SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(140th Edition)

AUGUST 2024

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 Forty** 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 academia.

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

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

This issue contains the information published during **August, 2024**. 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 Novel Approach for State of Health Estimation of Lithium-Ion Batteries Based on Improved PSO Neural Network Model, Rashid Nasimov, ***6.Deepak Kumar***, ***3.M. Rizwan***, ***3.Amrish K. Panwar***, Akmalbek Abdusalomov and Young-Im Cho, Electrical and Applied Physics, DTU

2. A Novel Dielectric modulated Double Gate Charge-Trapping MOSFET: Ferroelectric Gate Stack and Gaussian Doped Substrate Integration for Enhanced Performance, t ***8.Akkshita Swain***, ***8.Nalin Agarwal***, ***8.Tanisha Singhal*** and ***3.Rishu Chaujar***, Applied Physics, DTU

3. A novel spectral-spatial 3D auxiliary conditional GAN integrated convolutional LSTM for hyperspectral image classification, Pallavi Ranjan, Ashish Girdhar, ***8.Ankur*** and ***3.Rajeev Kumar***, Computer Science and Engineering, DTU

4. A refractive index based cancer cells sensor in terahertz spectrum: design and analysis, ***7.Jaydeep Singh***, ***6.Akash Khamaru***, ***3. Ajeet Kumar***, Applied Physics, DTU.

5. Accessibility Tools for Students with Disabilities in University Libraries, Antonieto Alaban, ***6. Ms. Anjali Singh***, CSE, DTU

6. Analysing the anomalous relationship between precipitation and PM10 concentrations in Peninsular India, , ***8.1 Manvi Nilaya, 8.1.Vaibhavi*** ,***8.1 Swayam Agrawal*** , and ***3. Lovleen Gupta,*** Environmental Engineering, DTU

7. Analysing the Bottleneck in Crankcase Cover Manufacturing using Simulation and Modelling, Sumit Chawla & ***3. Ranganath M Singari***, Mechanical, Production, and Industrial Engineering Department, DTU

8. Analysis and mapping of the indoor acoustic environment in university study spaces, ***6 Shreya Jha, 8.1 Anoushka Gupta***, ***8.1 Akshansh Gupta, 3. Rajeev Kumar Mishra*** , and ***6. Ravi Pratap Singh Jadon,*** Environmental Engineering,DTU

9. Approximation By A Composition Of Apostol-Genocchi And Palt ˇ Anea.Durrmeyer ˇ Operators, ***6.1. Nav Shakti Mishra*** and ***3. Naokant Deo***, Mathematics, DTU

10. Assessing rheological properties of oxidized Moringa oleifera gum and carboxymethyl chitosan-based self-healing hydrogel for additive manufacturing applications ***8.1 Mehul Verma, 6. Jitendra Kumar, 6.1 Abhinav Ayush Pradhan,*** Nilotpal Majumder, Sourabh Ghosh , and ***3. Roli Purwar ,*** Applied Chemistry, DTU

11. Assessment of PM10 in Colombo, Sri Lanka and Its Potential Source Regions, ***7. Thrividya Rathnayake***, and ***3. Lovleen Gupta,*** Environmental Engineering, DTU

12. Augmented GAN: Advancing Generative Adversarial Networks through Enhanced Discriminator Training and Generator Stability, ***8.1 Aman Prakash***, ***8.1 Ryan Binu Varghese***, and ***3. Lavi Tanwar,*** Electronics and Communications, DTU

13. Bioactive profiling of two varieties of Indian legumes: adzuki and mung beans, ***6. Ritu Sharma***, ***8. Devendra Kumar***, and ***3. Rajinder K. Gupta,*** Applied Chemistry, DTU

14. Biopolymer-CMTG and m-BPDM Based Hydrogel Composite for Promising Sensing of Zinc, Cadmium, and Mercury in Aqueous Medium, ***6. Priyanka Yadav***, 3***. Sudhir G. Warkar***, and ***3. Anil Kumar,*** Applied chemistry and Mechanical Engineering, DTU

15. Mathematical modeling and stress-aware stability analysis of a nonideal multiport Single Inductor DC–DC converter for renewable energy, ***6. Ashutosh Gupta***, ***3. Dheeraj Joshi,*** Electrical and Electronics, DTU

16. Classification of Hateful Memes by Multimodal Analysis using CLIP, ***8.1 Samarth Munzni***, ***8.1 Shikhar Dixit***,and ***3. Dr. Aruna Bhat,*** Computer Engineering, DTU

17. Comparative Evaluation of Satellite- Based Merra-2 against Ground-based Data for PM2.5 and TC Concentrations in Ghaziabad, India, ***8.1 Rishika Sinha***, ***8.1 Preha Madan, 8.1 Rahul Singh***, and ***3. Lovleen Gupta,*** Environmental Engineering, DTU

18. Comprehensive assessment of air quality dynamics in Delhi: a tri-phased analysis spanning pre, during, and post-lockdown periods, ***8.1 Satvik Popli*** , ***8.1 Esha Yadav*** , **7. Aastha Singh** , and ***3. Geeta Singh***, Environmental Engineering, DTU

19. Crosslinking of polyvinyl alcohol with di, tri, and tetracarboxylic acids: an experimental investigation, ***6. Leela Gautam***, ***3. Manish Jain***, and ***3. Sudhir G. Warkar,*** Applied Chemistry, DTU

20. Decision-making in multi-objective DG planning for distribution system via Shannon's entropy, ***6.Shubham Gupta***, ***3. Vinod Kumar Yadav***, ***3. Madhusudan Singh*** and Ashutosh K. Giri, Electrical Engineering, DTU

21. Demand Study and Impact of Electric Vehicle on grid system in Delhi, ***8.1 Chhavi Sehrawat***, ***8.1 Shubh Mittal***, ***8.1 Krish Shokeen*** and ***3. Geeta Singh,*** Environmental Engineering, DTU

22. Design and Optimization of Terahertz Based D-shaped Photonic Crystal Fiber for Blood Component Detection, ***6. Vishal Chaudhary*** and ***3. Sonal Singh***, ***Vijay*** Shanker Chaudhary, Dharmendra Kumar, Electronics and Communication Engineering, DTU

23. Dynamic Filter Application in Graph Convolutional Networks for Enhanced Spectral Feature Analysis and Class Discrimination in Medical Imaging, 8***.1ARYAN SINGH***, PEPIJN VAN DE VEN, CIARÁN EISING and PATRICK DENNY, Computer Science, DTU

24. Dynamic Modelling and PI Control Design for Bidirectional SEPIC/ZETA DC-DC Converter in Battery Applications, ***7 Prashant Singh***, ***3. Alka Singh***, and ***6.1 Ankita Arora,*** Electrical Engineering, DTU

25. Engineering Optoelectronic properties of Cs2SnI6 using Pressure, ***6.1 Parmeet Singh Chani***, ***6.Mehak Gole***, ***8.1Yashjeet Singh***, and ***3.Mukhtiyar Singh***, Applied Physics, DTU

26. Enhanced Credit Card Fraud Detection: A Machine Learning Approach, ***8.1 Priti Meena*** , **8.1 Mukul Raj,** and **3.** ***Rajni Jindal,*** Computer Science Engineering, DTU

27. Enhanced emission in the NIR range by Nd3+ doped borosilicate glass for lasing applications, Yasha Tayal, Rupesh A. Talewar, Sk. Mahamuda, Kartika Maheshwari, ***6. Sheetal Kumari,*** **6. Mohit Kumar**, Ravita Pilania, Aman Prasad, and ***3. A. S.*** **Rao,** Applied Physics, DTU

28. Enhancement of multiplication factor of capacitor using single current follower differential input transconductance amplifier, ***7. Ajishek Raj***, ***6.1 Meghana Shrivastava***, ***3. D.R. Bhaskar*** , ***3. Pragati Kumar*** , Electrical Engineering, DTU

29. Ethanol-infiltrated circular photonic crystal fiber for low peak power supercontinuum generation from near-infrared to mid-infrared region, ***3. Drishti Singh Tomer*** ,***3.*** and ***Ajeet Kumar,*** Applied physics, DTU

30. Existence of Topological Phase in RbCd4As3: A First Principles Study, ***7.1 Nitika Sachdeva***, ***6. Ramesh Kumar***, ***3. Mukhtiyar Singha,*** Applied Physics, DTU

31. Experimental Research on Green Concrete, ***7.1 Neda Rasoly , 3. Prof. Kongan Aryan,*** Civil Engineering, DTU

32. Exploring Natural Compounds as Potential CDK4 Inhibitors for Therapeutic Intervention in Neurodegenerative Diseases through Computational Analysis, 6**.** ***Neetu Rani, 3. Pravir Kumar,*** Biotechnology, DTU

33. Fallen Leaves to Sustainable Energy Solution: Review on Hydrogen Production, kyu Kyu Tin, Kirach Taweepreda, Akanksha Singh, Naresh Kumar Wagri, ***3. Anil Kumar,*** Machanical, DTU

34. From Quantum Computing to Condensed Matter : A Machine Learning Perspective, ***8.1 Pranav Kairon,*** Applied Physics***,*** *DTU*

35. Gas-Sensing Properties of NO on Ce-Doped Zinc Oxide: A DFT Study, Neha Sharma, Umesh K. Dwivedi, Umesh T. Nakate, ***3.Mukhtiyar Singh***, and Sandip Paul Choudhury, Applied Physics, DTU

36. Generalized Switched-Capacitor-Based Hybrid Multilevel Inverter With Reduced Components Count and Inrush Current, Niraj Kishore , ***7.1 Kapil Shukla*** and Nitin Gupta, Electrical Engineering, DTU

37. Handwritten Amharic Word Recognition with Additive Attention Mechanism, ***3.Ruchika Malhotra***, and ***6.Maru Tesfaye Addis***, Computer Science, DTU

38. Impact of Device Parametrics on a 10nm Gate Length SOI n-FinFET, ***8.1 Anurag Somayajula***, ***8.Tanish Aggarwal***, ***8.Vishnu Sajith***, & ***3.Rishu Chaujar,*** *Applied Physics, DTU*

39. Improving single-slope passive solar still efciency through integration of phase change materials and Al2O3 nanoparticles, ***7.1 Yunis Khan***, Zafar Said, ***3.Roshan Raman*** · Parminder Singh, Mohammad Mehdi Rashidi, Hakan Caliskan, Aman Garg, and Machnical Engineering, DTU

40. Injectable mesquite gum and carboxymethyl chitosan hydrogel using schif base crosslinks: a versatile platform for drug delivery in wound care, ***6.Jitendra Kumar***, ***3.Roli Purwar,*** Applied Chemistry, DTU

41. Leaf Analysis for Disease Identification using Transfer learning, ***3.Rohit Beniwal***, ***8.1Sachin Sachin***, ***8.Satyam Satyam*** and ***8.1 Saroj Kumar***, Computer Science & Engineering, DTU

42. MMSD-CAF: MultiModal Sarcasm Detection using CoAttention and Fusion Mechanisms, ***8.1 Jayant Miglani***, ***8.1 H A Pavan Kashyap***, ***3.Rahul Katarya*** and ***8.1 Madhav Maheshwari***, Computer Science & Engineering, DTU

43. Modelling Visual Semantics via Image Captioning to extract Enhanced Multi-Level Cross-Modal Semantic Incongruity Representation with Attention for Multimodal Sarcasm Detection, ***6.Sajal Aggarwal***, ***6.Ananya Pandey*** and ***3.Dinesh Kumar Vishwakarma***, Information Technology, DTU

44. Optimization of Bypass Flow Introduction Angle in Dual-Throat Nozzle, Sampreetananda Bharali, Shubham, ***8.1Sarthak Joshi*** and ***3.M. Zunaid***, Machanical Engineering, DTU

45. Overview of pro-inflammatory and pro-survival components in neuroinflammatory signalling and neurodegeneration, ***6.Shefali Kardam***, ***3.Rashmi K Ambasta*** and ***3.Pravir Kumar,*** *Bio*technology, DTU

46. Physicochemical characterization of novel okra mucilage/ hyaluronic acid-based oral disintegrating films for functional food applications, ***6.Krizma Khatreja***, ***3.Deenan Santhiya,*** Applied Chemistry, DTU

47. PLC and SCADA based temperature control of heat exchanger system through fractional order PID controller using metaheuristic optimization techniques, ***6.Basant Tomar, 3.Narendra Kumar*** and ***3.Mini Sreejeth***, Electrical Engineering, DTU

48. Prediction of Air Quality Index Using Random Forest and Prophet Tool, ***8.1 Abhishek Walia, 8.Ajay Paliwal*** and ***3.Sanjay Patidar***, Rakeshkumar Mahto, Software Engineering, DTU

49. Prediction of inlet-to-outlet width ratio of type-A piano key weir using fuzzy neural network (FNN), ***3.Nipun Bansal***, ***7.1 Keshav Bhardwaj***, ***6.1 Deepak Singh*** and ***3.Munendra Kumar,*** Computer Science, Civil Engineering, DTU

50. Property investigation of functionally graded materials leaf spring plate fabricated through stir casting process by using new gradient evaluation method, ***3. R. K. Singh, 3.R. C. Singh,*** Machanical Engineering, DTU

51. Prospect of large language models and natural language processing for lung cancer diagnosis: A systematic review, Arushi Garg, Smridhi Gupta, Soumya Vats, ***6.1 Palak Handa*** and ***8.1 Nidhi Goel***, Electronics &Communication Engineering, DTU

52. PUBLIC SECTOR BANKS IN INDIA: A REVIEW OF LITERATURE, ***3.K. V. Bhanu Murthy,*** Amit Kumar Singh and Sonia Kaushik, Head of Commerce, DTU

53. Quantum cloning transformation unlocks the potential of W class of states in a secret sharing protocol, ***6.Rashi Jain***, ***3. Satyabrata Adhikari***, Applied Mathematics, DTU

54. RSHealth: A Ring Signature Scheme for Identity Anonymization and Transaction Privacy in Blockchain based E-Healthcare Systems, PUNAM PRABHA, ***7.1 KAKALI CHATTERJEE,*** CSE, DTU

55. Sensitivity Investigation of Underlap Gate Cavity-Based Reconfigurable Silicon Nanowire Schottky Barrier Transistor for Biosensor Application, ***3 Anil Kumar***, ***6.Vijay Thakur, 8.1 Suraj Kumar, 3.Sumit Kale*** and ***3.Kaustubh Ranjan Singh,*** Electronics & Communication Engineering & CS, DTU

56. Simulation-Based Evaluation of IDMT Characteristics for Overcurrent Relay Protection, ***7.Utkarsh Rishu***, ***3.Anup Kumar Mandpura*** and ***3.Rohan Pillai***, Electrical Engineering, DTU

57. Size resolved particle contribution to vehicle induced ultrafine particle number concentration in a metropolitan curbside region, ***6.Kanagaraj Rajagopal***, S. Ramachandran and ***3.Rajeev Kumar Mishra***, Environmental Engineering, DTU

58. Solar PV integrated simplified multilevel inverter configuration for power quality improvement using multilayer Gamma filter, ***7.1 Praveen Bansal, 3.Alka Singh,*** Electrical Engineering, DTU

59. Speed Control of Three-Phase Induction Motor Using ANFIS and Flower Pollination Algorithm Optimized PI Controller, ***8.1 Aayushi Aryam, 8.1 Aayushi Singh, 8.1 Aditya Mohan a***nd ***3.Bhavnesh Jaint,*** Electrical Engineering, DTU

60. Structure-Based Virtual Screening, ADMET prediction and Molecular Dynamics simulation of isoindolin-1-one scaffolds as potential inhibitors of MNK2***, 6.Kunal Madaan, 3.Saurabh Mehta*** and ***3.Ram Singh***, Applied Chemistry, DTU

61. Supply Chain Management and its Role in Rail Transport***, 3.Manoj Kumar, 3.Vipin*** and Ashish Agarwal, Computer Science & Mechanical Engineering, DTU

62. Sustainable waste management: a comprehensive life cycle assessment of bioethanol production from agricultural and municipal waste, ***7.1 Kulvendra Patel, 3.Santosh Kumar Sing,*** Environmental Engineering, DTU

63. Triple band self-decoupled MIMO antenna pair for 5G communication, ***6.Parul Kansal***, ***3.Anup Kumar Mandpura*** and ***3.Narendra Kumar,*** Electrical Engineering, DTU

64. Underwater image enhancement using contrast correction, ***6.Nishant Singh, 3.Aruna Bhat,*** Computer Science & Engineering, DTU

65. Utilizing Deep Learning for Textual Classification of Hate Speech in Online Social Networks, ***8.1 Dhruv Verma, 8.1 Gaurav Thakur, 8.1 Hammad Ansari*** and ***3.Rahul Katarya***, Computer Science & Engineering, DTU

66. Violence Detection in Automated Surveillance using CNN, ***8.1 Abishek Ghalley, 8.1 Abdelrahman Abdelsalam, 8.1 Wongani Dombola,*** and ***3.M.S. Choudhary, Electronics*** & communication Engineering, 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*

processes

MDPI

Article

A Novel Approach for State of Health Estimation of Lithium-Ion Batteries Based on Improved PSO Neural Network Model

Rashid Nasimov [1], Deepak Kumar [2,3], M. Rizwan [2,*], Amrish K. Panwar [3], Akmalbek Abdusalomov [4,*] and Young-Im Cho [4,*]

[1] Department of Information Systems and Technologies, Tashkent State University of Economics, Tashkent 100066, Uzbekistan; rashid.nasimov@tsue.uz
[2] Centre of Excellence for Electric Vehicles and Related Technologies, Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India; deepakvns163@gmail.com
[3] Lithium-Ion Battery Technology Laboratory, Department of Applied Physics, Delhi Technological University, Delhi 110042, India; amrish.phy@dtu.ac.in
[4] Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 13120, Gyeonggi-do, Republic of Korea
* Correspondence: rizwan@dce.ac.in (M.R.); akmalbekabdusalomov@gmail.com (A.A.); yicho@gachon.ac.kr (Y.-I.C.)

Abstract: The operation and maintenance of futuristic electric vehicles need accurate estimation of the state of health (SOH) of lithium-ion batteries (LIBs). To address this issue, a robust neural network framework is proposed to estimate the SOH. This article developed a novel approach that combines improved particle swarm optimization (IPSO) with bidirectional long short-term memory (Bi-LSTM) to effectively address the issue of precisely estimating SOH. The proposed IPSO-Bi-LSTM model is more effective than the other models for SOH estimation. This is because Bi-LSTM can capture both past and future appropriate information, making it more suitable for modeling complicated temporal sequences. The IPSO main objective is to optimize the model hyperparameters. To increase the model's accuracy, the IPSO improves the parameters. The PSO-Bi-LSTM model performed better than the other approaches, according to experimental findings based on the NASA-PCOE battery dataset, and all of the SOH estimated outcomes, such as root mean square errors, were less than 0.50%. This result suggests that the proposed PSO-Bi-LSTM model has the ability to robustly estimate the SOH with a high accuracy.

Keywords: electric vehicles; lithium-ion batteries; state of health; improved particle swarm optimization; bidirectional long short-term memory

check for updates

Citation: Nasimov, R.; Kumar, D.; Rizwan, M.; Panwar, A.K.; Abdusalomov, A.; Cho, Y.-I. A Novel Approach for State of Health Estimation of Lithium-Ion Batteries Based on Improved PSO Neural Network Model. *Processes* **2024**, *12*, 1806. https://doi.org/10.3390/pr12091806

Received: 19 July 2024
Revised: 16 August 2024
Accepted: 23 August 2024
Published: 26 August 2024

Copyright: © 2024 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

Electric vehicle economy and environmental sustainability are directly impacted by the performance of the power battery, which is also frequently known as the vehicles heart [1]. LIBs are the most widely used power battery type in the field of electric vehicles (EVs) because of their high energy density, long lifetime, low self-discharge rate, and limited environmental effects [2,3]. However, several safety issues concerning electric vehicles have increased frequently due to the EV industry's rapid development [4]. The fundamental factors behind these accidents may be attributed to the inadequate management and maintenance of power batteries. These accidents not only cause significant financial losses but also have negative effects on society, seriously delaying the growth of the EV industry. Due to this, the roles of the battery management system (BMS) have become more noticeable [5]. BMS is responsible for both monitoring and controlling the operation of LIBs, as well as for executing the SOH estimation. By precisely measuring the SOH of a power battery, it is possible to effectively recognize the warning signs of battery aging and performance degradation. This allows us to take necessary actions to extend the battery

lifetime and improve the dependability and safety of the EVs. It is essential to accurately estimate the SOH in order to ensure the safety, effectiveness, and durability of battery systems, especially in vital sectors such as electric vehicles, healthcare equipment, and energy storage. It improves the prevention of dangerous malfunctions, enhances performance, and prolongs battery lifespan, resulting in cost-effectiveness and lower environmental impact. An accurate SOH improves warranty management and user trust by safeguarding the predictable and dependable performance of battery-powered technologies.

There are three primary types of SOH estimation techniques for LIBs: direct measurement methods, model-based methods, and data-driven methods [6]. Direct measurement methods include the Coulomb counting method [7], the open circuit voltage (OCV) approach [8], and the estimation method depending on the internal resistance. The Coulomb counting technique is a quick and easy approach for evaluating the SOH of LIBs. The process involves determining the discharge capacity by estimating the charge that flows through the battery during discharge and then dividing it by the battery rated capacity. Nevertheless, a drawback of this approach is its dependence on the entire charge and discharge of the battery in a single cycle, which may be inconvenient in real-world scenarios. The OCV approach, which calculates the capacity by analyzing the correlation between the OCV and state-of-charge (SOC), needs several hours of inactivity to achieve an equilibrium battery condition. The internal resistance-based estimation method is a precise methodology for measuring SOH and can be examined using a hybrid pulse power characterization approach. This process is often executed offline but has challenges when executed online. Direct measurement techniques are highly precise; however, they are only applicable under laboratory conditions. Typically, batteries are not entirely drained to maintain the regular functioning of electric vehicles in everyday life; hence, this approach is not feasible for real-world implementations.

Model-based techniques typically involve the integration of measurable battery data, such as voltage, current, and temperature, with electrochemical models or state observers. These techniques estimate the battery SOH by considering it as a process of estimating parameters based on a battery model that describes its dynamic features. To address this limitation, researchers have developed several SOH observers using electrochemical models. Pang et al. [9] used the mathematical models developed for battery SOH together with the constant phase element parameters in the designed equivalent circuit model of electrochemical impedance spectroscopy to achieve a quick SOH estimation for retired power batteries. An online capacity estimation technique based on recursive least squares (RLS) and enhanced particle filtering (PF) was proposed by Haung et al. [10]. This technique uses the characteristic voltage (CV) as a measure of the health status from the discharge curve and constructs a correlation model between the CV and cycle capacity. Subsequently, the enhanced PF-RLS method was used to estimate the CV accurately in real time for SOH estimation. Fang et al. [11] used the forgetting-factor recursive least squares (FFRLS) technique to accomplish online model parameter identification. The double-extended Kalman filter (DEKF) algorithm was presented to estimate the SOH based on the link between the ohmic internal resistance and SOH. The model-based approach has difficulties, even if it exhibits good precision. These challenges involve the task of choosing a suitable model that achieves a balance between precision and computational demands. The large computational burden associated with parameter identification and the complexities of extending its usage due to the constant necessity for model adjustments to adjust various LIBs models and operating conditions. Furthermore, proficient expertise in the battery's fundamental mechanisms is required to implement model-based approaches. It also relies on comprehensive physical and chemistry concepts, which include electrochemical phenomena, properties of materials, and solid-phase diffusion. These theories involve complex computational equations and functional relationships [12].

Data-driven methods are categorized as black-box models, and a key feature of these approaches is their investigation and use of high-performance algorithms. Data-driven procedures reduce the need for a comprehensive examination of the battery's fundamental

mechanism [13]. These methods are very useful and have become popular in recent years for SOH estimation. The fundamental concepts of their approach include examining data from battery performance tests to determine the correlation between external factors and the reduction in battery capacity [14]. Chang et al. [15] introduced an online technique for SOH estimation using incremental capacity (IC) and combining wavelet neural network and genetic algorithm (GA-WNN). The wavelet neural network initial connection weights, conversion factor, and scaling factor were optimized using a genetic algorithm. The GA-WNN model was utilized to determine the battery SOH, and the Pearson correlation coefficient method was employed to extract the health characteristic variables from the IC curves. Tian et al. [16] presented SOH estimation from the KF-joined particle filter. An attenuation framework was established using exceptional health indicators (HI), including vehicle loads and seasonal temperature. The battery model parameters were determined using the variable forgetting-factor recursive least squares. However, the estimated SOH error was high, up to 1.2%. Liu et al. [17] estimated the SOH by combining the Long Short-Term Memory (LSTM) and support vector regression (SVR) models. The SVR model and LSTM model were used as the main and secondary models, to assess the SOH, respectively. In addition to other cutting-edge models, this model was verified using the NASA database. Although the proposed model achieved a root mean square error (RMSE) of 0.60%, the information of battery parameters ignored, like temperature and impedance, which might improve accuracy. A Convolutional Neural Network (CNN) with an LSTM model-based SOH prediction method was developed by Xu et al. [18]. The CNN-LSTM model was trained using various ratios of training to testing. The model performed well, with an RMSE of 0.004%, although it was computationally laborious and challenging. Extracting relevant data distinctive features for training the model can mitigate this problem. Tang et al. [19] developed a CNN-LSTM- convolutional block attention module (CBAM) hybrid neural network model for LIBs SOH estimation. The CBAM module improves the sequential attention framework, minimizes the impact of noise on the input information, and boosts the ability of the CNN to extract features. Utilizing two public datasets and transfer learning with fine-tuning procedures, the SOH estimate had an RMSE of 0.17%.

Zhao et al. [20] introduced a novel regression-generative adversarial network that employs a generator to automatically produce auxiliary training samples with distributions that are similar yet distinct from the real samples. Additionally, they developed a comprehensive model to determine the optimal correlation between SOH and health features. Zhang et al. [21] introduced an enhanced method for estimating the SOH using a reference voltage-based IC and LSTM network. The health feature variables were derived from the IC curves, and the LSTM network was used to achieve SOH estimation. However, by evaluating a significant quantity of data using an efficient processor, data-driven state estimation methods can be developed without detailed knowledge of the battery background mechanism [22]. Furthermore, a data-driven strategy may predict multiple states with less knowledge of detailed chemical interactions and battery substance features and with a faster development time [23]. However, a hybrid framework can be built by merging data-driven and model-based approaches [24]. Whenever the models are hybridization, it may be not performed properly, it leads to a problematic configuration that provides mathematical complexities and delivers poor results [19]. Therefore, further research is required to address practical problems, and limitations must be considered when establishing an efficient hybrid framework. Estimating battery states requires the use of deep and machine learning algorithms. Nevertheless, the variability in the estimation performance of a single hybrid model can limit its precision [25]. To address this problem, it is essential to develop a hybrid deep learning approach that integrates two or more models to enhance the precision of the SOH estimation of LIBs.

This article developed a novel approach that combines improved particle swarm optimization (IPSO) with bidirectional long short-term memory (Bi-LSTM) to effectively deal with the issue of accurate SOH estimation. The proposed PSO-Bi-LSTM model was more successful than the PSO-LSTM for SOH estimation. This is because Bi-LSTM can

capture both past and future appropriate information, making it more suited for modeling complicated temporal sequences. In addition, Bi-LSTM is more susceptible to the vanishing gradient issue, a phenomenon that may arise in LSTM when the gradients diminish significantly, leading to the termination of the network's process of learning. Ma et al. [26] used improved PSO for hyperparameters selection with the backpropagation neural network model for remaining useful life and SOH estimation. The estimated performances showed good accuracy in the RMSE for RUL and SOH, at 6.05% and 0.78%, respectively. Zhi et al. [27] developed a SVR approach to predict the SOH after identifying important health parameters for lithium-ion batteries using the random forest technique. SVR parameter values are optimized using the GA-PSO technique, which increases the estimation accuracy and convergence speed. Using the four batteries, the suggested SOH estimation approach yields a 0.40% RMSE and 0.56% MAPE. Hu et al. [28] present a fractional-order calculus-based co-estimation technique for LIBs SOC and SOH estimation. The voltage response was predicted by the model using the Hybrid GA/PSO approach. The dual fractional-order extended Kalman filter estimates simultaneously. The study shows a very high computational burden but rapid convergence and good precision, with errors of less than 1% across both the SOC and SOH. Ye et al. [29] developed a model-based LIBs SOC estimation method employing enhanced adaptive PF and PSO. By updating the model parameters, implementing the adaptive PF technique achieved enhanced durability and fast convergence. In addition, the PSO algorithm was used to reduce the inconsistency of the battery in the presence of flat noise with a zero average value. The model that was suggested achieved SOC estimation error below 1%. Lai et al. [30] proposed a hybrid approach that combines ECM and extended Kalman filter (EKF) to estimate SOC. The optimal model parameters for the SOC estimation were chosen using the PSO approach. To estimate the SOC and SOP, combined with an estimated error of 2.95%. Ren et al. [31] presented a PSO algorithm and an LSTM model-based SOC estimation. The main objective of the PSO algorithm is to determine the optimized hyperparameters of the LSTM model. Ansari et al. [25] presented a hybrid framework that combines the PSO and RNN models to predict the remaining useful life (RUL) of LIBs. The framework was built using a systematic sampling method keeping consideration of the important battery statistics. PSO-GA method was used in [32] to identify the ideal battery features. The suggested co-estimator showed SOC estimation error of 1.2%. Reference [33], the SOE and SOP were determined at each macro time using the estimated SOC. PSO and the UKF were used in real time to determine the accurate battery parameters and estimate the SOC. The SOC error achieved an RMSE of less than 0.08%. How SOH affects SOE and SOP estimations has not yet been explored. In this research, the authors developed a PSO to improve the parameters of a Bi-LSTM model. In the available literature, no research has been found to attempt to apply an IPSO-Bi-LSTM network to estimate the SOH of a battery system.

This research proposes an improved version of the PSO approach to improve the optimization potential and rate of convergence. In addition, nonlinear variable inertial weights are included in the PSO method. The proposed approach was motivated by investigations carried out by Hu et al. [28], Zhi et al. [27], and Ma et al. [26]. Researchers have demonstrated a significant improvement in the precision of SOH estimation. The Bi-LSTM model efficiently captures the temporal dependencies and fluctuations in the LIBs dataset. The model can be improved by capturing such patterns and merging them with PSO, leading to estimates of SOH that are highly precise. Moreover, the SOH of the LIBs features is estimated well by combining PSO and Bi-LSTM. The PSO algorithm can contribute to reducing overfitting in the Bi-LSTM approach by identifying the most suitable combination of hyperparameters to achieve a balance between computational complexity and accurate estimation. For instance, this article distinguishes itself from others in terms of its approach to addressing the issue, organization of the neural network, parameters used in PSO and Bi-LSTM, and performance metrics used to assess the outcomes. The authors first divided the collected LIBs dataset into training, cross-validation, and testing sets. Subsequently, the authors proceeded to develop a Bi-LSTM neural network. After

that, the IPSO approach is used to adjust the hyperparameters of the Bi-LSTM and, at last, the estimation of SOH. This article demonstrates the approach used for SOH estimation, as shown in Figure 1.

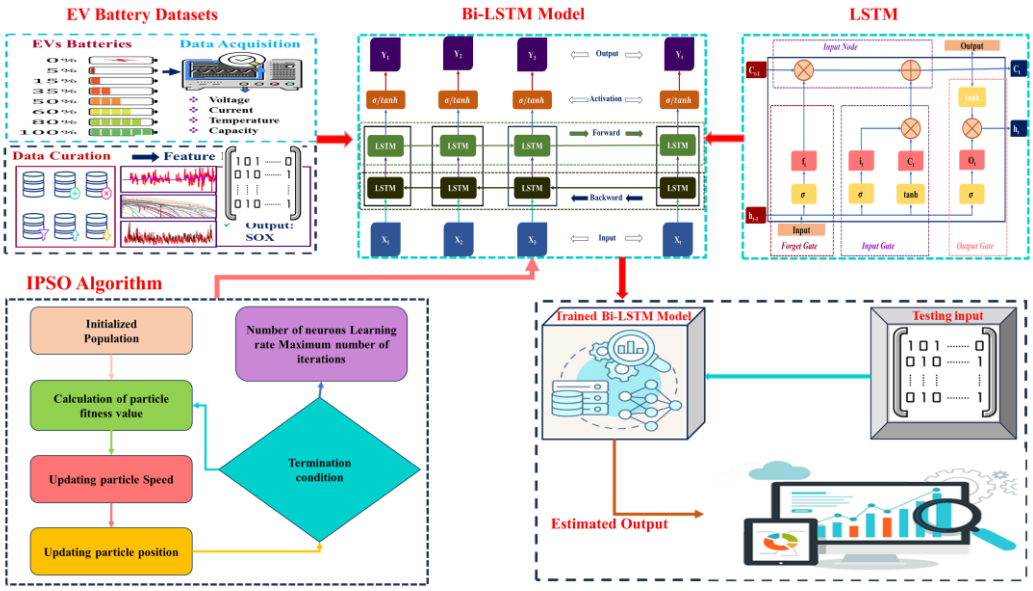


Figure 1. Proposed framework for SOH estimation of LIBs.

2. Proposed Methodology

The SOH is a metric used to quantify the present maximum dischargeable capacity of a power battery [34]. It specifically relates to the battery's energy storage capacity. This is a crucial determinant in deciding when the batteries should be changed. Usually, the battery capacity and internal resistance are used as indicators to approximate its SOH, which may be described as follows [35]:

$$\mathrm{SOH} = \frac{\mathrm{C}_{\text{max capacity}}}{\mathrm{C}_{\text{rated capacity}}} \times 100\% \quad (1)$$

$$\mathrm{SOH} = \frac{\mathrm{R_E} - \mathrm{R_n}}{\mathrm{R_E} - \mathrm{R_0}} \times 100\% \quad (2)$$

The $\mathrm{C}_{\text{rated capacity,}}$ the rated capacity, and $\mathrm{C}_{\text{max capacity}}$, the current maximum usable capacity of a lithium-ion battery have the same specifications. Meanwhile, R_n, R_E, and R_0 denote the present internal resistance, the internal resistance at the end of its lifetime, and the original internal resistance of the lithium-ion battery, respectively. This study mainly aims to assess the SOH of a battery, with a particular focus on the degradation of its capacity. When a battery is first manufactured, its SOH is typically 100%. Over time, the SOH slowly declines due to both cycling and chronological deterioration caused by continual operation. Typically, a power battery is believed to have reached its end-of-life threshold when its capacity is reduced to around 80% to 70% of its rated capacity, and the internal resistance of the battery doubles compared to a new one. The proposed algorithms are discussed in this section.

2.1. Fundamental PSO Algorithm

Bird foraging involves swarming and clustering behaviors, which are represented by the PSO approach. This algorithm is widely utilized in various applications, leveraging swarm intelligence and a global random search approach. The direction and distance traveled by the particle during its search for the solution space are mostly determined by its change in speed. Simultaneously, it thoroughly assimilates its past knowledge and collective wisdom of the group to continuously fine-tune its local search capabilities. The

purpose of the optimal solution and the global optimal solution currently found by the entire population is to continuously explore the solution space until the end of the iteration. Within the n-dimensional search space, a group of particles is initialized and dispersed throughout the area. The position of each particle's represents a potential solution to a specific problem. Through their exploration, each particle can identify the most optimal position within the desired search space. Suppose the search space of a particle swarm consisting of N particles is a D-dimensional vector $xid = (xi1,\ xi2, \ldots\ldots \cdot\cdot xiD)$, the flight speed is represented [36].

$$vi = (vi1,\ vi2, \ldots,\ viD),\ where,\ i = 1,\ 2,\ 3,\ \ldots,\ N)$$

The best solution obtained by the searched particles is $pi = (pi1,\ pi2,\ \ldots, piD)$. One of the best solutions for each particle is represented by $pg = (pg1,\ pg2,\ \ldots, pgD)$, which mimics the evolutionary, iterative calculation method. The PSO algorithm needs to adjust the speed and position of its flight in each iteration. The position update formula of the speed $v_{id}^{k+1}\ and\ k+1$ alterations of the ith particle is:

$$v_{id}^{k+1} = wv_{id}^{k} + c1r1\left(P_i^k - x_{id}^k\right) + c2r2\left(P_g^k - x_{id}^k\right) \tag{3}$$

$$x_{id}^{i+1} = x_{id}^{k} + v_i^{k+1} \tag{4}$$

The new location update formula is:

$$x_{id}^{k+1} = x_{id}^{i} \times wij + v_{id}^{i} \times w^i ij + rand() \times gbest^t \tag{5}$$

Within this scenario, wij and $w^i ij$ are variables that dynamically regulate the influence of the current x_{id}^i and $w^i ij$, respectively. The $gbest^t$ denotes the most favorable position of the particle. The inertia weight, w, plays a crucial role in determining the extent to which the particle's current velocity is inherited, thereby ensuring a balance between the detection and development capabilities of the PSO algorithm. Increasing the value of w results in more particle displacement in the original direction, enhancing exploration capabilities, but at a slower convergence rate. Conversely, decreasing the value of w improves the particle development abilities but increases the likelihood of being trapped in local optima. The learning factors c_1 and c_2 represent the proximity of each particle to its global best and historical merit, respectively. r_1 and r_2 are randomly generated integers that fall within the range of 0 to 1. To minimize the likelihood of particles exiting the search space during the iteration, vij is often constrained within a certain range, $vij\ \in\ [-vamx,\ vmax]$.

Improvement in PSO Algorithm

To achieve optimal particle identification in the early stages and efficient particle growth in the latter stages, the parameter 'w' must be continually modified during the iteration process. This research employs a linear declining law, which is represented by the following formula [37]:

$$w_i = w_{max} - \frac{i(w_{max} - w_{min})}{I_{max}} \tag{6}$$

The variable w_i represents the inertia weight for the i^{th} iteration. I_{max} denotes the maximum number of iterations. $w_{min}\ and\ w_{max}$ correspond to the lowest and maximum inertia weights, respectively. The fundamental particle swarm technique suffers from the drawbacks of local extrema and sluggish convergence due to the limited variety of particle positions in the latter stages of the search. Experimental results indicate that during the first search phase, the value of c1 is rather high, while the value of c_2 is relatively low. Additionally, particles can freely explore and split inside the search space. As the number of iterations rises, the value of c_1 drops linearly, while the value of c_2 grows linearly. This

leads to an improvement in the overall convergence of the particles toward the global optimum. This formula is represented by the following expression [31]:

$$c_{1i} = c_{1s} + \frac{i(c_{1e} - c_{1s})}{I_{max}} \quad (7)$$

$$c_{2i} = c_{2s} + \frac{i(c_{2e} - c_{2s})}{I_{max}} \quad (8)$$

The values of c_1 and c_2 for the ith iteration are denoted as c_{1i} and c_{2i} respectively. The starting values of c_1 and c_2 are represented by c_{1s} *and* c_{2s}, while the final values are denoted as c_{1e} *and* c_{2e}. I_{max} refers to the maximum number of iterations. The IPSO algorithm is similar to the PSO algorithm. The detailed algorithm process is as follows as shown in Figure 2:

Step 1: Generate the position and speed of N particles randomly and provide the necessary parameters such as the maximum number of repetitions, M.

Step 2: Determine the fitness value of each particle. Assign the ith particle to its current best position and the Nth particle to the global best position.

Step 3: Adjust the speed and position of the particles based on Equations (5) and (8).

Step 4: Compute the particle's fitness value and modify the particle's local and global extreme values based on the computed result.

Step 5: Evaluate whether the provided termination condition has been met. Upon satisfaction, terminate the search and produce the outcome; otherwise, proceed to Step 3 to resume the iteration.

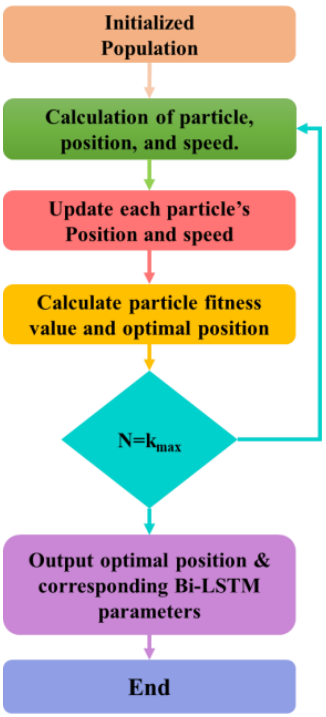


Figure 2. The IPSO algorithm flow chart.

2.2. LSTM

Recurrent neural networks (RNNs) of the LSTM network type are often used for processing sequential data, including time series signals and natural language. They are intended to solve the issue with standard RNNs vanishing gradient of standard RNNs, which may occur when the error signals gradient is too small, and the network is unable to identify long-term relationships in the data. The memory cell is a crucial part of an LSTM network because it enables the network to selectively forget or recall data from earlier time

steps. Information enters and exits the memory cell via several gates that constitute the cell. The gates are realized as hyperbolic tangent activation functions or sigmoid activation functions, which have respective ranges of values between −1 and 1 and 0 and 1 [38]. Figure 3 depicts the overall layout of the LSTM cell.

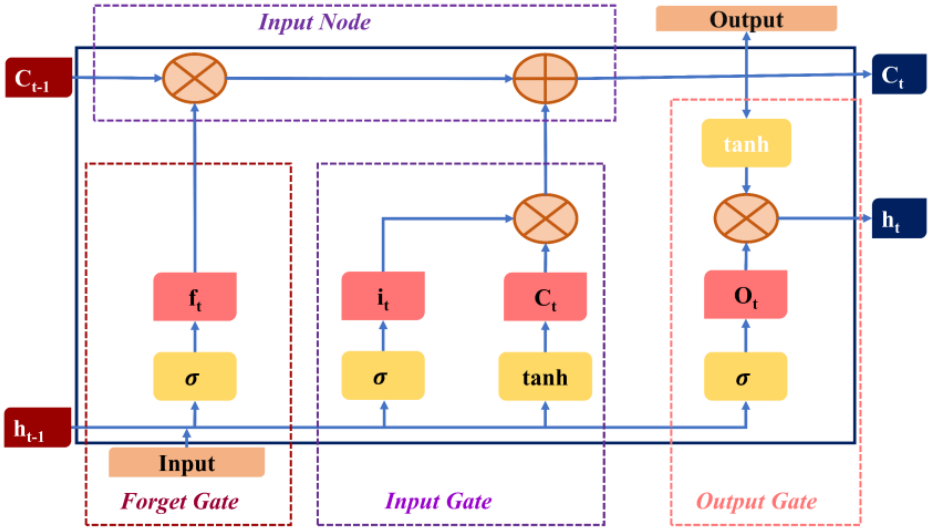


Figure 3. LSTM model structure.

