[1] V. Kumar, M. Antil, D. Kumar and N. Minocha, "Importance of ayurvedic medicine in modern lifestyle: A keynote review study", International Journal of Advanced Educational Research, vol. 1, no. 4, pp. 31-33, 2016.

[2] B. News and I. News, "Union Budget 2022: Explained in charts - Times of India", The Times of India, 2022. [Online]. Available: https://timesofindia.indiatimes.com/business/india-business/union-budget-2022-explained-in-charts/articleshow/89282022.cms.

[3] "Private Healthcare in India: Boons and Banes", Institut Montaigne, 2022. [Online]. Available:https://www.institutmontaigne.org/en/blog/private-healthcare-india-boons-and-banes: :text=India has a total of, all of India's health infrastructure

[4] Y. Wang, M. McKee, A. Torbica and D. Stuckler, "Systematic Literature Review on the Spread of Health-related Misinformation on Social Media", Social Science & Medicine, vol. 240, p. 112552, 2019. Available: 10.1016/j.socscimed.2019.112552..

[5] Jiwaji.edu, 2022. [Online]. Available: http://www.jiwaji.edu/pdf/ecourse/pharmaceutical/Indian system of medicine.pdf

[6] J. Laping, "Traditional medicine" in India decline and future perspectives", Ancient Science of Life, vol., no. 1, pp. 6 – 11, 2022. Available: https://www.researchgate.net/publication/224898814_Traditional medicine in India decline_and future perspectives.

[7] Survey Method - Research-Methodology", Research-Methodology, 2022. [Online]. Available: https://research-methodology.net/research-methods/survey-method/

[8] India: internet penetration rate 2021 — Statista", Statista, 2022. [Online]. Available: https://www.statista.com/statistics/792074/india-internet-penetration-rate/.

[9] T. Staff, "Harvard study finds yoga and meditation reduce healthcare cost by 43%", The News Minute, 2022. [Online]. Available: https://www.thenewsminute.com/article/harvard-study-finds-yoga-and-meditation-reduces-healthcare-cost-43-36026. [Accessed: 12- Mar- 2022].

[10] Pastor-Satorras, R. and Vespignani, A., 2001. Epidemic Spreading in Scale-Free Networks. Physical Review Letters, 86(14), pp.3200-3203.

[11] M. Newman, "The Structure and Function of Complex Networks", SIAM Review, vol. 45, no. 2, pp. 167-256, 2003. Available: 10.1137/s003614450342480.

[12] M. Newman, "Threshold Effects for Two Pathogens Spreading on a Network", Physical Review Letters, vol. 95, no. 10, 2005. Available: 10.1103/physrevlett.95.108701.

[13] Gross, T., D'Lima, C. and Blasius, B., 2006. Epidemic Dynamics on an Adaptive Network. Physical Review Letters, 96(20).

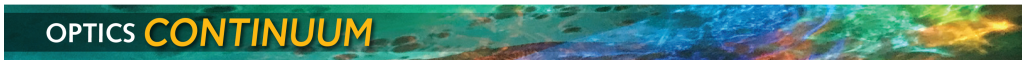
[14] X. Liu and C. Liu, "Information Diffusion and Opinion Leader Mathematical Modeling Based on Microblog", IEEE Access, vol. 6, pp. 34736-34745, 2018. Available: 10.1109/access.2018.2849722.

[15] Jin, W. Wang, and S. Xiao, "An SIRS model with a nonlinear incidence rate", Chaos, Solitons & Fractals, vol. 34, no. 5, pp. 1482-1497, 2007. Available: 10.1016/j.chaos.2006.04.022.

[16] KERMACK, W. and MCKENDRICK, A., 1991. Contributions to the mathematical theory of epidemics—I. Bulletin of Mathematical Biology, 53(1-2), pp.33-55.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 14,2022 at 09:32:08 UTC from IEEE Xplore. Restrictions apply.