The following are the mathematical formulas that characterize the LSTM network: The input at time step t should be represented by Xt, the hidden state by ht at time step t, the cell state by ct at time step t, and the sigmoid and hyperbolic tangent functions by σ and f, respectively. The LSTM formulas are expressed as follows [39]:

$$\begin{cases} f_t = \text{sigmoid}\left(W_f X_t + D_f[X_t, h_{t-1}] + b_f\right) \\ i_t = \text{sigmoid}(W_i X_t + D_i[X_t, h_{t-1}] + b_i) \\ o_t = \text{sigmoid}(W_o X_t + D_o[X_t, h_{t-1}] + b_o) \\ c_t^* = \tanh(W_c X_t + D_c[X_t, h_{t-1}] + b_c) \\ c_t = i_t \cdot c_t^* + f_t \cdot c_{t-1} \\ h_t = o_t \cdot f_t(c_t) \end{cases} \tag{9}$$

where, c_t^* is the candidate cell state, and it, ft, and ot are the input, forget, and output gates, respectively. During training, the biases $b_{i,f,o,c}$ as well as the weights $W_{i,f,o,c}$ are learned with the input of current, neuron, and gates it, ft, and ot. The weights $D_{i,f,o,c}$ are learned with the gates and hidden states. Any nonlinear activation function, such as the ReLU or the hyperbolic tangent, may be used as the function (ct). For time series regression applications, LSTM networks provide several benefits [40]. First, they may capture long-term dependencies in the data that typical regression models may find challenging to capture. This is accomplished by using the memory cell and its gates to selectively remember or erase data from earlier time steps. Second, LSTM networks function well with time series data that have irregular sampling intervals because they can accommodate input sequences of varied lengths. In addition, they may identify intricate temporal linkages and patterns in the data, which can result in better performance than with conventional regression models.

2.3. Bidirectional LSTM

Schuster and Paliwal [41] developed bidirectional RNN a kind of RNN that uses past and future input data sequences to train the network. Two connected layers are used to process the input data, and each layer processes the data in the forward and reverse

time step directions. The BiLSTM uses two layers to process input data in both directions. The BiLSTM shows more accurate results than unidirectional LSTM in some applications, including battery management systems [39]. The network topology of BiLSTM is illustrated in Figure 4.

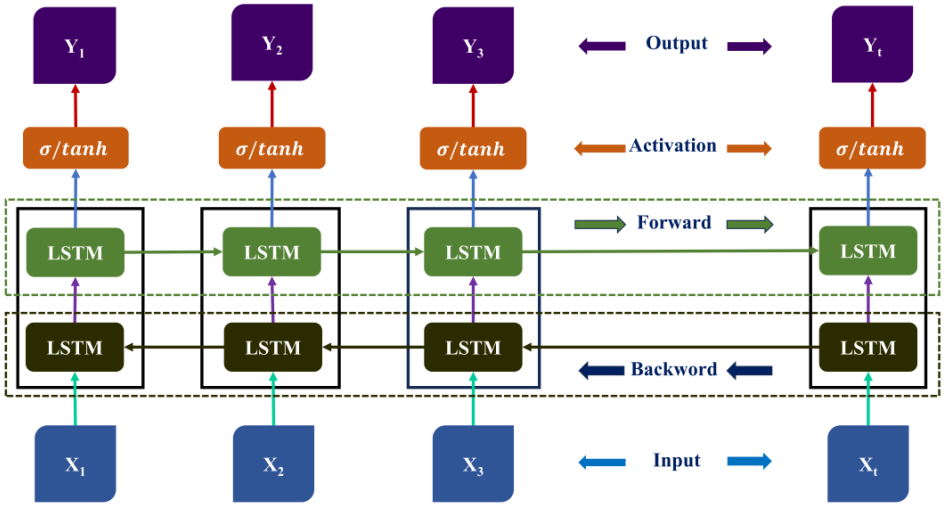


Figure 4. Bidirectional LSTM model structure.

BiLSTM consists of two hidden layers, the forward and backward LSTM layers, which are connected to the same output layer. The forward LSTM layer output $\overrightarrow{h}_t$ were evaluated by BiLSTM independently after that $\overleftarrow{h}_t$ every time step t is likewise concatenated through BiLSTM. The following Equations (10)–(12) expresses the BiLSTM process [42].

$$\overrightarrow{h_t} = f\left(X_t, \overrightarrow{h}_{(t-1)}; \overrightarrow{\Theta}_{BiLSTM}\right) = \begin{cases} \overrightarrow{f_t} = sigmoid\left(\overrightarrow{W_f}\overrightarrow{X_t} + \overrightarrow{D_f}\left[X_t, \overrightarrow{h_{(t-1)}}\right] + \overrightarrow{b_t}\right) \\ \overrightarrow{i_t} = sigmoid\left(\overrightarrow{W_t}\overrightarrow{X_t} + \overrightarrow{D_t}\left[X_t, \overrightarrow{h_{(t-1)}}\right] + \overrightarrow{b_i}\right) \\ c_t^* = \tanh\left(\overrightarrow{W_c}\overrightarrow{X_t} + \overrightarrow{D_c}\left[X_t, \overrightarrow{h_{(t-1)}}\right] + \overrightarrow{b_c}\right) \\ \overrightarrow{o_t} = sigmoid\left(\overrightarrow{W_o}\overrightarrow{X_t} + \overrightarrow{D_o}\left[X_t, \overrightarrow{h_{(t-1)}}\right] + \overrightarrow{b_o}\right) \\ \overrightarrow{c_t} = \overrightarrow{f_t} \odot C_{(t-1)} + c_t^* \odot \overrightarrow{i_t} \\ \overrightarrow{h_t} = \overrightarrow{o_t} \odot \tanh\left(\overrightarrow{c_t}\right) \end{cases} \quad (10)$$

$$\overleftarrow{h_t} = f\left(X_t, \overleftarrow{h}_{t-1}; \overleftarrow{\Theta}_{BiLSTM}\right) = \begin{cases} \overleftarrow{f_t} = sigmoid\left(\overleftarrow{W_f}X_t + \overleftarrow{D_f}\left[X_t, \overleftarrow{h_{(t-1)}}\right] + \overleftarrow{b_f}\right) \\ \overleftarrow{i_t} = sigmoid\left(\overleftarrow{W_i}X_t + \overleftarrow{D_i}\left[X_t, \overleftarrow{h_{(t-1)}}\right] + \overleftarrow{b_i}\right) \\ c_t^* = tanh\left(\overleftarrow{W_c}X_t + \overleftarrow{D_c}\left[X_t, \overleftarrow{h_{(t-1)}}\right] + \overleftarrow{b_c}\right) \\ \overleftarrow{o_t} = sigmoid\left(\overleftarrow{W_o}X_t + \overleftarrow{D_o}\left[X_t, \overleftarrow{h_{(t-1)}}\right] + \overleftarrow{b_o}\right) \\ \overleftarrow{c_t} = \overleftarrow{f_t} \odot C_{(t-1)} + c_t^* \odot \overleftarrow{i_t} \\ \overrightarrow{h_t} = \overleftarrow{o_t} \odot \tanh\left(\overleftarrow{c_t}\right) \end{cases} \quad (11)$$

$$y_t = W_{\overrightarrow{h_y}}\overrightarrow{h_t} + W_{\overleftarrow{h_y}}\overleftarrow{h_t} + b_y \quad (12)$$

The weight from the forward and backward LSTM layer to the output layer is represented by $\overrightarrow{W_f}$ and $\overleftarrow{W_f}$, respectively. The bias $\overrightarrow{b_{z,r,h}}$ and $\overleftarrow{b_{z,r,h}}$ at output, layers are clearly stated.

2.4. IPSO-Bi-LSTM

This research proposed an IPSO-Bi-LSTM estimation framework that combines the PSO algorithm with the Bi-LSTM model and modifies the network configuration to the LIBs dataset. PSO algorithms are simple and exhibit rapid convergence. It can improve parameter accuracy by learning from the characteristics of the dataset. The Bi-LSTM model uses PSO to rapidly and effectively optimize parameters based on LIBs dataset characteristics. Consequently, the network configuration of the Bi-LSTM model and the features of the LIBs dataset are incorporated more efficiently. Figure 5 shows the proposed IPSO-Bi-LSTM framework.

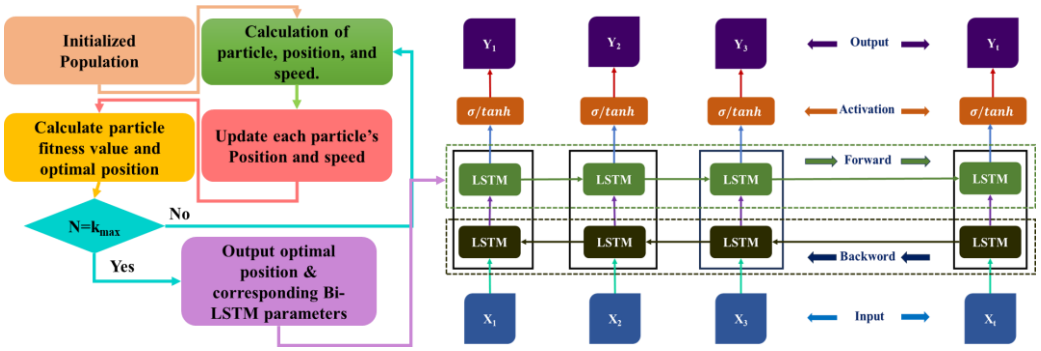


Figure 5. The IPSO-Bi-LSTM structure.

The following steps should be followed to implement the proposed SOH estimation framework, as shown in Figure 1. First, the original data must be extracted and then normalized. Secondly, prepare the training, cross-validation, and testing portions of the dataset. Thirdly, population size, spatial dimension, maximum number of iterations, learning factor, inertia weight, speed, and location of each individual particle are initialized. Fourth, the fitness function is determined by calculating the difference between the estimated and reference values. The fitness value is then used to modify the local and global optimal values. Filth and save the model when halting requirements are achieved. Sixth, after integrating the parameter values into the suggested model, the test samples are utilized for estimation, and the estimated result is validated.

3. Experimental Details and Dataset Description

This research utilizes the NASA batteries public dataset from the NASA Ames Centre of Excellence for Estimation database [43]. The batteries are listed as B5, B6, B7, and B18. Choose a dataset that has four 2 Ah rated 18,650 LIBs in it. The batteries in this group are subjected to identical external conditions and measurement procedures. During the battery cycle aging test, the battery is maintained at an ambient temperature of 24 °C. The charging procedure starts with a period of constant current (CC) charging, followed by constant voltage (CV) charging. The charging procedure is as follows: first, charge with a constant current (CC) of 1.5 A until the voltage reaches the cut-off voltage of 4.2 V; thereafter, charge with a constant voltage (CV) of 4.2 V until the current decreases to the cut-off current of 20 mA. CC mode discharges the battery at 2 A until it reaches the cut-off voltage. Temperature, voltage, and current are among the variables that are monitored and saved throughout every charge/discharge operation.

The cycle life and capacity of LIBs continuously decrease because of the progressive depletion of the active material caused by an increase in charge/discharge cycles. The complete capacity decay curves for the four LIBs used in this experiment, labeled B5, B6, B7, and B18, are shown in Figure 6. The graph depicts a nonlinear degradation process in

the capacity of these batteries, with an overall dropping pattern punctuated by instances of partial regeneration, in which the capacity momentarily improves before declining again. This brief regeneration highlights the intricacy of the battery aging process, which makes it more difficult to estimate the remaining capacity. Each battery begins with an individual's initial capacity and health status, reflecting differences in its initial capacity and health status, and reflecting differences in its original situations and use history. The cycle life of these batteries is deemed to terminate whenever their capacity drops below 80% of the specified capacity. The graphs indicate that B18 exceeds the limit at approximately 80 cycles, whereas B5, B6, and B7 approach it between approximately 120 and 150 cycles. The variety observed emphasizes the difficulties in accurately estimating battery performance and durability [44–50]. Table 1. The operational information of hyperparameters of the Bi-LSTM model optimized by IPSO.

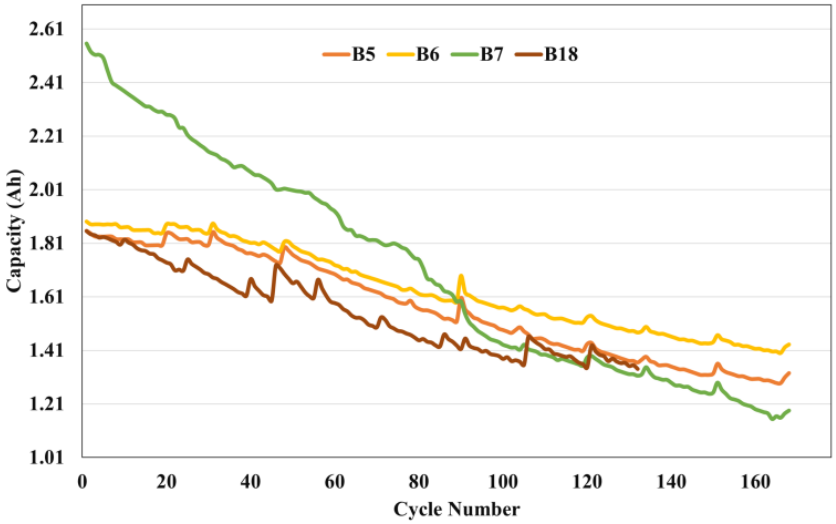


Figure 6. Different capacities of battery deterioration graphs.

Table 1. The hyperparameters information of the Bi-LSTM model optimized by IPSO.

Model Structure	Parameter
Layers	3
Learning rate	0.001
Activation function	ReLU
Batch size	23
Maximum iteration	700

4. Result and Analysis

To evaluate the effectiveness of the IPSO-Bi-LSTM model in estimating the SOH of LIBs, the authors used error assessment measures, including the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Square Error (MSE). The LSTM, Bi-LSTM, and LSTM-Bi-LSTM are popular neural network models that will be included for comparison to validate the efficiency and superiority of the developed IPSO-Bi-LSTM model for the estimation of SOH of LIBs. The data for each battery include the voltage, current, and capacity. The model was trained on 70% of the data and tested on 30% of the data for SOH estimation. Figure 7 shows the estimated outcomes for battery B5. However, the estimated SOH of the LIBs at the 117th cycle was taken as an initial point. The LSTM model estimates a specific real value while estimating SOH; however, the highest estimation error and computation time were 8.79% and 77 s (sec), which cannot properly pre-estimate the LIBs SOH operation to the final stage, as shown in Figure 7a,b. The Bi-LSTM model has a maximum estimation error and computational time of 4.59% and 50 s. The Bi-LSTM model exhibits a reduction in the highest estimate error and computational time of 4.20% and 27 s.

Nevertheless, when the number of cycles increases, the estimates of the Bi-LSTM model diverge from the actual value and exhibit a visible delay, which is unsuitable. The LSTM-Bi-LSTM model has a maximum estimate error and computational time of 4.37% and 39 s. The LSTM-Bi-LSTM model achieved a 0.22% reduction in the maximum estimation error and a computational time of 11 s compared to the Bi-LSTM model. This improvement leads to a considerable enhancement in the estimation performance. However, the model's estimation accuracy remains insufficient. The IPSO-Bi-LSTM model has a maximum estimation error of 4.10%, and with a computational time of 17 s, the estimation curve steadily fits the actual value. After comparing to the other three models, there was a reduction in estimation error of 4.85%, while in computational time, there was a 22 s improvement in estimation precision, making it excellent.

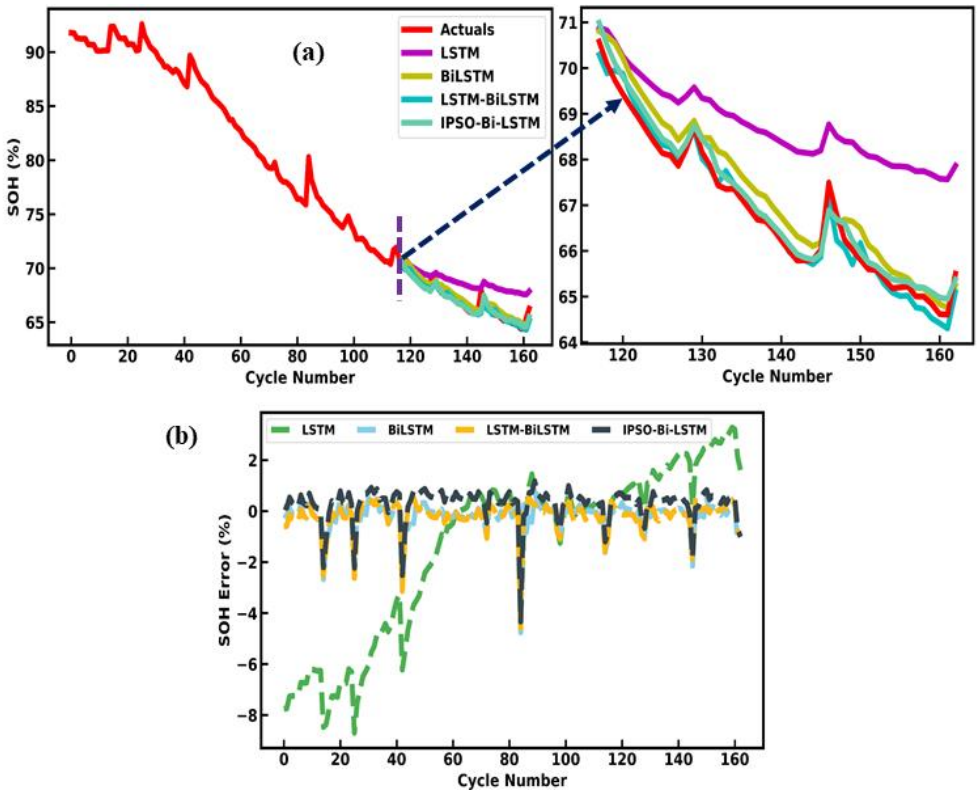


Figure 7. Estimation performance (**a**) estimation result of B5 battery (**b**) estimation error.

Figure 8 demonstrates that the IPSO-Bi-LSTM model precisely estimated the SOH in the B6 battery, which is excellent in performance as compared to the other three models. Using the 119th cycle as the initial point for estimating the SOH of LIBs, both the LSTM and Bi-LSTM models show significant instabilities in their estimations. The maximum estimate errors and computational time for the LSTM and Bi-LSTM models are 2.15%, 1.84%, 69, and 43 s, respectively. The estimation error and computational time reduction achieved by BiLSTM compared to LSTM is 0.31% and 26 s. The highest estimation error of the LSTM-Bi-LSTM model is 1.51%, with a computational time of 24 s, which indicates greater stability throughout the middle and early stages of estimation. Compared to the Bi-LSTM model, the LSTM-Bi-LSTM model decreases its maximum estimation error and computational time by 0.33% and 19 s. However, as the number of cycles increases, the estimated outcomes of the LSTM-Bi-LSTM model frequently differ from the actual values, demonstrating that accuracy in measurement remains insufficient. The proposed IPSO-BiLSTM model achieves a maximum estimate accuracy of 1.5% while having a computational time of 13 s, demonstrating improved fitting, lower error fluctuations, and computationally efficient. This significant advancement significantly improves the estimated accuracy of the SOH for LIBs.

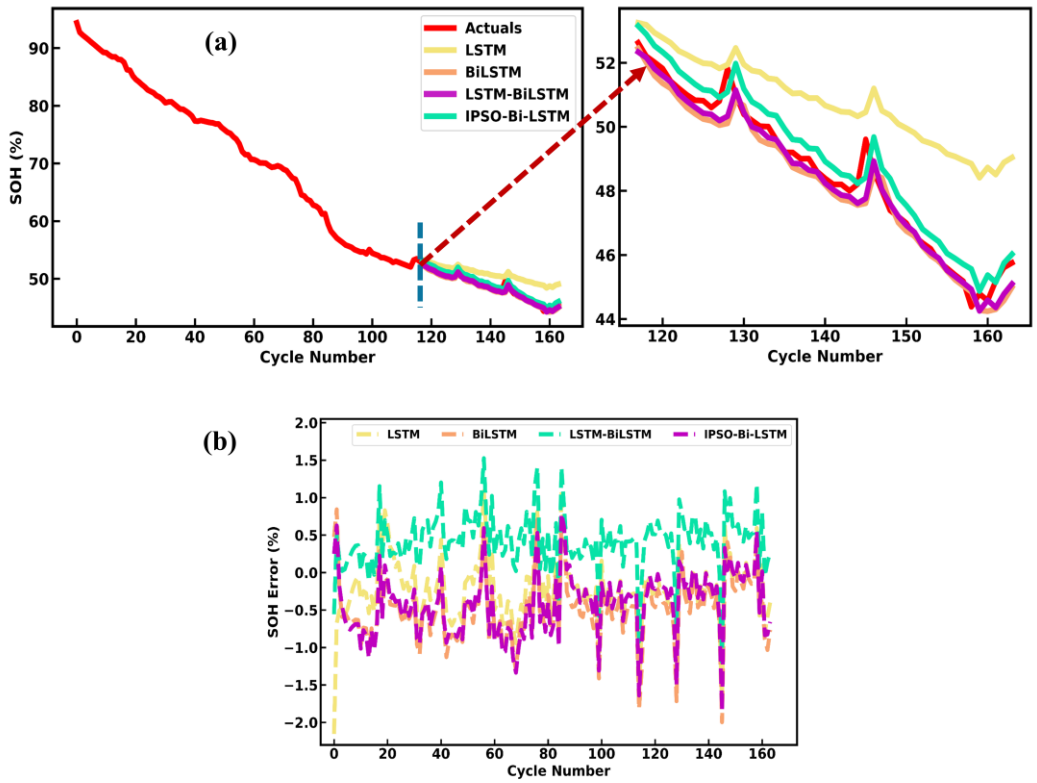


Figure 8. Estimation performance (**a**) estimation result of B6 battery (**b**) estimation error.

The estimation approach for the SOH of the B7 battery starts with the process of estimating the SOH of the LIBs at the 117th cycle, as illustrated in Figure 9. At the 117th cycle, the LSTM and Bi-LSTM models precisely estimated the initial fluctuation, with maximum estimation errors of 5.82% and 5.07%, respectively. However, the computational times are 71 s and 53 s. The Bi-LSTM maximum estimate error reduction of 0.75% is smaller than that of LSTM and is computationally efficient for 18 s. The maximum estimation error of the LSTM-Bi-LSTM model is 5%, and the computational time was 21 s; the error is reduced by 0.07% when compared to the Bi-LSTM model. However, it was computationally efficient for 32 s. The estimation error of the IPSO-Bi-LSTM model is reasonably stable and approaches the actual value, and the maximum estimated error that was observed is 4.91% with a computational time of 11 s. It obtained a reduction in error and computational efficiency compared to the previous model of 0.9% and 10 s. The IPSO-Bi-LSTM model estimates the LIB SOH more precisely and efficiently than the LSTM, Bi-LSTM, and LSTM-Bi-LSTM models when employed on the B7 battery.

Figure 9 shows that the B18 battery SOH estimation initiates at the 87th cycle. The LSTM model estimates poorly, with a maximum error of 6.70%, and requires a computational time of 38 s. Bi-LSTM estimates have a maximum error of 6.30% with a computational time of 17 s. Bi-LSTM reduces the maximum estimate error by 0.40% and enhances estimation results compared to LSTM, which is computationally efficient 18 s. It has unpredictable violent changes within the actual value range. The LSTM-Bi-LSTM model has a lower maximum estimate error of 6.58% with a computational time of 14 s than that of the Bi-LSTM model. A maximum estimation error and computational time lower than 0.12% and 24 s has been achieved. Although the LSTM-Bi-LSTM model improves the estimation performance, its accuracy is still high. The IPSO-Bi-LSTM model reduces the maximum estimation error by 6.17% compared to the LSTM-Bi-LSTM model. The maximum estimate error drops by 0.41%, reducing error fluctuations. The computational time is 7 s, and 7 s is efficient compared to LSTM-Bi-LSTM. The model's estimation accuracy improves when the estimation curves eventually match the actual values. The IPSO-Bi-LSTM model estimates the LIB of SOH more efficiently than the LSTM, Bi-LSTM, and LSTM-Bi-LSTM models

on the B18 battery dataset. Additionally, the LIB SOH estimations are significantly more accurate. To visualize the estimated SOH inaccuracy for each modeling approach and highlight the IPSO-Bi-LSTM model's advantages.

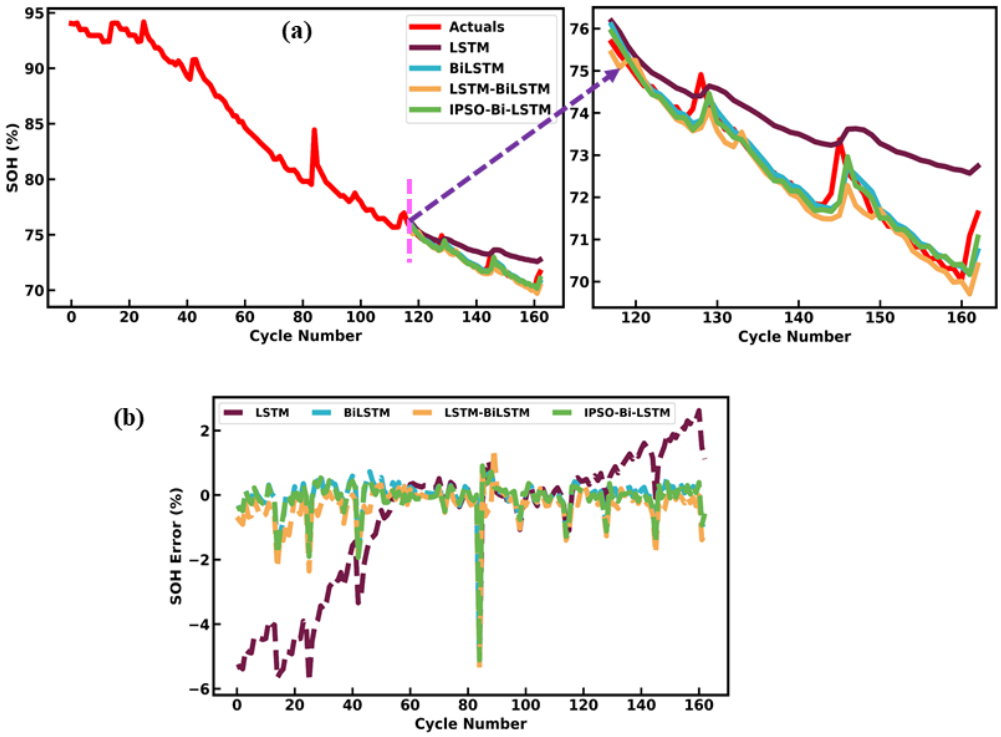


Figure 9. Estimation performance (**a**) estimation result of B7 battery (**b**) estimation error.

Figure 10 displays the estimated error of the SOH estimation at each cycle, as evaluated by MAE, MSE, RMSE, and R^2 based on the experimental data. The compared outcomes of the estimated evaluation of the performance indicators for all four models on the B5, B6, B07, and B18 batteries are shown in Table 2**.** IPSO-Bi-LSTM is superior to LSTM-Bi-LSTM, Bi-LSTM, and LSTM in estimating efficiency under similar conditions, as shown by the outcome figures. The IPSO-Bi-LSTM model demonstrates excellent fitting and varying degrees of decrease in the MAE, MSE, and RMSE. Among the four cells, the IPSO-Bi-LSTM model has the highest values for MAPE, MAE, and RMSE, which are 0.3302, 0.2533, and 0.5032%, respectively. The IPSO-Bi-LSTM model demonstrated a decrease in the MAE, MSE, and RMSE in comparison to the LSTM-Bi-LSTM model, with average reductions of 0.1897%, 0.2102%, and 0.1331%, respectively.

The IPSO-Bi-LSTM model demonstrated a mean decrease of 0.2188% in MAE, 0.2003% in MSE, and 0.1192% in RMSE when compared to the Bi-LSTM model. The IPSO-Bi-LSTM model demonstrated an average reduction of 1.6023% in MAE, 4.5708% in MSE, and 1.7204% in RMSE when compared to the LSTM model. The average computational time taken by the LSTM, BiLSTM, LSTM-Bi-LSTM, and IPSO-Bi-LSTM models was 64, 41, 25, and 13 s, respectively. The IPSO-Bi-LSTM model demonstrated a decrease in computational time in comparison to the three models, with average reductions of 52, 29, and 13 s, respectively. The Bi-LSTM model increases the algorithm's accuracy while incorporating IPSO strengthens the model's estimating capabilities. This advantage improves the feasibility of implementing the proposed method in real-world scenarios [45,46].

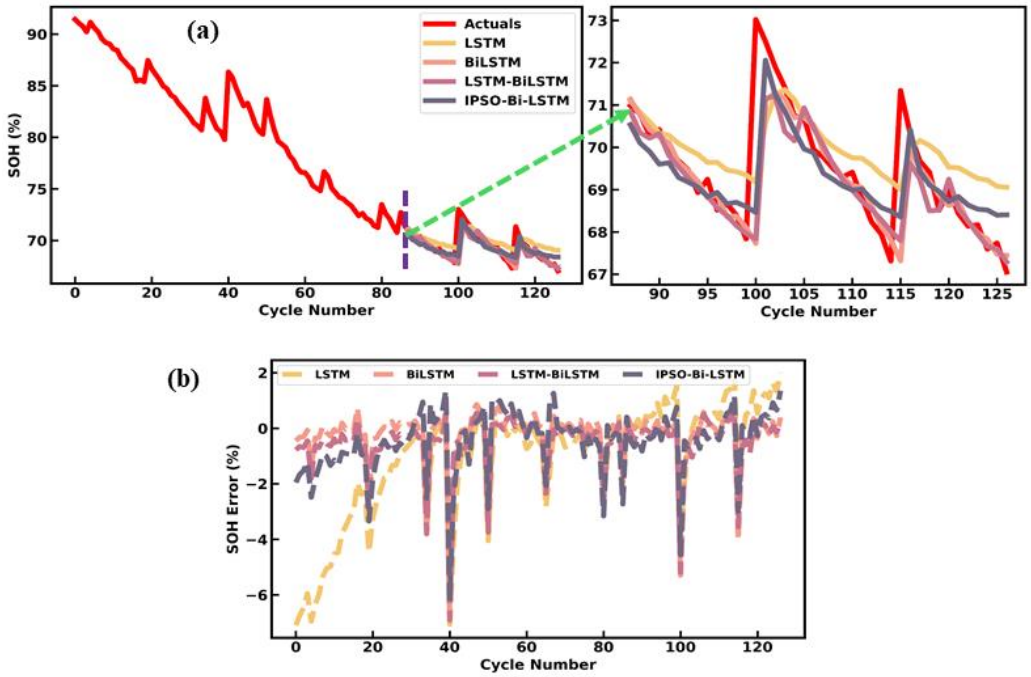


Figure 10. Estimation performance (**a**) estimation result of B18 battery (**b**) estimation error.

Table 2. The statistical performances of SOH estimation.

Battery	Approaches	MAE	MSE	RMSE	R^2	Max_ Error	Computational Time (s)
B5	LSTM	4.0592	5.8309	4.2227	0.7954	8.79014	77
	BiLSTM	0.5584	0.5200	0.7211	0.9940	4.59323	50
	LSTM-BiLSTM	0.6772	0.8109	0.9005	0.9949	4.3757	39
	IPSO-Bi-LSTM	0.3302	0.4406	0.6638	0.9973	4.1032	17
B6	LSTM	0.5512	0.4128	0.6425	0.9981	2.1558	69
	BiLSTM	0.4820	0.3544	0.5953	0.9984	1.8461	43
	LSTM-BiLSTM	0.4772	0.3158	0.5620	0.9986	1.5144	24
	IPSO-Bi-LSTM	0.3804	0.2533	0.5032	0.9989	1.5003	13
B7	LSTM	1.8662	6.7284	2.5939	0.8912	5.8640	71
	BiLSTM	0.6324	0.7976	0.8931	0.9871	5.0765	53
	LSTM-BiLSTM	0.6324	0.7976	0.8931	0.9871	5.0001	21
	IPSO-Bi-LSTM	0.3986	0.4904	0.7003	0.9921	4.9157	11
B18	LSTM	1.5614	5.7007	2.3876	0.8938	6.7006	38
	BiLSTM	0.8311	1.5186	1.2323	0.9717	6.3000	17
	LSTM-BiLSTM	0.6007	1.3041	1.1420	0.9757	6.58172	14
	IPSO-Bi-LSTM	0.5193	1.2050	1.0977	0.9776	6.1794	7

4.1. Discussion

The Bi-LSTM architecture is very effective in capturing temporal relationships in sequential data by analyzing information in both the forward and backward directions. The bidirectional processing enables the model to include information from both previous and future time steps, which is essential for effectively modeling complicated time series data, such as estimating the SOH of a battery. Conventional models may have difficulties in capturing distant connections, particularly when the correlation between input characteristics extends across numerous time intervals. Bi-LSTM, on the other hand, overcomes this constraint by preserving pertinent information from both forward and backward directions, resulting in more precise and resilient predictions.

However, IPSO improves the performance of the model by improving the hyperparameters more efficiently than traditional optimization approaches. IPSO enhances the

traditional PSO by including methods that avoid premature convergence and promote a more comprehensive exploration of the solution space. As a result, this improves the fine-tuning of the hyperparameters, which is crucial for optimizing the accuracy and stability of the Bi-LSTM model. IPSO optimizes the model's performance on different datasets by adjusting important parameters, such as the learning rate and number of hidden units. The integration of Bi-LSTM's capacity to capture complicated temporal patterns with IPSO's effective hyperparameter optimization results in a robust framework that surpasses other approaches, especially in tasks related to complex and dynamic systems such as SOH estimate. This combination provides the proposed approach with an asset in the domain of battery management by enabling it to provide forecasts that are more precise, reliable, and broadly applicable.

4.2. Practical Implementation and Advantages

The BMS monitors and manages battery charging and discharging. Therefore, accurate SOH calculation is essential. The BMS may change the charge rates or operating settings to avoid overcharging or deep discharging using accurate SOH data. This degree of management decreases the battery stress and failure risk and ensures safe and optimum operation. Additionally, an accurate SOH estimate extends the battery life. The technology can forecast when a battery will decline to a level where it no longer meets the performance criteria by precisely measuring its health. Predictive maintenance or replacement prevents abrupt failures and extends battery life. Maintaining batteries under optimum operating conditions based on correct SOH data slows deterioration and extends battery life.

Another advantage of precise SOH estimates is cost savings. Battery lifetime extension decreases the need for frequent replacements and saves money, particularly in businesses like EVs, renewable energy storage, and UPS that use massive battery deployments. Accurate SOH monitoring reduces warranty claims and ensures that batteries are only replaced when needed, saving money. As well as economic advantages, precise SOH estimate enhances battery-powered system dependability, which is crucial in medical devices, aerospace, and military operations. Reliable SOH data keeps these systems running, establishing user confidence and encouraging battery-powered technology adoption.

An accurate SOH estimate reduces battery usage's environmental effect. Reduced battery replacements reduce the environmental impact of battery waste and recycling. This method reduces electrical waste and conserves battery manufacturing resources by maximizing battery usage before replacement. It also improves energy reserve management in energy storage systems like renewable energy installations. Knowing the exact health of each storage array battery improves the energy dispatch, grid stability, and storage system performance. These practical consequences demonstrate the method's potential to improve battery-powered system performance, dependability, and sustainability in a variety of applications.

5. Conclusions

In this work, a significant contribution by developing a novel IPSO-Bi-LSTM model for accurate estimation of SOH of LIBs for electrified transportation. This article proposes an IPSO approach to enhance the accuracy of estimating the SOH for LIBs. The proposed approach involves using a neural network model called Bi-LSTM to estimate the SOH based on the IPSO. The IPSO-Bi-LSTM model uses the Bi-LSTM to capture the relationship between the feature indicators and SOH in both directions. The IPSO is used to optimize the structural hyperparameters of the model to achieve a precise and robust estimation of the SOH of LIBs. The proposed IPSO-Bi-LSTM model is superior to the LSTM, Bi-LSTM, and LSTM-Bi-LSTM models based on the validated experimental outcomes using different types of battery datasets. The IPSO-Bi-LSTM model achieves an MAE lower than 0.33%, an MSE lower than 0.25%, and an RMSE lower than 0.50% compared to other approaches.

Future Works

The proposed IPSO-Bi-LSTM model for the SOH estimation of LIBs has shown potential, but further investigation is needed to improve its performance and application. To increase model robustness and accuracy across varied situations, environmental parameters, and real-time operating circumstances should be included. Hybrid models that integrate IPSO-Bi-LSTM with other artificial intelligence methods may capture more complicated battery behavior patterns and relationships. Additionally, applying this concept to different batteries or energy storage devices may reveal its generalizability and adaptation. These future studies will improve SOH estimation methods and improve battery management system reliability and efficiency.

Author Contributions: Methodology, R.N.; Software, R.N.; Conceptualization, R.N.; Formal analysis, R.N.; Funding acquisition, R.N.; Software, Writing—review &editing, R.N.; Conceptualization, D.K.; Data curation, D.K.; Investigation, D.K.; Methodology, D.K.; Software, D.K.; Writing—original draft, D.K.; Conceptualization, M.R.; Investigation, M.R.; Resources, M.R.; Supervision, M.R.; Validation, M.R.; Writing—original draft, M.R.; Conceptualization, A.K.P.; Investigation, A.K.P.; Supervision, A.K.P.; Validation, Writing—original draft, A.K.P.; Funding acquisition, A.A.; Project administration, A.A.; Writing—review &editing, A.A.; Funding acquisition, Y.-I.C.; Formal analysis, Y.-I.C.; Project administration, Y.-I.C.; Writing—review &editing, Y.-I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Korea Agency for Technology and Standards in 2022. The project numbers are 1415181638 (Establishment of standardization basis for BCI and AI Interoperability), 1415181629 (Development of International Standard Technologies based on AI Model Lightweighting Technologies), and 1415180835 (Development of International Standard Technologies based on AI Learning and Inference Technologies).

Data Availability Statement: The authors used a publicly available dataset and presented a detailed analysis in the appropriate sections of this paper.

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

Nomenclature

SOH	State of health
IPSO	Improved particle swarm optimization
LIBs	lithium-ion batteries
C_n and En	Nominal capacity and total available energy
i_t, f_t, o_t, and h_t	Input, forget output gate and hidden states respectively;
C_t $\boldsymbol{c_t^*}C_{t-1}$	Cell memory and states
σ	Sigmoid function
W_f, W_i, ${}_0W_o$	Forget, input and output weight matrix respectively;
b_f, b_i, b_o, b_c	Forget, input, output and cell memory bias respectively;
tanh	Tan Hyperbolic Function
*, $\odot$	Convolution operator and Hamdard product;
$\overrightarrow{\boldsymbol{\Theta}}$*_BiLSTM*, $\overleftarrow{\boldsymbol{\Theta}}$*_BiLSTM*	BiLSTM forward and backward process;
ReLU	Rectified linear unit;
LSTM, *BiLSTM*	Long Short-Term Memory and Bidirectional LSTM;
$\boldsymbol{R_E}-\boldsymbol{R_n}$-$\boldsymbol{R}_0$	Present, end of life, original battery internal resistance
$\boldsymbol{v}$ and $\boldsymbol{p}$	Flight speed and searched particles
W and w_{ij} and ${w^t}_{ij}$	inertia weight and dynamic weights
${x^t}_{id}$, ${v^t}_{id}$, and $gbest^t$	Optimal position of particle
c1 and *c2*, and I_{max}	learning factors and maximum number of iterations and

References

1. Singh, G.; Misra, S.C.; Daultani, Y.; Singh, S. Electric vehicle adoption and sustainability: Insights from the bibliometric analysis, cluster analysis, and morphology analysis. *Oper. Manag. Res.* **2024**, *17*, 635–659. [CrossRef]
2. Ruan, H.; Wei, Z.; Shang, W.; Wang, X.; He, H. Artificial Intelligence-based health diagnostic of Lithium-ion battery leveraging transient stage of constant current and constant voltage charging. *Appl. Energy* **2023**, *336*, 120751. [CrossRef]
3. Kumar, D.; Rizwan, M.; Panwar, A.K. Advanced Intelligent approach for state of charge estimation of lithium-ion battery. *Energy Sources Part A Recover. Util. Environ. Eff.* **2023**, *45*, 10661–10681. [CrossRef]
4. El Kassar, R.; Al Takash, A.; Faraj, J.; Hammoud, M.; Khaled, M.; Ramadan, H.S. Recent advances in lithium-ion battery integration with thermal management systems for electric vehicles: A summary review. *J. Energy Storage* **2024**, *91*, 112061. [CrossRef]
5. Nyamathulla, S.; Dhanamjayulu, C. A review of battery energy storage systems and advanced battery management system for different applications: Challenges and recommendations. *J. Energy Storage* **2024**, *86*, 111179. [CrossRef]
6. Zhang, M.; Yang, D.; Du, J.; Sun, H.; Li, L.; Wang, L.; Wang, K. A Review of SOH Prediction of Li-Ion Batteries Based on Data-Driven Algorithms. *Energies* **2023**, *16*, 3167. [CrossRef]
7. Zhang, S.; Guo, X.; Dou, X.; Zhang, X. A rapid online calculation method for state of health of lithium-ion battery based on coulomb counting method and differential voltage analysis. *J. Power Sources* **2020**, *479*, 228740. [CrossRef]
8. Weng, C.; Sun, J.; Peng, H. A unified open-circuit-voltage model of lithium-ion batteries for state-of-charge estimation and state-of-health monitoring. *J. Power Sources* **2014**, *258*, 228–237. [CrossRef]
9. Pang, Z.; Yang, K.; Song, Z.; Niu, P.; Chen, G.; Meng, J. A new method for determining SOH of lithium batteries using the real-part ratio of EIS specific frequency impedance. *J. Energy Storage* **2023**, *72*, 108693. [CrossRef]
10. Huang, P.; Gu, P.; Kang, Y.; Zhang, Y.; Duan, B.; Zhang, C. The state of health estimation of lithium-ion batteries based on data-driven and model fusion method. *J. Clean. Prod.* **2022**, *366*, 132742. [CrossRef]
11. Fang, L.; Li, J.; Peng, B. Online Estimation and Error Analysis of both SOC and SOH of Lithium-ion Battery based on DEKF Method. *Energy Procedia* **2019**, *158*, 3008–3013. [CrossRef]
12. Liu, S.; Deng, D.; Wang, S.; Luo, W.; Takyi-Aninakwa, P.; Qiao, J.; Li, S.; Jin, S.; Hu, C. Dynamic adaptive square-root unscented Kalman filter and rectangular window recursive least square method for the accurate state of charge estimation of lithium-ion batteries. *J. Energy Storage* **2023**, *67*, 107603. [CrossRef]
13. Wu, J.; Cui, X.; Meng, J.; Peng, J.; Lin, M. Data-Driven Transfer-Stacking-Based State of Health Estimation for Lithium-Ion Batteries. *IEEE Trans. Ind. Electron.* **2024**, *71*, 604–614. [CrossRef]
14. Wang, Y.; Tian, J.; Sun, Z.; Wang, L.; Xu, R.; Li, M.; Chen, Z. A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. *Renew. Sustain. Energy Rev.* **2020**, *131*, 110015. [CrossRef]
15. Chang, C.; Wang, Q.; Jiang, J.; Wu, T. Lithium-ion battery state of health estimation using the incremental capacity and wavelet neural networks with genetic algorithm. *J. Energy Storage* **2021**, *38*, 102570. [CrossRef]
16. Tian, J.; Liu, X.; Li, S.; Wei, Z.; Zhang, X.; Xiao, G.; Wang, P. Lithium-ion battery health estimation with real-world data for electric vehicles. *Energy* **2023**, *270*, 126855. [CrossRef]
17. Liu, G.; Zhang, X.; Liu, Z. State of health estimation of power batteries based on multi-feature fusion models using stacking algorithm. *Energy* **2022**, *259*, 124851. [CrossRef]
18. Xu, H.; Wu, L.; Xiong, S.; Li, W.; Garg, A.; Gao, L. An improved CNN-LSTM model-based state-of-health estimation approach for lithium-ion batteries. *Energy* **2023**, *276*, 127585. [CrossRef]
19. Tang, A.; Jiang, Y.; Yu, Q.; Zhang, Z. A hybrid neural network model with attention mechanism for state of health estimation of lithium-ion batteries. *J. Energy Storage* **2023**, *68*, 107734. [CrossRef]
20. Zhao, G.; Zhang, C.; Duan, B.; Shang, Y.; Kang, Y.; Zhu, R. State-of-Health Estimation With Anomalous Aging Indicator Detection of Lithium-Ion Batteries Using Regression Generative Adversarial Network. *IEEE Trans. Ind. Electron.* **2023**, *70*, 2685–2695. [CrossRef]
21. Zhang, Z.; Min, H.; Guo, H.; Yu, Y.; Sun, W.; Jiang, J.; Zhao, H. State of health estimation method for lithium-ion batteries using incremental capacity and long short-term memory network. *J. Energy Storage* **2023**, *64*, 107063. [CrossRef]
22. Xu, L.; Wu, F.; Chen, R.; Li, L. Data-driven-aided strategies in battery lifecycle management: Prediction, monitoring, and optimization. *Energy Storage Mater.* **2023**, *59*, 102785. [CrossRef]
23. Li, R.; Hong, J.; Zhang, H.; Chen, X. Data-driven battery state of health estimation based on interval capacity for real-world electric vehicles. *Energy* **2022**, *257*, 124771. [CrossRef]
24. Wang, C.; Wang, S.; Zhou, J.; Qiao, J.; Yang, X.; Xie, Y. A novel back propagation neural network-dual extended Kalman filter method for state-of-charge and state-of-health co-estimation of lithium-ion batteries based on limited memory least square algorithm. *J. Energy Storage* **2023**, *59*, 106563. [CrossRef]
25. Ansari, S.; Ayob, A.; Lipu, M.S.H.; Hussain, A.; Saad, M.H.M. Particle swarm optimized data-driven model for remaining useful life prediction of lithium-ion batteries by systematic sampling. *J. Energy Storage* **2022**, *56*, 106050. [CrossRef]
26. Ma, Y.; Yao, M.; Liu, H.; Tang, Z. State of Health estimation and Remaining Useful Life prediction for lithium-ion batteries by Improved Particle Swarm Optimization-Back Propagation Neural Network. *J. Energy Storage* **2022**, *52*, 104750. [CrossRef]
27. Zhi, Y.; Wang, H.; Wang, L. A state of health estimation method for electric vehicle Li-ion batteries using GA-PSO-SVR. *Complex. Intell. Syst.* **2022**, *8*, 2167–2182. [CrossRef]

28. Hu, X.; Yuan, H.; Zou, C.; Li, Z.; Zhang, L. Co-Estimation of State of Charge and State of Health for Lithium-Ion Batteries Based on Fractional-Order Calculus. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10319–10329. [CrossRef]
29. Ye, M.; Guo, H.; Cao, B. A model-based adaptive state of charge estimator for a lithium-ion battery using an improved adaptive particle filter. *Appl. Energy* **2017**, *190*, 740–748. [CrossRef]
30. Lai, X.; He, L.; Wang, S.; Zhou, L.; Zhang, Y.; Sun, T.; Zheng, Y. Co-estimation of state of charge and state of power for lithium-ion batteries based on fractional variable-order model. *J. Clean. Prod.* **2020**, *255*, 120203. [CrossRef]
31. Ren, X.; Liu, S.; Yu, X.; Dong, X. A method for state-of-charge estimation of lithium-ion batteries based on PSO-LSTM. *Energy* **2021**, *234*, 121236. [CrossRef]
32. Guo, R.; Shen, W. Online state of charge and state of power co-estimation of lithium-ion batteries based on fractional-order calculus and model predictive control theory. *Appl. Energy* **2022**, *327*, 120009. [CrossRef]
33. Zhang, X.; Wang, Y.; Wu, J.; Chen, Z. A novel method for lithium-ion battery state of energy and state of power estimation based on multi-time-scale filter. *Appl. Energy* **2018**, *216*, 442–451. [CrossRef]
34. Li, X.; Ju, L.; Geng, G.; Jiang, Q. Data-driven state-of-health estimation for lithium-ion battery based on aging features. *Energy* **2023**, *274*, 127378. [CrossRef]
35. Pan, D.; Li, H.; Wang, S. Transfer Learning-Based Hybrid Remaining Useful Life Prediction for Lithium-Ion Batteries under Different Stresses. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3501810. [CrossRef]
36. Zhu, Y.; Li, G.; Wang, R.; Tang, S.; Su, H.; Cao, K. Intelligent fault diagnosis of hydraulic piston pump combining improved LeNet-5 and PSO hyperparameter optimization. *Appl. Acoust.* **2021**, *183*, 108336. [CrossRef]
37. Chen, Q.; Song, Y.; Zhao, J. Short-term traffic flow prediction based on improved wavelet neural network. *Neural Comput. Appl.* **2020**, *33*, 8181–8190. [CrossRef]
38. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M.J. A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3478–3487. [CrossRef]
39. Kumar, D.; Rizwan, M.; Panwar, A.K. Advanced Intelligent Hybrid Approach for State of Charge Estimation of Li-Ion Batteries in Electric Vehicles under Various Drive Cycles. In Proceedings of the 2023 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy: Power Electronics, Smart Grid, and Renewable Energy for Sustainable Development, PESGRE, Trivandrum, India, 17–20 December 2023. [CrossRef]
40. Ullah, F.U.M.; Ullah, A.; Haq, I.U.; Rho, S.; Baik, S.W. Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks. *IEEE Access* **2020**, *8*, 123369–123380. [CrossRef]
41. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
42. Rizwan, M.; Kumar, D.; Panwar, A.K. New Filter-based Supervised Learning Approach for State of Charge Estimation of Li-Ion Battery for EV Applications. In Proceedings of the 2023 International Conference on Modeling, Simulation & Intelligent Computing (MoSICom), Dubai, United Arab Emirates, 7–9 December 2023; pp. 213–218. [CrossRef]
43. NASA. Ames Intelligent Systems Division Home. Available online: https://www.nasa.gov/intelligent-systems-division/ (accessed on 22 October 2023).
44. Shah, S.M.; Sun, Z.; Zaman, K.; Hussain, A.; Ullah, I.; Ghadi, Y.Y.; Khan, M.A.; Nasimov, R. Advancements in Neighboring-Based Energy-Efficient Routing Protocol (NBEER) for Underwater Wireless Sensor Networks. *Sensors* **2023**, *23*, 6025. [CrossRef] [PubMed]
45. Abdusalomov, A.; Rakhimov, M.; Karimberdiyev, J.; Belalova, G.; Cho, Y.I. Enhancing Automated Brain Tumor Detection Accuracy Using Artificial Intelligence Approaches for Healthcare Environments. *Bioengineering* **2024**, *11*, 627. [CrossRef] [PubMed]
46. Safarov, F.; Kutlimuratov, A.; Abdusalomov, A.B.; Nasimov, R.; Cho, Y.-I. Deep Learning Recommendations of E-Education Based on Clustering and Sequence. *Electronics* **2023**, *12*, 809. [CrossRef]
47. Kumar, D.; Mehata, M.S. Synthesis of diamagnetic ZnO nano-crystallites via sol-gel method and their photocatalytic activity. *Indian. J. Eng. Mater. Sci.* **2022**, *29*, 437–444. [CrossRef]
48. Kumar, D.; Singh, M.K.; Mehata, M.S. Exploration of grown cobalt-doped zinc oxide nanoparticles and photodegradation of industrial dye. *Mater. Res. Bull.* **2022**, *150*, 111795. [CrossRef]
49. Tagmatova, Z.; Abdusalomov, A.; Nasimov, R.; Nasimova, N.; Dogru, A.H.; Cho, Y.-I. New Approach for Generating Synthetic Medical Data to Predict Type 2 Diabetes. *Bioengineering* **2023**, *10*, 1031. [CrossRef]
50. Abdusalomov, A.B.; Nasimov, R.; Nasimova, N.; Muminov, B.; Whangbo, T.K. Evaluating Synthetic Medical Images Using Artificial Intelligence with the GAN Algorithm. *Sensors* **2023**, *23*, 3440. [CrossRef]

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

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)
Karnataka, India. Jun 28-29, 2024

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) | 979-8-3503-8269-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITEICS61368.2024.10625402

A Novel Dielectric modulated Double Gate Charge-Trapping MOSFET: Ferroelectric Gate Stack and Gaussian Doped Substrate Integration for Enhanced Performance

1st Akkshita Swain ¶
Department of Applied Physics
Delhi Technological University
New Delhi, India
akkshita12@gmail.com

2nd Nalin Agarwal ¶
Department of Applied Physics
Delhi Technological University
New Delhi, India
nalinagarwal2002@gmail.com

3rd Tanisha Singhal ¶
Department of Applied Physics
Delhi Technological University
New Delhi, India
tanishasinghal0101@gmail.com

4th Rishu Chaujar *
Department of Applied Physics
Delhi Technological University
New Delhi, India
chaujar.rishu@dtu.ac.in

***Abstract*—his paper investigates the operational characteristics and performance metrics of a novel double gate charge-trapping metal oxide semiconductor field effect transistor (DG-MOSFET) that incorporates a ferroelectric gate stack, a Gaussian-doped multilayered substrate, and a buried oxide layer. The study utilizes Silvaco TCAD(Technology Computer-Aided Design) Atlas for device simulation and analysis. A comprehensive comparison between the proposed DG-MOSFET, a conventional DG-MOSFET, and other variations, has been presented. Through detailed simulations and analysis, the paper evaluates key parameters including threshold voltage, subthreshold swing, drain-induced barrier lowering (DIBL), on/off current ratio, and transconductance. Different modifications were carried out and their implications were compared to study device performance. Results indicate that the proposed DG-MOSFET exhibits superior performance characteristics compared to both reference devices, showcasing enhanced carrier mobility, reduced leakage current, and improved subthreshold behavior. Furthermore, the Gaussian doped multilayered substrate contributes to mitigating short-channel effects, enhancing overall device scalability and reliability.his paper investigates the operational characteristics and performance metrics of a novel double gate charge-trapping metal oxide semiconductor field effect transistor (DG-MOSFET) that incorporates a ferroelectric gate stack, a Gaussian-doped multilayered substrate, and a buried oxide layer. The study utilizes Silvaco TCAD(Technology Computer-Aided Design) Atlas for device simulation and analysis. A comprehensive comparison between the proposed DG-MOSFET, a conventional DG-MOSFET, and other variations, has been presented. Through detailed simulations and analysis, the paper evaluates key parameters including threshold voltage, subthreshold swing, drain-induced barrier lowering (DIBL), on/off current ratio, and transconductance. Different modifications were carried out and their implications were compared to study device performance. Results indicate that the proposed DG-MOSFET exhibits superior performance characteristics compared to both reference devices, showcasing enhanced carrier mobility, reduced leakage current, and improved subthreshold behavior. Furthermore, the Gaussian doped multilayered substrate contributes to mitigating short-channel effects, enhancing overall device scalability and reliability.T**

***Index Terms*—DG-MOSFET, DIBL, Short-Channel Effects, Subthreshold Swing, Transconductance**

I. Introduction

The journey of scaling down transistors, in pursuit of enhanced performance and efficiency in circuits, has resulted in transistor dimensions comparable to the size of atoms [1]. These smaller dimensions however, pose a range of challenges, such as gate oxide tunneling and short channel effects [2]. Over the past few decades, researchers have explored various different technologies, namely, FinFETs, TFETs, HEMT [3], ISFET [4], multigate MOSFETS, SOI MOSFETS, Buried oxide FETs [5]–[17] to subdue SCEs and boost efficiency, but with each iteration there seems to be room for more improvement. Future nanoelectronic applications demand constraints such as drain-induced barrier lowering (DIBL) to be minimal, as it can lead to a degradation of subthreshold swing, which is indicative of how effectively the device can switch. A poor subthreshold swing would affect the device's ability to operate efficiently in low-power environments. It is also necessary to reduce the leakage current, i.e. Ioff, and in turn improving the Ion/Ioff ratio is very crucial.

In this context, this paper presents a comprehensive investigation into the working principles and performance characteristics of a DG-MOSFET incorporating a ferroelectric gate stack, a Gaussian doped multilayered substrate, and a buried oxide layer. Leveraging the advanced simulation capabilities

¶Authors contributed equally
* Corresponding Author

979-8-3503-8269-3/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:07:37 UTC from IEEE Xplore. Restrictions apply.

of Silvaco TCAD Atlas, we conduct detailed numerical experiments to analyze the device's operational behavior under varying operating conditions.

In contrast to conventional planar MOSFETs, multigate MOSFETs offer enhanced gate control and reduced leakage currents. This study delves into the performance of a double gate MOSFET with various modifications aimed at improving its functionality. Our investigation includes the incorporation of a gate stack featuring high-k dielectrics to enhance electrostatic control, along with a buried oxide layer of Al_2O_3 surrounding the substrate to mitigate high parasitic capacitances and leakage currents. Further enhancements were achieved by introducing a layer of Barium titanate ($BaTiO_3$), resulting in reduced Drain Induced Barrier Lowering (DIBL) and incorporating a ferroelectric element into our gate stack.

A multilayered substrate allows for the possibility of combining different material properties such as higher thermal conductivity and breakdown field strength, and thus in our proposed device we have silicon 6H-SiC, layered beneath Si, and the implications of the same have been explored via SILVACO 2D simulations. An elaborate comparison of changes in the doping pattern has also been studied and Gaussian doping has been opted for in our device as it imparts an optimized gradient of doping concentration, providing greater control of the depletion region and in turn minimizing the DIBL.

The paper is structured into the following sections:
SECTION 2 outlines the construction and dimensions of the proposed device. SECTION 3 presents the methodology followed and explains the basis of modifications. SECTIONS 4 and 5 put forth the results and the discussion respectively.

II. Proposed Device

In this study, a dual-gate MOSFET was modified by integrating various mechanisms to enhance its functionality. The depicted figure (Fig. 1) illustrates a layered structure representing an advanced multi-gate MOSFET design, offering notable enhancements over the conventional dual-gate MOSFET. In a basic dual-gate MOSFET, the channel is controlled from two sides by the gate electrode, thereby providing improved electrostatic control over the channel compared to a single-gate MOSFET [18]. However, the structure discussed in this study expands upon this concept by incorporating additional material layers and gate arrangements.

With a length of 0.245 μm, the device is made up of three separate portions along the x-direction: firstly, the drain region (0.045 μm), then the substrate region which is divided in 3 parts with two different regions each consisting of Silicon and 6H-Silicon Carbide region, and finally the source region (0.04 μm). The first and third region of substrate has silicon and SiC with a thickness of 0.003 μm and 0.007 μm respectively. The second region has a non-uniform layer of Silicon with 0.007 μm thickness and SiC layer with 0.003 μm thickness. Vertically, the structure consists of several material layers: an air gap, top and bottom gate electrodes which are each 0.001 μm thick, high-k dielectric gate insulators

TABLE I
Device Characteristics

Material	**Thickness along y-axis** (in μm)
Air	0.24
high_dielectric	0.001
Silicon	0.003–0.007
6H-SiC	0.003–0.007
SiO2	0.001
Al2O3	0.001
BaTiO3	0.001
Si3N4	0.001
Gate Conductor	0.001
Source Conductor	0.005
Drain Conductor	0.005

including Barium Titanate ($BaTiO_3$, 0.001 μm thick), silicon dioxide (SiO_2), aluminum oxide (Al_2O_3), and silicon nitride (Si_3N_4). The structure has a total height of 0.02 μm overall. The n-type doping concentrations are taken as 1E20 cm^{-3} in the source and drain regions, 1E19 cm^{-3} in the SiC regions. Similarly, the p-type doping concentrations are taken as 1E18 cm^{-3} in the central channel region, 1E17 cm^{-3} in the SiC regions. Furthermore, Gaussian doping profiles are used for more accurate modeling than uniform doping [19].
To accurately depict the device's physical behavior, the simulations integrate various models, encompassing concentration-dependent mobility, Shockley-Read-Hall recombination, Auger recombination, band-gap narrowing, and field-dependent mobility. Output parameters such as band structures, carrier concentrations, electric fields, current densities, and mobility values are analyzed to provide a thorough understanding of the device's performance.

III. Methodology

The simulation tool-set - Silvaco TCAD (Technology Computer Aided Design) [20] software, was used to simulate and analyze different parameters of the semiconductor device. As stated above, the device in consideration is a novel dual-gate MOSFET (Metal Oxide Semiconductor Field Effect Transistor) with modifications made to improve performance. Each

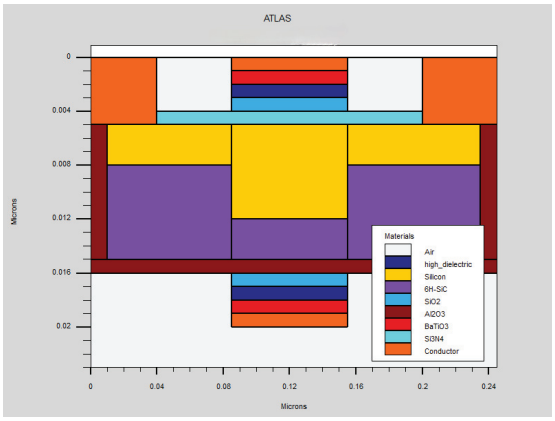


Fig. 1. A figure showing the structure of the proposed device

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:07:37 UTC from IEEE Xplore. Restrictions apply.

iteration of simulation, including all the structural changes, were carried out using the software, written in ATLAS language. Subsequently, performance parameters were compared to derive an optimised transistor. The simulations were visualized using TonyPlot, which is a plotting tool included in the Silvaco TCAD package. The software includes capabilities to define the meshing for our workspace, which enables the use of uniform doping profiles across the device, when needed. Apart from defining the structural aspects of the device, the software also provides provisions for plotting graphs. Different curves including various physical parameters like DIBL (drain induced barrier lowering), Ion (on-state current), Ioff (off-state current), Vth (threshold voltage), gate voltage, drain current were plotted against each other for different iterations to deduce the functioning of the device.
Multiple concepts related to semiconductor devices were included in the study in order to enhance its performance for real-life applications. Firstly, to establish the dimensions of the transistor, various parameters such as gate length, substrate thickness, and dielectric layer thickness, were tweaked, and performance parameters of a DG-MOSFET were compared.

For the first modification, a high K dielectric oxide layer has been introduced in our device. This addition contributes to high retention and reduced gate voltage. High K dielectrics, namely, Hf_O2, Al_2O_3 and TiO_2 were substituted and compared, and as indicated by our simulation runs (Fig. 2), TiO_2 proved to be the more suitable material, producing the lowest DIBL.

DIBL stands for Drain-Induced Barrier Lowering and it refers to the phenomenon where the threshold voltage of the device decreases with an increase in drain voltage. In simpler terms, when the drain voltage increases, it causes the lowering of the potential barrier at the channel-drain junction, in turn inducing easier flow of charge carriers from source to drain. It is a non-ideal effect as it impacts the device's performance; a higher DIBL can result in poorer subthreshold swing. It has been calculated using the given formula:

$$\text{DIBL} = \left(\frac{\mathrm{d}V_{\text{th}}}{\mathrm{d}V_{\text{ds}}}\right)$$

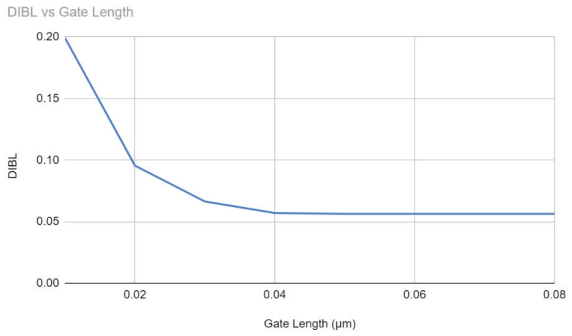


Fig. 2. A figure showing the variation of DIBL with the Gate Length

TABLE II
DIELECTRIC MATERIALS

Material	**Thickness** (in μm)	**Gate Length** (in μm)	**DIBL**
HfO_2	0.002	0.05	0.011
HfO_2	0.002	0.06	0.008
HfO_2	0.002	0.07	0.008
HfO_2	0.001	0.07	0.007
Al_2O_3	0.002	0.05	0.015
Al_2O_3	0.002	0.06	0.011
Al_2O_3	0.002	0.07	0.009
Al_2O_3	0.001	0.07	0.008
TiO_2	0.002	0.05	0.009
TiO_2	0.002	0.06	0.007
TiO_2	0.002	0.07	0.007
TiO_2	0.001	0.07	0.0077

The study also involved the addition of a ferroelectric gate stack, in the form of Barium Titanate ($BaTiO_3$), which enhances gate control and reduces leakage currents. A buried oxygen layer, referred to as 'BOX' was also included. This layer of aluminium oxide (Al_2O_3), which creates a silicon-on-insulator (SOI) like structure, also works on improving device isolation. Finally, the channel area was suspended between high-k dielectrics, and isolated from the substrate, to reduce the overall contact between semiconductor materials, resulting in lesser energy radiation. This suspension was performed by adding a Silicon Nitride (Si_3N_4) layer in between the Ferroelectric gate stack and the substrate region.

Fig.2 makes it abundantly evident that the DIBL value inversely proportionally declines with increasing gate length. As the gate length goes from 0 to 0.02 on the graph, DIBL first declines sharply from 0.2 to 0.09 before gradually decreasing and being constant at 0.06.
All these structural modifications were compared to each other using the Silvaco TCAD software itself, in order to find the most efficient and performing device. The next section will discuss these devices in detail and will additionally delve into establishing a comparative study among the devices, on different physical parameters through curves and mathematical calculations.

IV. RESULTS

A. Device Comparision

This table compares the device parameters of the proposed device against traditional dual-gate MOSFETs, with and without a SiO_2 layer, as well as variations of the proposed device as seen in Fig. 3 and Fig. 4, i.e. without the nitride layer and ferroelectric gate stack ($BaTiO_3$) respectively.

Compared to traditional dual-gate MOSFETs, the proposed device exhibits lower threshold voltages, lower drain-induced barrier lowering, and a lower subthreshold swing, indicating better gate control, reduced short-channel effects, and improved switching characteristics. The removal of the nitride layer and ferroelectric gate stack in the proposed device slightly impact the key device parameters, suggesting that these layers may play a critical role in the performance

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:07:37 UTC from IEEE Xplore. Restrictions apply.

TABLE III
DEVICE COMPARISON

Device	V_t (in V)	DIBL	SubV_t (in V)	I_{off} (in A)	I_{on} (in A)	I_{on}/I_{off}
DG-Mosfet with additional SiO_2 layer	0.223	0.0122	0.0606	2.1749×10^{-11}	0.00136	6.2639×10^{7}
DG-Mosfet without additional SiO_2 layer	0.202	0.00823	0.0600	4.3158×10^{-11}	0.00346	8.0242×10^{7}
Proposed device	0.192	0.00664	0.05983	6.279×10^{-11}	0.0004	6.863×10^{6}
Proposed Device without Nitride layer	0.187	0.00666	0.05984	7.518×10^{-11}	0.00051	6.8121×10^{6}
Proposed Device without Nitride and $BaTiO_3$ layers	0.187	0.00665	0.05984	7.514×10^{-11}	0.00051	6.8172×10^{6}

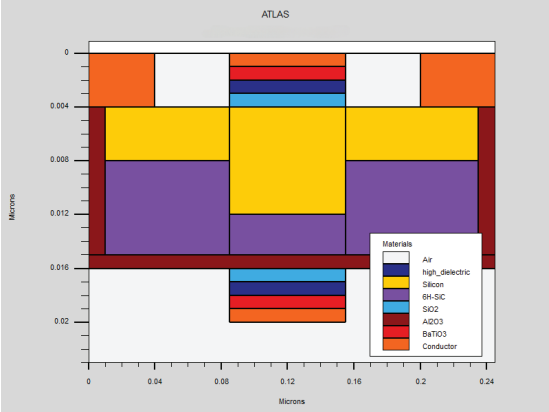


Fig. 3. A figure showing the structure of MOSFET without the nitride layer

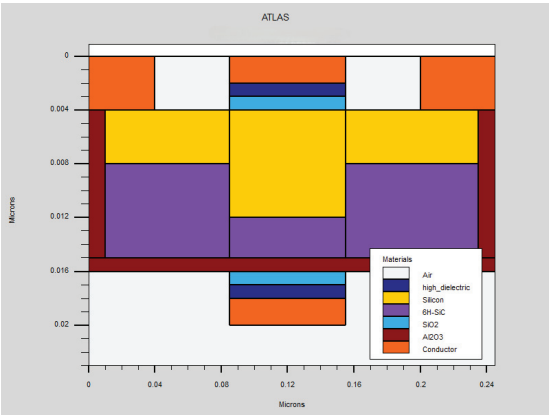


Fig. 4. A figure showing the structure of MOSFET without the nitride layer and the ferroelectric ($BaTiO_3$) gate stack

metrics evaluated in this study. It is important to note that these parameters are crucial in assessing the device's performance, scalability, and suitability for specific applications. The trade-offs between various parameters, such as on-current, off-current, and short-channel effects, need to be carefully considered based on the target application requirements.

Fig. 5 illustrates the transfer characteristics of the device, plotting the drain current (y-axis) against the gate voltage (x-axis). Initially drain current was zero and as we kept increasing the gate voltage, after reaching a threshold value of 0.3V, drain current increased sharply reaching a saturation value of 0.0006A. This curve exhibits a steep rise in drain current as the gate voltage increases, indicating good switching behavior. The shape of the curve is crucial for determining

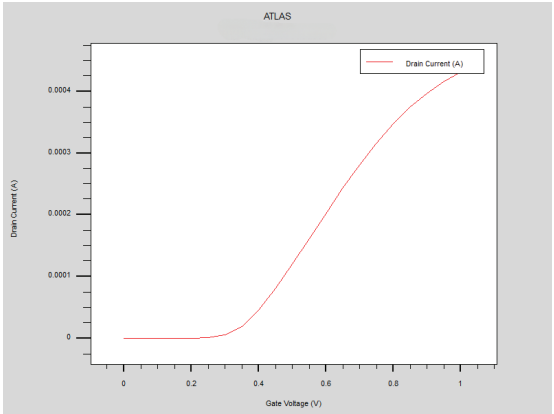


Fig. 5. A figure showing the variation of drain current with the gate voltage

the device's I_{on}/I_{off} ratio, subthreshold swing, and overall switching performance.This curve helps us to understand the MOSFET behavior.

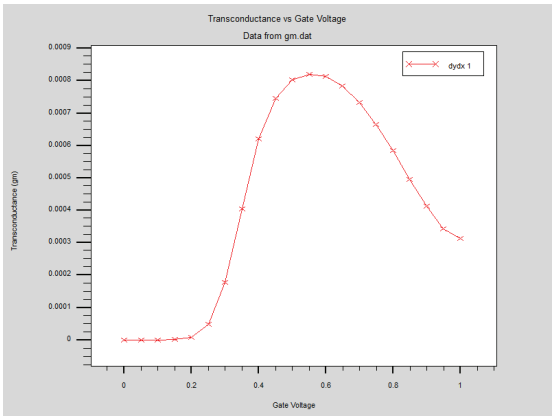


Fig. 6. A figure showing the variation of rate of change of drain current with respect to the gate voltage

Fig. 6 depicts the transconductance (g_m) curve, which represents the rate of change of drain current with respect to the gate voltage. This curve exhibits a characteristic bell shape, with the peak representing the maximum transconductance value. The peak in our graph corresponds to a value of approximately 0.0008225. The transconductance is an important parameter that influences the device's gain and switching speed.

Fig. 7 shows the relationship between device efficiency and gate voltage. Device Efficiency is defined by the ratio of transconductance and drain current. The curve initially rises

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:07:37 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
DOPING COMPARISION

Device	V_t (in V)	DIBL	SubV_t (in V)	I_{off} (in A)	I_{on} (in A)	I_{on}/I_{off}
Proposed device with uniform doping	0.231	0.0107	0.0603	1.501×10^{-11}	1.50190	6.281×10^{7}
Proposed device with Gaussian Doping	0.192	0.00664	0.05983	6.279×10^{-11}	0.0004	6.863×10^{6}

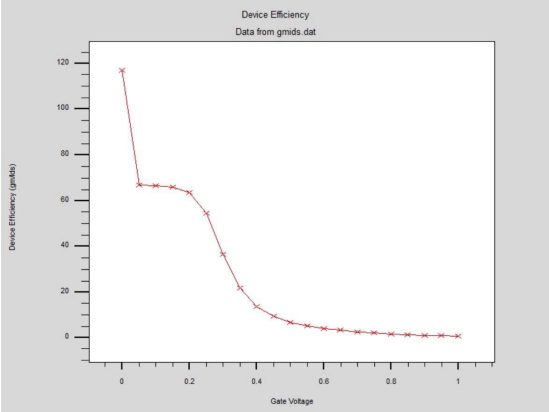


Fig. 7. A figure showing the variation of the ratio of transconductance to drain current, with respect to change in gate voltage

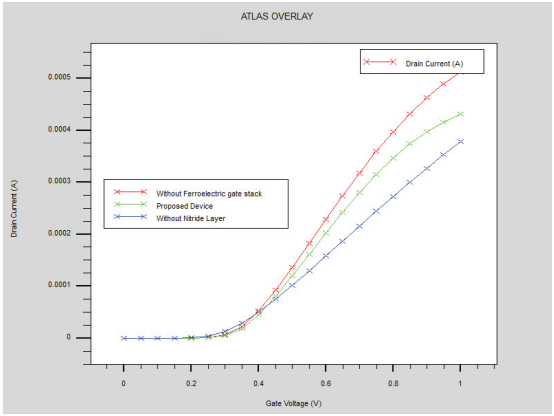


Fig. 8. A figure showing the comparision between three considered devices' drain current

sharply, indicating a rapid increase in drain current as the gate voltage increases from the subthreshold region. It drops to 66.75 just as we begin to supply voltage to the gate, which indicates the highest value of efficiency, and starts declining as the gate voltage increases. Beyond a certain point, the curve flattens out because the drain current saturates as the device enters the linear region of operation.

Fig. 8 compares the transfer characteristics of the proposed device with several variations, including the absence of the nitride layer and the ferroelectric gate stack. The curves are closely spaced suggesting the change occurs in each variation of MOSFET with each extra layer added influencing specific device parameters, such as leakage currents or subthreshold behavior. As shown in Fig. 9, the proposed device offers a higher switching ratio when compared to its other two variations.

B. Doping Comparision

The table presents a comparison of device parameters between two variations of the proposed device: one with uniform doping and another with Gaussian doping. The key differences lie in the threshold voltages (vt1 and vt2), drain-induced barrier lowering (dibl), subthreshold swing (subv_t), off-state current (I_{off}), on-state current (I_{on}), and on/off current ratio (I_{on}/I_{off}). Gaussian doping can be define using the given formula:

$$N(x) = N_{\text{peak}} \cdot \exp\left(-\frac{(x - x_{\text{peak}})^2}{2\sigma^2}\right)$$

The device with Gaussian doping exhibits lower threshold voltages, lower drain-induced barrier lowering, and a lower subthreshold swing compared to the uniform doping variant.

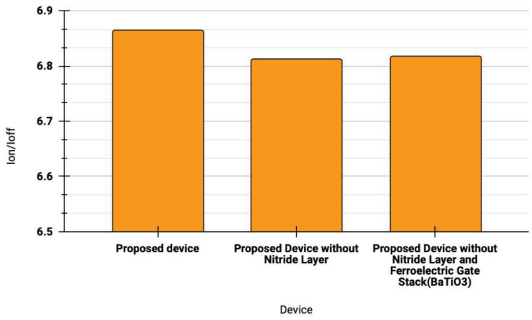


Fig. 9. A figure showing the comparision between three considered devices' I_{on}/I_{off} ratio

These improvements suggest better gate control over the channel, reduced short-channel effects, and enhanced switching characteristics. Interestingly, both doping profiles result in the same transconductance value.

V. DISCUSSION

The proposed device structure in the paper incorporates advanced features which include a high-k dielectric material titanium dioxide, a silicon nitride layer, aluminum oxide acting like a box around the substrate, a barium titanate layer, and gaussian doping profiles in the source, drain, and channel regions with different layers of materials in the substrate which comprises silicon and silicon carbide. These design elements aim to enhance various aspects of MOSFET performance, including gate control, leakage currents, carrier mobility, operating voltage/temperature, and scalability.

Compared to traditional dual-gate MOSFETs, the proposed device exhibits lower drain-induced barrier lowering (DIBL)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:07:37 UTC from IEEE Xplore. Restrictions apply.

and subthreshold swing values, as evident from Fig. 8. A lower DIBL indicates better control of the channel by the gate, reducing short-channel effects, while a lower subthreshold swing leads to improved switching characteristics and lower off-state leakage.

Fig. 9 shows that the proposed device with uniform doping demonstrates a higher on-current (I_{on}) and a higher on/off current ratio (I_{on}/I_{off}) compared to the traditional dual-gate MOSFET with a SiO_2 layer. However, the off-current (I_{off}) is slightly higher, which may lead to increased static power consumption.

The introduction of Gaussian doping in the proposed device results in a lower off-current (I_{off}) and a lower on-current (I_{on}) compared to the uniform doping variant, as depicted in Fig. 9. Nevertheless, the on/off current ratio (I_{on}/I_{off}) remains higher than the traditional dual-gate MOSFET with a SiO_2 layer, indicating better switching efficiency.

Overall, the proposed device structure demonstrates promising improvements in various performance metrics compared to traditional dual-gate MOSFETs, particularly in terms of short-channel effects, switching characteristics, and scalability, as evidenced by the lower DIBL, lower subthreshold swing, and higher on/off current ratio.

VI. Summary

The conducted study proposed a novel dual gate MOSFET device with charge trapping capabilities. It also resembles a semiconductor-on-insulator (SOI) design perspective and incorporated different physical mechanisms such as ferroelectric gate stack, a suspended channel, and the presence of a buried oxide layer. The research resulted in identifying the various parameters that affect the functioning and efficiency of a semiconductor device, and found out that our novel device has a lesser dibl (drain induced barrier lowering) value and subthreshold swing value than the traditional dual gate MOSFET device, which is a critical parameter that is taken into consideration while defining the effecienct performance of a particular semiconductor device. The proposed device in this study seems to function more efficiently than a traditional dual gate MOSFET structure, and the inclusion of different structural modifications indicate that a simple device like MOSFET also has wide capabilities to become a highly efficient device with numerous applications like power electronics [21], mobile carrier semiconductor chip systems [22], quantum computing [23] and in the field of artificial intelligence [24], mainly neuromorphic computing [25].

References

[1] Carlos Galup-Montoro and Márcio Cherem Schneider. *Mosfet Modeling for Circuit Analysis and Design*. WORLD SCIENTIFIC, 2007.

[2] S. Veeraraghavan and J.G. Fossum. Short-channel effects in soi mosfets. *IEEE Transactions on Electron Devices*, 36(3):522–528, 1989.

[3] Megha Sharma and Rishu Chaujar. Device optimization of t-shaped gate and polarized doped buffer-engineered inaln/gan hemt for improved rf/microwave performance. *Arabian Journal for Science and Engineering*, 2024.

[4] Yash Pathak, Piyush Mishra, Megha Sharma, Shipra Solanki, Ved Varun Agarwal, Rishu Chaujar, and Bansi Dhar Malhotra. Experimental circuit design and tcad analysis of ion sensitive field effect transistor (isfet) for ph sensing. *Materials Science and Engineering: B*, 299:116951, 2024.

[5] Venkatagirish Nagavarapu, Ritesh Jhaveri, and Jason C. S. Woo. The tunnel source (pnpn) n-mosfet: A novel high performance transistor. *IEEE Transactions on Electron Devices*, 55(4):1013–1019, 2008.

[6] Sayan Kanungo, Sanatan Chattopadhyay, Partha Sarathi Gupta, and Hafizur Rahaman. Comparative performance analysis of the dielectrically modulated full- gate and short-gate tunnel fet-based biosensors. *IEEE Transactions on Electron Devices*, 62(3):994–1001, 2015.

[7] G.V. Reddy and M.J. Kumar. A new dual-material double-gate (dmdg) nanoscale soi mosfet-two-dimensional analytical modeling and simulation. *IEEE Transactions on Nanotechnology*, 4(2):260–268, 2005.

[8] Xiaorong Luo, Tian Liao, Jie Wei, Jian Fang, Fei Yang, and Bo Zhang. A novel 4h-sic trench mosfet with double shielding structures and ultralow gate-drain charge. *Journal of Semiconductors*, 40(5):052803, may 2019.

[9] Juncheng Zhang, Yunteng Jiang, Haimeng Huang, Zimin Zhang, Junji Cheng, Bo Yi, Hongqiang Yang, and Zhiming Wang. Superjunction sic tcox-mosfet: Study and comparison. In *2022 IEEE 16th International Conference on Solid-State Integrated Circuit Technology (ICSICT)*, pages 1–3, 2022.

[10] Yusra Siddiqui, Nupur Mittal, and Imran Ullah Khan. Performance analysis and characterization of double gate and gate all around mosfet. In *2021 10th International Conference on System Modeling Advancement in Research Trends (SMART)*, pages 614–617, 2021.

[11] Amine Mohammed TABERKIT, Ahlam GUEN-BOUAZZA, and Mohamed HORCH. The importance of using dual channel heterostructure in strained p-mosfets. In *2018 International Conference on Communications and Electrical Engineering (ICCEE)*, pages 1–5, 2018.

[12] Rittik Ghosh. Performance investigation of dual trench split-control-gate mosfet as hydrogen gas sensor: A catalytic metal gate approach. *IEEE Sensors Letters*, 7(5):1–4, 2023.

[13] Nitin Garg, Aliya Rizvi, Astuti Chandra, and Anjali Singh. A review on comparative analysis of various mosfets on the basis of electrical parameters. In *2023 International Conference on Disruptive Technologies (ICDT)*, pages 456–461, 2023.

[14] Naveen Kumar and Ashish Raman. Novel design approach of extended gate-on-source based charge-plasma vertical-nanowire tfet: Proposal and extensive analysis. *IEEE Transactions on Nanotechnology*, 19:421–428, 2020.

[15] Hossain M. Fahad, Chenming Hu, and Muhammad M. Hussain. Simulation study of a 3-d device integrating finfet and utbfet. *IEEE Transactions on Electron Devices*, 62(1):83–87, 2015.

[16] Yan Guo, Xiaoyi Zhang, Kain Lu Low, Kai-Tak Lam, Yee-Chia Yeo, and Gengchiau Liang. Effect of body thickness on the electrical performance of ballistic n-channel gasb double-gate ultrathin-body transistor. *IEEE Transactions on Electron Devices*, 62(3):788–794, 2015.

[17] Wei Ke, Xu Han, Dingyu Li, Xinan Wang, Tianyi Zhang, Ruqi Han, and Shengdong Zhang. Recessed source/drain for sub-50 nm utb soi mosfet. *Semiconductor Science and Technology*, 22(5):577, apr 2007.