Research Article Vol. 1, No. 12 / 15 Dec 2022 / Optics Continuum 2426



Voltage-tunable surface-enhanced Raman scattering substrates based on electroactive polymeric membranes containing plasmonic nanoparticles

Yashna Sharma,[1,2,6] Enes Cakmak,[3] Hsiang-kuo Yuan,[4] Roberto Garcia,[5] Dale Batchelor,[5] Tuan Vo-Dinh,[4,7] Tushar Ghosh,[3,8] and Anuj Dhawan[1,6,9]

[1]*Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, 110016, India*
[2]*Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, 110042, India*
[3]*Department of Textile Engineering, Chemistry, and Science, NC State University, Raleigh, NC 27606, USA*
[4]*Fitzpatrick Institute for Photonics, Duke University, Durham, NC 27708, USA; Department of Chemistry, Duke University, Durham, NC 27708, USA*
[5]*Department of Materials Science and Engineering, NC State University, Raleigh, NC 27695, USA*
[6]*Equal contributors.*
[7]*tuan.vodinh@duke.edu*
[8]*tghosh@ncsu.edu*
[9]*adhawan@ee.iitd.ac.in*

Abstract: We describe voltage-controlled surface-enhanced Raman scattering (SERS) substrates in which the SERS-signals can be actively modulated by applying voltage. These SERS-substrates employ a dielectric electroactive polymer (D-EAP) membrane with a pair of electrically-actuated active regions. When these regions are simultaneously activated, they produce an in-plane contractile strain in the regions of the D-EAP where SERS dye-coated nanoparticles are placed. We demonstrate that SERS-signals from dye-coated silver nanoparticles, deposited on the D-EAP membrane, increases by ~100% upon application of an actuating voltage. Upon removal of the voltage, actuated active-areas move towards their original positions, leading to a decrease in the SERS-signals.

© 2022 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

1. Introduction

Over the past few decades, surface-enhanced Raman scattering (SERS) has emerged as a highly promising technique for the detection of small concentrations of chemical and biological agents [1–6]. When chemical or biological analyte molecules are placed in close proximity of metallic nanostructured surfaces, the Raman signals are substantially enhanced, thereby enabling extremely sensitive detection of these analytes using SERS. The increase in the Raman cross-section of the coupled molecule-substrate complex is attributed to electromagnetic (EM) enhancement and chemical enhancement. While chemical enhancement results from the charge transfer between the analyte molecules and the nanostructured substrate, EM enhancement results from an increase in localized optical fields at electromagnetic hotspots (i.e., regions of high electromagnetic fields) in the nanostructured substrate [2,7] at plasmon resonance wavelengths. The electromagnetic enhancement of the optical E-field occurs at both the emission and excitation wavelengths. Consequently, the total electromagnetic enhancement of the Raman signals is approximated to be proportional to the fourth power of optical E-field enhancement in the electromagnetic hotspot where the analyte is present [1,2].

#455460 https://doi.org/10.1364/OPTCON.455460
Journal © 2022 Received 8 Mar 2022; revised 7 May 2022; accepted 17 May 2022; published 18 Nov 2022

OPTICS CONTINUUM

SERS offers an efficient way of detecting trace amounts of biochemical species and consequently, it has been widely used in sensing applications. However, producing a reliable, easy-to-fabricate and robust SERS substrate with tunable SERS activity has remained a challenge. Many tunable SERS substrates developed in the past have shown tunability of the SERS-activity by changing some parameter of the substrate during fabrication [8,9] although reports on post-fabrication active tunability of the SERS activity have been limited. Recently, active tunability of SERS substrates has been achieved via thermal activation of a temperature-sensitive polymer film [10], by mechanical stretching of plasmonic nanostructures on elastomeric membranes [11], or by opto-thermal actuation of the substrate [12]. In addition, an electromechanically-tunable, suspended two-wire plasmonic nanoantenna was employed by Hecht et al. [13] to achieve tuning of the plasmon resonance wavelengths by the application of voltage. However, active tunability of SERS signals — by the application of voltage — has not been demonstrated so far. In this work, we describe a simple but effective strategy to develop an active SERS substrate, in which the SERS signal from a cluster of dye-coated plasmonic nanoparticles can be tuned and increased by over 100% upon the application of an actuating voltage. The advantage of using the field-controlled SERS signal tunability is that we can potentially develop a closed-loop fully programmable system to control such devices.