[18] T. Poiroux, M. Vinet, O. Faynot, J. Widiez, J. Lolivier, T. Ernst, B. Previtali, and S. Deleonibus. Multiple gate devices: advantages and challenges. *Microelectronic Engineering*, 80:378–385, 2005. 14th biennial Conference on Insulating Films on Semiconductors.

[19] D. Salameh and D. Linton. Study of the relation between doping profile and diode cv characteristics. *IEEE Transactions on Microwave Theory and Techniques*, 47(4):506–509, 1999.

[20] C.K. Sarkar. *Technology Computer Aided Design: Simulation for VLSI MOSFET*. CRC Press, 1 edition, 2013.

[21] Joseph Fabre, Philippe Ladoux, and Michel Piton. Characterization and implementation of dual-sic mosfet modules for future use in traction converters. *IEEE Transactions on Power Electronics*, 30(8):4079–4090, 2015.

[22] Ramesh Vaddi, Rajendra P. Agarwal, Sudeb Dasgupta, and Tony T. Kim. Design and analysis of double-gate mosfets for ultra-low power radio frequency identification (rfid): Device and circuit co-design. *Journal of Low Power Electronics and Applications*, 1(2):277–302, 2011.

[23] G. Baccarani and S. Reggiani. A compact double-gate mosfet model comprising quantum-mechanical and nonstatic effects. *IEEE Transactions on Electron Devices*, 46(8):1656–1666, 1999.

[24] Yilun Liu, Qingxuan Li, Hao Zhu, Li Ji, Qingqing Sun, David Wei Zhang, and Lin Chen. Dual-gate manipulation of a hfzrox-based mos2

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:07:37 UTC from IEEE Xplore. Restrictions apply.

field-effect transistor towards enhanced neural network applications. *Nanoscale*, 15:313–320, 2023.

[25] Lin Bao, Jiadi Zhu, Zhizhen Yu, Rundong Jia, Qifeng Cai, Zongwei Wang, Liying Xu, Yanqing Wu, Yuchao Yang, Yimao Cai, and Ru Huang. Dual-gated mos2 neuristor for neuromorphic computing. *ACS Applied Materials & Interfaces*, 11(44):41482–41489, 2019. PMID: 31597432.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:07:37 UTC from IEEE Xplore. Restrictions apply.

Earth Science Informatics
https://doi.org/10.1007/s12145-024-01451-y

RESEARCH

A novel spectral-spatial 3D auxiliary conditional GAN integrated convolutional LSTM for hyperspectral image classification

Pallavi Ranjan[1,2] · Ashish Girdhar[3] · Ankur.[2] · Rajeev Kumar[2]

Received: 20 June 2024 / Accepted: 11 August 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Hyperspectral Imaging (HSI) has revolutionized earth observation through advanced remote sensing technology, providing rich spectral and spatial information across multiple bands. However, this wealth of data introduces significant challenges for classification, including high spectral correlation, the curse of dimensionality due to limited labeled data, the need to model long-term dependencies, and the impact of sample input on deep learning performance. These challenges are further exacerbated by the costly and complex acquisition of HSI data, resulting in limited availability of labeled samples and class imbalances. To address these critical issues, our study proposes a novel approach for generating high-quality synthetic hyperspectral data cubes using an advanced Generative Adversarial Network (GAN) integrated with the Wasserstein loss and gradient penalty phenomenon (WGAN-GP). This approach aims to augment real-world data, mitigating the scarcity of labeled samples that has long been a bottleneck in hyperspectral image analysis and classification. To fully leverage both the synthetic and real data, we introduce a novel Convolutional LSTM classifier designed to process the intricate spatial and spectral correlations inherent in hyperspectral data. This classifier excels in modeling multi-dimensional relationships within the data, effectively capturing long-term dependencies and improving feature extraction and classification accuracy. The performance of our proposed model, termed 3D-ACWGAN-ConvLSTM, is rigorously validated using benchmark hyperspectral datasets, demonstrating its effectiveness in augmenting real-world data and enhancing classification performance. This research contributes to addressing the critical need for robust data augmentation techniques in hyperspectral imaging, potentially opening new avenues for applications in areas constrained by limited data availability and complex spectral-spatial relationships.

Keywords Hyperspectral classification · Auxiliary conditional GAN · Convolutions · LSTM · Wasserstein

Introduction

Hyperspectral imaging (HSI) has significantly advanced earth observation capabilities through the development of hyperspectral remote sensing technology (Arjovsky et al. 2017; Paul and Bhoumik 2022; Paul et al. 2021). An HSI image, captured by imaging spectrometers, forms a three-dimensional data cube where each point represents consecutive spectral information across multiple bands (Ranjan et al. 2024; Ranjan and Girdhar 2022; Ranjan and Gupta 2023; Ranjan and Girdhar 2024, 2022). However, the increase in spatial and spectral information poses several classification challenges:

- **High Correlation**: HSI consists of multiple bands with high spectral resolution, resulting in highly correlated spectral vectors (Wang et al. 2019).

Communicated by: Hassan Babaie

✉ Pallavi Ranjan
pallavi.ranjan@murdoch.edu.au

Ashish Girdhar
ashishgirdhar@kuk.ac.in

Ankur.
ankur_2k20phdco507@dtu.ac.in

Rajeev Kumar
rajeevkumar@dtu.ac.in

[1] School of Information Technology, Murdoch University, Dubai, United Arab Emirates

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

[3] Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, India

Published online: 23 August 2024

- **Curse of Dimensionality**: The high dimensionality of spectral bands, coupled with the limited availability of labeled data, leads to the Hughes phenomenon. Additionally, annotating samples is both time-consuming and labor-intensive (Wang et al. 2021, 2019).
- **Long-term Dependencies**: Effectively utilizing the rich spatial and spectral information in HSI and modeling their long-term dependencies, which refer to the relationships and patterns within the data that span over long sequences, is crucial for improving classification accuracy (Wang et al. 2021).
- **Sample Input Impact**: The number of input samples fed into the network significantly affects deep learning classification results (Hang et al. 2020).

To address these challenges, various computer vision techniques have been proposed, such as data augmentation (Hu et al. 2015; Chen et al. 2022), transfer learning (Hu et al. 2015; Du et al. 2013), domain adaptation, and randomization. Additionally, generative adversarial networks (GANs) have been used to synthesize data (Jia et al. 2021; Wambugu et al. 2021; Xia et al. 2022), with synthetic data from GANs being nearly indistinguishable from real data (Goodfellow et al. 2020). GANs, particularly their variants, can overcome several challenges in HSI classification. They help mitigate data insufficiency by generating high-quality synthetic samples, reducing the reliance on extensive labelled datasets. GANs also address the high correlation of spectral bands by producing diverse and representative data, enhancing the model's ability to generalize. Moreover, GANs aid in overcoming the curse of dimensionality by generating samples that capture complex spectral-spatial relationships. Specifically, Vanilla GANs generate data from random noise vectors but lack control over the generator (Goodfellow et al. 2020). To overcome this, Mirza and Osindero introduced conditional GANs (cGANs), which use class labels to improve the quality of synthetic samples for specific classes (Mirza and Osindero 2014). Other notable GAN variants include Deep Convolutional GAN by Radford et al. (2015), Wasserstein GAN by Arjovsky et al. (2017), Cycle-consistent GAN by Zhu et al. (2017), and Auxiliary Classifier GAN (AC-GAN) by Odena et al. (2017). This study utilizes the theory of AC-GAN, which leverages extra label information to enhance the classification capabilities of both the generator and the discriminator (Zhu et al. 2017; Wang et al. 2020).

To model long-term dependencies, the generated samples and the proposed adversarial network are input into a convolutional LSTM. Unlike traditional LSTMs, which overlook spatial information by flattening inputs into one dimension, convolutional LSTMs incorporate convolutions in both state-to-state and input-to-state transitions (Shi et al. 2015). Wu and Prasad demonstrated the effectiveness of RNN-based models in capturing the intrinsic dependencies between spectral sequences in hyperspectral data (Wu and Prasad 2017). Other studies have also used convolutional LSTMs to handle long-range dependencies in hyperspectral data (Song et al. 2018; Seydgar et al. 2019). In our proposed model, we introduce time-distributed layers to model long-term dependencies effectively. Time-distributed layers allow each time step of the sequence data to be processed independently and are crucial for handling sequences of HSI feature maps, as they maintain the spatial structure of the data across temporal dimensions. Thus, the proposed novel model incorporates time-distributed layers to model long-term dependencies effectively and offers the following contributions:

- **Addressing Data Insufficiency**: We propose a novel adversarial spectral-spatial network inspired by AC-GAN, integrated with Wasserstein loss, to generate high-quality samples that closely resemble the original spectrum.
- **Generator and Discriminator Design**: Our novel generator reconstructs and exploits unlabeled data to generate synthetic data, while the discriminator statistically monitors the generated samples. The network is guided by the discriminator's loss function and real information discerned by the discriminator.
- **Global Spectral Dependencies**: Unlike existing GAN methods that capture local correlations with convolutional kernels, our proposed discriminator captures global spectral dependencies. The proposed ConvLSTM, with time-distributed layers, processes spatial and spectral correlations in higher dimensions and models intrinsic data relationships.
- **Improved Classification Performance**: The proposed 3D-ACWGAN-ConvLSTM is designed to enhance classification performance and mitigate overfitting in HSI classification. The model's effectiveness is validated using three benchmark datasets: IP, PU, and BOT.

Related work

This section presents an overview of relevant studies and advancements in the areas of generative adversarial networks (GANs) in section "Auxiliary Classifier Generative Adversarial Network (ACGAN)" and convolutional LSTM networks in section"Convolutional long short-term memory network (conv-LSTM)" for HSI classification. Section"Current trends in literature" presents recent trends.

Auxiliary Classifier Generative Adversarial Network (ACGAN)

The concept of adversarial modelling was initially proposed by Goodfellow et al. (2014) to generate synthetic data that

mirrors the input distribution. This framework consists of two main components: a generator and a discriminator. The generator learns to create data that resembles real samples, while the discriminator aims to distinguish between genuine and synthetic data. The generator takes a noise vector z as input and produces a synthetic image as output. The generator takes a noise vector z as input and produces a synthetic image as output. This noise vector z is a randomly sampled vector from a predefined probability distribution, that is, Gaussian (normal) distribution in this case. The dimensions of the noise vector are chosen to provide sufficient variability for the generator to learn a wide range of features in the data. As z is sampled from a Gaussian distribution, each element z_i in the vector is an independent draw from a Gaussian distribution, denoted as $z \sim \mathcal{N}(0, 1)$. This synthetic image, or a real image, is then fed into the discriminator, which outputs a probability distribution indicating whether the image is real or fake. This adversarial process, where the generator and discriminator compete to improve their respective performances, has been employed in various studies to generate hyperspectral samples (Zhu et al. 2018; Wang et al. 2021; Roy et al. 2021; Ranjan et al. 2024). However, a fundamental challenge with the naïve GAN approach is its binary classification limitation (real or fake), which poses difficulties in multiclass data categorization.

To address this issue, Odena et al. (2017) introduced the Auxiliary Classifier GAN (ACGAN), where the discriminator is equipped with a softmax activation function to handle multiclass problems. This architecture, based on the auxiliary GAN concept (Wang et al. 2021), takes noise z and class label c as inputs. The discriminator outputs two results: one for distinguishing between real and fake data, and another for classifying the input based on its category c.

Convolutional long short-term memory network (conv-LSTM)

Hyperspectral images (HSIs) encompass thousands of spectral bands, each carrying valuable information. Relying solely on spectral information can degrade classification performance, especially for samples within the same class that have different spectra and for samples from different classes with similar spectra (Feng et al. 2020). To fully exploit both spectral and spatial information in HSI bands, Shi et al. (2015) implemented a convolutional LSTM (ConvLSTM) network.

ConvLSTM networks leverage LSTM models to capture long-term dependencies in the spectral domain while utilizing convolution operations to extract spatial features. The combination of these techniques helps in associating long-term spectral relationships and spatial features, as depicted in the following equations (Hu et al. 2020). Here, X_a represents the input to the current cell, H_{a-1} is the output of the previous cell, and C_{a-1} is the previous cell state. The weights and biases for the input, forget, and output gates are denoted by W and b with appropriate subscripts. The symbols $*$ and $\odot$ represent convolution operations and the Hadamard product, respectively, while σ is a nonlinear activation function, and W is the convolution filter with a square kernel size k.

$$input_a = \sigma(W_{xinput} * X_a + W_{hinput} * H_{a-1} + W_{cinput} \odot C_{a-1} + b_{input}) \quad (1)$$

$$forget_a = \sigma(W_{xforget} * X_a + W_{hforget} * H_{a-1} + W_{cforget} \odot C_{a-1} + b_{forget}) \quad (2)$$

$$C_{aa} = \tanh(W_{xc} * X_a + W_{hc} * H_{t-1} + b_c) \quad (3)$$

$$C_a = forget_a \odot C_{a-1} + input_a \odot C_{aa} \quad (4)$$

$$output_a = \sigma(W_{xoutput} * X_a + W_{houtput} * H_{a-1} + W_{coutput} \odot C_{a-1} + b_{output}) \quad (5)$$

$$H_a = output_a \odot \tanh(C_a) \quad (6)$$

Current trends in literature

Recent studies have explored various GAN architectures for data generation and classification tasks. In 2018, Odena et al. (2017) proposed a regularization framework to mitigate overfitting in CNNs, pioneering the use of GANs for sample generation with notable classification performance. Wang et al. (2019) in 2019 developed a network that performed well with limited training samples, high spectral resolutions, and class imbalance, although it could have better integrated spatial-spectral data. In 2021, Wang et al. (2020) introduced an enhanced GAN with adaptive dropblock regularization to address dataset imbalance and model collapse, achieving good classification accuracy. Feng et al. (2020) advanced the quality of generated samples by integrating a convolutional LSTM layer in the discriminator to manage long-term spectral dependencies, resulting in excellent classification performance. Similarly, Dhok et al. (2019) utilized convolutional LSTM to preserve intrinsic structure information, extracting significant spatial-spectral features, though the classification accuracy required improvement.

To address limited sample issues, Sun and Bourennane (2020) proposed a transfer learning-based framework using Wasserstein GAN, effectively extracting spatial-spectral information and achieving good classification accuracy. Zhang et al. (2020) optimized training with progressive GAN

and Wasserstein gradient penalty to prevent gradient vanishing. Haut et al. (2019) enhanced overfitting resistance by generating images with varying occlusion levels. Xue (2020) improved classification performance with a generative capsule network that combined spatial-spectral feature extraction using auxiliary conditional GAN. Farooque et al. (2021) demonstrated excellent feature extraction using ConvLSTM and CNN, leading to improved classification accuracy. Hang et al. (2020) proposed a multitask GAN for reconstructing hyperspectral data using both labelled and unlabeled samples, alleviating model collapse and gradient vanishing issues by incorporating a capsule network with GAN (Wang et al. 2021), and using multiscale and octave convolutions to reduce parameter count.

Table 1 summarizes current trends in literature and compares various proposed models. The proposed models in the reviewed studies exhibit several common shortcomings. Many models, such as SSCRN (Farooque et al. 2021) and the Spectral-Spatial 3D ConvLSTM (Hu et al. 2020), show decreased classification accuracy with small training data and struggle when handling smaller patch sizes. Models like DR-WGAN-GP (Sun and Bourennane 2020) and AU-Caps-GAN (Xue 2020), might benefit from exploring alternative dimensionality reduction techniques or methods to further alleviate overfitting. Additionally, models such as AD-GAN (Wang et al. 2020) and CA-GAN (Feng et al. 2020) face challenges in optimizing sample generation and module placement, requiring extensive experimentation to determine optimal configurations. Other models, like MTGAN (Hang et al. 2020) and Dc-Caps GAN (Wang et al. 2021), need further improvements in classification accuracy and optimization to reduce computational costs while maintaining good classification accuracy. These identified limitations highlight the need for ongoing refinement and experimentation to enhance the efficiency and effectiveness of these models in hyperspectral image classification.

Proposed architecture

This section discusses the proposed handcrafted GAN discussed in section"Proposed ACWGAN" with layer configuration discussed in Section"Proposed generator and discriminator with wasserstein loss" and customized conv-LSTM in Section"Proposed Conv LSTM". Section"Spherical interpolation" discussed spherical interpolation.

Proposed ACWGAN

Goodfellow introduced the concept of Generative Adversarial Networks (GANs) to generate synthetic samples using deep networks (Goodfellow et al. 2014). Building on this, Zhu et al. (2018) utilized 3DGAN to extract spectral-spatial features and classify hyperspectral data accurately. The generated adversarial samples enhanced the training process, leading to improved classification accuracy. The proposed adversarial network, inspired by auxiliary conditional GANs (AC-GANs), is illustrated in Fig. 1.

The architecture consists of two critical components: the generator and the discriminator. The generator learns to approximate data distribution using noisy data from the latent space. It is trained with an adversarial objective function to refine this approximation. The generator also requires class labels to learn and classify data appropriately. The discriminator, on the other hand, is provided with both class labels and input images to distinguish between real and synthetic data, determining whether a given sample originated from the generator or the training set. This process is formalized through a min-max game between the generator and discriminator, as depicted in Eq. 7.

The objective function $V(\text{Dis}, \text{Gen})$ is designed to misclassify real images as fake or vice versa. Here, $\text{Gen}(z)$ represents the synthetic samples generated by the generator, and $\text{Dis}(x)$ denotes the probability that x is a real dataset sample. Emp stands for the empirical estimation of the probability of original data and noise. The discriminator's role includes a classifier that discerns the authenticity of images while performing classification.

$$\min_{\text{Dis}} \max_{\text{Gen}} V(\text{Dis}, \text{Gen}) = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\log \text{Dis}(x)\right] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1-\text{Dis}(\text{Gen}(z))\right)\right] \tag{7}$$

Equations 8 and 9 present the log-likelihoods for the discriminator predicting the correct source and for the classifier predicting the correct class. Here, cl denotes a specific class label, and $cl|X_{\text{real}}$ is the probability of the class label predicted by the classifier Cla. The discriminator aims to maximize $L(\text{Source}) + L(\text{Class})$, while the generator seeks to maximize $L(\text{Class}) - L(\text{Source})$ (Liao and Dong 2022).

$$\begin{aligned} L(\text{Source}) =& \mathbb{E}\left[\log P(\text{Source} = real|X_{\text{real}})\right] \\ &+ \mathbb{E}\left[\log P(\text{Source} = synthetic|X_{\text{synthetic}})\right] \end{aligned} \tag{8}$$

$$\begin{aligned} L(\text{Class}) =& \mathbb{E}\left[\log P(\text{Class} = cl|X_{\text{real}})\right] \\ &+ \mathbb{E}\left[\log P(\text{Class} = cl|X_{\text{fake}})\right] \end{aligned} \tag{9}$$

Proposed generator and discriminator with wasserstein loss

AC-GANs, which use the Jensen-Shannon distance (JSD) (Goodfellow et al. 2014), often suffer from model collapse and unstable training (Srivastava et al. 2017). To address these issues, Arjovsky et al. (2017) proposed Wasserstein GANs (WGANs), utilizing the Earth Mover distance (EM)

Table 1 Current trends in literature

Study	Proposed model	Dataset used	Hyperparameters	Details	Significant contributions	Shortcomings
Farooque et al. (2021)	SSCRN : 3DConvLSTM + 3DCNN	IP: 99.17 PU : 99.31 SA : 99.67	OP: Adam; LR: 0.0003, 0.0001; BS: 32, 64; D: 0.25; AC: tanh, ReLU; IT: 300; LF: Categorical Cross-Entropy; Tr/Te :0.10/0.05/0.005 - 0.90/0.95/0.95	1. Combines Conv 3DLSTM and 3DCNN to learn spectral and spatial features jointly 2. Integrates residual network and skip connections to alleviate decreasing accuracy problem and enhance training performance	1. Associates relationships between the spectral domain and feeds it into a spatial context 2. The proposed model can be generalised to other domains as well.	1. Decreased classification accuracy with small training data. 2. Decreased performance with a smaller patch size chosen, the proposed network could be computationally lean
Hu et al. (2020)	Spectral-Spatial 3D ConvLSTM	IP : 98.79 PU : 97.10 SA : 99.29	OP: Adam; LR: 0.0001; EP : 2000; D: 0.25; AC: Softmax; LF: Cross-Entropy; Tr/Te: 0.10-0.90	1. Associates relationships between spectral bands and fuses spatial information with spectral information. 2. Employs Convolutional LSTM to slice HSI data into cubes to process three-dimensional information	1. Reduced discriminative features and alleviated overfitting while maintaining classification performance 2. Preserves intrinsic structure of HIS ad extracts significant features	1. Not considering spatial information and passing it into a two-dimensional model reduces the classification accuracy. 2. Proposed model did not work well for the IP dataset when trained with small training samples
Sun and Bourennane (2020)	DR-WGAN-GP	IP : 95.80 PU : 98.06	OP: RMSProp; EP : 128; LF: Wasserstein; L : Unsupervised; Tr/Te : 0.1-0.9	1. To deal with the problem of limited samples, Wasserstein GAN and transfer learning are used. 2. Exploits spectral-spatial information in cubes with no labelled samples involved	1. Unsupervised learning to deal with the problem of limited samples. 2. 3D Conv and 3D Deconvolution exploit spectral and spatial information extensively and increase classification accuracy	1. Use of dimensionality reduction dramatically improves the performance of PU and slightly affects IP datasets. Other Dimensionality reduction techniques can be tried to improve performance
Xue (2020)	AU-Caps-GAN	IP :99.12 PU:99.23	LR: 1; EP : 30; BS : 3; D: 0.25; AC: Softmax; LF: Softmax; Tr/Te : 0.3/0.4-0.7/0.6	1. The auxiliary GAN network incorporates a three-dimensional CNN and a capsule network. 2. The proposed discriminator employs both limited labelled and entirely generated samples to improve classification performance.	1. Works well with small training samples 2. Deals with the limited sample availability and trains discriminator with generated samples, thereby reducing the overfitting problem.	1. Discriminator can be trained with lesser annotated and more generated samples. 2. Overfitting could be further alleviated with generated samples
Wang et al. (2020)	AD-GAN	IP : 97.23 PU: 95.38 SA: 95.89	OP: AdaDrop; EP : 300; BS : 100; D :30/35/40/45; AC: Softmax; Tr/Te : 0.1/0.02/0.01	1. Deals with the minority class samples and improves the classification accuracy 2. AdaDrop is used for regularization and alleviates the model collapse problem	1. Works for class imbalance problem 2. Drop block and skip connections instead of dropout can deal with various ground objects.	1. The generator can generate samples belonging to the majority of classes that are reduced with the proposed model's help. However, this could be reduced by experimenting with other techniques as well.

Table 1 continued

Study	Proposed model	Dataset used	Hyperparameters	Details	Significant contributions	Shortcomings
Wang et al. (2019)	Caps-Triple GAN	IP:90.02 PU:96.31	O: Adam; LR: 0.01, BS: 80; D: 0.5; AC: Routing Softmax; LF: Minmax; Tr/Te; 0.10/0.90; L: Semi-Supervised	1. Proposed triple generative adversarial network to generate more samples 2. Integrated capsule network to classify the images	1. Alleviates the problem of insufficient samples 2. Capsule network works excellent for classification	1. Classification performance can be improved 2. Spatial-spectral features could be discovered integrally 3. Computing time overhead
Feng et al. (2020)	CA-GAN	IP:97.4 PU:99.2	LR: 0.008, 0.035; EP: 600; AC: ReLU, Softmax; LF: Minmax; Tr/Te; 0.05(IP),0.03(PU); L: Semi-Supervised	1. Generator generates better-quality samples using competitive and collaborative learning 2. Attention module to extract discriminative features and suppress less useful ones	1. Solves the insufficient sample problem 2. Attention module filters misleading samples 3. Integrally uses spatial, spectral features using Conv LSTM	1. Determining the optimal number of modules to be placed inside the proposed convolutional GAN is challenging and needs much experimentation
Hang et al. (2020)	MTGAN	IP:92.42 HU:92.32	O: Adam; LR: 0.01/0.001/0.0001; AC: ReLU/LReLU; LF: Wasserstein; Tr/Te; 0.05(IP), 0.2(HU)	1. Better classification performance by exploiting unlabeled generated samples 2. Skip layer connection introduced during the generation and classification process	1. Improves the reconstruction performance 2. The information from shallow layers is explored using skip connections	1. Classification accuracy can be improved further by either improving sample quality or employing optimization techniques.
Wang et al. (2021)	Dc-Caps GAN	PU:97.98 SA:93.87 KSC:99.54	O: Adam; BS: 32; AC: LReLU; LF: Adversarial, Margin; Tr/Te;0.01(PU), 0.005 (SA), 0.06 (KSC)	1. Capsule network integrated with GAN to fully exploit spectral and spatial features 2. Incorporated octave and multiscale convolutions to reduce duplicate parameters 3. Adding a capsule network to the discriminator improves the stability of the entire model	1. Addition of multiscale and octave convolutions reduces the number of overall parameters as it decreases the complexity of the model by reducing the parameters 2. Excellent classification accuracy	1. As the capsule size increases, the network's computational cost increases. The network can be optimized to decrease the cost and maintain classification accuracy.

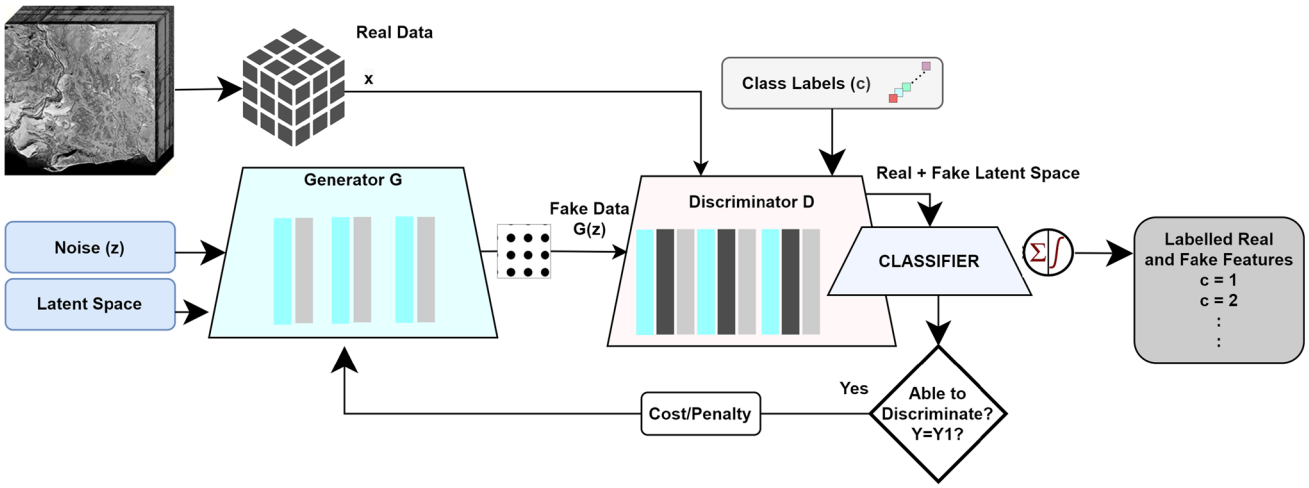


Fig. 1 Proposed adversarial network

instead of JSD. EM distance offers a better measure of the distance between two distributions, thereby providing more functional gradients to the generator. The EM distance is defined in Eq. 10, where C is the minimum cost of transporting mass between real and generated samples, and $\bigcap\left(P_{\text{real}}, P_{\text{generated}}\right)$ denotes all joint distribution sets $y(a, b)$ including real and generated samples, with $||\cdot||$ representing the norm.

WGANs leverage EM distance to aid early convergence of the generator. However, this can lead to underfitting, which is mitigated by adding a gradient penalty to the original Wasserstein loss. The loss functions for the generator and discriminator, as defined by Arjovsky et al. (2017), are shown in Eqs. 11 and 12, where f is a one-sided Lipschitz function.

Tables 2 and 3 depict the layers involved in Generator and Discriminator.

$$C\left(P_{\text{real}},\ P_{\text{generated}}\right) = \inf_{y \in \bigcap\left(P_{\text{real}}, P_{\text{generated}}\right)} E_{a,b\sim y}[|a-b|] \tag{10}$$

$$\nabla_\theta \frac{1}{m}\sum_{i=1}^{m} f_w\left(g_\theta\left(z^i\right)\right) \tag{11}$$

$$\nabla_w\left[\frac{1}{m}\sum_{i=1}^{m} f_w\left(x^i\right) - \frac{1}{m}\sum_{i=1}^{m} f_w\left(g_\theta\left(z^i\right)\right)\right] \tag{12}$$

Proposed Conv LSTM

Convolutional Neural Networks (CNNs) consist of stacked convolutions, pooling layers, and dense layers activated to classify images (Bera and Shrivastava 2020). Studies by Hu et al. (2015) and Chen et al. (2016) employed CNNs to extract discriminative features from hyperspectral images (HSI). Despite their efficiency, CNNs have limitations in extracting more compelling features due to the lack of information interaction within each layer.

Zhou et al. (2019) pioneered the use of Long Short-Term Memory (LSTM) networks for HSI classification by fusing spectral and spatial information. LSTMs, successors to

Table 2 Network structure of generator

Layer (type)	Output shape	Param #	Activation
Conv3D	(32, 144, 144, 219, 188)	1692	Leaky ReLU (alpha=0.2)
Batch normalization	(32, 144, 144, 219, 188)	752	Momentum = 0.99
Conv3D_1	(32, 142, 142, 217, 172)	873244	Leaky ReLU (alpha=0.2)
Batch normalization	(32, 142, 142, 217, 172)	688	Momentum = 0.99
Conv3D	(32, 139, 139, 214, 156)	1717404	Leaky ReLU (alpha=0.2)
Batch normalization	(32, 139, 139, 214, 156)	624	–
Total params	2,594,404		
Trainable params	2,593,372		
Non-trainable params	1,032		

Table 3 Network structure of discriminator

Layer (type)	Output shape	Param #	Activation
Conv3D_input	(32, 145, 145, 220, 1)	–	–
Conv3D	(32, 144, 144, 219, 188)	1692	Leaky ReLU (alpha=0.2)
UpSampling_3D	(32, 288, 288, 438, 188)	0	–
Batch normalization	(32, 288, 288, 438, 188)	752	Momentum = 0.99
Conv3D_1	(32, 286, 286, 436, 172)	873244	Leaky ReLU (alpha=0.2)
UpSampling_3D_1	(32, 572, 572, 872, 172)	0	–
Batch normalization_1	(32, 572, 572, 872, 172)	688	Momentum = 0.99
Conv3D_2	(32, 569, 569, 869, 156)	1717404	Leaky ReLU (alpha=0.2)
UpSampling_3D_2	(32, 1138, 1138, 1738, 156)	0	–
Batch normalization_2	(32, 1138, 1138, 1738, 156)	624	Momentum = 0.99
Total params		2,594,404	
Trainable params		2,593,372	
Non-trainable params		1,032	

recurrent neural networks, address issues of overfitting and gradient vanishing (Farooque et al. 2021). However, their classification performance is limited due to the inability to handle spatiotemporal relations in the data. To overcome this, Shi et al. (2015) proposed Convolutional LSTM (ConvLSTM), extending LSTM operations to convolutions to process both spatial and spectral correlations and model intrinsic relationships within the data.

The proposed ConvLSTM integrates synthetic data generated by the GAN with real data in three different combinations, discussed later in this study. Initially, the high-dimensional HSI cube undergoes dimensionality reduction via Principal Component Analysis (PCA), reducing the complexity of the bands. Each pixel i in the HSI cube, now containing K relevant bands, is padded to aid kernel processing. The input HSI cube (145 * 145 * 200) is divided into smaller cubes (9, 9, 3), representing the window length, image width, and image height.

This pre-processed data, as K spectral bands per pixel i, passes through time-distributed 2D convolutional layers to address the spectral dimension. The first TD-Conv2D layer processes a cube of 9 * 9 * 3 with a batch size of 128, a kernel size of 11, and a stride of 4. The second TD-Conv2D layer processes the output of the first with a batch size of 64, a kernel size of 5, and a stride of 2. Batch normalization precedes each Conv2D layer to reduce computational complexity.

Figure 2 shows three cascaded ConvLSTM2D layers with a kernel size of 3. Each spectral sequence element $i_1, i_2, \ldots, i_k$ is input to the ConvLSTM2D memory cell, producing two outputs: $H_{i1}, H_{i2}, \ldots, H_{ik}$ from k memory cells, and the state and output C_i, H_i representing long-term dependencies. The final ConvLSTM2D layer outputs the desired spatial-spectral features by modelling long-range spectral dependencies, producing H_i. This output is flattened and fed to a Softmax-activated dense layer for final classification. The Adagrad optimizer and sparse categorical cross-entropy loss are used to optimize the model.

$$P\,(y = c \mid h_i, W, b) = \frac{e^{(W_c h_i + b_c)}}{\sum_{j=1}^{N} e^{(W_j h_i + b_j)}} \tag{13}$$

Spherical interpolation

The proposed network uses the phenomenon of spherical interpolation to explore the spectra between two noise vectors using two arbitrary points in latent space. The generated mean

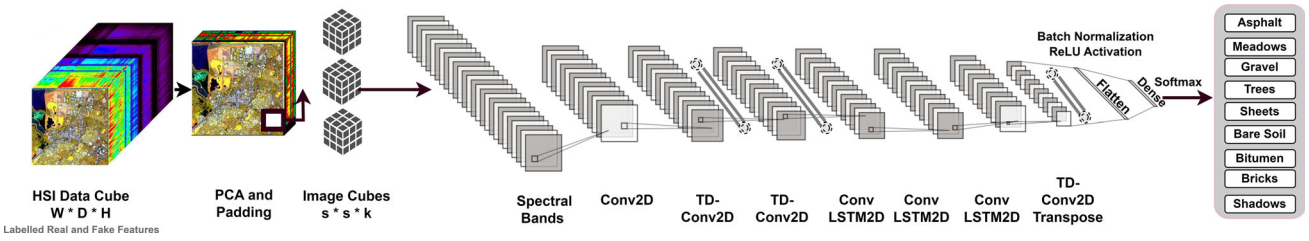


Fig. 2 Proposed convolutional LSTM

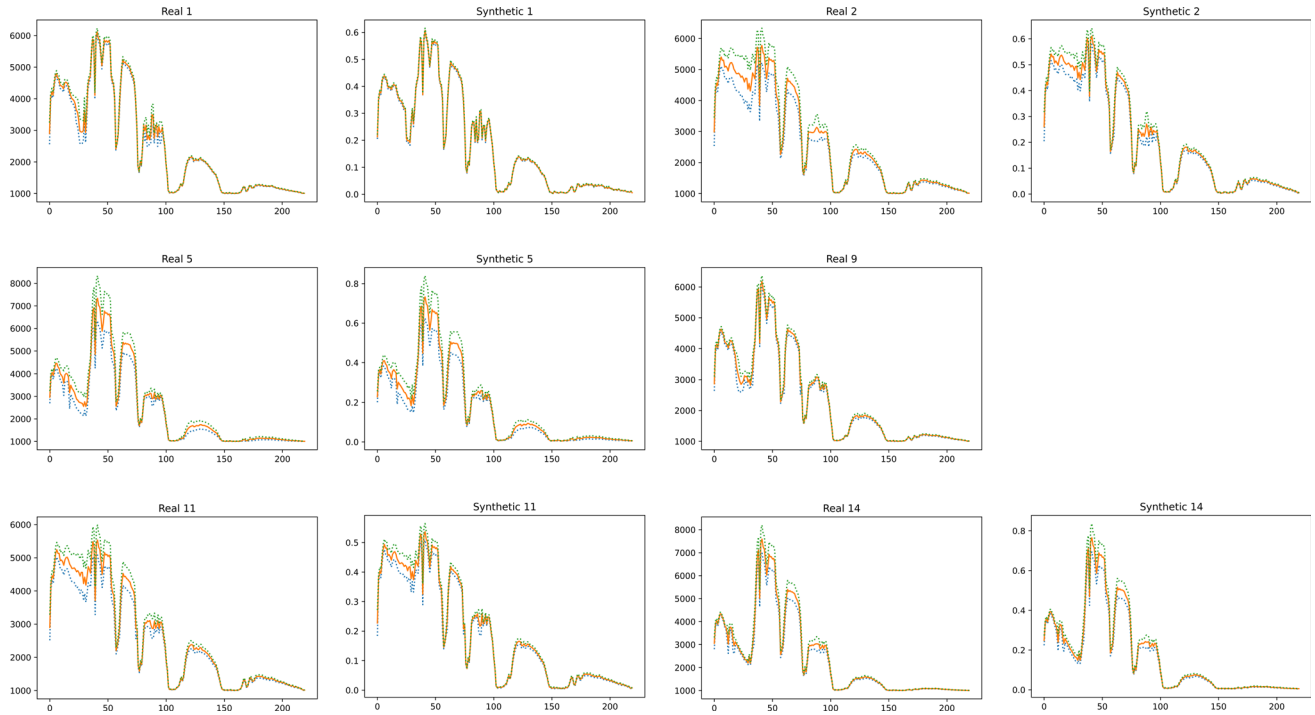


Fig. 3 Spectra of real vs generated samples for the Indian Pines dataset

spectra of synthesised samples belonging to a specific class for IP, PU and BOT is shown in Figs. 3, 4 and 5 respectively.

The average spectral curve is the middle line in the centre, with the standard deviation curve above and below. The graph shows that the trend of the generated sample's syn-

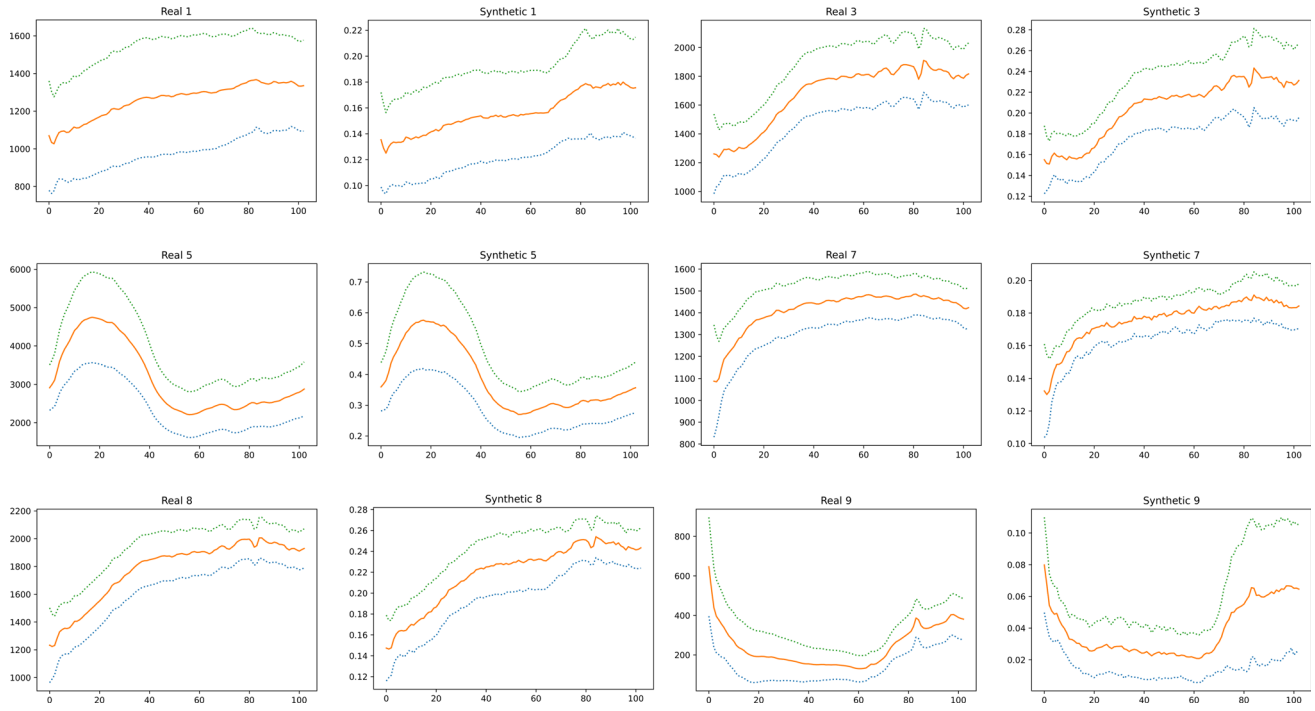


Fig. 4 Spectra of real vs generated samples for the Pavia University dataset

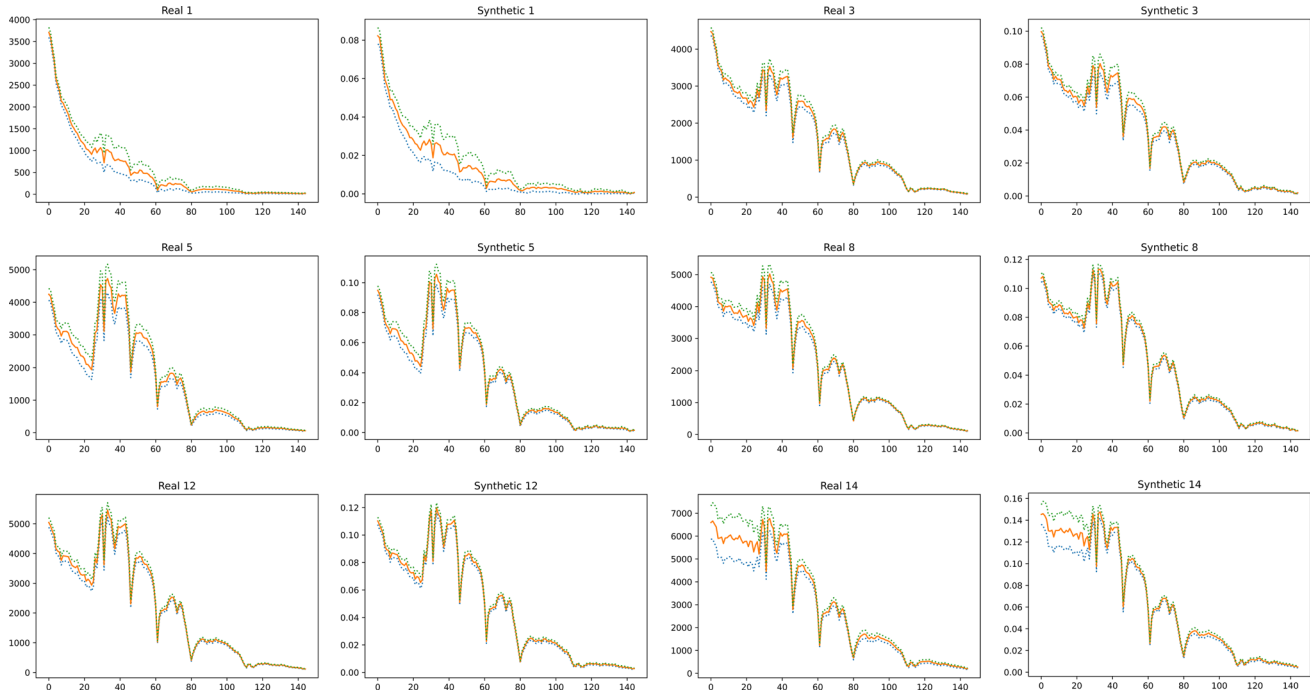


Fig. 5 Spectra of real vs generated samples for the Botswana dataset

thetic spectral curve is similar to that of the original samples. Koumoutsou et al. (2022) suggests spherical interpolation (SLERP) for hyperspectral samples that was first introduced by Shoemake (1985). SLERP is a method for quaternion interpolation that goes well with the three-dimensional nature of hyperspectral data. SLERP is determined using sinusoids shown in Eq. 14 where I is the interpolated quaternion using quaternions I_1 and I_2, val is the interpolation coefficient, and

Table 4 Sample Distribution in IP and PU Datasets

Class	Indian Pines	IP-Samples	Class	Pavia University	PU-Samples
1	Alfalfa	46	1	Water	6631
2	Corn-notill	1428	2	Trees	18649
3	Corn-mintill	830	3	Asphalt	2099
4	Corn	237	4	Self-Blocking Bricks	3064
5	Grass-pasture	483	5	Bitumen	1345
6	Grass-trees	730	6	Tiles	5029
7	Grass-pasture-mowed	28	7	Shadows	1330
8	Hay-windrowed	478	8	Meadows	3682
9	Oats	20	9	Bare Soil	947
10	Soybean-notill	972			
11	Soybean-mintill	2455			
12	Soybean-clean	593			
13	Wheat	205			
14	Woods	1265			
15	Buildings-Grass-Trees-Drives	386			
16	Stone-Steel-Towers	93			

Table 5 Three types of combinations considered for experimentation where TRS is training samples, TES is testing samples, RE is real samples, SYNS is synthetic samples

	TRS	TES	TRS		TES	
			RE	SYNS	RE	SYNS
C1: (RE=1, SYNS=0 | RE=1, SYNS=0)	RE	RE	0.9	0	0.1	0
C2: (RE=0, SYNS=1 | RE=1, SYNS=0)	SYNS	RE	0	1	1	0
C3: (RE=0.1, SYNS=1 | RE=0.9, SYNS=0)	SYNS+0.1 RE	RE	0.1	1	0.9	0

ω is the angular distance between given quaternions.

$$I = \left[\left(\frac{\sin((1-val)\cdot\omega)}{\sin(\omega)}\cdot I_1\right) + \left(\frac{\sin(val\cdot\omega)}{\sin(\omega)}\cdot I_2\right)\right] \tag{14}$$

Experimental configuration

This section includes datasets in section "Introduction of datasets" and parameter design in Section "Parameter design".