In this paper, we employ electroactive polymers (EAPs) to achieve the active tunability of the SERS signal. Electroactive polymers constitute a class of synthetic macromolecules that undergo a dimensional change in response to an applied electric potential [14]. Of all EAPs, dielectric EAPs (D-EAP) exhibit the largest actuation strain upon exposure to an electric-field, efficiently coupling input electrical energy and output mechanical energy [14]. The electromechanical response of the D-EAPs is initiated by the field-induced charge separation that generates a compressive (Maxwell) normal stress, $\sigma_M=\varepsilon\varepsilon_o E^2$, on the film surfaces along the transverse direction (see Fig. 1). The variables ε_0 and ε represent the permittivity of free space and dielectric constant of the elastomer, respectively, and E denotes the magnitude of applied actuating electric-field. Additionally, repulsive like-charges that accumulate along both electrode surfaces also act to stretch the film in its plane.

The actuated media in this research is a commercially available acrylic elastomer (VHB 4905, 3M Co., Minneapolis, MN), appropriately prestrained with two pairs of deformable electrodes strategically placed on the opposite sides of the film segments AA1 and AA2 in order to form the active areas (See Fig. 1). Prestraining of D-EAPs is essential to reduce the voltage needed to create the electric-field necessary for actuation [14]. The deformable electrodes were deposited on opposite sides of the active areas AA1 and AA2. Carbon black (CB) filled polydimethylsiloxane (PDMS) elastomer composite was sprayed, using a stencil mask, to create the stretchable and uniformly thin electrodes (See details of Experimental Methods in Supplement 1). A voltage applied between the opposite electrodes, i.e., between E1a and E1b, and between E2a and E2b — results in the actuating electric-field necessary to activate the active areas AA1 and AA2, respectively (See Fig. 1). The synthesized SERS dye-coated silver nanoparticles (See details of synthesis of nanoparticles in Supplement 1) were drop-coated on the passive area of the D-EAP film, spanning the two active areas, labeled PA in Fig. 1. The active areas are positioned and activated to work in consort upon application of an actuating voltage, such as to generate an in-plane expansive areal strain in the active areas. The isochoric deformation of these segments (AA1 and AA2) results in a contractile strain in the passive middle portion, PA, of the D-EAP membrane (See Fig. 1). Consequently, the spacing betwen the silver nanoparticles is expected to decrease with an increase in the actuation voltage applied to the active areas, which can lead to an increase in the electromagnetic (EM) enhancement in the corresponding electromagnetic hotspots. In this paper, we have, thus, demonstrated voltage-controlled actuation of a dielectric electroactive polymer membrane containing SERS dye-coated plasmonic nanoparticles to achieve voltage-tunable, active, and reversible modulation of SERS signals.

OPTICS CONTINUUM

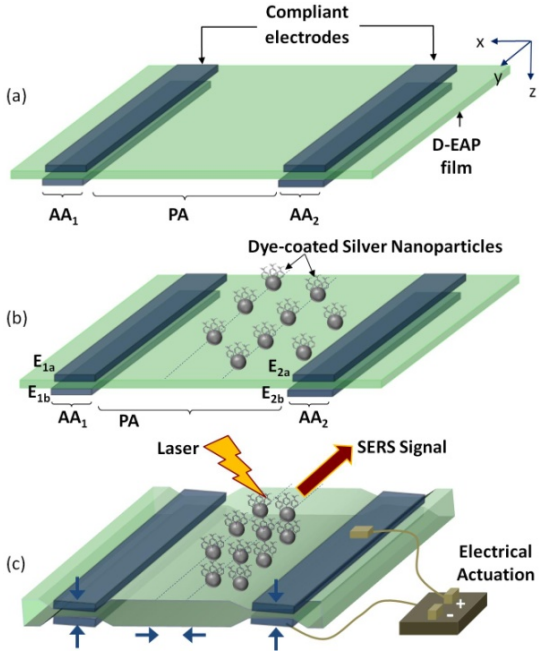


Fig. 1. (a) Schematic showing a dielectric electro-active polymeric (D-EAP) membrane with two pairs of compliant deformable electrodes to form the active areas— AA1 and AA2. PA represents the passive region (PA) of the D-EAP membrane, spanning the two active areas. (b) SERS dye-coated Ag nanoparticles are drop-coated on the passive area (PA) of the D-EAP film when no actuating field is applied. (c) Due to an actuating voltage, an in-plane expansive strain in the active areas, AA1 and AA2 results in a contractile strain in the passive region, PA.