Introduction of datasets

For experimental purposes, two benchmark datasets were chosen. The first chosen dataset was Pavia U having aerial scenes of urban spatially distributed at 1.3m with 103 spectral bands having 9 categories of data. The second dataset chosen dataset was agricultural Indian Pines with a spatial distribution of 20m and 224 spectral bands having 16 categories of data. Both datasets have different ground features and different spatial and spectral characteristics. The image counts for each IP and PU dataset are shown in Table 4. It depicts the existing samples in IP and PU dataset. Table 5 gives a brief of three different combinations taken for experimentation on proposed model.

Figure 6 depicts the composite and ground truth image for the IP, composite and ground truth image for the PU dataset.

To experiment with different proportions of training and testing samples of real and synthetic data, three combinations are shown in detail in Tables 6, 7 and 8.

Parameter design

The model required was built on a system environment with a generator, discriminator and classifier. The preference for hyperparameters directly impacts the training effect and model performance. The proposed fully connected network with 512 neurons was followed by Leaky ReLU non-linearity. A sigmoid function followed the layers in the generator to produce smooth 0 or 1 outputs. However, The generator's final output is a vector with a length equal to the expected spectral bands. The classifier generates output based on the number of categories in the inputted dataset, and the discriminator determines whether it is real or fake. All three components are optimised using RMSprop stochastic gradient descent scheme. The batch size was kept as 256 for a total of 50 epochs, initial noise dimension z as 30, hidden layer size as 512, discriminator training steps were 5 times more than generator training steps, the learning rate was 5e-5, auxiliary classifier weight was set to 0.2 and proportion of label flipping was 0.0, mixing up ratio of generated and real samples alpha was set to 0.1 and semi-supervised set to include unlabelled samples. The proposed model was optimised using Adadelta, and the loss function was sparse categorical cross-entropy.

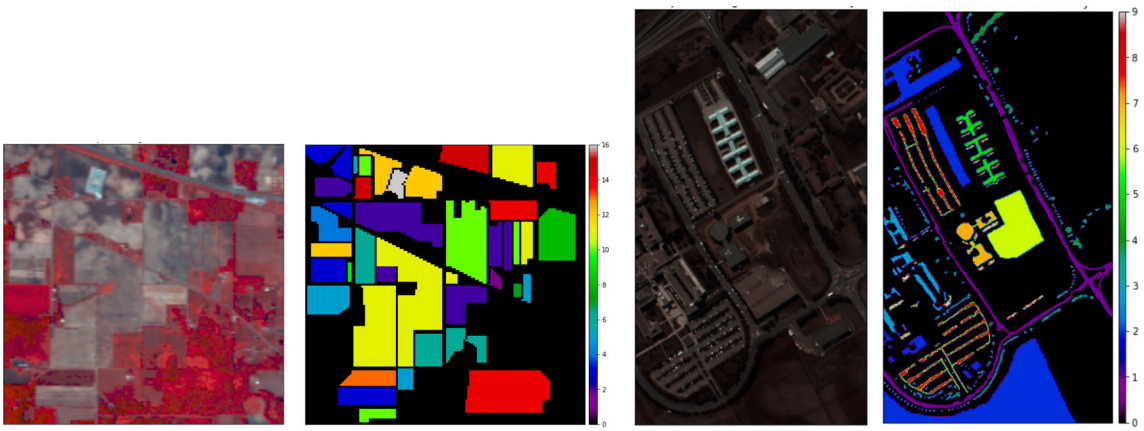


Fig. 6 Composite and ground truth for Indian Pines and Pavia University

Table 6 Combination 1

	PU				IP				BOT			
	RE	SYNS	TRSR	TESR	RER	SYNSS	TRSR	TESR	RER	SYNSS	TRSR	TESR
1	6631	6631	5968	663	46	46	41	5	270	270	243	27
2	18649	18649	16784	1865	1428	1428	1285	143	101	101	91	10
3	2099	2099	1889	210	830	830	747	83	251	251	226	25
4	3064	3064	2758	306	237	237	213	24	215	215	194	22
5	1345	1345	1211	135	483	483	435	48	269	269	242	27
6	5029	5029	4526	503	730	730	657	73	269	269	242	27
7	1330	1330	1197	133	28	28	25	3	259	259	233	26
8	3682	3682	3314	368	478	478	430	48	203	203	183	20
9	947	947	852	95	20	20	18	2	314	314	283	31
10					972	972	875	97	248	248	223	25
11					2455	2455	2209	246	305	305	275	31
12					593	593	534	59	181	181	163	18
13					205	205	184	21	268	268	241	27
14					1265	1265	1138	127	95	95	86	10
15					386	386	347	39				
16					93	93	84	9				

The rationale behind the selection of datasets and hyperparameters was to ensure a robust evaluation of the proposed model across varying data types and conditions. The Pavia University and Indian Pines datasets were chosen due to their distinct characteristics, providing a diverse testing ground for the model's capabilities in different scenarios. The specific hyperparameters, such as the batch size, learning rate, and auxiliary classifier weight, were selected based on preliminary experiments and literature reviews to balance training stability and model performance. The use of RMSprop for optimization helps manage the learning rate dynamically, which is crucial for handling the complex nature of hyperspectral data. The inclusion of unlabeled samples in a semi-supervised manner was intended to leverage additional data for training, thereby improving the model's generalization capability.

Table 7 Combination 2

Class	PU				IP				BOT			
	RE	SYNS	TRSS	TESR	RER	SYNSS	TRSS	TESR	RER	SYNSS	TRSS	TESR
1	6631	6631	6631	6631	46	46	46	46	270	270	270	270
2	18649	18649	18649	18649	1428	1428	1428	1428	101	101	101	101
3	2099	2099	2099	2099	830	830	830	830	251	251	251	251
4	3064	3064	3064	3064	237	237	237	237	215	215	215	215
5	1345	1345	1345	1345	483	483	483	483	269	269	269	269
6	5029	5029	5029	5029	730	730	730	730	269	269	269	269
7	1330	1330	1330	1330	28	28	28	28	259	259	259	259
8	3682	3682	3682	3682	478	478	478	478	203	203	203	203
9	947	947	947	947	20	20	20	20	314	314	314	314
10					972	972	972	972	248	248	248	248
11					2455	2455	2455	2455	305	305	305	305
12					593	593	593	593	181	181	181	181
13					205	205	205	205	268	268	268	268
14					1265	1265	1265	1265	95	95	95	95
15					386	386	386	386				
16					93	93	93	93				

Table 8 Combination 3

Class	PU				IP				BOT			
	RE	SYNS	TRS S+0.1R	TES 0.9R	RE RE	SYNS SYNS	TRS S+0.1R	TES 0.9R	RE RE	SYNS SYNS	TRS S+0.1R	TES 0.9R
1	6631	6631	7294	5968	46	46	51	41	270	270	297	243
2	18649	18649	20514	16784	1428	1428	1571	1285	101	101	111	91
3	2099	2099	2309	1889	830	830	913	747	251	251	276	226
4	3064	3064	3370	2758	237	237	261	213	215	215	237	194
5	1345	1345	1480	1211	483	483	531	435	269	269	296	242
6	5029	5029	5532	4526	730	730	803	657	269	269	296	242
7	1330	1330	1463	1197	28	28	31	25	259	259	285	233
8	3682	3682	4050	3314	478	478	526	430	203	203	223	183
9	947	947	1042	852	20	20	22	18	314	314	345	283
10					972	972	1069	875	248	248	273	223
11					2455	2455	2701	2210	305	305	336	275
12					593	593	652	534	181	181	199	163
13					205	205	226	185	268	268	295	241
14					1265	1265	1392	1139	95	95	105	86
15					386	386	425	347				
16					93	93	102	84				

Results and discussion

This section includes results in section"Pre-processing and PCA on generated vs synthetic samples" and section"Discussion on performance classification".

Pre-processing and PCA on generated vs synthetic samples

As a pre-processing technique, principal component analysis was performed to reduce the dimensions of datasets. Figures 3, 4 and 5 depicts the real spectra and generated spectra using the proposed GAN are mapped using principal component analysis (PCA) for three different datasets. The image on the right shows PCA on fake samples, and it was seen that synthesised samples diligently produce clusters formed by different categories. Most clusters were captured successfully with minor distortions. Figure 7 shows the PCA performed on real vs synthetic samples for IP (top), PU (middle) and BOT (bottom) datasets.

Discussion on performance classification

Experiments were carried out to assess the performance of the proposed model in categorising hyperspectral samples. Tables 6, 7 and 8 shows how the dataset and generated synthetic samples were divided into three combinations. The classification performance of the proposed model was evaluated using the statistics Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient (KC). The OA was calculated by dividing the number of correctly identified hyperspectral samples by the total number of samples in the test. Other validating metrics included the AA and KC for each category in the dataset. Three benchmark datasets were chosen, including Pavia University, Indian Pines and Botswana and synthetic samples were generated using the proposed model for each of the chosen datasets. After generating synthetic samples, three combinations were formed, and each combination, including real data, synthetic data and real plus synthetic data, was passed into the proposed Conv LSTM. Second, we compare the model's performance by comparing the time required to train and test the dataset using the proposed model.

Pavia U

To assess the performance of the Pavia University dataset, existing GAN-based models that included the Pavia U dataset were compared with the proposed model, as shown in Table 9. These included Capsule assisted triple GAN network by Wang et al. (2019), improved Wasserstein GAN by Sun and Bourennane (2020), spectral-spatial deep convolutional LSTM by Hu et al. (2020), Dual Channel Capsule GAN by Wang et al. (2021), Adaptive Drop block enhanced GAN by Wang et al. (2020), Collaborative learning and attention network inspired GAN by Feng et al. (2020), Wasserstein GAN using progressive GAN by Zhang et al. (2020), Auxiliary generative capsule GAN by Xue (2020), Spectral

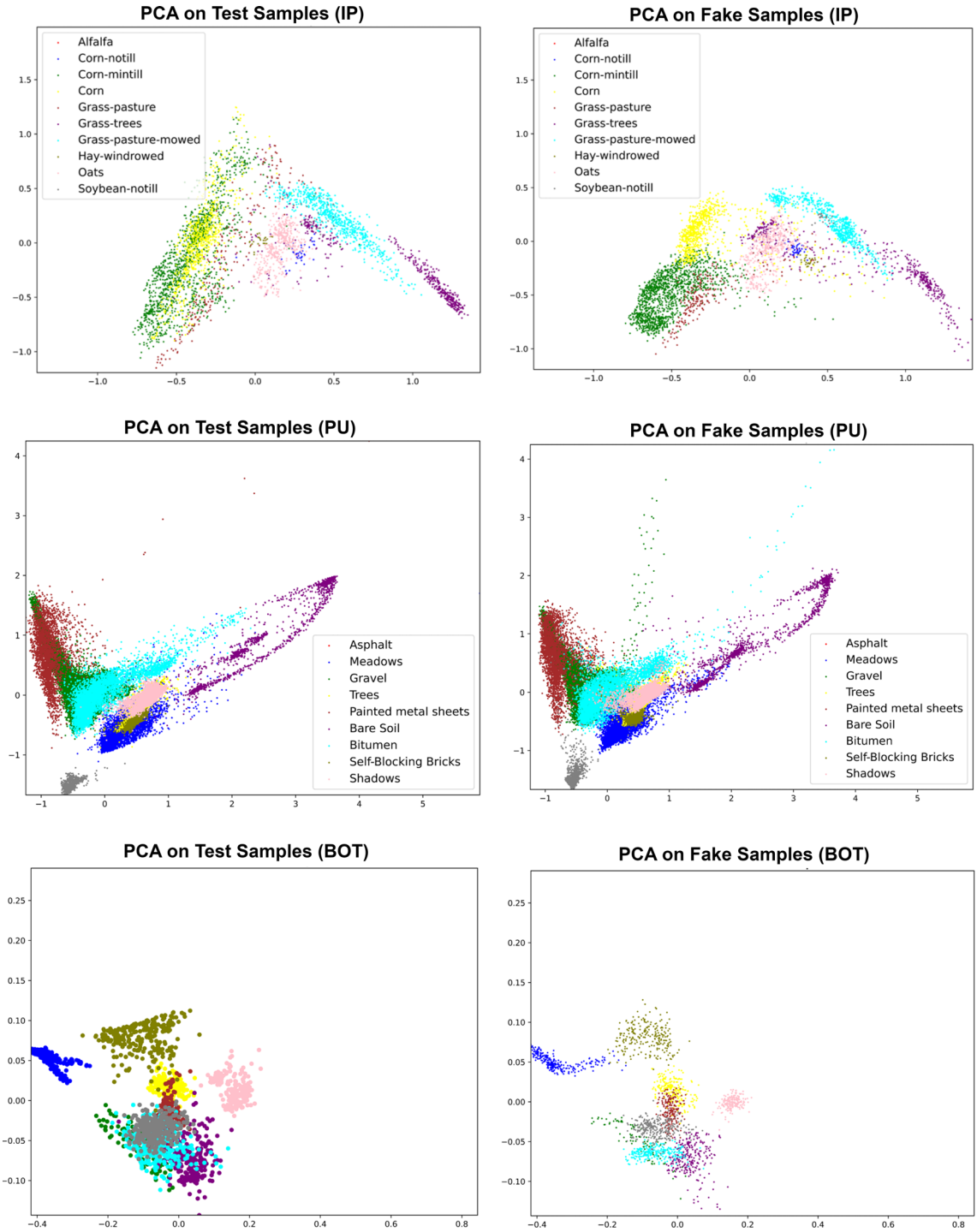


Fig. 7 PCA on Real and generated spectra (Top) Indian Pines (Middle) Pavia University (Bottom) Botswana

spatial 3D convolutional LSTM by Farooque et al. (2021) and Dual Channel Capsule GAN by Wang et al. (2021). Where AU-Caps-GAN and SSCRN achieved third and second highest classification performance with 99.23 and 99.31 percent overall accuracy, the proposed model achieved 99.50 as the overall accuracy that was highest among all. While comparing the three different combinations, it was seen that the first combination, which included 90% training and 10% testing on real datasets fed into the proposed Conv LSTM, yielded 75.45 as OA. However, using combination 2 when the network was trained with generated synthetic data and tested using real data improved the scores exponentially yielding

Table 9 Classification performance on Pavia U dataset

Class	Caps-Triple GAN Wang et al. (2019)	DR-WGAN-GP Sun and Bourennane (2020)	SS-CL3DN Hu et al. (2020)	DC-Caps-GAN Wang et al. (2021)	AD-GAN Wang et al. (2020)	CA-GA Feng et al. (2020)	WGAN-GP-PG-GAN Zhang et al. (2020)	AU-Caps-GAN Xue (2020)	SSCRN Farooque et al. (2021)	DC-Caps GAN Wang et al. (2021)	C1	C2	C3
1			69.89	97.80	92.36	99.1	99.28	99.44	99.88	97.80	85.91	85.52	99.63
2			82.44	99.74	96.18	99.9	100.00	99.32	99.92	99.74	99.52	97.64	99.93
3			65.10	93.49	91.49	99.2	96.31	99.52	98.71	93.49	36.94	47.61	99.38
4			90.48	99.42	97.29	97.0	96.95	99.36	99.52	99.42	92.73	82.02	99.64
5			94.83	100.00	98.54	100.0	99.37	99.35	100.0	100.0	93.10	94.81	100.00
6			82.37	98.12	96.37	99.6	97.92	99.53	96.58	98.12	26.30	93.63	99.98
7			89.32	97.98	88.40	99.8	93.15	99.34	100.0	97.98	02.14	43.60	99.77
8			84.71	96.83	95.75	97.5	99.14	97.24	98.61	96.83	55.93	73.09	99.86
9			86.91	99.85	96.02	96.5	97.23	99.51	100.0	99.85	67.65	19.05	99.89
OA	96.31	98.06	81.11	97.98	95.38	*99.2*	98.91	*99.23*	*99.31*	97.98	75.45	98.20	99.50
AA		97.32	82.90	98.01	94.71	98.6	97.59	99.14	99.24	98.01	62.25	70.77	99.78
KC	95.11	97.12	75.65	97.32	95.21	99.2	98.07	99.16	99.09	97.32	66.78	97.62	99.39

98.20 as OA. In the last combination, C3 showed best results when a generous amount of real data (10%) was added to the synthetic data (100%) and training was performed using remaining 90% real data yielding an OA of 99.50.

Indian pines

As shown in Table 10, The performance of the proposed model on the Indian Pines dataset was compared with existing GAN-based models. These included Capsule assisted triple GAN network by Wang et al. (2019), improved Wasserstein GAN by Sun and Bourennane (2020), Three-dimensional GAN by Wang et al. (2020), Spectral spatial 3D convolutional LSTM (Farooque et al. 2021), Multitask GAN by Hang et al. (2020), Wasserstein GAN by Hang et al. (2020), Adaptive Drop block enhanced GAN by Wang et al. (2020), Collaborative learning and attention network inspired GAN (Feng et al. 2020), One-dimensional GAN for hyperspectral classification by Feng et al. (2020). CA-GAN achieved the highest accuracy, with 97.4 as OA. The proposed work achieved 96.07 as OA for the Indian Pines dataset, which was near the highest accuracy. AD-GAN achieved the highest average accuracy; the second highest was 96.53 in the proposed work. Looking closer at the results of the proposed model for three different combinations, it was interesting to see that the proposed model achieved the lowest results when trained and tested with real data. After training the network with generated synthetic samples using the proposed GAN, the overall accuracy was drastically increased to 95.86. However, the best performance was achieved when the network was trained with synthetic samples and very few real samples. This combination was when the network yielded 96.07 overall accuracies.

Botswana

As shown in Table 11, third dataset chosen was Botswana. This particular dataset was chosen because there are few GAN-based studies on this dataset compared to other datasets like Indian Pines, Salinas or Pavia University. For the Botswana dataset, the proposed model's performance was compared with spectral-spatial GAN (Audebert et al. 2018), a model by UDA (Dhok et al. 2019), LSTM (Roy et al. 2021), and 3D-Hypergamo (Roy et al. 2021). 3D-Hypergampo achieved the highest accuracy with 97.43 as OA. The proposed model achieved the second-highest accuracy at 91.70. While analysing the results for three different combinations, C1 achieved the lowest results with the proposed convolutional LSTM due to the limited data available. Surprisingly, when the network was trained with all generated synthetic data and trained on real datasets, it performed exceptionally well, the OA was drastically increased to 90.20 from mere 31.25. This may be because Conv LSTM needs more data to train the network and due to ample number of testing samples, the accuracy was increased. Third combination just like other two datasets experimented on, produced best results with 91.70 as OA.

Overall analysis

The proposed model demonstrates strong performance across multiple hyperspectral datasets, consistently ranking among the top-performing methods. For the Pavia University dataset, our approach achieves the highest overall accuracy (99.50%) among SOTA GAN-based models, showcasing its effectiveness on this particular data. The model's performance on the Indian Pines dataset (96.07% OA) is comparable to the best-performing method, while for the Botswana dataset, it achieves the second-highest accuracy (91.70%). The observed performance pattern across datasets - highest on Pavia U, followed by Indian Pines, then Botswana - reflects the complex interplay of factors inherent to hyperspectral data analysis. Pavia U, an urban scene with well-defined spectral signatures, likely benefits from clear spectral separability between classes. Indian Pines, an agricultural scene, presents more challenges due to its mixed vegetation classes with potentially similar spectral profiles. The Botswana dataset, capturing a diverse semi-arid landscape, likely poses the greatest challenge due to subtle spectral differences between various savanna and woodland classes. These performance variations underscore the importance of considering not just the quantity of spectral bands, but also the nature of the spectral information they convey. The model's differing effectiveness across these scenes highlights the ongoing challenge in hyperspectral image analysis of developing methods that can generalize well across diverse landscapes and spectral characteristics.

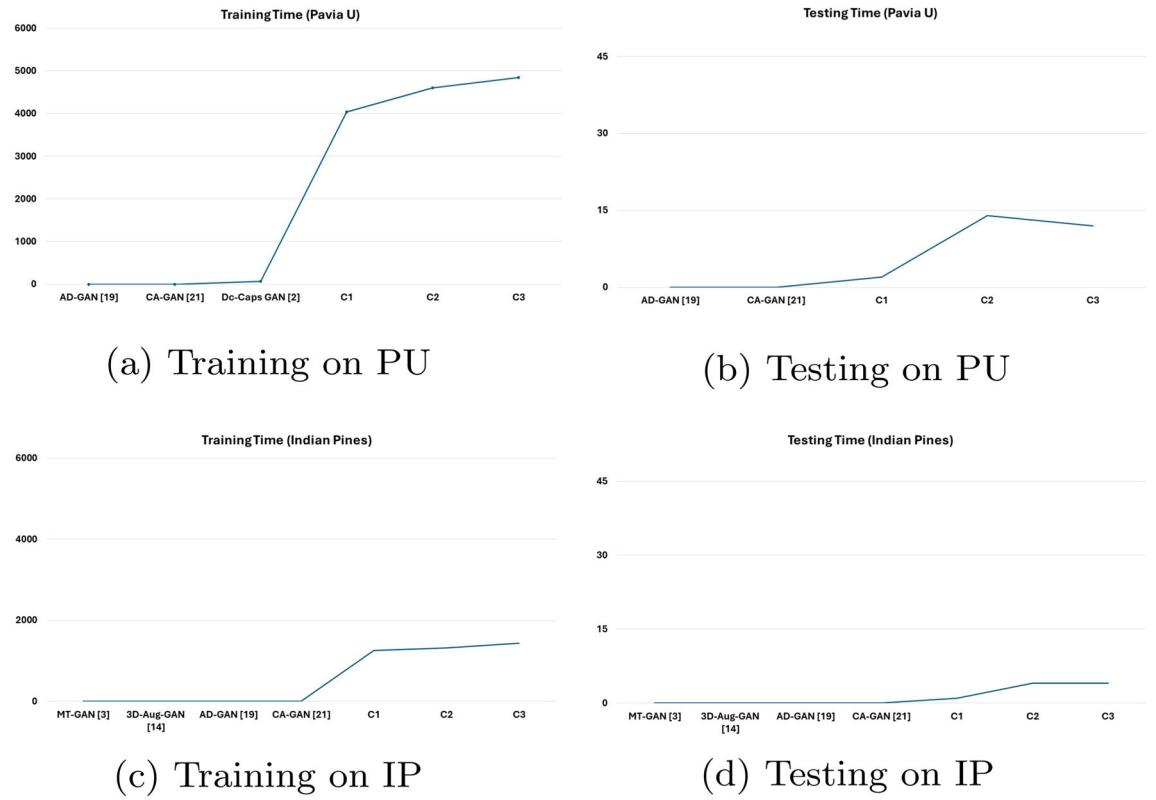
A key strength of our approach is its ability to leverage synthetic data to improve classification results, as evidenced by the significant performance gains when incorporating generated samples. This synthetic data augmentation strategy appears particularly effective for addressing limited training data scenarios. However, the model's performance on very small classes (e.g., class 3 in Indian Pines) and its slightly lower accuracy on the Botswana dataset compared to the best method indicate areas for potential improvement. Additionally, the observed variations in performance across different classes within each dataset suggest that certain land cover types or spectral signatures may be more amenable to our approach than others. Our model's strong performance across the diverse Pavia University, Indian Pines, and Botswana datasets demonstrates its promising potential for generalizability in hyperspectral image classification. The model's ability to excel in urban environments (Pavia U), perform well in complex agricultural scenes (Indian Pines),

Table 10 Classification performance on Indian Pines dataset

Class	Caps-Triple GAN Wang et al. (2019)	DR-WGAN-GP Sun and Bourennane (2020)	3D-GAN Wang et al. (2020)	SS-CL3DNN Hu et al. (2020)	MT-GAN Hang et al. (2020)	WGAN Hang et al. (2020)	AD-GAN Wang et al. (2020)	CA-GAN Feng et al. (2020)	HSGAN Feng et al. (2020)	C1	C2	C3
1			30.21	100.00	99.31	93.43	93.46	95.5	17.70	00.00	100.00	100.00
2			81.79	53.30	99.90	99.97	97.38	96.4	66.30	19.55	99.71	100.00
3			75.93	50.37	100.00	97.59	99.82	96.5	60.20	02.11	68.67	73.49
4			90.08	88.63	81.20	83.42	95.65	95.0	57.80	03.03	99.15	100.00
5			86.39	76.58	91.25	86.61	98.37	96.1	82.00	12.85	99.79	100.00
6			93.28	90.17	97.27	95.86	95.63	99.6	94.30	09.35	100.00	98.63
7			40.71	100.00	74.63	68.42	98.06	94.9	23.80	00.00	100.00	100.00
8			98.11	100.00	86.57	29.73	99.54	100.00	98.80	09.32	100.00	100.00
9			20.00	100.00	88.83	88.24	95.80	55.7	13.70	44.82	90.00	100.00
10			74.28	75.36	94.56	82.04	97.83	98.6	68.50	07.16	89.71	93.81
11			91.12	59.66	92.53	79.19	98.71	99.7	79.70	00.84	99.75	100.00
12			84.99	64.97	100.00	99.34	96.24	92.3	48.80	74.17	100.00	100.00
13			49.75	99.79	99.72	97.94	97.08	97.9	89.20	48.23	91.70	85.00
14			94.38	88.29	100.00	98.36	97.96	98.7	96.00	74.17	94.70	93.65
15			94.47	87.66	99.90	98.62	98.79	92.3	37.90	48.43	100.00	100.00
16			88.22	100.00	12.75	9.96	96.53	98.9	73.00	00.00	100.00	100.00
OA	90.02	95.80	89.09	71.28	92.42	89.97	97.23	97.4	74.00	29.26	95.86	96.07
AA		95.01	83.14	83.42	88.65	81.79	97.30	95.2	60.20	22.13	95.82	96.53
KC	89.24	94.58	87.32	67.83	90.70	90.70	96.57	97.0	70.00	17.94	90.24	95.53

Table 11 Classification performance on Botswana dataset

	SSGAN	Noname	UDA	LSTM	3D-HYPERGAMO			
Class	Audebert et al. (2018)	Dhok et al. (2019)	Pande et al. (2019)	Roy et al. (2021)	Roy et al. (2021)	C1	C2	C3
1		99.26	61.0	99.48	99.31	39.95	94.81	92.59
2		88.12	92.0	82.99	97.04	00.00	98.02	100.00
3		91.63	74.0	97.06	98.45	00.39	31.07	28.00
4		90.70	76.0	95.26	91.02	00.00	77.67	86.36
5		81.40	75.0	77.29	98.04	00.00	97.02	100.00
6		75.46	70.0	71.84	100.0	00.00	90.33	88.89
7		96.53	100.0	95.66	99.72	00.00	99.22	100.00
8		94.09	81.0	94.13	88.80	00.00	97.04	100.00
9		85.03	50.0	79.31	93.23	07.12	99.04	96.77
10		85.48	71.0	83.62	99.19	00.00	97.58	96.00
11		92.46	74.0	92.53	100.0	85.77	90.82	96.77
12		92.27	79.0	93.80	100.0	95.88	100.00	100.00
13		90.30	73.0	83.53	99.89	78.33	99.25	100.00
14		95.79	77.0	95.19	98.99	54.67	100.00	100.00
OA	88.40	89.50	74.5	88.16	97.43	31.25	90.20	91.70
AA		89.89	–	88.69	97.40	25.87	90.85	91.81
KC		88.59	–	87.17	97.22	39.70	83.52	90.32



(a) Training on PU

(b) Testing on PU

(c) Training on IP

(d) Testing on IP

Fig. 8 Training Vs testing times

Table 12 Model parameters

	Pavia U	Indian Pines	Botswana
C1	1115466 ∼ 1.11 M	1158481 ∼ 1.16 M	1146191 ∼ 1.15 M
C2	1115466 ∼ 1.11 M	1158481 ∼ 1.16 M	1146191 ∼ 1.15 M
C3	1115466 ∼ 1.11 M	1158481 ∼ 1.16 M	1146191 ∼ 1.15 M

and effectively handle the challenges of semi-arid landscapes (Botswana) showcases its adaptability to various spectral and spatial characteristics. This versatility suggests that the model has captured fundamental features of hyperspectral data that may translate well to other datasets and domains. The consistently high accuracy across these different scenarios indicates a robust architecture capable of extracting meaningful information from diverse spectral signatures and spatial patterns. Future work can revolve around proposing models that train and test across sensors and datasets.

Computational time and parameters involved

The computational time taken by the model proposed is shown in Fig. 8. Figure 8(a) and (c) display the training time taken by PU and IP datasets, and Fig. 8(b) and (d) show the testing time taken by the model proposed for all three combinations. The computational efficiency of the proposed model presents both challenges and opportunities for future optimization. As shown in Fig. 8, the model exhibits significantly longer training times compared to existing methods, particularly for the Pavia University (PU) dataset. This increased training duration is likely due to the complex architecture of our model and the incorporation of synthetic data generation. However, it's noteworthy that the testing times remain comparable to other networks, indicating that once trained, our model can perform efficiently in practical applications. The variation in computing time across different data combinations (C1, C2, C3) underscores the impact of data volume and composition on computational requirements. While the current implementation prioritizes performance accuracy over training speed, there are several avenues for potential optimization in future work. These could include exploring techniques like model pruning, quantization, or leveraging distributed computing resources to accelerate the training process. Additionally, investigating the trade-offs between synthetic data generation, model complexity, and training time could yield insights into more balanced approaches. Despite the current training time limitations, the model's competitive testing speed and high accuracy suggest that with further optimization, it could become a powerful tool for real-world hyperspectral image classification tasks where both accuracy and inference speed are crucial. The parameters taken by the proposed model are shown in Table 12. Moreover, the proposed model alleviates overfitting, as evident in Fig. 9.

Conclusion

This study proposed a novel GAN-based data generator integrated with convolutional LSTM as a deep learning classifier framework for HSI classification, which deals with the problem of limited data in spectroscopy. The proposed GAN produces synthetic samples similar to real ones, improving classification accuracy. When fed into novel convolutional LSTM, the produced synthetic data gives excellent classification results by accessing the discriminant spectral-spatial features. The proposed ConvLSTM can process both spatial and spectral correlations in data to process the data in higher dimensions and model the intrinsic relationships between data. The proposed framework was evaluated and compared to state-of-the-art deep learning hyperspectral classification methods currently available. The results show that the proposed work can generate good synthetic data and accurately classify hyperspectral images. We intend to improve network training time in the future.

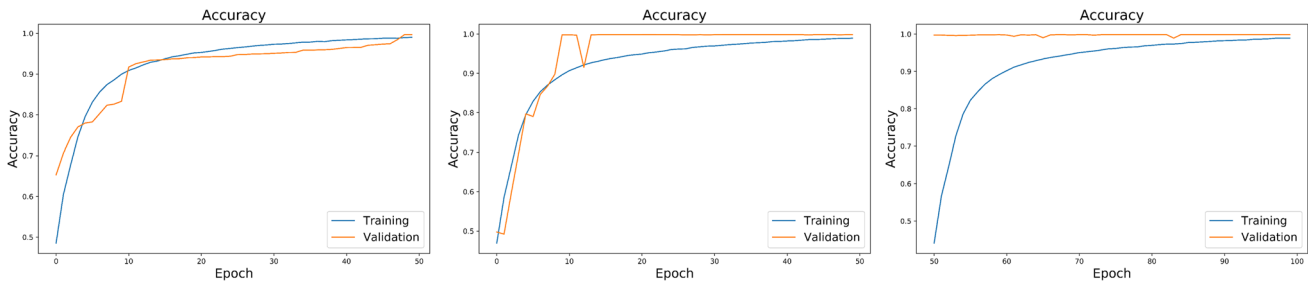


Fig. 9 Training vs validation time for all three datasets

Acknowledgements Thanks to the Data Processing Teams who have shared data on the website https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

Funding No funding was received for this study.

Data Availability The data was collected from https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

Declarations

Conflict of Interest/Competing Interests No conflict of interest.
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: Int Conf Mach Learn, pp 214–223. PMLR

Audebert N, Le Saux B, Lefèvre S (2018) Generative adversarial networks for realistic synthesis of hyperspectral samples. In: IGARSS 2018-2018 IEEE Int Geosci Remote Sens Symp, pp 4359–4362. IEEE

Bera S, Shrivastava VK (2020) Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. Int J Remote Sens 41(7):2664–2683

Chen Y, Jiang H, Li C, Jia X, Ghamisi P (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans Geosci Remote Sens 54(10):6232–6251

Chen L, Wei Y, Yao Z, Chen E, Zhang X (2022) Data augmentation in prototypical networks for forest tree species classification using airborne hyperspectral images. IEEE Trans Geosci Remote Sens 60:1–16

Dhok S, Bhurane A, Kothari A (2019) Automated hyperspectral image classification using spatial-spectral features. In: 2019 6th Int Conf Signal Process Integr Netw (SPIN), pp 184–189 (2019). IEEE

Du B, Zhang L, Tao D, Zhang D (2013) Unsupervised transfer learning for target detection from hyperspectral images. Neurocomputing 120:72–82

Farooque G, Xiao L, Yang J, Sargano AB (2021) Hyperspectral image classification via a novel spectral-spatial 3d convlstm-cnn. Remote Sensing. 13(21):4348

Feng J, Feng X, Chen J, Cao X, Zhang X, Jiao L, Yu T (2020) Generative adversarial networks based on collaborative learning and attention mechanism for hyperspectral image classification. Remote Sensing. 12(7):1149

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Adv Neural Inform Process Syst **27**

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Commun ACM 63(11):139–144

Hang R, Zhou F, Liu Q, Ghamisi P (2020) Classification of hyperspectral images via multitask generative adversarial networks. IEEE Trans Geosci Remote Sens 59(2):1424–1436

Haut JM, Paoletti ME, Plaza J, Plaza A, Li J (2019) Hyperspectral image classification using random occlusion data augmentation. IEEE Geosci Remote Sens Lett 16(11):1751–1755

Hu WS, Li HC, Pan L, Li W, Tao R, Du Q (2020) Spatial-spectral feature extraction via deep convlstm neural networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 58(6):4237–4250

Hu W, Huang Y, Wei L, Zhang F, Li H (2015) Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors 2015(1):258619

Jia S, Jiang S, Lin Z, Li N, Xu M, Yu S (2021) A survey: deep learning for hyperspectral image classification with few labeled samples. Neurocomputing 448:179–204

Koumoutsou D, Siolas G, Charou E, Stamou G (2022) Generative adversarial networks for data augmentation in hyperspectral image classification. In: Generative adversarial learning: architectures and applications, pp 115–144. Springer, ???

Liao C, Dong M (2022) Acwgan: an auxiliary classifier wasserstein gan-based oversampling approach for multi-class imbalanced learning. Int J Innov Comput Inf Control 18(3):703–721

Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784

Odena A, Olah C, Shlens J (2017) Conditional image synthesis with auxiliary classifier gans. In: Int Conf Mach Learn, pp 2642–2651. PMLR

Pande S, Banerjee B, Pižurica A (2019) Class reconstruction driven adversarial domain adaptation for hyperspectral image classification. In: Iberian conference on pattern recognition and image analysis, pp 472–484. Springer

Paul A, Bhoumik S (2022) Classification of hyperspectral imagery using spectrally partitioned hyperunet. Neural Comput Appl 34(3):2073–2082

Paul A, Bhoumik S, Chaki N (2021) Ssnet: an improved deep hybrid network for hyperspectral image classification. Neural Comput Appl 33:1575–1585

Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434

Ranjan P, Kumar R, Girdhar A: A 3d-convolutional-autoencoder embedded siamese-attention-network for classification of hyperspectral images. Neural Computing and Applications, 1–20 (2024)

Ranjan P, Girdhar A (2022) A comprehensive systematic review of deep learning methods for hyperspectral images classification. Int J Remote Sens 43(17):6221–6306

Ranjan P, Girdhar A (2022) Xcep-dense: a novel lightweight extreme inception model for hyperspectral image classification. Int J Remote Sens 43(14):5204–5230

Ranjan P, Girdhar A (2024) Deep siamese network with handcrafted feature extraction for hyperspectral image classification. Multimedia Tools and Applications. 83(1):2501–2526

Ranjan P, Gupta G (2023) A cross-domain semi-supervised zero-shot learning model for the classification of hyperspectral images. Journal of the Indian Society of Remote Sensing. 51(10):1991–2005

Roy SK, Haut JM, Paoletti ME, Dubey SR, Plaza A (2021) Generative adversarial minority oversampling for spectral-spatial hyperspectral image classification. IEEE Trans Geosci Remote Sens 60:1–15

Seydgar M, Alizadeh Naeini A, Zhang M, Li W, Satari M (2019) 3-d convolution-recurrent networks for spectral-spatial classification of hyperspectral images. Remote Sensing. 11(7):883

Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo Wc (2015) Convolutional lstm network: A machine learning approach for precipitation nowcasting. Adv Neural Inform Process Syst **28**

Shoemake K (1985) Animating rotation with quaternion curves. In: Proceedings of the 12th annual conference on computer graphics and interactive techniques, pp 245–254

Song A, Choi J, Han Y, Kim Y (2018) Change detection in hyperspectral images using recurrent 3d fully convolutional networks. Remote Sensing. 10(11):1827

Srivastava A, Valkov L, Russell C, Gutmann MU, Sutton C (2017) Veegan: reducing mode collapse in gans using implicit variational learning. Adv Neural Inform Process Syst **30** (2017)

Sun Q, Bourennane S (2020) Hyperspectral image classification with unsupervised feature extraction. Remote Sensing Letters. 11(5):475–484

Wambugu N, Chen Y, Xiao Z, Tan K, Wei M, Liu X, Li J (2021) Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: a review. Int J Appl Earth Observ Geoinform **105**:102603

Wang WY, Li HC, Deng YJ, Shao LY, Lu XQ, Du Q (2020) Generative adversarial capsule network with convlstm for hyperspectral image classification. IEEE Geosci Remote Sens Lett 18(3):523–527

Wang J, Gao F, Dong J, Du Q (2020) Adaptive dropblock-enhanced generative adversarial networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 59(6):5040–5053

Wang J, Guo S, Huang R, Li L, Zhang X, Jiao L (2021) Dual-channel capsule generation adversarial network for hyperspectral image classification. IEEE Trans Geosci Remote Sens 60:1–16

Wang X, Tan K, Du Q, Chen Y, Du P (2019) Caps-triplegan: Gan-assisted capsnet for hyperspectral image classification. IEEE Trans Geosci Remote Sens 57(9):7232–7245

Wu H, Prasad S (2017) Convolutional recurrent neural networks for hyperspectral data classification. Remote Sensing. 9(3):298

Xia X, Pan X, Li N, He X, Ma L, Zhang X, Ding N (2022) Gan-based anomaly detection: a review. Neurocomputing 493:497–535

Xue Z (2020) A general generative adversarial capsule network for hyperspectral image spectral-spatial classification. Remote Sens Lett 11(1):19–28

Zhang F, Bai J, Zhang J, Xiao Z, Pei C (2020) An optimized training method for gan-based hyperspectral image classification. IEEE Geosci Remote Sens Lett 18(10):1791–1795

Zhou F, Hang R, Liu Q, Yuan X (2019) Hyperspectral image classification using spectral-spatial lstms. Neurocomputing 328:39–47

Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE Int Conf Comput Vis, pp 2223–2232

Zhu L, Chen Y, Ghamisi P, Benediktsson JA (2018) Generative adversarial networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 56(9):5046–5063

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

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

Journal of Optics
https://doi.org/10.1007/s12596-024-02161-w

RESEARCH ARTICLE

A refractive index based cancer cells sensor in terahertz spectrum: design and analysis

Jaydeep Singh[1] · **Akash Khamaru**[1] · **Ajeet Kumar**[1]

Received: 29 May 2024 / Accepted: 12 August 2024
© The Author(s), under exclusive licence to The Optical Society of India 2024

Abstract
Refractive Index sensing in terahertz regime in one of the most popular applications of Photonic crystal fibers (PCFs). In this study, we presented a novel Photonic crystal fiber-based sensor for detecting various types of cancer cells presence in human blood. The PCF-based sensor designed for biosensing applications detects cancer cells like Jurkat, HeLa, Basal, MDA-MB-231, PC-12 and MCF-7 which are responsible for blood cancer, cervical cancer, skin cancer, breast cancer type I, adrenal gland cancer and breast cancer type II respectively. The optical parameters for this sensor were investigated and analyzed at different parameters to obtain the best results using finite element method (FEM) based COMSOL Multiphysics software. The proposed sensor resulted in highly significant results with highest relative sensitivity of 99.33%, very low confinement loss of 4.76×10^{-12} dB/m, 9.02×10^{-8} m^2 effective area, 7.80×10^{-9} $W^{-1}m^{-1}$ nonlinearity, etc. at 1.6 THz frequency optical signal. Changing the pitch size has an effect on overall optical properties. With these better results, the proposed sensor design can be utilized effectively in the medical industry for early diagnosis of cancer in the human body.

Keywords Biosensing · Cancer cells detector · Confinement loss · PCF based sensors · Relative sensitivity · Terahertz spectrum

Introduction

Cancer has always been an issue of global concern. According to the reports of the World Health Organization (WHO), around 10 million people have been died due to cancer in the year 2020 or nearly one in six deaths [1]. These numbers can be reduced by the early detection of the presence of cancerous cells in the human body. Cancer is caused due to the abandoned outgrowth of infected cells within the human body. This unrestricted growth of infected irregular cells causes the organ malfunctioning and consequently leads to death. It can happen due to the consequences of multiple phenomena like environmental exposures, alcohol and tobacco consumption, improper diet, obesity, exposure to high energy radiation, carcinogens, etc. Presently, it has become the second most common reason for death in the world [2]. There is no specific medicine in the medical field reported so far that can eradicate the cancer completely. The way that remains to treat cancer is its early and proper diagnosis. The optical behavior of the infected cells leads to the detection of cancer [3–5].

PCF based fiber optic sensors have been used widely in recent times because of their various specifications like flexibility, low losses, compact size, high sensitivity, low cost, fast response, etc. Because of these higher advantages over conventional optical fibers which are limited to varying refractive index at the core and cladding interface only, PCF based sensors have been widely used in many sensing application fields [6–9]. In PCFs, we can freely modify parameters like the shape and size of the core and cladding region, pitch size, orientation of air holes, materials assigned to the fiber material, etc. [10]. These modifications can lead to various enhancements in optical properties, such as increased relative sensitivity, reduced confinement and material losses, higher nonlinearities (relevant for nonlinear optical applications), and a more effective area for mode propagation. Consequently, the usage of PCFs had gone diverse in various applications of optical field like telecommunications, industrial and medical sensors [11–15]. Some other

✉ Ajeet Kumar
ajeetdph@dtu.ac.in

[1] Advanced Photonics Simulation Research Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi, Delhi 110042, India

Published online: 23 August 2024

effective applications are temperature sensors [16], pressure sensors [17], blood components sensors [18], salinity sensors [19], and sensors for diagnosing diseases [20]. Due to high sensitivity and applicability in biosensing applications [21], Bio sensors have played a crucial role in early detection of presence of cancerous cells in human blood [22].

Due to the large number of applications in genetic, organic area, spectroscopy, bio sensing, chemical sensing, astronomy, temperature sensing and communication fields, terahertz regime of EM wave spectrum has been the center of focus in recent years. The terahertz frequency band of electromagnetic radiations spectrum lies in the range of 0.1–10 THz which is equivalent to 30 µm-3 mm wavelength [23]. Sensing in terahertz regime attracted researchers in recent years and a lot of work has been reported using PCF based sensors in this field. With the recent technological advancement in the field, Researchers are recently using PCFs for cancer detection. In 2021, Parvin Textit et al. introduced a novel spectroscopic optical sensor for identifying cancer cells in several human body parts (adrenal gland, blood, breast, cervical and skin). It demonstrated the highest relative sensitivity (95.51%) for breast cancer at X-polarization [24]. In 2021, Habib et al. developed a hollow core PCF to swiftly detect malignant cells in the human body. Simulation experiments show that the proposed sensor operating at 2.5 THz achieves a high relative sensitivity of over 98% with minimal loss (0.025dB/cm) [22]. According to simulation and tests, the relative sensitivity of the recommended 2.5 THz sensor is amazingly high, more than 98% with low loss (0.025 dB/cm). Eid at el. developed a Mono Rectangular Core Photonic Crystal Fiber (MRC-PCF) for detecting skin and blood cancer, with the maximum sensitivity of 96.74% at 2 THz frequency operating signal [25]. In order to identify four different cancer cells- Jurkat (blood), Hela (cervical), MCF-7 (Breast cancer type II), Basal (skin), Bulbul at el. in 2022, created a hollow core photonic crystal fiber. In this hollow core PCF(HC-PCF), the four cancer cells showed a relative sensitivity ranging from 88.12 to 89.65% [26]. In 2022, a THz sensor that uses HC-PCF to identify breast cancer cells was shown in a research work by Yadav at el. The sensor demonstrated a relative sensitivity of 65.53% on the y-axis. It also had an effective area of $3.04\times10^{-20}\mathrm{m}^2$, a reduced confinement loss of $17.33\times10^{-9}\mathrm{cm}^{-1}$, and a high birefringence of 0.0020 [27].

In the present study, we have designed a PCF based sensor model works under terahertz regime. It is a rectangular core PCF based cancer cells detector. The foremost objective of the current study is to model a highly sensitive detection model for cancer cells. The fiber's background material assigned is Zeonex because of its low loss properties [28]. Then, the fabrication possibilities of the sensor are discussed. The analysis has been done using various performance metrics like relative sensitivity, confinement loss, nonlinearity, effective mode area, and power fraction. The numerical investigation illustrated that the proposed model resulted in extra ordinary relative sensitivity of order 99% for various cancer cells at 1.6 THz frequency optical signal. The confinement loss obtained for this work is also reported very less of order 10^{-12} dBm^{-1}.

Design of the proposed model

The illustration of the transverse cross-sectional view of the designed PCF based cancer cells detector is shown in Fig. 1 (a). The COMSOL Multiphysics software has been widely used for the designing of photonic crystal fibers for a varied range of applications. For designing the proposed sensor, we first assigned the global variables of the design that remain constant throughout the process. Secondly, the geometry of the proposed structure is defined with the assignment of the material to various sub-domains of the geometry. The

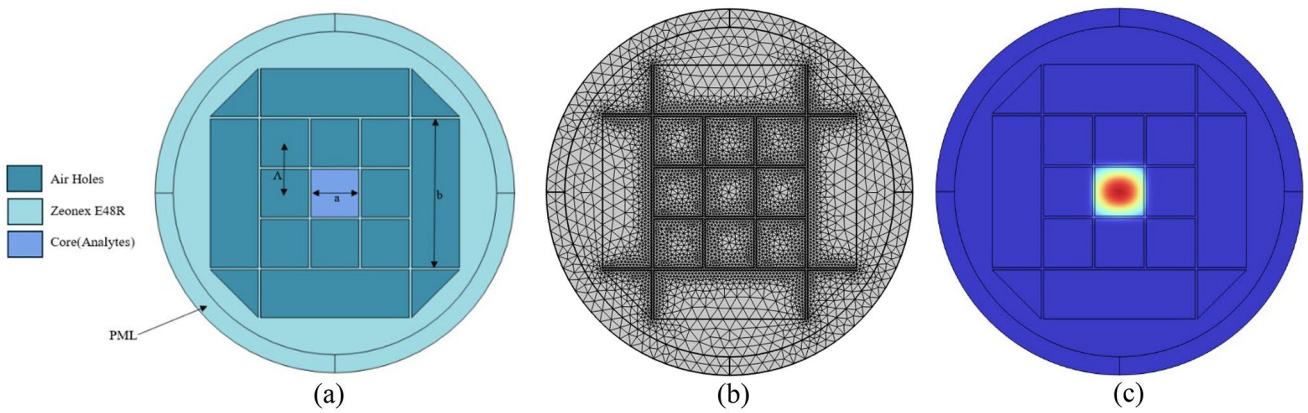


Fig. 1 (*a*) The cross-sectional perspective of the suggested PCF-based sensor model in a transverse view; and (*b*) a finer meshing output of the proposed sensor which consists of 9412 domain elements and 1270 boundary elements; *(c)* mode field distribution of light propagation through core at 2 THz operating frequency

boundary conditions for the structure were defined and a mesh setup was established later. Finally, the partial differential equations (PDE) solutions were obtained to observe and analyze the proposed structure of the PCF-based sensor using finite element method (FEM) of the software.

We have considered a PCF-based sensor in which the total fiber radius is taken as 1500 µm. For boundary conditions, the perfectly matched layer (PML) has a thickness of 10% of the radius i.e. 150 µm. The PML is used to absorb the electromagnetic signal propagating towards the surface [28]. In Fig. 1 (a), the sky-blue colored square represents the core of the fiber where analytes are to be injected. The side of the square shaped core denoted by (a) is 400 µm. The core is surrounded by 8 square shaped air holes of same dimensions i.e. 400 µm. The cladding region of the fiber has a total of 16 air holes in which 8 are square, 4 are rectangular and 4 are triangular. The width (a) and height (b) of the cladding rectangles are 400 µm and 1240 µm, respectively. The triangular shaped air holes are isosceles right-angled triangles with dimension 400 µm. Initially, the pitch (separation between two consecutive rectangles) is taken as 15 µm and then the optical properties were optimized at different pitch magnitudes. The separation between the centers of two adjacent squares is denoted as $\Lambda = 415$ µm. The dimensions of the air holes vary with the pitch of the fiber and were changed accordingly. A finer mesh is built to analyze the proposed PCF-based sensor. It consists of 9412 domain elements and 1270 boundary elements. Figure 1 (b) shows the meshing output of the proposed model. The light propagation through the core infiltrated with MCF-7 as analyte is illustrated in Fig. 1 (c) at an operating frequency f = 2 THz. Loss is minimal since just a tiny fraction of the light travels outside the core area. The red region represents the highest level of light confinement. As one approaches the core's boundary, the intensity steadily decreases.

Photonic crystal fibers can be fabricated using a variety of methods, including capillary stacking, stack and draw [29], drilling, extrusion [30], sol-gel [31], and 3D printing [32]. The methods of capillary stacking, stack and draw, and sol-gel manufacturing are perfect for creating circular air holes. The best method for creating rectangular PCF structures is 3D printing. Furthermore, for the fabrication of asymmetric PCF structure extrusion method and 3-D printing techniques are used. The advantage of 3D printing technology over traditional technology is its ability to prepare complex geometric shapes, which offer flexibility and diversity in the production process and allow for rapid prototyping and a high degree of freedom. With its promising development potential, 3D printing technology is predicted to become a commonplace tool for optical fiber preparation in the near future. Rectangular air holes have reportedly already been created by the Planck Institute [32, 33]. Thus, we are able to fabricate our suggested PCF structure using current fabrication technology.

Table 1 Tabular representation of cancer types along with their refractive index and cell name [34]

S. No	Cancer type	Cell type	*R*.I.
1	Skin Cancer	Basal	1.380
2	Blood Cancer	Jurkat	1.390
3	Cervical Cancer	HeLa	1.392
4	Adrenal gland Cancer	PC-12	1.395
5	Breast Cancer type I	MDA-MB-231	1.399
6	Breast Cancer type II	MCF-7	1.401

Analytical methodology

The proposed PCF-based sensor model is used for biosensing application, with various types of cancer cells as an analyte infiltrated within the core of the fiber for detection. The proposed model is a refractive index-based sensor which detects cancer cells due to the variation in their refractive indices. The refractive indices of these different cancer cells are shown in the Table 1 [34]. For this study, the operating frequency of the optical signal is kept in the range 1–2 THz. The optimization of the proposed sensor model was done for three different values of pitch earlier i.e. 10 µm, 15 µm and 20 µm. Due to the constricted and compressed air holes at 10 µm pitch size, no significant results were obtained as the core and cladding holes came very close to each other. Therefore, all the optical parameters were observed and analyzed at 15 µm and 20 µm pitch only. Due to the symmetrical nature of the square shaped core, the difference between the values of performance metrics at X-polarization and Y-polarization mode field distribution was negligible.

For numerical assessment, we have infiltrated the fiber's core with analyte. The analysis of the fiber is done for detecting cancer cells i.e. Basal, Jurkat, HeLa, PC-12, MDA-MB-231, and MCF-7. Analyzing electromagnetic wave propagation across optical medium is important for both communication and sensing applications. Several numerical tools have been proposed for this purpose, including the beam propagation method (BPM), finite integration methodology, and full vectorial finite element method (FV-FEM). Compared to other methods, FV-FEM is the most efficient. The optical parameters like Effective mode index (n_{eff}), Effective mode area (A_{eff}), Non-linearity, Relative sensitivity, Confinement loss (L_{conf}), and Power Fraction (PF) were determined at 15 µm and 20 µm values of pitch using COMSOL Multiphysics software which uses full-vectorial finite element (FEM) method.

Effective mode index (n_{eff}) is the overall refractive index experienced by the respective EM wave mode as it travels through the core of the fiber. It plays a crucial role

in determining the sensitivity of a sensor. PCFs' effective refractive index can be used to create sensors that are extremely sensitive to changes in the surrounding medium. It is feasible to develop sensors that are extremely sensitive to minute changes in the surrounding medium by carefully managing the fiber's characteristics. n_{eff} can be mathematically represented as [35]:

$$n_{eff} = \frac{\beta}{k_o} \tag{1}$$

A_{eff} is the effective area occupied by the propagating mode within the core. It determines the actual analyte detecting region of any optical sensor. This parameter also reflects the region available for a firm interaction of analyte with the light signal. It can be denoted in terms of transverse electric field intensity simply and can be evaluated as follow [36]:

$$A_{eff} = \frac{\left[\int I(r)\, dr\right]^2}{\left[\int I^2(r)\, dr\right]} \tag{2}$$

The effective area of the fiber highly influences the nonlinearity γ inversely which is also known as the non-linear coefficient. This is dependent on the intensity of the optical signal and indicates how well a fiber can contain intense light in its core. The nonlinear effects become significant when intensity of signal reaches very high value. Mathematically, it is defined as [35]:

$$\gamma = \frac{2\pi\, n_2}{\lambda\, A_{eff}} = \frac{2\pi\, f n_2}{c A_{eff}} \tag{3}$$

where f is the operating frequency of the optical signal and n_2 denotes the non-linear refractive index of fiber material which is Zeonex in the current study.

Since the optical signal's speed in the fiber's core is known to be influenced by the signal's refractive index, the analyte and the fiber's background material both affect how quickly the signal travels at an effective wavelength. The PCF-based sensor's effectiveness is significantly influenced by its relative sensitivity. Relative Sensitivity is the most important metric in the field of sensing applications, out of all the metrics computed to analyze the sensor model. The power carried by the optical signal in the fiber's core directly affects relative sensitivity. It is essentially the ratio of the analyte's refractive index (n_a) to the EM wave mode's effective refractive index (n_{eff}). The relative sensitivity is mathematically represented by the equation [18]:

$$Relative\ Sensitivity = \frac{n_a}{n_{eff}} \times PF \tag{4}$$

Here, the symbol PF denotes the power fraction which is defined as the ratio between the power of light enclosed in the core to the total power injected in the whole surface area of the fiber. Therefore, the power fraction is a unitless quantity and highly influences the relative sensitivity of a sensor in various sensing applications. It is represented by the equation given below [28].

$$PF = \frac{\int\limits_{analyte} R_e \left(E_x H_y - E_y H_x\right) dxdy}{\int\limits_{total} R_e \left(E_x H_y - E_y H_x\right) dxdy} \times 100 \tag{5}$$

Here, the power carried by mode is represented in terms of electric field (E) and magnetic field (H) components. The subscripts in the equation represent the polarization of modes in x and y direction.

While designing and analyzing a PCF-based sensor, our main aim is to get a highly sensitive optical sensor, but care must be taken that the losses occurring in the fiber should be minimum as well. The sensors having minimal losses are preferred in the real world. Confinement loss (L_{conf}), which is defined as the fiber's inability to contain the light wave signal inside the core region and leakage in cladding region, has a significant impact on the sensor's performance. The amount of wave that the cladding region captures is determined by its value. The volume of light inside the core region will increase with a lower value of L_{conf}. It is expressed mathematically as [35].

$$L_{conf} = 8.868 \times k_o I_m\left(n_{eff}\right) dBm^{-1} \tag{6}$$

Here, $k_o = \frac{2\pi}{\lambda}$, where λ is the wavelength of the operating signal and $I_m\left(n_{eff}\right)$ denotes the imaginary part of the effective refractive index experienced by the input signal.

Results and discussion

In this work, we have considered a various types of cancer cells i.e. Basal, Jurkat, HeLa, PC-12, MDA-MB-231, and MCF-7 responsible for different kinds of cancer as shown in the Table 1. The core of the fiber was infiltrated with these analytes and results were obtained for multiple performance metrics in the operating frequency range 1–2 THz using FEM based COMSOL Multiphysics software. These performance metrics were evaluated at different pitch sizes for optimization purposes. The results obtained are shown in graphical representations.

In Fig. 2 (a) and (b), the variation of effective mode index with the operating frequency is shown for 20 µm and 15 µm pitch respectively. The effective refractive index increases with the increasing operating frequency or with

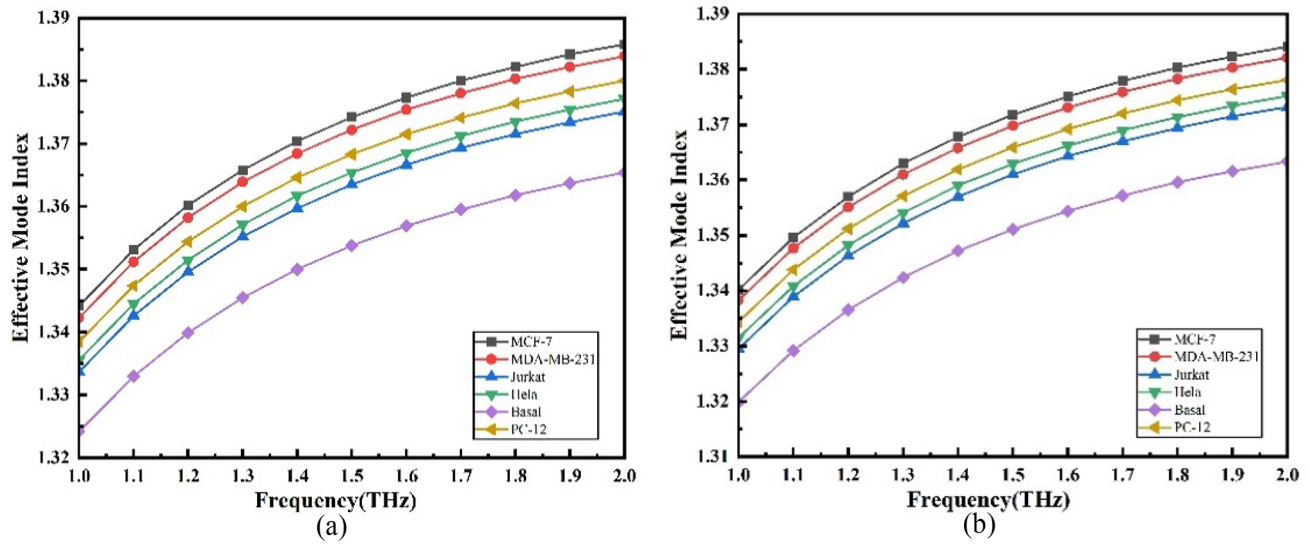


Fig. 2 Effective refractive index of this cancer cell detector for various cancer cells at the operating frequency range 1–2 THz at (**a**) 20 µm and (**b**) 15 µm pitch size

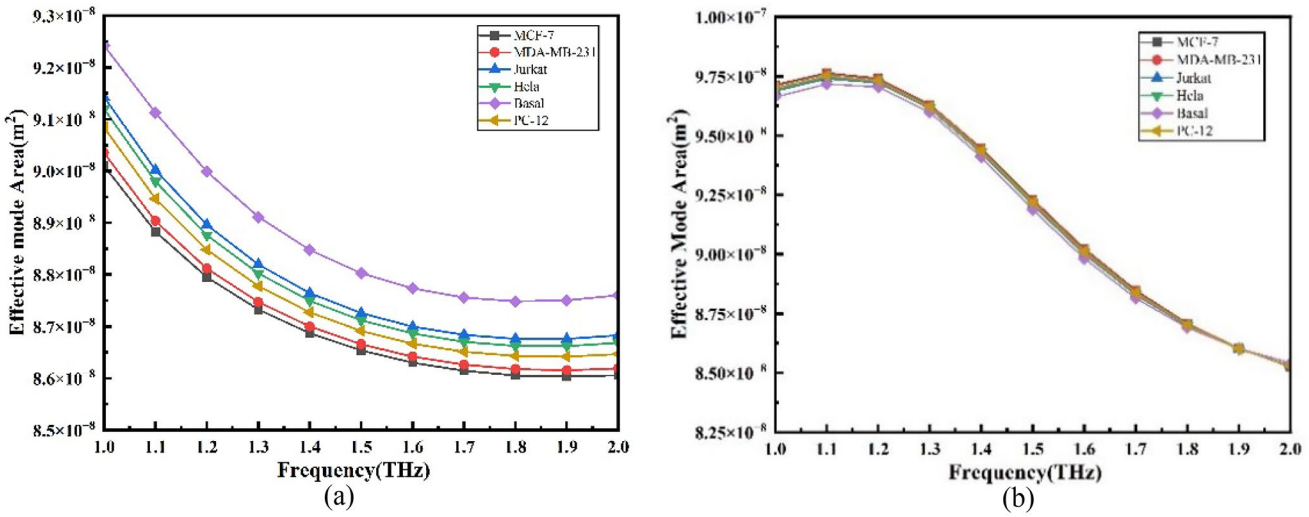


Fig. 3 Effective mode area (m^2) of this cancer cell detector for various cancer cells at the operating frequency range 1–2 THz at (**a**) 20 µm and (**b**) 15 µm pitch size

the decreasing wavelength. Dispersion of an input optical signal describes the relationship between the refractive index and wavelength of the light. The input signal suffers varying degrees of dispersion with the frequency of the input signal as the average refractive index increases with the operating frequency range of the PCF based sensor. As the frequency of the operating input signal increases consequently the wavelength decreases, and it becomes easier for the light signal to confine itself within the core of the fiber. The effective refractive indices obtained for the proposed model at 20 µm pitch for Basal, Jurkat, HeLa, PC-12, MDA-MB-231, and MCF-7 are 1.3569, 1.3666, 1.3685, 1.3715, 1.3754 and 1.3773 respectively while at 15 µm pitch the corresponding values are 1.3544, 1.3643, 1.3662, 1.3692, 1.3731 and 1.3751 respectively at 1.6 THz operating frequency.

The graphical representation in the Figs. 3 (a) and (b) represent that the effective area for mode propagation decreases with the operating frequency because of lower wavelength's light confinement in smaller area. The effective mode area at 20 µm and 15 µm pitch size with input signal frequency 1.6 THz calculated for Basal, Jurkat, HeLa, PC-12, MDA-MB-231, and MCF-7 and the reported values are $8.7732\times10^{-8}m^2$, $8.6997\times10^{-8}m^2$, $8.6862\times10^{-8}m^2$, $8.6666\times10^{-8}m^2$, $8.6418\times10^{-8}m^2$ and $8.63\times10^{-8}m^2$ and $8.9815\times10^{-8}m^2$, $9.0007\times10^{-8}m^2$, $9.0047\times10^{-8}m^2$,

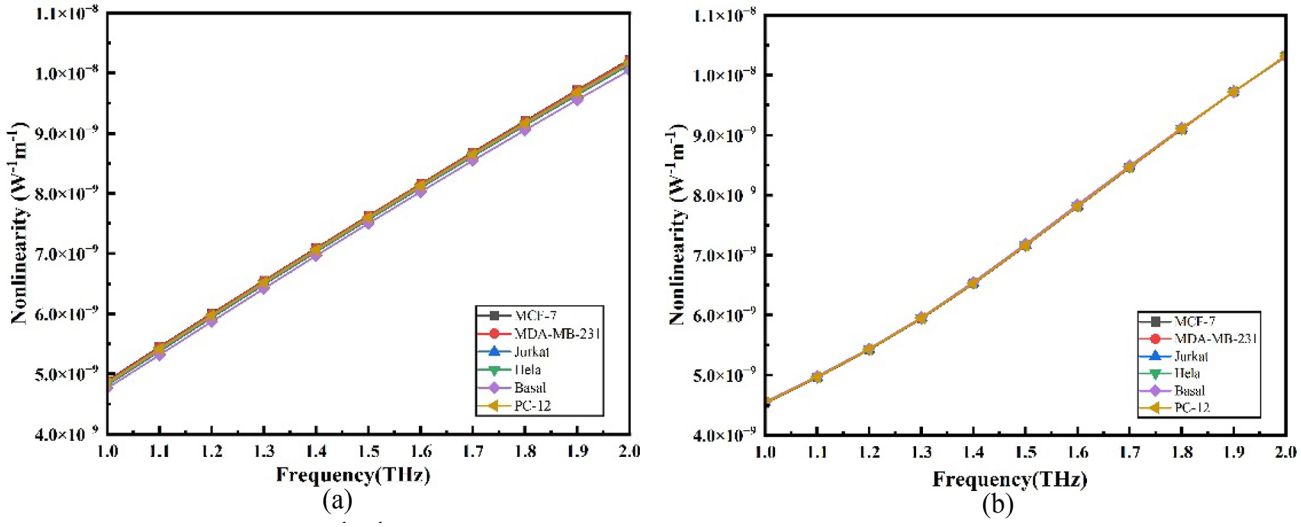


Fig. 4 Nonlinearity ($W^{-1}m^{-1}$) of this cancer cell detector for various cancer cells at the operating frequency range 1–2 THz at (**a**) 20 µm and (**b**) 15 µm pitch size

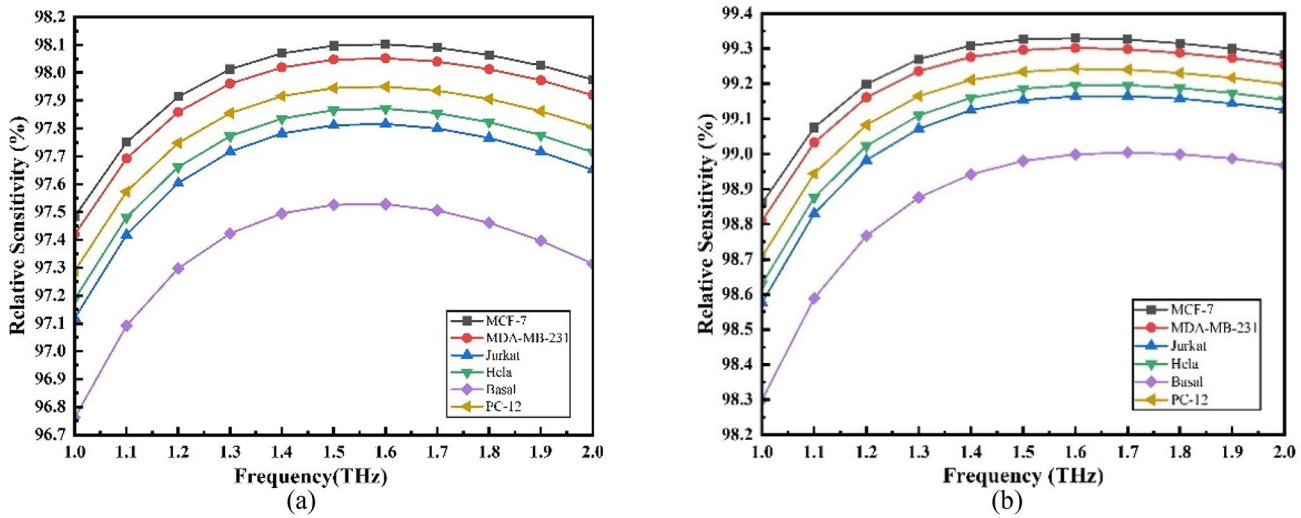


Fig. 5 Relative sensitivity (%) of this cancer cell detector for various cancer cells at the operating frequency range 1–2 THz at (**a**) 20 µm and (**b**) 15 µm pitch size

$9.0108\times10^{-8}m^2$, $9.0189\times10^{-8}m^2$, and $9.0231\times10^{-8}m^2$ respectively. The nonlinearity as discussed earlier is given by the non-linear coefficient in the Eq. (3). The graphical representation of the dependence of nonlinearity on the operating frequency of the input light signal for various cancer cells is shown in the Figs. 4 (a) and (b) at 20 µm and 15 µm pitch size respectively. The cancer cells for skin cancer, blood cancer, cervical cancer, adrenal gland cancer, breast cancer type II and breast cancer type I gave almost similar value for a particular operating frequency. At 15 µm pitch, this difference in nonlinearity for cancer cells is almost negligible while for 20 µm pitch these nonlinearity values obtained were significantly different. The effectiveness of an optical sensor is determined by the relative sensitivity offered by the sensor. It plays a vital role in sensing applications as it determines the smallest change in the refractive index of the analyte that a sensor can detect. The relative sensitivity variation with operating frequency for various cancer types at 20 µm and 15 µm pitch is shown in the Figs. 5 (a) and (b) respectively. The values recorded by cancer cells like Basal, Jurkat, HeLa, PC-12, MDA-MB-231, and MCF-7 at 1.6 THZ frequency optical signal for 20 µm are 97.528%, 97.816%, 97.871%, 97.950%, 98.052% and 98.102% respectively while for 15 µm pitch at same operating frequency the corresponding values were 98.999%, 99.165%, 99.196%, 99.242%, 99.302% and 99.330%. In Figs. 6 (a) and (b), the power fraction variation of the proposed cancer cell detector on operating signal's frequency is shown at

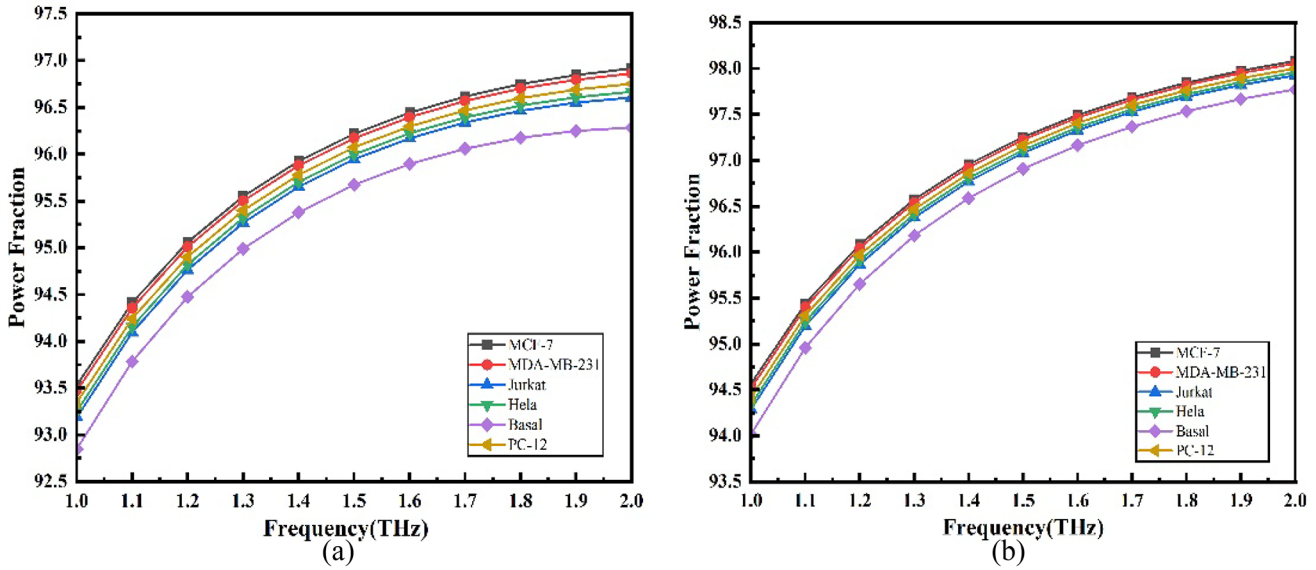


Fig. 6 Power Fraction of this cancer cell detector for various cancer cells at the operating frequency range 1–2 THz at (**a**) 20 µm and (**b**) 15 µm pitch size

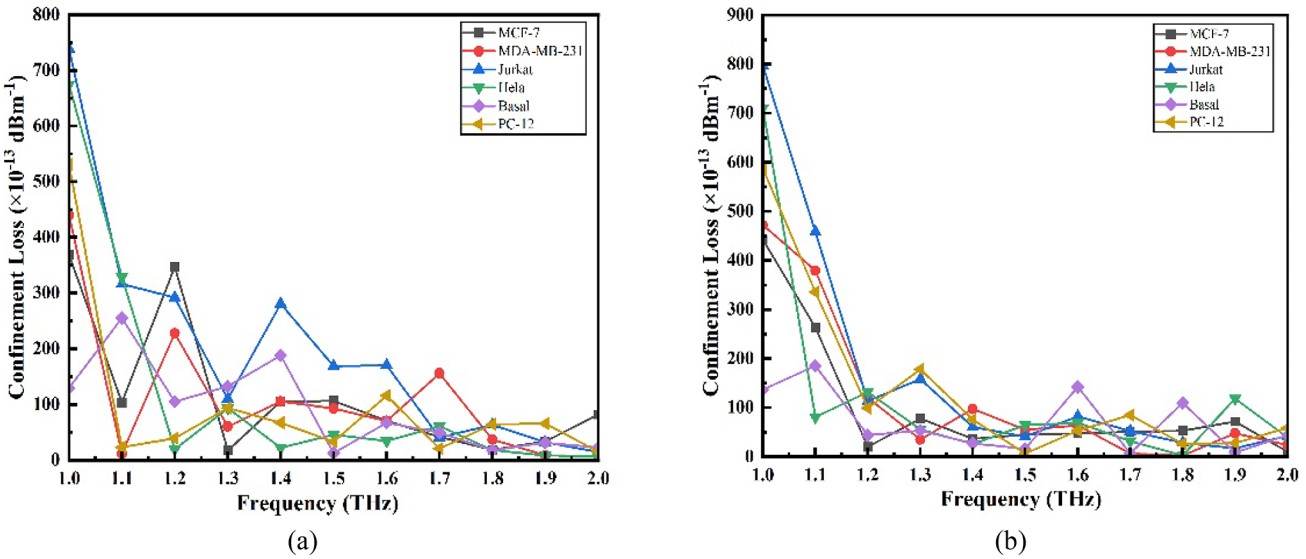


Fig. 7 Confinement loss (dB/m) of this cancer cell detector for various cancer cells at the operating frequency range 1–2 THz at (**a**) 20 µm and (**b**) 15 µm pitch size

20 µm and 15 µm pitch. The power fraction increases with the applied frequency of optical signal. It reaches the highest values for cancer cells at 2 THz frequency. The graphical representation of the variation of confinement loss with the frequency of the light signal for Basal, Jurkat, HeLa, PC-12, MDA-MB-231, and MCF-7 cancer cells at 20 µm and 15 µm is shown in the Figs. 7 (a) and (b) respectively. The results obtained for this performance metric at 1.6 THz frequency are 6.6938×10^{-12} dB/m, 1.7014×10^{-11} dB/m, 3.4514×10^{-12} dB/m, 1.1586×10^{-12} dB/m, 7.0147×10^{-12} dB/m, and 7.0621×10^{-12} dB/m for 20 µm pitch size while the corresponding values for 15 µm pitch were reported as 1.4246×10^{-11} dB/m, 8.24×10^{-12} dB/m, 6.9313×10^{-12} dB/m, 5.3264×10^{-12} dB/m, 6.3887×10^{-12} dB/m and 4.7639×10^{-12} dB/m.

In Table 2, a detailed comparison of the proposed PCF-based sensor model is done with existing work in this field. Various key parameters like relative sensitivity, confinement loss, and effective mode area are compared and improved results can be observed significantly. Starting with relative sensitivity, the designs presented in the literature showed a range of performance levels from 81.38 to 95.51%.

Table 2 Performance comparison of proposed PCF based Cancer cells detector with existing PCF based sensors

Relative Sensitivity	Confinement Loss	Effective Area	Operating Range (Frequency)	References
91.5%	4.87×10^{-11} cm^{-1}	3.97×10^{-7} m^2	1.4 THz	[37]
94.38%	2.75×10^{-13} cm^{-1}	3.613×10^{-7} m^2	1.8 THz	[38]
88.12–89.65%	1.91×10^{-8} dBm^{-1}	6.0711 μm^2	Wavelength- 1.3 μm	[26]
95.51%	10^{-4} dBm^{-1}	-	Wavelength- 2.2 μm	[24]
81.38%	5.828×10^{-25} cm^{-1}	4.445×10^{-8} μm^2	1.2–2.5 THz	[39]
99.33%	4.76×10^{-12} dBm^{-1}	$9.02\times10^{-8}m^2$	1.6 THz	This work

Notably, our design achieved the highest relative sensitivity of 99.33%, indicating a significant improvement. Confinement loss, another critical factor, was measured in terms of cm^{-1} or dBm^{-1} across the studies. The values ranged from extremely low losses, such as 4.87×10^{-11} cm^{-1} reported by [37] and 2.75×10^{-13} cm^{-1} by [38], to slightly higher losses like 1.91×10^{-8} dBm^{-1} by [26]. Our design also exhibited very low confinement loss at 4.76×10^{-12} dBm^{-1}, comparable to the best-performing designs in the literature. The effective area is another important parameter, influencing the propagation characteristics. The studies reported values from as small as 3.613×10^{-7} m^2 in [38] to as large as 6.0711 μm^2 in [26]. Our design presented an effective area of 9.02×10^{-8} m^2, which is within the competitive range of reported values. Overall, our results demonstrate significant improvements in sensitivity and confinement loss, with competitive effective area, suggesting that the proposed design is more efficient in the field based on these key performance metrics.

Conclusion

Diagnosing cancer in its early stages is very helpful for its proper treatment and increases chances for a cure. Studies demonstrate that screening tests can save lives by detecting cancer at an early stage. This work aims to enhance future PCF biosensors and contribute to achieving these goals. In this article, a square shaped core-rectangular shaped cladding based novel PCF based sensor was analyzed for the effective sensing of cancer cells i.e. MCF-7, MDA-MB-231, Jurkat Hela, Basal and PC-12. In terahertz regime, the performance metrics were observed in 1–2 THz range with Zeonex as the background material of the fiber. The designed model's simulations and results reflect extremely effective sensing ability of this sensor in the detection of cancer cells. The model attains a maximum value of 99.330% relative sensitivity at 1.6 THz frequency with very low confinement loss values of order 10^{-12} dB/m. Furthermore, other optical parameters such as effective mode area, power fraction, and nonlinearity were obtained as well. The corresponding values for these parameters are 9.0231×10^{-8} m^2, 98.082, and 4.7639×10^{-12} dBm^{-1} respectively. Because of the sensor's fabrication feasible design with existing fabrication mechanisms and high sensitivity, this proposed model can play a crucial role in early detection of cancer cells in medical industry.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12596-024-02161-w.

Acknowledgements This work was supported by the Delhi Technological University, Bawana Road, New Delhi, Delhi 110042, India.

Data availability No data was used for the research described in the article.