2. Results and discussion

First, the effect of the applied voltage (applied in consort to the electrode pairs in active areas AA1 and AA2) on the in-plane actuation strain in the D-EAP membrane was evaluated for various values of the pre-actuation spacing, 'S' as shown in Fig. 2(a). The in-plane actuation strain (in the x-direction) in the passive segment (PA) of the D-EAP membrane as a function of applied voltage is plotted in Fig. 2(a) for the pre-actuation spacing values of 1 mm, 2 mm, and 5 mm. It was found that, for the proposed substrate, a higher actuation strain in the passive area is achieved for lower pre-actuation spacing between the active areas AA1 and AA2. Hence, a value of 1 mm was chosen as the pre-actuation spacing between the active areas for all the measurements carried out in this paper.

The silver nanoparticles were synthesized using a seed-mediated method published previously (See details in experimental methods, Supplement 1). Subsequently, the silver nanoparticle solution and the aqueous solution of the SERS-active dye CVP (Cresyl Violet Perchlorate) were mixed in a 1:1 (v/v) ratio. A 2 mM concentration of CVP was employed to prepare the solution containing dye-coated silver nanoparticles. These dye-coated silver nanoparticles were then drop-coated (in the passive region) onto the prestrained polymeric film suspended on a plastic frame.

The ability of the SERS substrate to modulate the SERS signals from the dye-coated silver nanoparticles drop-coated on the D-EAP membrane was experimentally demonstrated by applying a voltage across the deformable electrodes. A portable Raman system (See Details of Experimental

OPTICS CONTINUUM

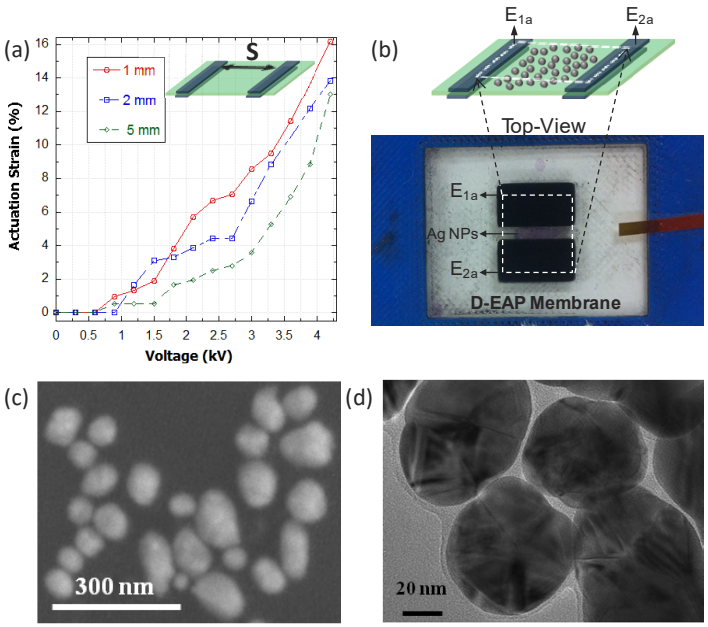


Fig. 2. (a) Effect of pre-actuation spacing, 'S', between conformable electrodes pairs on the actuation strain in the passive segment of the D-EAP membrane, (b) Top view of blue plastic frame on which D-EAP membrane was suspended. Electrodes were deposited on front (E1a, E2a) and back ($E1_b$, $E2_b$) of the film. Ag nanoparticles were drop-coated on the D-EAP membrane between the two electrode pairs. (c) SEM image of Ag nanoparticles on the D-EAP membrane and (d) TEM image of Ag nanoparticles before drop-coating on D-EAP membrane.

Methods in Supplement 1) with a 785-nm excitation wavelength was employed to measure the SERS-signals from the cluster of SERS dye-coated silver nanoparticles placed on the D-EAP membrane, as the voltage applied across the electrodes was reversibly varied from 0 kV to ~4.5 kV. The experimental set-up employed for the SERS measurements is shown in Fig. S1 in Supplement 1. A close view of the electrodes deposited on the D-EAP membrane is shown in Fig. 2(b). Figure 2(c) shows the Scanning Electron Microscopy (SEM) image for the silver nanoparticles deposited on the D-EAP membrane. The Transmission Electron Microscopy (TEM) image of the nanoparticles employed for SERS measurements is shown in Fig. 2(d) (See Fig. S2 in Supplement 1 for more SEM/TEM images).

The SERS data was collected from the passive segment of the D-EAP membrane at various operating voltages, using CVP (Cresyl Violet Perchlorate) as the SERS dye (See Figs. 3(a)–(b)). As the voltage was increased from 0.0 kV to 4.5 kV, the baseline corrected SERS signal (i.e., the height of the Raman peak) at 593.71 cm^{-1} increased from approximately 1400 to about 2800 counts (See Fig. 3(b)) which is an increase of ~ 100%. As the applied voltage is increased, the in-plane expansive areal strain in the active areas increases. The isochoric deformation of the active segments results in an increase in the contractile strain in the passive middle portion of the D-EAP membrane. Consequently, the silver nanoparticles dropped on the passive portion of the membrane moved closer as the applied voltage was increased. It is well known that a decrease in the gap between plasmonic nanoparticles leads to an increase in the optical electric-field (E-field) enhancement between the nanoparticles.[12] Thus, it is expected that the optical E-field enhancement between the silver nanoparticles would increase on increasing the applied voltage. Further, as the electromagnetic enhancement in the Raman signal is approximately proportional to

OPTICS CONTINUUM

the fourth power of the optical E-field enhancement [2], the electromagnetic SERS enhancement shown in Fig. 3(b) shows an increase with the applied voltage. Therefore, it can be concluded that a 100% increase in the Raman signals obtained in our experiment is primarily attributable to the reduction of spacing between the Ag nanoparticles on the substrate due to the contractile strain in the passive segment of the D-EAP membrane on application of voltage. When the applied voltage was brought back to 0.0 kV, the SERS signal from the substrate reduced back to 1850 counts. On removal of the applied potential, the strain in the passive and active segments of the D-EAP is restored to a value close to its original level, leading to an increase in the inter-particle spacing and hence, lowering of the SERS signal as shown in Figs. 3(a)–(c). The lack of complete recovery is due to the hysteretic loss in D-EAP film due to its viscoelastic behavior. It must be noted that the viscoelastic behaviour of dielectric electoactive polymers depends on the coupled phenomena of creep and hysteresis [15,16].