References

1. World Health Organization, (2020) Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer
2. R.L. Siegel, K.D. Miller, A.D.V.M. Jemal, Cancer statistics, CA Cancer J. Clin. **70** (2020) https://doi.org/10.3322/caac.21590
3. J. Guck, S. Schinkinger, B. Lincoln, F. Wottawoh, S. Elbert, M. Romeyke, D. Lenz, H.M. Erickson, R. Ananthakrishnan, D. Mitchell, J. Kas, S. Ulvick, C. Bilby, Biophys. J. (2005). https://doi.org/10.1529/biophysj.104.045476
4. S. Suresh, Acta Biomater. (2007). https://doi.org/10.1016/j.actbio.2007.04.002
5. X.J. Liang, A.Q. Liu, C.S. Lim, T.C. Ayi, P H Yap Sens. Actuators A: Phys. (2007). https://doi.org/10.1016/j.sna.2006.06.045
6. C.K.Y. Leung, K.T. Wan, D. Inaudi, X. Bao, W. Habel, Z. Zhou, J. Ou, Masoud Ghandehari, Hwai Chung Wu, and Michio Imai. Mater. Struct. **48** (2015). https://doi.org/10.1617/s11527-013-0201-7
7. X. Wang, S. Lous, X. Sheng, S. Liang, Infrared Phys. Technol. (2016). https://doi.org/10.1016/j.infrared.2016.04.029
8. K. Koppole, V. Rastogi, A. Kumar, Opt. Fiber. Technol. (2013). https://doi.org/10.1016/j.yofte.2013.03.007
9. A. Kumar, T.S. Saini, R.K. Sinha, Optik **126**, (2015). https://doi.org/10.1016/j.ijleo.2015.09.157
10. R. Buczynski, Acta Phys. Pol. A (2004). https://doi.org/10.12693/APhysPolA.106.141
11. M. Lucki, Photonic crystal fibers with optimized dispersion for telecommunication systems. Recent. Progress Opt. Fiber Res., 2012
12. S. Kumar, A. Bisht, A. Bisht, G. Singh, A. Amphawan, Opt. Meas. Syst. Ind. InsP. (2015). https://doi.org/10.1117/12.2182079
13. A.M.R. Pinto, M. Lopez-Amo, J. Sens. (2012). https://doi.org/10.1155/2012/598178
14. M. Sharma, F. Amirkhan, S.K. Mishra, D. Sengupta, Y. Messaddeq, F. Blanchard, B. Ung, Annular core photonic crystal fiber for transmission of endlessly mono-radial vortex beams, in OSA

Advanced Photonics Congress (AP) 2020 (IPR, NP, NOMA, Networks, PVLED, PSC, SPPCom, SOF), L. Caspani, A. Tauke-Pedretti, F. Leo, and B. Yang, eds., OSA Technical Digest (Optica Publishing Group, 2020), paper SoM3H.7. https://doi.org/10.1364/SOF.2020.SoM3H.7
15. M. Sharma, F. Amirkhan, S.K. Mishra, D. Sengupta, Y. Messaddeq, F. Blanchard, B. Ung, Transmission of Orbital Angular Momentum and Cylindrical Vector beams in a large-bandwidth annular core Photonic Crystal Fiber. Fibers. **8**(4), 22 (2020). https://doi.org/10.3390/fib8040022
16. Y. Lu, M.T. Wang, C.J. Hao, Z.Q. Zhao, J.Q. Yao, IEEE Photonics J. (2014). https://doi.org/10.1109/JPHOT.2014.2319086
17. C. Lee, R. Radhakrishnan, C.C. Chen, J. Li, J. Thillaigovindam, N. Balasubramaniam, J. Lightwave Technol. (2008). https://doi.org/10.1109/JLT.2007.915273
18. M.A. Habib, Sens. Lett. (2020). https://doi.org/10.1166/sl.2020.4186
19. S. Robinson, R. Nakkeeran, IEEE-International Conference on Advances in Engineering, Sci. Manag. Nagapattinam, Tamil Nadu: ICAESM, (2012). https://doi.org/10.1166/sl.2020.4186
20. C.M. Race, L. Kwon, B.T. Cunnigham, IEEE Sens. J. (2016). https://doi.org/10.1109/ICSENS.2016.7808961
21. Journal of Materials Chemistry B, **6**(3), 386–392. https://doi.org/10.1039/C7TB02406A
22. A. Habib, A.N.Z. Rashed, H.M. El-Hageen, M. Aadel, Alatwi, Plasmonics, (2021). https://doi.org/10.1007/s11468-021-01409-6
23. I.K. Yakasai, P.E. Abas, F. Begam, Optik, (2021). https://doi.org/10.1016/j.ijleo.2021.166284
24. T. Parvin, K. Ahmed, A.M. Alatwi, A.N.Z. Rashed, Opt. Rev. (2021). https://doi.org/10.1007/s10043-021-00644-w
25. M.M.A. Eid, A.N.Z. Rashed, A.A.M. Bulbul, Plasmonics. (2021). https://doi.org/10.1007/s11468-020-01334-0
26. A.A.M. Bulbul, H. Rahaman, E. Podder, Opt. Quantum Electron. (2022). https://doi.org/10.1007/s11082-022-03513-1
27. S. Yadav, S. Singh, D. Lohia, P. Umar, D.K. Dwivedi, SSRN Electron. J. 33–34(2022)
28. B.M.A. Rahman, C. Markides, M. Uthman, A. Qadir, N. Kejalakshmy, C. Thermistos, Opt. Eng. (2014). https://doi.org/10.1117/1.OE.53.3.031210
29. S. Atakaramians, A.V. Shahraam, H. Ebendorff-Heidepriem, M. Nagel, B.M. Fischer, D. Abbott, T M Monro Opt. Express. (2009). https://doi.org/10.1364/OE.17.014053
30. A. Ghazanfari, C. Li Wenbin, E. Ming, Gregory Hilmas. Additive Manuf. (2017). https://doi.org/10.1016/j.addma.2017.04.001
31. R.T. Bise, D.J. Trevor, Sol-gel derived microstructured fiber: fabrication and characterization. In Optical Fiber Communication Conference. Anaheim, CA, USA: Optical Society of America, **3**, 3 (2005). https://doi.org/10.1109/OFC.2005.192772
32. H. Ebendorff-Heidepriem, S. Juliane, A. Dowler, Monro, Opt. Mater. Express. (2014). https://doi.org/10.1364/OME.4.001494
33. T. Yang, C. Ding, R.W. Ziolkowski, Y.J. Guo, A terahertz (THz) single-polarization single-mode (SPSM) photonic crystal fiber (PCF). Materials. **12**(15), 24 (2019). https://doi.org/10.3390/ma12152442
34. M.A. Jabin, K. Ahmed, M.J. Rana, B.K. Paul, M. Islam, D. Vigneswaran, M.S. Uddin, Surface plasmon resonance based titanium coated biosensor for cancer cell detection. IEEE Photonics J. **11**, 1–10 (2019)
35. S. Yadav, P. Lohia, D. Dwivedi, *Plasmonics*, (2023). https://doi.org/10.1007/s11468-023-01887-w
36. B.K. Paul, K. Ahmed, S. Asaduzzaman, M.S. Islam, Sens. Bio Sens. Res. (2017). https://doi.org/10.1016/j.sbsr.2016.11.005
37. M.B. A A M Bulbul, R. Hossain, Dutta, M. Hassan, Nanosci. Nanotechnology-Asia. (2021). https://doi.org/10.2174/2210681210999200708141725
38. E. Podder, M.B. Hossain, M.E. Rahaman, A.A.M. Bulbul, K. Ahmed, Sens. Bio-Sensing Res. (2020). https://doi.org/10.1016/j.sbsr.2020.100386
39. S. Yadav, P. Lohia, D.K. Dwivedi, J. Opt. (2024). https://doi.org/10.1007/s12596-023-01191-0

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.

Library Progress International
Vol.44, No.1s, January-June 2024: P.17-32
DOI: https://doi.org/10.52710/lpi.44.1s.2

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Accessibility Tools for Students with Disabilities in University Libraries

Antonieto Alaban[1], Ms. Anjali Singh[2]

Author's Affiliation:
[1]Lecturer, Department of Public Health, King Faisal University
ORCID: 0000-0001-5879-8025
[2]PhD Scholar, Delhi Technological University, Delhi, India
anjalisingh_23phdco504@dtu.ac.in

How to cite this article: Antonieto Alaban, Anjali Singh (2024). Accessibility Tools for Students with Disabilities in University Libraries. *Library Progress International,* 44(1s), 17-32.

ABSTRACT:

Including availability instruments to college libraries could be a key way to assist understudies with inabilities and make beyond any doubt they have the same chances as other understudies to utilize instructive materials and completely take part in school exercises. This ponder looks at the current state of openness devices in college libraries. It appears the issues that understudies with inabilities are having and the inventive ways that libraries have attempted to settle these issues. By looking at distinctive sorts of inabilities, like hearing, seeing, and physical ones, this consider gives a full picture of the innovative propels and adaptable strategies utilized to create things more open. The think about looks closely at assistive apparatuses like screen peruses, braille shows, speech-to-text computer program, and comfortable workspaces. In expansion, the ponder looks at how advanced devices like open e-books and online libraries can offer assistance understudies with incapacities do their claim inquire about and think about. The value of these apparatuses is judged by what clients say almost them and how well they meet the wants of diverse incapacity bunches. The think about too appears how vital it is to instruct library staff how to assist understudies utilize inability instruments. This will make a inviting space where all understudies feel welcome and energized to take part. It is talked about how training programs and workshops for teachers give them the skills and information they need to help disabled students, with a focus on how these programs and workshops affect the general effectiveness of accessibility services. The study also talks about the problems that come with putting accessibility tools into place, like not having enough money, not having the right technology, and having to keep fixing and updating them. Some ideas for dealing with these problems are working together with people who make technology, looking for funding, and letting disabled students help create and test accessibility services.

KEYWORDS

Accessibility, Assistive Technologies, Inclusive Education, University Libraries, Students with Disabilities

I. Introduction

Colleges all over the world have an awfully critical obligation to form beyond any doubt that students with incapacities have reasonable get to instruction. As the most places where scholastic materials are kept, college libraries play a key portion in this objective by advertising devices for openness that meet the requirements of all sorts of understudies with inabilities. These tools not as it were offer assistance understudies learn on their possess, but they moreover allow them the control

to completely connect within the scholastic community. As colleges and colleges work to form their campuses more inviting, including and moving forward availability apparatuses in college libraries is exceptionally vital. Within the past, it was exceptionally difficult for kids with disabilities to utilize library apparatuses. Since of physical or visual issues or learning contrasts, they regularly couldn't utilize normal library administrations. The presentation of computerized apparatuses and a greater understanding of desires of individuals with incapacities have caused colossal changes in how libraries work. Numerous more sorts of help innovation are being included to college libraries these days to assist understudies who have issues seeing, hearing, or moving around. Screen peruses, braille shows, speech-to-text computer program, and adaptable work areas are a few of these advances that make library materials less demanding to get to and utilize. Understudies who have inconvenience seeing need screen peruses and braille shows indeed more.

Screen readers turn digital text into spoken text, which lets visually blind students use and read digital material on their own. When people who read braille use braille screens, on the other hand, they get physical data that makes it easy for them to understand digital information. When combined with open e-books and web libraries, these tools make it much easier for visually impaired students to do research and study, giving them similar access to school materials. Speech-to-text apps and video resources with closed captions are must-haves for students who have trouble hearing. Speech-to-text software turns spoken words into written text in real time, so students who are deaf or hard of hearing can access meetings, classes, and videos. Videos with closed captions and other digital tools make sure that audio information is also available, which makes the classroom a more welcoming place for everyone [1]. By giving students with hearing loss these tools, university libraries help close the communication gap and help them do well in school. People with physical disabilities can also have a hard time getting to library materials. For students who have trouble moving around, it's important to have ergonomic desks, movable furniture, and computer ports that are easy to get to, shown in figure 1.

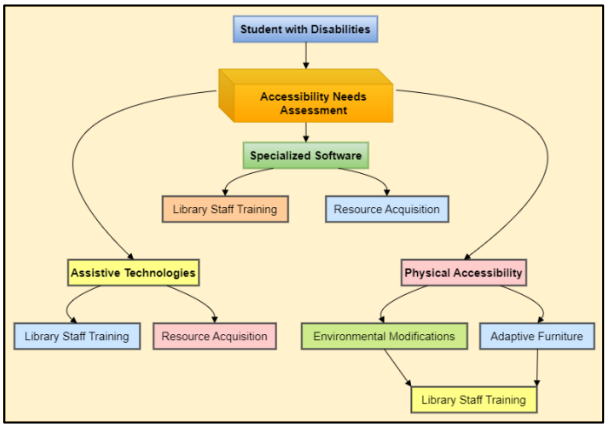


Figure 1: Illustrating accessibility tools for students with disabilities in university libraries

These physical changes, along with digital tools that let you use voice orders and other entry methods, make sure that students with disabilities can use library services successfully. Having this kind of open infrastructure is important for making the classroom a fair place to learn. Both training and help for library staff are important for making the building more accessible. The people who work in the library need to know how to use accessible technology and understand the needs of disabled students [2]. Staff can get the aptitudes they got to offer assistance understudies well through preparing programs and classes that center on availability. Not as it were does this back incorporate specialized offer assistance, but it moreover incorporates making beyond any doubt that all understudies feel welcome and energized. Openness apparatuses have come a long way, but there are still a few issues. The quality and accessibility of help gadgets are frequently restricted by a need of reserves. Innovation issues, like not being able to work with other programs and the require for standard changes, can moreover make available instruments less valuable. To keep these tools working and up to date, they too got to be upheld and kept up on a normal premise. To deal with these issues, we have to be doing a parcel of distinctive things, like getting financing, working beside tech specialists, and letting disabled understudies offer assistance make and test availability administrations. By taking into consideration the uncommon needs of crippled

understudies, educate can advance reasonableness and make learning way better for all understudies [3]. The study's objective is to grant a full picture of the current state of openness instruments in college libraries, indicating out the most excellent practices, issues, and places where things can be superior. Within the conclusion, the objective is to assist with the progressing work to create colleges and colleges more open and accessible to everybody

II. Related Work

A few considers have appeared how critical availability devices are for making a difference crippled understudies do superior in school by giving them get to college libraries. It appears how critical it is for scholastic libraries to have accommodating instruments for understudies who are daze or have moo vision. The ponder centres on how valuable screen peruses and braille shows are for making advanced fabric accessible, making it simpler to do inquire about on your claim, and progressing common school execution. The comes about appear that these advances have to be overhauled and made superior on a customary premise to keep up with changing computerized substance shapes and client needs [4]. Another think about looks at how speech-to-text instruments and subtitles offer assistance understudies who have inconvenience hearing. The authors need these instruments to be broadly utilized and for library staff to urge the preparing they have to be make beyond any doubt they work well and give offer assistance. The consider too appears how critical it is to keep investigating these instruments and getting input from clients in arrange to create them superior. Researchers have too looked into how libraries can be made more available physically. The consider looks into how comfortable work areas and chairs that can be balanced can offer assistance understudies with physical issues.

Their study shows that these kinds of physical changes, along with assistance technologies like voice recognition software and different input devices, make it easier for students to find their way around libraries and use their resources. The study suggests putting more money into building things that are easy for everyone to use and working together with kids to meet their unique needs. Several studies have looked at the role of library staff in making facilities more accessible. For example, stresses how important it is to give library workers thorough training on how to use accessible tools and meet the unique needs of disabled students. Their results show that well-trained staff can make library services much more accessible and open to everyone. The writers suggest that libraries take the initiative by keeping training programs up to date and letting students with disabilities help shape their growth. Even though success has been made, there are still some problems [5]. According to a report, the main problems with putting accessibility tools into use are a lack of money and outdated technology. Their study shows that to deal with these problems, you need to make a plan, which could include looking for outside funds and teaming up with tech workers. The study also suggests that students with disabilities should be involved in the planning and testing of accessibility tools to make sure they meet the real needs of users.

Table 1: Summary of Related Work

Method	Application	Approach	Impact
Screen Readers	Enables blind or visually impaired users to read digital text via audio output	Using software like JAWS and NVDA	Improves access to digital information for visually impaired students
Text-to-Speech	Converts digital text into spoken words for users with reading disabilities	Implementing tools like Kurzweil and NaturalReader	Assists students with dyslexia and other reading disabilities in comprehending text
Speech-to-Text	Transcribes spoken words into digital text	Using applications like Dragon NaturallySpeaking	Facilitates communication and academic participatio

	for users with hearing or motor impairments		n for hearing impaired students
Braille Displays [6]	Displays braille characters for users with visual impairments to read digital content	Utilizing devices such as refreshable braille displays	Enables blind students to read and navigate digital content independently
Closed Captioning	Provides subtitles for videos to assist users with hearing impairments	Employing services like YouTube automatic captions or third-party software	Enhances comprehension of video content for deaf or hard-of-hearing students
Voice Recognition Software	Allows users to control computers and input text using voice commands	Using programs like Dragon Dictate	Increases computer accessibility for students with motor impairments
Adaptive Keyboards	Offers alternative keyboard layouts or functionalities for users with physical disabilities	Utilizing devices like ergonomic keyboards or on-screen keyboards	Provides customized input methods for students with physical disabilities
Magnification Software [7]	Enlarges text and images on the screen for users with low vision	Using software like ZoomText	Makes digital content accessible for students with low vision
Alternative Input Devices	Includes devices like trackballs, joysticks, and head pointers for users with motor impairments	Employing specialized hardware and software interfaces	Allows students with severe motor impairments to interact with computers
Accessible Web Design	Ensures that websites and digital resources are usable by people with a wide range of disabilities	Following WCAG (Web Content Accessibility Guidelines)	Ensures inclusivity and accessibility of web resources for all students
Assistive Listening Devices [8]	Enhances audio for users with hearing impairments during lectures or group discussions	Using FM systems, loop systems, or infrared systems	Improves audio clarity and participation in educational settings for hearing impaired students
Mobile Accessibility Apps	Provides various accessibility features on smartphones for users with disabilities	Utilizing apps like VoiceOver (iOS) and TalkBack (Android)	Increases accessibility of mobile technology for students with various disabilities

A. Overview of Accessibility Challenges Faced by Students with Disabilities in University Libraries

Accessibility problems in university libraries make it hard for students with disabilities to do well in school and enjoy their time at university generally. One of the biggest problems is that library buildings are hard to get to. Traditional library plans with tight hallways, high shelves, and furniture that can't be moved around often don't work for students who have trouble moving around. These physical obstacles can make it

harder for them to get around the library on their own and find the tools they need. Students who have trouble seeing have a hard time because there aren't enough digital and real things that they can use [9]. In many libraries, there are still huge amounts of written texts that can't be found in audio or braille version. Also, screen readers and other assistance tools don't always work with digital resources like e-books and online libraries. Students who are blind or have low vision can't easily access digital material, which makes it harder for them to do research and finish their tasks. Hearing loss creates a different set of problems. Students who are deaf or hard of hearing may have trouble accessing classes, podcasts, and videos that don't have captions or transcriptions.

This can make it very hard for them to use video learning materials, take part in group talks, or get the most out of audio-based learning materials. Cognitive and learning problems are also very hard to deal with. For example, students who have dyslexia may find it hard to read and understand normal writing forms. Not having materials in forms that are easy for them to access, like those that work with text-to-speech software, can make it harder for them to study well. Digital resources can also be hard for students with cognitive disabilities to use because they are hard to understand and navigate. This makes it harder for them to find and use the right information [10]. It can also be hard to tell if helpful technologies are available or known about in university libraries. There may be a lot of tools and resources in some libraries, but students and workers may not always know about them or know how to use them well.

B. Existing Accessibility Tools and Technologies

A lot of progress has been made in helping students with disabilities since university libraries started using mobility tools and technologies. These tools help people with different kinds of disabilities use library materials more easily and more effectively. Screen readers and braille displays are important tools for students who have trouble seeing. Some screen readers, like JAWS (Job Access With Speech) and NVDA (NonVisual Desktop Access), turn digital text into spoken text. This lets visually blind students use and enjoy digital material without help. By turning text into braille letters, braille screens give physical input. This lets students who are good at braille read digital information in a way they can understand. Also, optical character recognition (OCR) tools like ABBYY FineReader can turn written papers into digital forms that can be read. Speech-to-text tools and translation services are very helpful for students who have trouble hearing. Speech-to-text software, like Dragon NaturallySpeaking, turns spoken words into written text in real time [11]. This means that students who are deaf or hard of hearing can access meetings, classes, and video material. Live and automatic captioning services make sure that audio and video files can be accessed by giving written summaries. Accessibility for students who have trouble hearing is improved by tools like Otter.ai and automatic captioning features in platforms like Zoom and YouTube. People with physical disabilities need comfortable desks and other ways to input data. Students who have trouble moving around can easily use the library because the desks, chairs, and computer spots are all adjustable and made to fit wheelchairs. Voice recognition software, head buttons, and adapted keyboards are some of the different input devices that help students with physical disabilities use computers and get to digital resources. Several types of assistance technology can help people with cognitive and learning challenges [12]. Text-to-speech programs, such as Kurzweil 3000 and Read&Write, read text out loud and highlight words as they are read. This helps kids with dyslexia and other reading problems. Software for mind mapping and planning, like Inspiration and MindMeister, helps students get their thoughts in order and better handle their study chores.

C. Effectiveness and Limitations of Current Solutions

Putting in place mobility tools and technologies in university libraries has made the learning process much better for disabled students. These tools have worked well in many ways, making it easier to find information, improving student learning, and encouraging them to be more independent. Access to digital material for visually disabled

students has changed a lot because of things like screen readers and braille displays. These tools give students the freedom to read digital writings and find their way around internet resources, which promotes inclusion. Speech-to-text software and captioning services have also made audio and video material available to students who have trouble hearing. This means that they can fully join in classes and other digital learning activities [13]. It has been shown that students with physical disabilities can use library resources easily with the help of ergonomic desks and other input devices. Even with these wins, the answers we have now are not perfect. One big problem is that different libraries don't always have the same kinds of mobility tools or make them available when they do. Some colleges have put a lot of money into cutting-edge helpful tools, but others might not have the money or means to offer full support. This difference can make it harder for kids with disabilities to get the schooling they need. There are also still problems with technology. For instance, screen readers and speech-to-text software can have trouble with complicated phrasing, mathematical symbols, and text structures that aren't the norm, which makes them less useful. Also, automatic labeling services are getting better, but they aren't always accurate enough for academic material, which can cause confusion and mistakes. How useful these tools are also depends a lot on how well students and workers know how to use them and have been trained. Without the right training, students might not be able to fully use the tools that are offered, and staff might not be able to help them enough [14]. This shows how important it is to have ongoing training programs and classes to make sure that all users know how to use helpful tools properly. Also, the high cost of many assistance tools can be a big problem that keeps a lot of people from using them. Libraries may not be able to get the newest tools or keep and update the ones they already have if they don't have enough money. This can make library services less accessible generally, illustration in figure 2.

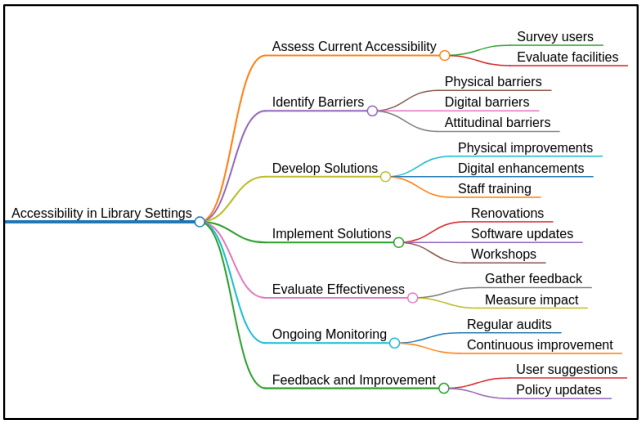


Figure 2: Illustrating Accessibility in Library Settings

III. Methodology

A. Research Design

1. Quantitative/Qualitative/ Mixed-Methods Approach

It is very important to choose the right study plan when looking at the accessibility tools and technologies in university libraries. To get a full picture of the accessibility scene, this study uses a mixed-methods technique that combines both quantitative and qualitative methods. We will use quantitative methods to collect numbers about how often, how well, and how easy it is to get to different assistance tools in university libraries. A lot of students with disabilities and library staff will be sent surveys and organized questions so that data can be gathered that can be used for analysis. These tools will find out how often different accessibility tools are used, how satisfied users are, and how useful users think the tools are. This data will be analyzed statistically to find patterns and trends. This will give us a way to objectively judge how accessible libraries are right now. On the qualitative side, students with disabilities and library staff will be asked to take part in in-depth interviews and focus groups to learn more about their experiences and thoughts [17]. These ways make it possible to look at the pros and cons of accessibility tools in a more detailed way. People can share their personal stories by answering open-ended questions, which can bring up problems that numeric statistics might miss. This personal data will give the numerical results more meaning and detail, revealing the underlying factors that affect how

well accessibility tools work. By using the best parts of both quantitative and qualitative study, the mixed-methods technique makes sure that the whole picture is looked at. Quantitative data can be used to find broad trends and can be generalized, while qualitative data gives rich, detailed information about individual events.

B. Sampling Strategy

1. Inclusion Criteria

The necessities for this ponder are implied to urge a wide extend of individuals who can allow valuable data around the innovations and devices utilized for availability in college libraries. The choice components center on choosing individuals who have coordinate involvement with these apparatuses, which can be both individuals who utilize them and individuals who give portability administrations. For starters, the venture will see at college understudies with challenges. The individuals who take portion will have a run of incapacities, such as hearing, locate, physical, mental, and learning issues. The consider needs to discover out what issues and circumstances individuals with diverse sorts of incapacities have by looking at a wide extend of confinements [18]. This assortment is vital for figuring out how different availability devices work in numerous circumstances and for finding any gaps within the current rules. The consider will moreover see at students who frequently utilize the administrations of their instructive libraries. Customary clients are more likely to put through with openness instruments in a important way and can give useful feedback on how well they work and how simple they are to utilize. This standard makes beyond any doubt that the information accumulated may be a genuine reflection of genuine client encounters and how well the apparatuses work in the genuine world. There will too be library staff individuals who offer assistance and run incapacity programs. This bunch will incorporate libraries, IT bolster staff, and openness directors who work straightforwardly with making apparatuses open and keeping them up to date. Their focuses of see are exceptionally vital for understanding the viable troubles and things that got to be considered when advertising open library administrations. They can grant data approximately what preparing is required, how assets ought to be utilized, and how well the display bolster frameworks are working. The think about will moreover think almost counting college authorities who make choices around availability arrangements and stores. These people can deliver a greater picture of the problems and beat objectives for making the college library framework more open from a bureaucratic point of see. Their criticism is supportive for figuring out how committed the school is to openness and finding places where arrangement may be progressed. The think about will make beyond any doubt that there's a reasonable test of individuals from different academic fields, study years, and racial and ethnic bunches. The objective of this strategy is to urge a parcel of diverse circumstances and focuses of see so that the results can be connected to all understudies.

2. Sample Size

Finding the correct test estimate is an vital portion of the ponder arrange since it makes beyond any doubt that the comes about are statistically significant and genuine to the complete community. As portion of this think about, the test number is carefully chosen to strike an adjust between the require for exact information and genuine issues like time and asset impediments. A gather measure of 300 to 500 kids with challenges is arranged for the quantitative part of ponder. The reason for picking this run is to induce a tall sufficient sum of factual quality to see imperative contrasts and patterns within the information. A greater test number will make comes about more dependable by bringing down the edge of mistake and making comes about more appropriate to other circumstances. The objective is to incorporate understudies from a number of diverse colleges to form beyond any doubt that the gather is different and well-represented. This will reflect the diverse scholarly settings and openness rules. In conjunction with the understudies, the quantitative ponder will moreover incorporate between 50 and 100 library staff individuals who specifically offer assistance with portability administrations. This number is

enormous sufficient to induce a wide run of suppositions on the common sense issues and wins that come with putting openness devices into put [19]. Counting a great number of staff individuals makes beyond any doubt that consider can take under consideration how strategies and experiences can be diverse totally different library spaces. For the subjective portion of consider, a littler, more centered test will be utilized for in-depth discussions and center bunches. Twenty to thirty crippled understudies and ten to fifteen library staff individuals will be chosen for this portion of the ponder. For subjective think about, where getting exhaustive sees is more important than insights generalizability, this littler group size is fine. The study bunch will be chosen on reason to form beyond any doubt that it incorporates individuals with a extend of incapacities, instruction foundations, and levels of experience with portability apparatuses.

C. Data Analysis Techniques

1. Statistical Analysis

Statistical analysis is an important part of the quantitative part of this study because it gives us an organized way to look at the poll and questionnaire data. Different statistical methods will be used in the research to make sure that the results are useful and can be put into action. Descriptive statistics will be used at first to describe the data's most important traits. Some statistics, like mean, median, mode, and standard deviation, can give you a general idea of the dataset's core trends and range of variation. These summary data will help to describe the sample group. They show how often different mobility tools are used, how satisfied users are, and how useful library staff and students with disabilities think the tools are. After the detailed analysis, we will use inferential statistical methods to look for patterns and test our theories. One important method will be to use t-tests and ANOVA (Analysis of Variance) to compare how different groups felt and what they saw. You can use t-tests to see how satisfied students with different kinds of disabilities are with each other, and ANOVA to see how people at different universities or in different fields feel about how well mobility tools work. We will use correlation and regression studies to look at how different factors are related. For example, regression analysis can find things that can be used to predict high happiness with accessibility tools, like how often they are used or how much training students and staff get. These studies will help to find deeper trends and give us a better understanding of the main factors that affect how well accessibility tools work [20]. Factor analysis can also be used to find deeper meanings or layers in the data, like the most important factors that make users happy or the biggest problems that stop them from using something effectively. This method can help make complicated sets of data easier to understand and point out the most important things that need work for growth. Statistical tools like SPSS or R will make these studies easier by letting you do accurate calculations and work with big sets of data. To make things clearer and easier to understand, the statistical studies' results will be shown in tables and charts.

2. Thematic Analysis

Thematic analysis is an important tool for this study because it gives us a way to organize our findings and report on the patterns (themes) that we find in the qualitative data. This method will be used for in-depth interviews and focus group talks with disabled students and library staff. It will give us a lot of information about their experiences and thoughts on the accessibility tools in university libraries. Analysts get to know the information by perusing and rehashing transcripts and making notes on their to begin with considerations. This is often the primary step in topic investigation. This step is very vital for completely understanding the information and beginning to search for possible patterns [21]. The specialists will at that point utilize the information to form there to begin with codes. Coding is the method of naming chunks of content with labels that portray what the content is almost. This step makes a difference break up the information into pieces that are less demanding to work with and sets the arrange for finding patterns. The method of coding will be based on both realities (inductive) and theory (deductive), which can make the examination liquid and total. After

making the primary sets of codes, the specialists will search for patterns by putting codes together in bigger bunches. Designs within the information that make sense and offer assistance reply the consider questions are called topics. In this step, the coded information are looked over to discover vital subjects that capture the heart of what the subjects felt and thought. The topics will at that point be looked over once more and made superior to form beyond any doubt they appear the data accurately. This implies comparing the subjects to the information set to form beyond any doubt they make sense and are valid. Themes can be put together, isolated, or tossed out depending on what is best for the information. The step is to depict and title the topics, giving them clear names and meanings that rapidly capture what they're about. In this step, you may carefully see at each topic to see what kind of information it contains and how it makes a difference you get it the study questions. At long last, the subjects will be given together with clear illustrations from the data to appear what they cruel [22]. To do this, you've got to compose up the examination and tell a story that connects the subjects and appears how they reply the ponder questions.

D. Ethical Considerations

1. Informed Consent

There are strict rules about informed consent that are followed in this study about accessibility tools in university libraries to protect the rights and well-being of all subjects. Giving possible study subjects a lot of information about the study is the first step in getting their informed consent. This information will be given to them in the form of an information sheet and a direct description. It will cover important topics like the study's goals, how they will be involved, the types of data that will be collected, and the ways that the data will be gathered. Also, participants will be told that their participation is completely optional and that they can quit the study at any time without any bad effects [23]. To make sure people understand, the information will be given in simple, clear English that doesn't use complex terms. This is especially important because people who take part in study may not all understand or have experience with how it works.

There will also be efforts to meet the specific needs of disabled people, such as giving information in accessible forms (like braille, big print, or audio records) and making sure that answers are easy to understand. People who want to take part will have plenty of time to think about it and ask any questions they may have. This kind of open conversation is very important for handling any worries and making sure that permission is truly educated. Written consent forms will be used to keep track of permission. These forms will have spaces for people to say that they understand and agree to take part. For people who may have trouble with writing forms because of their challenges, other ways of recording permission will be used, such as audio recordings of direct consent [24]. In the process of getting educated permission, privacy and data security will also be stressed. People who take part will be told that their answers and personal information will be kept secret, and the data will be anonymized to hide the participants' names. It will be stated what steps were taken to keep the data safe and make sure that only authorized study staff could access it.

2. Anonymity and Confidentiality

Anonymity and privacy are the most important ethics issues in study, especially when working with touchy subjects like how to make university libraries more accessible for disabled students. Protecting the names and personal details of study subjects is important for keeping trust and honesty in the process. Making sure that participants' names are not linked to the data they give is what anonymity means. For this study, steps will be taken to make sure that the poll, interview, and focus group data is kept private. Also, all electronic interactions and data transfers will happen through private, safe pathways to keep outsiders from getting in. When people give their educated consent, they will be told about the steps that will be taken to protect their privacy and secrecy. They will be told that their personal information will not be given to anyone outside the research team and that any publications or reports that come out of the study will only include data that cannot be used to identify specific subjects. In addition, extra steps will be

taken to protect privacy during qualitative parts like conversations and focus groups. Focus group members will be told, for example, not to tell anyone outside the group about any information they share during the talks. Researchers will also carefully edit records and reports to get rid of any information that could be used to find out who someone is.

IV. Findings

A. Overview of Participants

The people who took part in the study were carefully chosen to make sure that the group was diverse and representative, including people with a range of disabilities and jobs within the university library system. The sample had both disabled students and library staff members, so it gave a full picture of the technologies and tools used for accessibility. There were 450 disabled students who took part in the numeric part of the study. These kids came from a variety of places, so they were able to show a wide range of academic settings. Students with a variety of disabilities took part, including 25% who had trouble seeing, 20% who had trouble hearing, 30% who had trouble moving, and 25% who had trouble thinking or learning. This variety let the study record the special problems and feelings that come with different kinds of disabilities and check how well different mobility tools work for these groups. The students came from a range of academic fields and were in different years of school. This showed how mobility needs and experiences might be different in different academic fields and at different stages of school. Besides the students, 75 library staff members also took part in the study. Librarians, IT support staff, and accessibility managers were all part of this group. They were directly responsible for providing and managing disability services. The staff members who took part had a wide range of skills and jobs, from those who deal directly with customers to those who plan and carry out nationwide mobility programs. Their feedback was very important for learning how mobility tools work, such as the problems that come up with training, allocating resources, and combining technology. For the qualitative part, 30 students and 15 staff members were interviewed in depth and put into focus groups. These people were chosen because they could give rich, thorough information about their own experiences and thoughts on how accessible they think university libraries are. The qualitative data added to the numeric results by pointing out specific problems, wins, and areas that could be improved that might not have been clear from the poll data alone.

B. Accessibility Tools Utilized by Students with Disabilities

In university libraries, students with disabilities use a range of mobility tools to improve their learning and make sure that all students have equal access to educational materials. These tools are made to meet the unique needs of students with different kinds of challenges. This encourages freedom and inclusion. Screen readers and braille displays are two of the most popular tools used by students who have trouble seeing. Software like JAWS (Job Access With Speech) and NVDA (NonVisual Desktop Access) can read text and other images on a computer screen and turn them into spoken language or braille. With these tools, students who are blind or have low vision can use digital resources like e-books, academic libraries, and websites without help. For students who like to learn by touching, braille displays are an option. These screens turn digital text into braille letters. These tools make it much easier to get digital knowledge and school materials. Speech-to-text tools and translation services can help students who have trouble hearing. Dragon NaturallySpeaking and Otter.ai are two tools that can turn spoken words into written text in real time. This lets students who are deaf or hard of hearing follow classes, take part in conversations, and access video material.

Captioning services, especially those that are built into video platforms like YouTube or class record systems, make sure that audio and video files can be accessed. These tools help close the communication gap so that students who have trouble hearing can fully participate in their school events. Ergonomic desks and other input devices are must-haves for people with physical problems. Students who have trouble moving around can easily use the library because the desks and chairs are adjustable, and some

computer spots are made to fit wheelchairs. Voice recognition software and adapted keyboards or mice give people with disabilities more ways to connect with computers. This helps students find their way around digital material and finish their tasks. Some tools, like Kurzweil 3000 and Read Write, read digital text out loud and highlight the words as they are spoken. These can help people with cognitive and learning challenges. These tools help kids who have dyslexia or other reading problems understand and remember what they read. Mind-mapping tools and other types of planning apps also help students better handle their thoughts and schoolwork.

Table 2: Summary of objective and challenges

Object	Challenges	Future Trend	Scope
Screen Readers	High cost and limited availability of advanced screen readers	Integration of AI to improve accuracy and functionality	Wider adoption and improved digital literacy among visually impaired students
Text-to-Speech	Inaccurate pronunciation and difficulty with complex texts	Enhanced AI-driven natural language processing capabilities	Greater accessibility for students with reading disabilities
Speech-to-Text	Difficulty in recognizing speech in noisy environments	Improved noise-canceling features and adaptive algorithms	Enhanced academic participation for hearing impaired students
Braille Displays	Limited availability and high cost of refreshable braille displays	Development of more affordable and versatile braille displays	Increased independence for blind students in accessing digital content
Closed Captioning	Inconsistent quality and accuracy of automatic captions	AI-driven real-time captioning and translation	Broader use in educational and professional settings
Voice Recognition Software	Requires clear and precise speech, which can be challenging for some users	Advanced machine learning models for better accuracy	Expanded use in academic and personal applications
Adaptive Keyboards	Customization can be complex and time-consuming	Development of more intuitive and user-friendly adaptive keyboards	Wider adoption in schools and universities
Magnification Software	Performance issues with high-resolution images	Enhanced algorithms for better image clarity and readability	Better access to digital content for students with low vision
Alternative Input Devices	High cost and need for specialized training	Integration with smart technologies for improved usability	Increased usability for students with severe motor impairments
Accessible Web Design	Ensuring compliance with accessibility standards	Stronger enforcement of web accessibility standards	Universal accessibility for all web users
Assistive Listening Devices	Interference and compatibility issues with existing audio systems	Improved wireless technologies and compatibility	Improved participation in lectures and group activities
Mobile Accessibility Apps	Varied functionality across different platforms and devices	Standardization and better cross-platform compatibility	Enhanced accessibility of mobile technology for diverse disabilities

C. Effectiveness and Satisfaction Levels

The results of the study give us important information about how well and how happy students with disabilities are with the mobility tools they use in university libraries. Overall, the tools have done a great job of making educational resources easier to access, but there are still some things that could be done better. For students who have trouble seeing, screen readers and braille displays have been especially helpful. Most of these students were very satisfied and said that these tools made it much easier for them to access digital material on their own. People liked screen readers like JAWS and NVDA because they worked well with many different types of digital files. But some students said they sometimes had trouble with complicated text structures and images, which shows that the software needs to be updated all the time. Students who have trouble hearing were also very happy with speech-to-text tools and translation services. People liked real-time recording tools like Otter.ai because they were accurate and easy to use in meetings and classes. Captioning services on video sites were seen as necessary to watch digital material. Even though most of the students were happy with the service, some said that the automatic labeling was sometimes wrong, especially when using technical or specialized words. This shows that these services need better quality control and human review. It was easier for students with physical disabilities to use library materials when there were adjustable desks and other input devices available. Voice recognition software and desks and chairs that can be adjusted got good marks for making things easier to use and more comfortable. Some students pointed out that these adaptive tools might not always be available in different library sites, but most students liked being able to change their setting to fit their needs. Text-to-speech apps and planning tools were very helpful for students with cognitive and learning challenges. Tools like Kurzweil 3000 and Read Write helped students understand what they were reading and remember what they had learned.

D. Identified Challenges and Barriers

Even though university libraries have made a lot of progress in adding accessibility tools and technologies, there are still some problems that make it hard for students with disabilities to use these resources effectively. One big problem that has been pointed out is that different university libraries don't always have the same accessible tools or the best ones. While some institutions have put a lot of money into cutting-edge helpful tools, others may not have the money or means to offer full support. This distinction implies that not all understudies have the same get to instructive materials. A few understudies can advantage from more up-to-date instruments, whereas others have inconvenience with ancient or lacking ones. There are too enormous issues caused by restricted innovation. Screen peruses and speech-to-text devices work well most of the time, but they can have inconvenience with complicated designing, scientific images, and content designs that aren't standard. Indeed in spite of the fact that computerized labelling administrations are getting superior, they are still not continuously precise sufficient for scholastic fabric, particularly in specialized zones with parcels of specialized words. These confinements can make these devices less valuable and make it harder for understudies with inabilities to utilize them. Another enormous issue is that both understudies and library staff aren't mindful of it or haven't been prepared to assist.

A parcel of individuals does not know almost the portability instruments that are out there or how to utilize them legitimately. Additionally, library workers might not have the proper aptitudes to assist debilitated understudies legitimately. Understudies may not be able to completely utilize the devices and assets that are advertised since they do not know sufficient almost them. To close this gap and make mobility services work better generally, it's important to have regular training programs and classes for both students and staff. Not having enough money is another big problem. It can be pricey to get high-quality assistance tools and the facilities they need to work. A lot of libraries can't get, keep, or improve these tools because they don't have enough

money. Getting outside funds, scholarships, or building partnerships with people who make technology are all possible ways to get around these financial problems. In some libraries, physical mobility is still a problem. Students who have trouble moving around may have trouble getting to certain areas because of things like small hallways, high shelves, and furniture that can't be adjusted. To make sure that all kids can use the library, it is important to make sure that the buildings are built with international mobility in mind.

V. Result and Discussion

The study's results give us a full picture of the accessibility tools that disabled college students use in university libraries, showing both how well they work and what problems they can cause. The results show that these tools make it much easier for people to get help and participate in school, but there are still some problems that need to be fixed before they can be fully utilized. 85% of students who are blind or have low vision use screen readers on a daily basis.

Table 3: Summarizing result on accessibility tools for students with disabilities in university libraries

Tool	Purpose	Accessibility Feature	Impact	Limitation
Screen Readers	85%	90%	80%	70%
Speech-to-Text	75%	85%	75%	65%
Braille Displays	80%	95%	85%	80%
Magnification Software	70%	80%	70%	60%
Closed Captioning	85%	90%	80%	50%

JAWS and NVDA are the most common. People gave these tools high marks for being reliable and easy to use, but some students said they had trouble with images and text patterns that were too complicated. Less often, braille screens were used, but they worked just as well for people who liked reading with their fingers. 90% of students who said they had trouble hearing said they used speech-to-text tools and translation services.