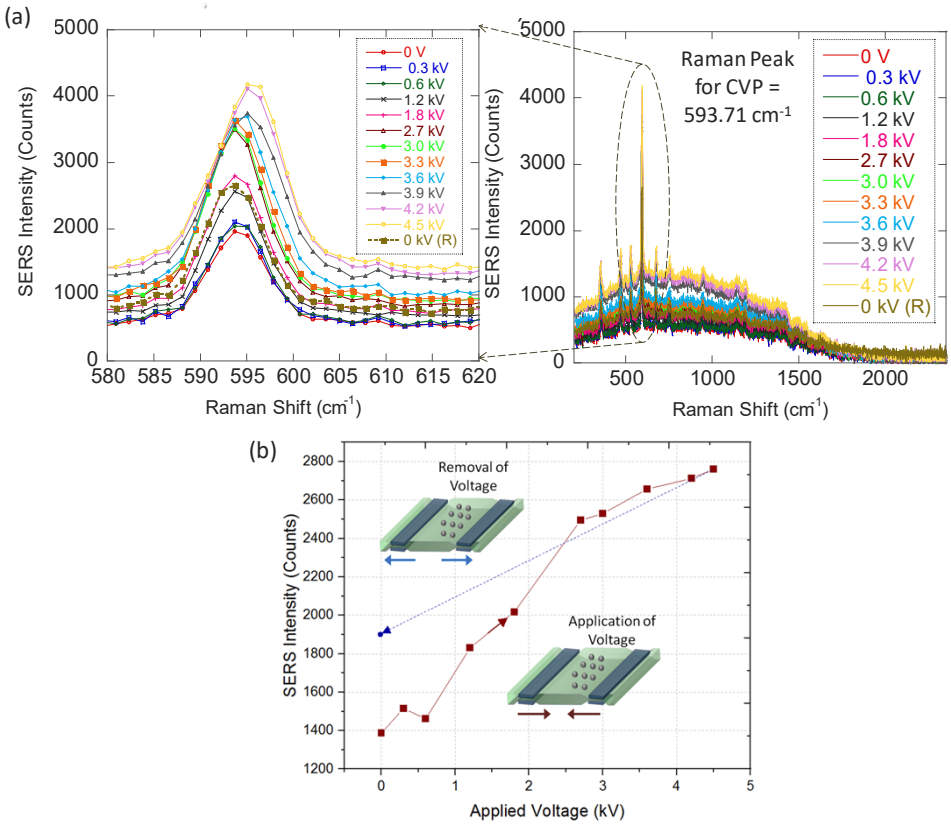


Fig. 3. (a) SERS spectra obtained from Ag nanoparticles on the D-EAP membrane for different actuation voltages, (b) Baseline corrected peak SERS signals (for the primary CVP peak at 593 cm^{-1}) versus actuation voltage, showing an increase in SERS signals with actuation voltage (red curve) and a decrease in SERS signals as the voltage difference is removed (blue curve).Additional results of SERS measurements are given in Supplement 1.

Figure 4 (Visualization 1) shows the behavior of the D-EAP membrane — with the silver nanoparticles drop-coated on the passive segment in between the electrodes — on the application and subsequent removal of the actuating voltage. It can be observed from Fig. 4 (Visualization 1) that as the applied voltage increases, the boundaries of the carbon black conformal electrodes can be seen expanding due to the areal strain produced in the active segments sandwiched in the electrodes. Consequently, the passive region with the drop-coated silver nanoparticles — seen

OPTICS CONTINUUM

as the dark spot in Fig. 4 — can be seen undergoing a contractile strain. Therefore, the silver nanoparticles sticking to the D-EAP membrane are expected to move closer and thus lead to higher SERS signals which agrees well with the results shown in Fig. 3.

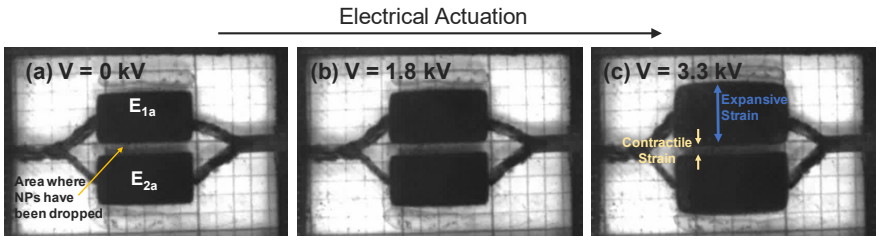


Fig. 4. Top-view images of the electrically-actuated plasmonic substrate on application of voltage: (a) 0 kV, (b) 1.8 kV and (c) 3.3 kV captured using a video camera. As the applied voltage increases, the boundaries of the carbon black conformal electrodes can be seen expanding due to the areal strain produced in the active segments. The passive region with the drop-coated silver nanoparticles can be seen undergoing a contractile strain as V is increased from 0 to 3.3 kV (Visualization 1).

In order to elucidate the effect of inter-particle gaps on the electromagnetic fields and SERS signals from a cluster of metallic nanoparticles, numerical modeling for a cluster of silver nanoparticles deposited on a D-EAP membrane, was carried out using Finite Difference Time Domain (FDTD) modeling (see Supplement 1 for details of Numerical Methods). We consider particles with diameters (D) of 60 nm and 75 nm in accordance with the average size of the nanoparticles in the acquired SEM and TEM images. The particles are assumed to be periodically arranged in the x- and y- directions with a gap, G, between them in the x-direction and a fixed inter-particle gap of 10 nm in the y-direction (see Fig. 5). As the inter-nanoparticle gap 'G' is expected to decrease with increase in the actuation strain (in the x-direction) on application of voltage to the electrodes, the effect of decreasing the gap (in the x- direction) on the EM SERS enhancement was simulated.

Figure 5 shows the change in the electromagnetic SERS enhancement (which is proportional to the fourth power of $|\mathbf{E}|/|\mathbf{E}_0|$ — where $\mathbf{E}$ is the E-field in the center of two nanoparticles and $\mathbf{E}_0$ is the incident E-field) as a function of the actuation strain (%). It can be seen from Fig. 5 that as the actuation strain is increased from 0 to 25% (i.e., the gap between the nanoparticles in the x-direction decreases from 10 nm to 7.5 nm), the SERS signals increase by ~ 200% and 154% from the zero-actuation strain values for D = 60 nm and D = 75 nm, respectively. This happens due to the strengthening of the dipole interactions between the nanoparticles on reducing the inter-particle gaps which results in an increase in the optical E-field enhancement. In our experiments, a pre-actuation spacing of 1 mm between the electrode pairs was chosen, which lead to a maximum actuation strain of 16% on application of voltage of 4.5 kV, as shown in Fig. 2(a). It can be seen from the simulated results in Fig. 5, that for an actuation strain of 16%, the SERS signals increase by ~ 96% and 76% from the zero-actuation strain values for D = 60 nm and D = 75 nm, respectively. As the nanoparticles dropped on the membrane had a distribution of particles close to the simulated values of diameters, these results agree well with the experimentally obtained value for the increase in the SERS signal of ~ 100% for an actuation voltage of 4.5 kV. The SERS enhancement versus the contractile actuation strain for the different values of diameters of the silver nanoparticles are given in Supplement 1 (See Supplementary Information, Note 4).

OPTICS CONTINUUM

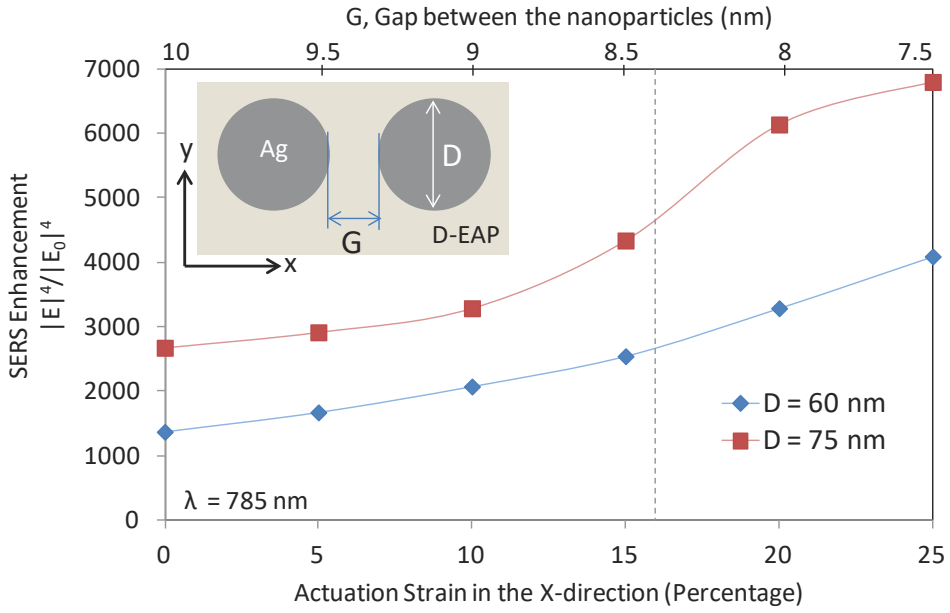


Fig. 5. SERS enhancement ($|E|^4/|E_0|^4$) versus the contractile actuation strain at a wavelength of 785-nm for periodic nanoparticle clusters with D = 60 nm and D = 75 nm on a D-EAP membrane. The pre-actuation gap between the nanoparticles in the x-direction is assumed to 10 nm and is reduced to 7.5 nm at an actuation strain of 25%.

3. Conclusion

In conclusion, we have proposed a voltage-tunable SERS substrate that allows active modulation of SERS-intensity over a broad range of actuating voltage. The SERS-intensity is tuned by varying an actuating voltage applied to the D-EAP substrate. These substrates are inexpensive and can be easily fabricated using drop-coating nanoparticles onto a variety of easily available D-EAP films types (acrylic, silicone, polyurethane, etc.) based on the application requirements. These substrates offer the possibility of large-area fabrication if multiple parallel electrode pairs, in the proposed and other geometries, are developed on a large-area polymeric film. These voltage-tunable SERS substrates can be employed for efficient chemical and biological sensing applications. Hence, the tunable SERS substrates proposed in this paper exhibit active modulation of the SERS-signals, as well as an increase of the SERS signals, by the application of an actuating voltage.

Funding. Impacting Research Innovation and Technology (RP03417G); Wallace H. Coulter Foundation; Duke University Exploratory Projects Funds.

Acknowledgments. Above all, A. D. and Y.S. would like to thank Lord Jesus for blessing this work. A.D. would also like to thank Ministry of Human Resource Development (MHRD) of the Government of India (Grant # RP03417G: IMPRINT program). T.V. would like to thank the Wallace Coulter Foundation, and the Duke University Exploratory Projects Funds for sponsoring this work.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this Letter are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. M. Moskovits, "Surface-enhanced spectroscopy," Rev. Mod. Phys. **57**(3), 783–826 (1985).

OPTICS CONTINUUM

2. A. Otto, A. Bruckbauer, and Y. X. Chen, "On the chloride activation in SERS and single molecule SERS," J. Mol. Struct. **661-662**, 501–514 (2003).
3. Y. C. Cao, J. Rongchao, and C. A. Mirkin, "Nanoparticles with Raman Spectroscopic Fingerprints for DNA and RNA Detection," Science **297**(5586), 1536–1540 (2002).
4. C. Zhang, Z. Li, S. Qiu, W. Lu, M. Shao, C. Ji, and Z. Li, "Highly ordered arrays of hat-shaped hierarchical nanostructures with different curvatures for sensitive SERS and plasmon-driven catalysis," Nanophotonics **11**(1), 33–44 (2021).
5. M. Sabzehparvar, F. Kiani, and N. S. Tabrizi, "Mesoporous-assembled TiO2-NiO-Ag nanocomposites with pn/Schottky heterojunctions for enhanced photocatalytic performance," J. Alloys Compd. **876**, 160133 (2021).
6. M. S. S. Bharati and V. R. Soma, "Flexible SERS substrates for hazardous materials detection: recent advances," OEA **4**(11), 210048 (2021).
7. E. You, Y. Chen, J. Yi, Z. Meng, Q. Chen, S. Ding, H. Duan, M. Moskovits, and Z. Tian, "Nanobridged rhombic antennas supporting both dipolar and high-order plasmonic modes with spatially superimposed hotspots in the mid-infrared," OEA **4**(12), 210076 (2021).
8. M. Kim, Y. W. Lee, D. Kim, S. Lee, S. R. Ryoo, D. H. Min, S. B. Lee, and S. W. Han, "Reshaping nanocrystals for tunable plasmonic substrates," ACS Appl. Mater. Interfaces **4**(9), 5038–5043 (2012).
9. H. H. Wang, C. Y. Liu, S. B. Wu, N. W. Liu, C. Y. Peng, T. H. Chan, C. F. Hsu, J. K. Wang, and Y. L. Wang, "Highly raman-enhancing substrates based on silver nanoparticle arrays with tunable sub-10 nm gaps," Adv. Mater. **18**(4), 491–495 (2006).
10. Y. Lu, G. L. Liu, and L. P. Lee, "High-density silver nanoparticle film with temperature-controllable interparticle spacing for a tunable surface enhanced Raman scattering substrate," Nano Lett. **5**(1), 5–9 (2005).
11. Y. L. Chiang, C. W. Chen, C. H. Wang, C. Y. Hsieh, Y. T. Chen, H. Y. Shih, and Y. F. Chen, "Mechanically tunable surface plasmon resonance based on gold nanoparticles and elastic membrane polydimethylsiloxane composite," Appl. Phys. Lett. **96**(4), 041904 (2010).
12. S. Simoncelli, E. M. Roller, P. Urban, R. Schreiber, A. J. Turberfield, T. Liedl, and T. Lohmuller, "Quantitative single-molecule surface-enhanced Raman scattering by optothermal tuning of DNA origami-assembled plasmonic nanoantennas," ACS Nano **10**(11), 9809–9815 (2016).
13. K. Chen, G. Razinskas, T. Feichtner, S. Grossmann, S. Christiansen, and B. Hecht, "Electromechanically tunable suspended optical nanoantenna," Nano Lett. **16**(4), 2680–2685 (2016).
14. R. Shankar, T. K. Ghosh, and R. J. Spontak, "Electroactive nanostructured polymers as tunable actuators," Adv. Mater. **19**(17), 2218–2223 (2007).
15. M. Bozlar, C. Punckt, S. Korkut, J. Zhu, C. Chiang Foo, Z. Suo, and I. A. Aksay, "Dielectric elastomer actuators with elastomeric electrodes," Appl. Phys. Lett. **101**(9), 091907 (2012).
16. G. Y. Gu, U. Gupta, J. Zhu, L. M. Zhu, and X. Zhu, "Modeling of viscoelastic electromechanical behavior in a soft dielectric elastomer actuator," IEEE Trans. Robot. **33**(5), 1263–1271 (2017).