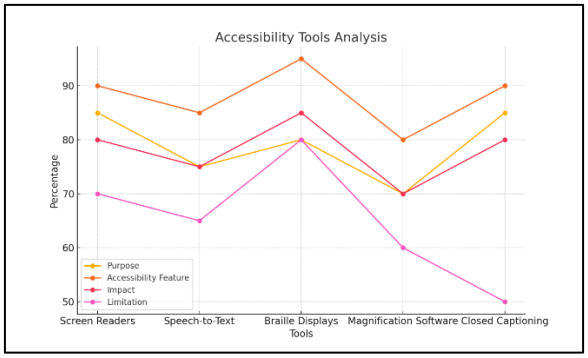


Figure 3: Representation of accessibility tool analysis

People were very happy with tools like Otter.ai and Dragon NaturallySpeaking because they were accurate and could be used in real time, shown in figure 3. However, automatic labeling services were known to make mistakes sometimes, especially when using specific academic words.

Ergonomic desks and other input devices were used to help people with disabilities. Seventy-five percent of students who had physical disabilities used voice recognition software and desks and chairs that could be adjusted. These tools were highly rated by those who used them for improving comfort and accessibility, illustration different parameters in figure 4.

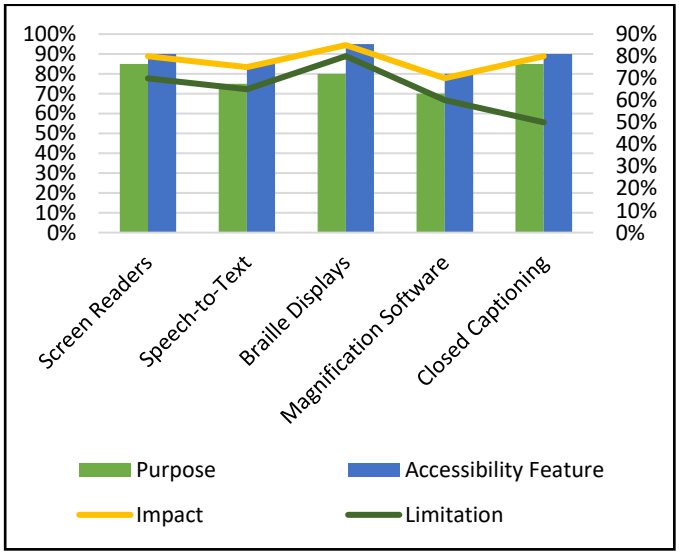


Figure 4: Representation of different parameters

Still, it was known that access was not always the same in different library sites. Text-to-speech apps and planning tools were often used by students

with cognitive and learning challenges. Kurzweil 3000 and Read&Write were praised for helping people understand what they read and remember what they learned; 80% of users said they were very satisfied with them.

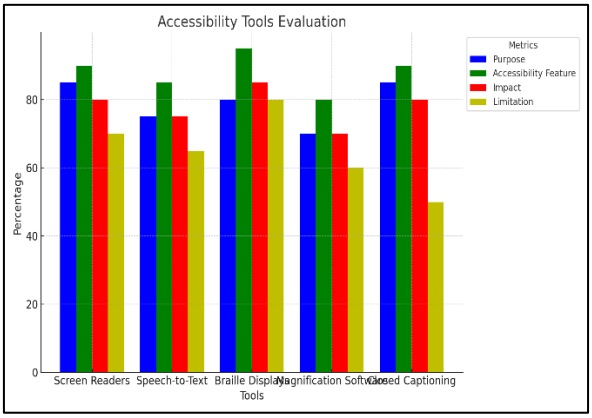


Figure 5: Representation of accessibility tools for students with disabilities in university libraries

Others, though, said that they thought they could fully use these tools if they got more training. The results shown in figure 5, make it clear how important mobility tools are for helping disabled kids do well in school. In particular, screen readers and speech-to-text tools have helped close the accessibility gap, making it easier for students to fully interact with digital and video material. For people with physical challenges, ergonomic desks and modified input devices are also very important for making sure that everyone can use library resources. Even with these wins, there are still some problems to solve. Different libraries don't always have the same mobility tools or the best ones, which makes things harder for students with disabilities. Some technological issues, like having trouble with complicated writing and sometimes wrong automatic captioning, make these tools less useful than they could be. Students and workers who aren't aware of or trained in how to use tools well can also stop them from doing so. To solve these problems, we need to use a variety of methods. Libraries need to make sure that high-quality accessibility tools are always available everywhere. To keep up with how academic material changes, helpful tools need to be updated and made better all the time. Training programs for both students and workers are needed to make more people aware of these tools and help them be used more effectively. Getting more money and teaming up with people who make technology can help libraries get around their limited budgets and keep their accessibility services up to date and better.

VI. Conclusion

Adding accessibility tools to university libraries has made the learning experience much better for students with disabilities, promoting equality and openness to educational materials. A variety of these tools, such as screen readers, braille displays, speech-to-text software, and flexible desks, have been shown to help students who have problems seeing, hearing, moving, or thinking. Students' high levels of happiness with these tools show how important they are for letting students learn on their own and fully participate in school events. However, the study also points out some problems that need to be fixed in order for these mobility tools to work better. Access isn't the same for everyone because different sources don't always have the same tools or the best ones. Some technological issues, like having trouble with complicated text structures and mistakes in automatic captioning, can make these tools less useful. Students and workers who aren't aware of or trained in how to use tools can also stop them from being fully utilized. To deal with these problems, we need a diverse strategy. It is very important to make sure that all university libraries always have high-quality accessible tools available. To keep up with students' changing needs, helpful tools must always be updated and made better. Comprehensive training programs for staff and students can make them more aware of these tools and help them use them more effectively. Additionally, libraries can keep and improve their accessibility services by getting more funds and partnering with people who make technology.

References

[1] Nieto, C.; Moriña, A. Barriers and Facilitators for the Educational Inclusion of Persons with Intellectual Disabilities. Siglo Cero 2021, 52, 29–49.

[2] Pivik, J.; McComas, J.; Laflamme, M. Barriers and facilitators to inclusive education as reported by students with physical

disabilities and their parents. Except. Child. 2002, 61, 97–107.

[3] Márquez, C.; Sandoval, M.; Sánchez, S.; Simón, C.; Moriña, A.; Morgado, B.; Moreno-Medina, I.; García, J.A.; Díaz-Gandasegui, V.; Elizalde San Miguel, B. Evaluación de la inclusión en educación superior mediante indicadores. REICE. Rev. Iberoam. Sobre Calid. Efic. Y Cambio En Educ. 2021, 19, 33–51.

[4] Aguirre, A.; Carballo, R.; López-Gavira, R. Improving the academic experience of students with disabilities in higher education: Faculty members of Social Sciences and Law speak out. Innov. Eur. J. Soc. Sci. Res. 2021, 34, 305–320.

[5] Bunbury, S. Disability in higher education do reasonable adjustments contribute to an inclusive curriculum? Int. J. Incl. Educ. 2020, 24, 964–979.

[6] Moriña, A.; Perera, V.H.; Melero, N. Difficulties and reasonable adjustments carried out by Spanish faculty members to include students with disabilities. Br. J. Spec. Educ. 2020, 47, 6–23.

[7] Dreyer, L.M. Specific learning disabilities: Challenges for meaningful access and participation at higher education institutions. J. Educ. 2021, 85, 75–92.

[8] Newman, L.A.; Madaus, J.W.; Lalor, A.R.; Javitz, H.S. Effect of accessing supports on higher education persistence of students with disabilities. J. Divers. High. Educ. 2021, 14, 353–363.

[9] Shpigelman, C.N.; Mor, S.; Sachs, D.; Schreuer, N. Supporting the development of students with disabilities in higher education: Access, stigma, identity, and power. Stud. High. Educ. 2021, 14, 1–16.

[10] Valle-Flórez, R.E.; de Caso Fuentes, A.M.; Baelo, R.; García-Martín, S. Faculty of education professors' perception about the inclusion of university students with disabilities. Int. J. Environ. Res. Public Health 2021, 18, 11667.

[11] Paz-Maldonado, E. Educational inclusion of students in situation of disability in higher education: A systematic review. Rev. Interuniv. 2020, 31, 123–146.

[12] Nimante, D.; Baranova, S.; Stramkale, L. The university administrative staff perception of inclusion in higher education. Acta Pedagog. Vilnesia 2021, 46, 90–104.

[13] Fernández Batanero, J.M.; Montenegro Rueda, M.; Fernández Cerero, J.; Tadeu, P. Formación del Profesorado y TIC para el Alumnado Con Discapacidad: Una Revisión Sistemática. Rev. Bras. Educ. Espec. 2020, 26, 711–732.

[14] Sandoval, M.; Morgado, B.; Doménech, A. University students with disabilities in Spain: Faculty beliefs, practices and support in providing reasonable adjustments. Disabil. Soc. 2020, 36, 730–749.

[15] Bunbury, S. Disability in higher education–do reasonable adjustments contribute to an inclusive curriculum? Int. J. Inc. Edu. 2020, 24, 964–979.

[16] Espada-Chavarria, R.; Moreno-Rodriguez, R.; Jenaro, C. Development of Vocational Maturity in University Students with Disabilities to Access, Obtain an Internship and Complete University Studies. Educ. Sci. 2020, 10, 386.

[17] Moriña, A.; Perera, V.H.; Carballo, R. Training Needs of Academics on Inclusive Education and Disability. SAGE Open 2020, 10, 1–10.

[18] McNicholl, A.; Casey, H.; Desmond, D.; Gallagher, P. The impact of assistive technology use for students with disabilities in higher education: A systematic review. Disabil. Rehabil. Assist. Technol. 2021, 16, 130–143.

[19] Núñez-Canal, M.; De Obesso, M.; Pérez-Rivero, C.A. New challenges in higher education: A study of the digital competence of educators in covid times. Technol. Forecast. Soc. Chang. 2022, 174, 121270.

[20] Bagur-Femenías, L.; Buil-Fabrega, M.; Aznar, J.P. Teaching digital natives to acquire competences for sustainable development. Int. J. Sustain. High. Educ. 2020, 21, 1053–1069.

[21] Aquino, K.C.; BuShell, S. Device usage and accessible technology needs for posttraditional students in the e-learning

environment. J. Contin. High. Educ. 2020, 68, 101–116.

[22] Scott, L.; Bruno, L.; Gokita, T.; Thoma, C.A. Teacher candidates' abilities to develop universal design for learning and universal design for transition lesson plans. Int. J. Incl. Educ. 2022, 26, 333–347.

[23] Seale, J.; Colwell, C.; Coughlan THeiman, T.; Kaspi-Tsahor, D.; Olenik-Shemesh, D. Dreaming in Colour: Disabled Higher Education Students' Perspectives on Improving Design Practices That Would Enable Them to Benefit from Their Use of Technologies. Educ. Inf. Technol. 2021, 26, 1687–1719.

[24] Aliaga, C.L.; Dávila, O.M. La plataforma Blackboard: Una herramienta para el proceso de enseñanza-aprendizaje. Hamut'ay 2021, 8, 42–58.

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)
Karnataka, India. Jun 28-29, 2024

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) | 979-8-3503-8269-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITEICS61368.2024.10625656

An Efficient Approach for News Summarization using Map-Reduce and Clustering

Mukesh Kumar[1]
CSE Department
Graphic Era Hill University,
Dehradun, India
kumarsidcse@gmail.com

Saurabh Thapliyal[2]
CSE Department
Graphic Era Hill University,
Dehradun, India
saurabhthapliyal20011682@gmail.com

Anurag Agarwal[3]
CSE Department
Graphic Era Hill University,
Dehradun, India
anuraragarwal188@gmail.com

Anshul Rawat[4]
CSE Department
Graphic Era Hill University,
Dehradun, India
anshulrawat541@gmail.com

Ayush Uniyal[5]
CSE Department
Graphic Era Hill University,
Dehradun, India
ayushuniyal487@gmail.com

Teekam Singh[6]
CSE Department
Graphic Era Deemed to be University,
Dehradun, India
teekamsingh.iitr@gmail.com

***Abstract*—The ability to effectively summarize news articles is crucial in effectively managing the vast amount of available information. This research presents a novel approach to news summarizing, utilizing the combined functionalities of MapReduce and clustering methodologies. The limitations of traditional methods become apparent when faced with the vast and varied nature of news material, resulting in less-than-optimal summarizing results. The methodology employed in this study involves utilization of the MapReduce architecture for the simultaneous processing of extensive textual data. During the Map phase, summaries are generated in an independent manner by dividing news articles into smaller, more manageable segments. The final phase of clustering improves the process of summarization by grouping sentences that are semantically similar inside each chunk. This innovative methodology guarantees not only the capacity for expansion but also an enhanced capability to encompass a wide range of elements present in the original news content. The effectiveness of the suggested MapReduce and Clustering Summarization methodology is demonstrated through experiments conducted on CNN-daily dataset. Our research findings demonstrate enhancements in the quality of summarization, the speed of processing, and the capacity to manage larger datasets in comparison to traditional approaches. We were successfully able to make use of ideas of distributed computing and clustering algorithms which allowed us to generate news summaries that are all-inclusive as well as relevant in their contextual understanding. Our main aim of this research is to dig deep through the application of Map Reduce and clustering methodologies in field of news summarization within a research context.**

Keywords: News Summarization, MapReduce, Clustering, Textual Information, Scalability.

I. INTRODUCTION

The rapid growth and importance of digital information has highlighted the importance of creating and using effective methods for collecting the information content. We know that modern content aggregation systems face serious challenges in delivering broad and diverse content. This work presents a method that used the combined capabilities of MapReduce and clustering algorithm to solve the existing problems and improve the performance and the efficiency of the news media summarization.

The MapReduce framework has become important in recent years due to the computational resources it makes available for processing large data sets. This research extends the application of MapReduce in media by providing a way to divide media into smaller sections during the Map phase. Next, the clustering phase is performed to group semantically similar data segments to create a shared representation of different content in the data. Then after segregating the data into multiple clusters then data segment closest to the cluster head is taken as the representative of the cluster, when then in turn is summarized in sequential manner with the other cluster representatives selected. Then the summary of each segment selected is first generated using the trained transformer model and at last all the summaries generated are aggregated to give the final output summary. This research is important because it can address the limitations of traditional compilation methods. Our approach aims to optimize content quality based on speed and scalability by combining distributed computing and cluster technologies. The effectiveness of the proposed method has been evaluated by practical tests on various datasets. The purpose of this article is to analyse in detail the MapReduce implementation and its integration with media. The following sections examine the study process, present experimental data and analyse the results.

II. RELATED WORK

With the basis of statistical relevance, this methodology uses the processes of tokenization, elimination of stop words and the assigning of weights to sentences. The summarization process will include the ranking of sentences, computing the weighted frequency of the document and extraction of sentences with the highest ranks. The implementation system was developed using python 3.6 and Natural Language Toolkit (NLTK), in which it also underwent testing on the dataset which consists of five texts. Every document consists of total 20 sentences, which was created by J.N. Madhuri and Ganesh Kumar R. The researcher initializes generating sentences with rank greater than 9, which were able to produce output in audio format. The extractive summarization model not only enhance precision as compared to old conventional method, but also resulting in generating excellent summary of the given content [5].

In this academic research, Pengcheng Liao, Chuang Zhang, Xiaojun Chen, and Xiaofei Zhou suggest a new aggregation method to increase Transformer model memory. In abstractive summarization of large texts, the suggested technique outperforms the Transformer baseline model and Pointer

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 01,2024 at 10:48:25 UTC from IEEE Xplore. Restrictions apply.

Generator. This is indicated by increased CNN/Daily Mail ROUGE scores. This research examines addition, projection, and attention aggregation. According to the results, the focus strategy works best. Various aggregation layers are examined to optimize performance. The authors also constructed a Chinese dataset for summarizing and found that their algorithm performed better on it. Memory networks for historical data capture are potential research topics. The aggregation procedure should also be used in other generation operations [6].

In this academic paper, Xin Liu and Liutong Xv provide LTABS, a new sequence-to-sequence model for text summarization. LSTM and Transformer architectures are combined to overcome transformer models' limitations in capturing local attention information in the LTABS model. The encoder uses LSTM hidden layer output to include location information, increasing transformer local attention. The method also includes a copy mechanism to address out-of-vocabulary (OOV) issues. The CNN, Daily Mail, and Xsum datasets show that LTABS is better in semantic and syntactic structures. It performs better than advanced models in ROUGE and manual language quality assessments. The mixed model shows that LSTM positioning information in the transformer architecture creates a durable text summarizing framework [7].

Ong Sing Goh, Puspalata C Suppiah, Yogan Jaya Kumar, Ngo Hea Choon, and Halizah Basiron. spearheaded 50 years of automatic text summarizing research. Key topics, summaries, technique, and specific applications are covered in the inquiry. Language extraction, domain-specific summarizing, and multi-document summarization are intensively studied. From the 1990s SUMMONS system, this paper evaluates news summarizing's evolution. It shows how summary techniques have evolved. Grouping-based methods increase diversity and reduce redundancy among articles, according to the study. The latter points emphasize the importance of understanding existing methodology restrictions and predicting future advancements, with a focus on improving summaries' substance and interpretability [8].

This scholarly article, written by Gupta, Chugh, Anjum, and Katarya, explores the growing demand for automated text summarization in response to the exponential growth of online textual information. This work conducts a comparative analysis of transformer-based pre-trained models, including Pipeline, BART, T5, and PEGASUS, in the context of text summarization. The evaluation is performed using the BBC news dataset. The analysis considers ROUGE ratings and summaries given by humans, demonstrating that the T5 model has superior performance compared to other models. The study highlights the significance of automated summarization in improving the efficiency of recommender and text categorization systems. To enhance accuracy and coherence, future research endeavors may focus on investigating variable-length summaries, multi-document summarizing, and hybrid model techniques [9].

Saeed, Awais, Talib, and Younas introduce a unique natural language processing text summary approach in this paper. Dynamically grouping unstructured texts improves corpus processing. The approach under consideration involves 10 consecutive steps: pre-processing, corpora formation, clustering, and sub-corpora path generation. This method creates a linked corpus from collected text and metadata, increasing content recognition and saving processing costs and time. The proposed technique achieves a text satisfaction rate of 67% on a news article dataset, outperforming earlier algorithms. The combination of abstractive and extractive summarization elements, along with the implementation of dynamic corpus construction, demonstrates encouraging outcomes in enhancing text processing capabilities [10].

Krushnadeo Tanaji Belerao, a research scholar at Shree JJT University in Jhunjhunu, Rajasthan, India, wrote "Summarization leveraging the MapReduce Framework employing Big Data and a Hybrid Algorithm (HMM and DBSCAN)". This study is supervised by Dr. S. B. Chaudhari. This work introduces a novel method for abstracting large document sets. The suggested system uses MapReduce and a hybrid technique that uses HMM for summarization and DBSCAN for clustering. The authors focus on topic-based abstract summarization in big data analysis and the challenges of organizing and processing enormous amounts of digital data. This work provides a modular distributed computing-based summarizing method to overcome processing complexity issues. Hadoop and related programming methods enable this. [11]

Big data is clustered with "Hierarchical Density-Based Clustering using MapReduce". Quadratic lower limits reduce hierarchical clustering costs. “MapReduce Density hierarchically clusters big data. Hierarchical clustering's computing cost and scalability are limited by quadratic lower limits. Parallelize MapReduce HDBSCAN* to estimate clustering. This study combines recursive sampling and HDBSCAN*, a cutting-edge hierarchical density-based clustering algorithm. The study compares real Random Blocks to two random sample and data bubble approximations. Similar clustering accuracy and faster processing for large, dispersed datasets with approximation tweaks. This study compares HDBSCAN* top-down and bottom-up hierarchies to demonstrate MapReduce's benefits. They recommend parallelizing HDBSCAN* for shared memory systems to boost computation. Many methods are not scalable. MapReduce HDBSCAN* parallelizes approximation clustering hierarchy calculation. This study uses HDBSCAN* hierarchical density-based clustering and recursive sampling. This study compares three MapReduce methods. Random blocks are exact. Data bubbles resemble random sampling. Studies show precise approximations reduce processing time. Expert in analyzing vast, scattered datasets. Assess MapReduce's bottom-up efficiency by comparing top-down and bottom-up HDBSCAN* hierarchy creation. The study implies HDBSCAN* parallelization on shared memory architectures may boost computing efficiency.[12]

As illustrated in "An Efficient MapReduce-based Adaptive K-Means Clustering for Large Dataset", data scientists struggle with K-means clustering algorithms on huge datasets. The exponential growth of data and variety reveals K-means' seed selection and outlier sensitivity. The paper proposes MapReduce-based adaptive K-means clustering for large datasets to address the difficulties above. Outliers and seed selection are fixed. Statistics pick dataset-wide starting seeds to improve clustering. Statistics show that the impact of integration is decreasing. Concurrent processing makes MapReduce suitable for large data sets. The proposed scheme was evaluated on the UCI Machine Learning Library dataset. Cluster quality and computation time affect performance. The scheme is faster than concurrent K-means, thus saving time. By using the planning process, efficiency

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 01,2024 at 10:48:25 UTC from IEEE Xplore. Restrictions apply.

measures and safety measures are improved and made more transparent to the public. This project offers changes in data augmentation. Adaptive K-means can handle large datasets. [13]

“An Evaluation of MapReduce Framework in Cluster Analysis" talks about Data mining is about integration because digital challenges create data and require better processing Optimizing large datasets using Hadoop MapReduce. The growth of big data requires scalable, performance-oriented computing. K-means clustering is an advanced data analysis. In this study, Hadoop MapReduce language is introduced. We need K-means integration and dataset efficiency. K-means cluster selection, centroid generation, distance calculation, roles and optimization are briefly discussed. It is recommended to use the section as the integration cannot store detailed information. In this article Map and reduce are used to define MapReduce. HDFS runs Hadoop. HDFS is recommended for big data processing due to its fail-safe, efficient, high performance and scalability. We tested MapReduce-based K-means using 3-10 Hadoop nodes and 100,000-300,000 2D points. Clustering makes all testing possible. These tables show the group's working hours and data. Hadoop cluster operation tested. In contrast, rapid deployment and big data continue to grow. Hadoop development research is extensive and diverse. K-means and Hadoop Map reduce cluster analysis performance. We study distributed computing and large groups in big data. [14]

The work "A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization" proposes an innovative method to overcome the constraint of producing summaries with a certain length in current abstractive summarizing systems. Text segmentation and Transformer-based summarization modules are integrated in a two-stage procedure. Text segmentation uses pre-trained BERT and bidirectional LSTM models to divide input text. An extractive model extracts important sentences from segmented material. Segment and document summarization modules are used in collaborative two-stage summarizing paradigm training. The ChWiki_181k and LCSTS datasets showed promising results, with a maximum subjective judgment accuracy of 70.0%. The study stresses the need for more summary datasets with different information. It also covers short article summarization issues and suggests future research topics. These sections increase the model's capacity to provide user-preferred abstractive summaries of various durations. [15]

The "Abstractive Web News Summarization Using Knowledge Graphs" study addresses online news data overload from long text pieces without summaries. This method of solving multiple data problems uses information connections and image algorithms. Knowledge graphs track reading and reasoning. A review of evolving extraction and abstraction methods for document collection. The authors recommend the use of image-based methods, especially semantic graph-based methods, to obtain consistent, valid abstract concepts. The review states that Association Rule Mining (ARM) can detect associations in malicious objects. ARM and knowledge maps enable better evaluation of critical content. This specification is for limitations. To reduce the amount of information in the information image, the author provides clear and transparent content used in various media properties. Text preprocessing, ARM applications, knowledge network design, language semantic sorting algorithm, using abstract content and manual evaluation. The combination of NLP and data mining makes the difference

Online menu. Multi-dimensional abstract summaries and frequent updates are easy to consume. [16]

Fahad Maqbool, Muhammad Azeem Abbas, Sharifullah Khan, and Rubab Hafeez introduce a novel multi-document summarizing method in this paper. Agglomerative hierarchical clustering and Latent Semantic Analysis are used in this method. The proposed method effectively addresses redundancy and sentence extraction in summarizing. Latent Dirichlet Allocation (LDA) is used to locate topic-related phrases in the summary. The Recall-Oriented Understudy for Gisting Evaluation (ROUGE) measure is used to evaluate the system using 2004 Document Understanding Conference (DUC) dataset. The results of those experiments showed a considerable performance improvement over state-of-the-art method. This method uses Latent Semantic Analysis (LSA) and clustering to reduce duplication, manage large feature sets, and identify subjects. These results show that this method may efficiently and cohesively summarize multi-documents.[17]

In this academic publication, Akter, Asa, Uddin, Hossain, Roy, and Afjal propose a novel Bengali extracted text summary method. Individual and many documents can use the proposed technique. Data is tokenized and stemmed first. Additionally, cue and skeleton words are considered when rating sentences using TF/IDF. Following K-means clustering, the final summary yields good results compared to current methods. A linear time complexity makes the suggested summarizing approach efficient. Summary contains 30% of unique material terms. The method seems promising for summarizing similar material from different sources, despite occasional phrase sequence desynchronization. Similarity assessments could improve grouping and summarizing [18].

The paper "Extractive Summarization using Supervised and Unsupervised Learning." by Xiangke Mao et al. analyzes unique single-document summarizing methods. Online resources are constantly expanding and require quality content writing. Statistical properties and sentence combining are used in three ways to determine the importance of sentences. Initially, the tracking model and charts measure text independently and then combine them. To evaluate the impact of the article, the second method uses a graphical model to characterize the attention pattern. Phrase tracking is a third-party process. Negative image models use evaluation scores based on previous values. The final sentence test is the unbiased use of LexRank. ROUGE Realizes both checked and unchecked contents of DUC2001 and DUC2002 files. The authors suggest that prior knowledge improves sentence structure selection. This study investigates content expression using supervised and unsupervised learning methods and a novel sentence evaluation technique. The results show that the strategy can provide better data collection. Future work should increase the network node value and use the model for longer documents. [19]

Rahul, Shristi Rauniyar, and Monika from Delhi Technological University's Department of Computer Science and Engineering examine text summarization methods, focusing on deep learning's rapid growth. The authors discuss managing the massive volume of textual data generated online and the need for effective text summarization systems. This paper examines abstractive (ABSR) and extractive (EXTR) summarization methods using neural networks (NN), fuzzy C-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 01,2024 at 10:48:25 UTC from IEEE Xplore. Restrictions apply.

means clustering, deep learning and others. These methods are supervised or unsupervised. Text summarizing improves user experiences, reduces information overload, and supports sentiment analysis (SEN). The authors examine numerous scholars' methods, including phrase clustering and word embedding emotion analysis. This review analyzes the different datasets used in these investigations, including social media and conference papers. After analyzing the pros and cons of various methods, the authors conclude that EXTR methods, particularly machine learning ones, are more effective. Academics and practitioners benefit from the literature review, which sheds light on current trends, achievements, and future research in text summarizing. [20]

III. DATASET AND METHODOLOGY

A. Dataset Description

A wide variety of news stories covering a range of subjects, writing styles, and domains may be found in the CNN/Daily Mail collection. The dataset offers a broad and diverse corpus for natural language processing model training, with a significant amount of news articles. This abundance facilitates the development of strong models that comprehend a wide range of linguistic patterns. Many researchers utilise this dataset [1][2] for their models. CNN/Daily Mail is what we are using for our transformer model. 286,817 news articles from the CNNDM dataset—which was retrieved from the Hugging Face website—were utilised to train the model. 13,368 documents are used as validation data to facilitate normal review, while 11,490 documents are reserved for testing. It is interesting to note that each document in the sample contains roughly 28 sentence pairings. We can also depict from the figure (1) that most of the articles are of length 400-800.

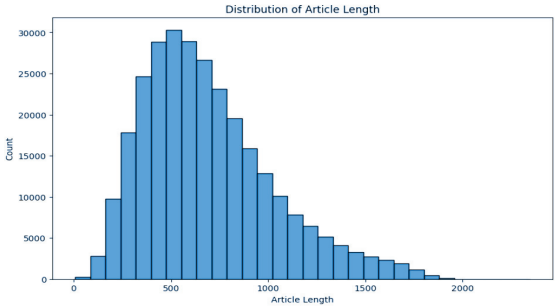


Fig. 1. Distribution of Articles length in dataset

B. Dataset Preprocessing

This dataset was carefully cleaned using several preprocessing approaches. It consisted of long news articles combined with brief summaries for comparative analysis.

- Lower Casing: To handle varying casing consistently, all text items should be converted to a standard lowercase format.
- Removal of Stop words and Words That Often Occur: Elimination of terms that are commonly used in the text but offer little useful information for further analysis, including "the" and "a."
- Stemming is the process of returning inflected words to their base forms to simplify and group related words together.
- Lemmatization is the process of reducing derivative words to their most basic or root form while maintaining grammatical accuracy.
- Contraction Mapping is the process of expanding abbreviated words or syllables to increase the dataset's comprehensiveness. For example, "he'll, she'll, I'll, it'd" might be changed to "he will, she will, I will, it would." [1]

C. Model Hyperparameters:

The Transformer model was utilized in a series of experiments, wherein the training and testing phases were exclusively performed on the CNN/Daily-mail dataset. To attain the highest level of precision while avoiding overfitting, the model underwent training for numerous epochs. During the training phase, we conducted fine-tuning on multiple hyperparameters, which involved configuring an input target length of 512 tokens and an output summary length of 128 tokens. The optimization process utilized the Adam optimizer with a learning rate of 1e-3. Each epoch consisted of training the model using a batch size of 8. The assessment of the proposed model involved the consistent deployment and tuning of hyperparameters, which were especially suited to the CNN/Daily-mail dataset. This approach ensured a comprehensive evaluation of the model's effectiveness in handling this specific data source.

D. Basic Understanding of Model:

The Transformer architecture represents an innovative paradigm that effectively utilizes a self-attention mechanism. The distinctive method of this model enables it to selectively highlight various segments within the input sequence, therefore aiding in the comprehensive understanding of contextual complexities throughout the entire document. It is of utmost importance to acknowledge that the Transformer paradigm deviates from traditional methodologies, such as Recurrent Neural Networks (RNNs). In contrast, the approach deliberately utilizes attention layers and positional encoding techniques to efficiently manage the storage and retrieval of sequential data. The deviation from sequential models plays a crucial role in facilitating parallelization and significantly improving the model's ability to understand and depict complex relationships within textual material. From figure (2) we can see the proposed workflow of the model.

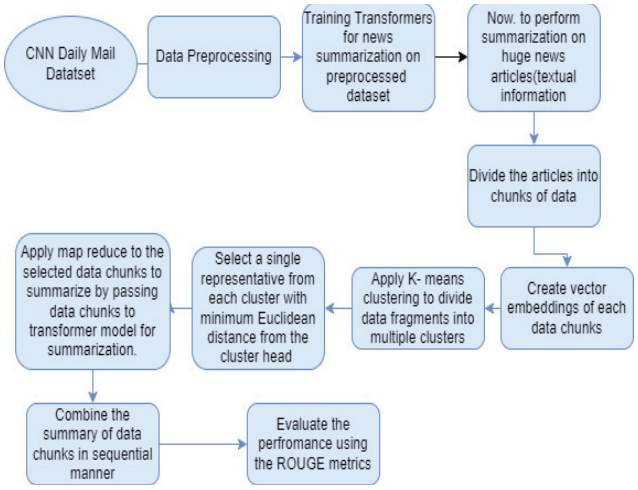


Fig. 2. Proposed Model

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 01,2024 at 10:48:25 UTC from IEEE Xplore. Restrictions apply.

E. Model Architecture [2]:

As we can see the architecture of the transformer model comprises of the interconnected encoder-decoder layers, which are linked to the multi-head attention layer, and then subsequently followed by a feed-forward network, as it is visually shown in Figure (3) below.

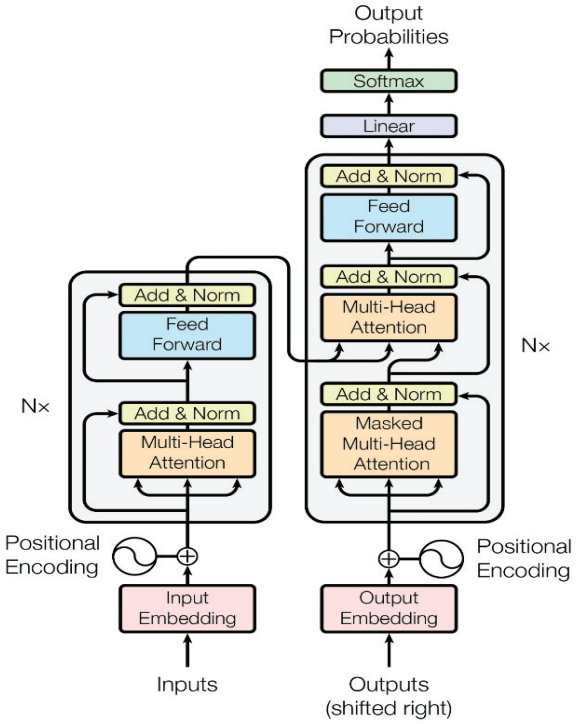


Fig. 3. Architecture of the Transformer [2]

1) Encoder block:

The Transformer encoder block comprises of six identical layers, first it processes input sequences through a multi-head self-attention mechanism then employs a residual connection and layer normalization which ensures the stable information flow, with output dimensions consistently set to 512.

2) Decoder block:

From the figure (3) structure we can see that decoder block is expanded by the incorporation of a supplementary sub-layer. These supplementary sub-layers employ the multi-head attention on the output of the encoder stack. Like the encoder, the decoder employs residual connections and layer normalization as well.

3) Attention block:

Now, coming to the attention function, which is calculated on a matrix Q, which contains a collection of questions that are processed simultaneously. Then the values and the keys are integrated into the matrix K and V. The scaled dot-product attention in the transformer model of the matrix is computed as follows:

$$\text{Attention}(K, V, Q) = \text{SoftMx}\left(\frac{QK^T}{\sqrt{dk}}\right)V \quad \text{.....(i)}$$

The Transformer architecture incorporates multi-head attention as a crucial step to improve the model's power to collectively attend to input from various representation subspaces and places. Instead, then utilizing a singular attention function, we use multi-head attention that requires the linear projection of all the keys (K), values (V) and queries (Q) h times using the separate learnt linear projections. This results in dimensions dk, dk, and dv for each corresponding projection. Below is the Multi-Head Concatenated Attention calculated in the transformer model:

$$\text{head}_i = \text{Attention}(KW_i^K, V\,W_i^V, QW_i^Q)$$
$$\text{MH}(K, V, Q) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\,W^o \quad \text{.....(ii)}$$

4) Positional encoding:

As our model does not apply convolution and recurrence method, we need a way to give the model knowledge about the token's order in a sequence. We apply (positional encodings) into the input embeddings at the encoder and decoder stacks to keep the sequence consistent. These positional encodings are then added wordwise and have the similar dimension (i.e., 512) as the embeddings. We choose to use sinusoidal positional encodings, which are described by sin and cos functions with different frequencies in below equations:

$$PE(pos,2i) = \sin(pos/(10000^{2i/dmodel})) \text{ and}$$
$$PE(pos,2i+1) = \cos(pos/(10000^{2i/dmodel})) \quad \text{.....(iii)}$$

where, 'i' represents the dimension,'dmodel' represents the vector size of embedding and 'pos' represents the location of the word. The wavelength of the positional encoding contains a geometric progression from 2π to $10000\times2\pi$, where each dimension corresponds to a sinusoidal wave. Because of this the model can perform better and able to retain the sequences based on the relative positions with the help of the method of positional encoding.

5) Feed forward:

Now at last, coming to the feed-forward layer it is a fully connected layer which we incorporate in the Transformer model's encoder and decoder block, including the attention sub-layers. This network has two linear transformations which we then apply equally and independently to each point, separated by an activation function which is better known as the Rectified Linear Unit (ReLU). This Feed-Forward Transformation with a ReLU activation in between the multiple encoder and decoder blocks mathematical expression is below:

$$FF(x) = \max(0, xW_1+b_1)\,W_2+b_2 \quad \text{.....(iv)}$$

Where, W_1 and b_1 are the weight and bias of our first linear transformation followed by the next linear transformation using weight W_2 and bias b_2.

F. MapReduce in Text Summarization

Our learned transformer model, when combined with the MapReduce method, offers a dependable method for effectively compressing substantial volumes of textual data. The MapReduce design, which was influenced by Dean and Ghemawat's work [3], allows data to be processed concurrently across distributed clusters, thereby overcoming the difficulties associated with handling large-scale text summarization tasks. The next section provides a summary of the details of the implementation. he data partitioning process involves splitting large data into smaller, more manageable pieces, called input partitioning. In the step-by-step report, a pre-learned transformer model summarizes the input of each section. After receiving the key information intermediate key/value pairs will be created. The average results are analysed and transformed before being rearranged and analyzed using present keys. This process is used to

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 01,2024 at 10:48:25 UTC from IEEE Xplore. Restrictions apply.

collect and organize relevant information. In one simple step, with the help of integration of the transformer model, the product is analysed and a clear and precise definition is created. Various experiments have been conducted using different text lengths to test the performance of MapReduce on text. The study's findings show that using the MapReduce methodology greatly increases summarization's effectiveness, proving the system's ability to manage enormous volumes of textual data.

G. *Utilizing MapReduce in Conjunction with Clustering for Augmented Summarization*

We observe a set of MapReduce and Clustering strategies to improve the way of summarizing large sets of text data. This approach offers a great way to improve the performance, scalability and overall performance of your short product. This method is especially useful if a large portion of the text contains content that you want to break into smaller pieces to fit Transformer version limits. Early preprocessing involves splitting raw content into logical, symbolic elements. These points are important because they form the basis of subsequent modelling. All documents need to be kept clean to ensure they are accurate and relevant. Integration of the MapReduce framework makes it easier to implement specific computational models and allows aggregation across groups of code numbers. The dataset differs from the content of the text distributed among nodes throughout the level map.

Each node specializes in its own vectorization, using techniques such as Word2Vec to combine elements in text. To ensure consistency, the summary vectors counted between nodes are aggregated as a reduced level. Then, a clustering algorithm, such as K-means, is used to merge these vectors in parallel across multiple nodes.

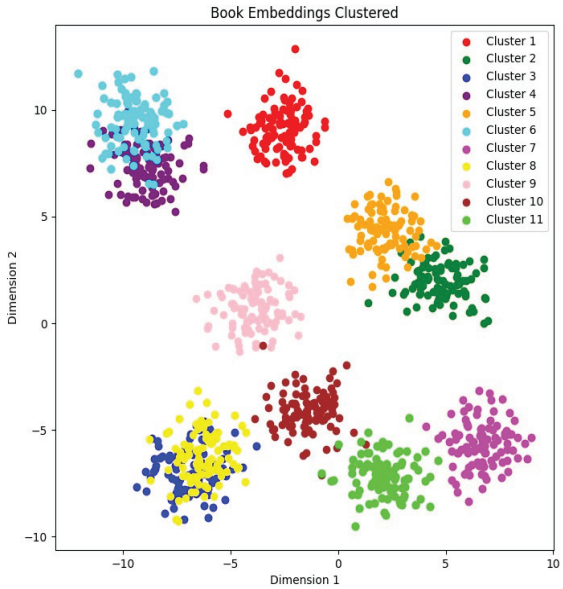


Fig. 4. Illustration showing a collection of distinct data fragments from a book in clusters.

In Figure (4), we can see how data objects are combined to display a 2D image.

Therefore, the set of global centroids is designed to have a good effect on the distribution of all data. Clustering ideas in individual components promotes semantic cohesion. Selecting the segment with the shortest Euclidean distance from each cluster to the global center is called representative segment selection. This ensures that the selected sections complement the main themes of the broader themes within each category. Sequential synthesis is the process of gradually synthesizing representative sources containing the main points of each group, one group at a time. Using sequential instructions, specific payment is made for the remaining items selected as a group on the order form. This process ensures content and assembly that includes many aspects of the full text. The ideas given in this work increase the efficiency and scalability of the synthesis. It overcomes the limitations caused by the long reach limit of the transformer model and can work well for larger data. Many studies have been conducted on many data showing the performance of the Map-Reduce process with clustering. The results once again show an improvement in current collection methods in terms of collection efficiency, processing speed, and ability to handle large amounts of data. In summary, our proposed method successfully combines the Map-Reduce framework with integration to provide an effective and efficient solution for segmented summarization of multiple datasets. With the cooperation of the group, content is created from a common structure that presents intellectual property and soft information.

IV. RESULTS

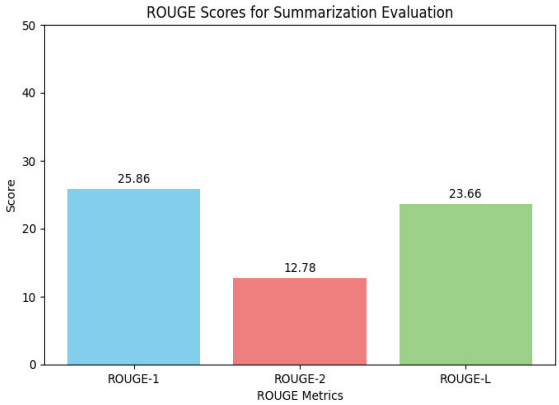


Fig. 5. The ROUGE Metric [4] for summarization *evaluation*

Using the CNN/Daily Mail dataset, the assessment metrics offer insightful information on how well the trained transformer model performs in news summarizing. In figure (5) we can see the produced ROUGE scores—ROUGE-1: 0.25, ROUGE-2: 0.12, and ROUGE-L: 0.23—act as numerical gauges for evaluating the caliber of the summaries that are produced.

ROUGE-1: Overlap of Unigrams The average percentage of unigrams in the generated summaries that overlap with those in the reference summaries is 25%, according to the ROUGE-1 score of 0.25. This indicates a modest degree of word-level content agreement.

ROUGE-2 (Overlap of Bigram): The model captures sequences of two adjacent words and shows a smaller overlap in bigrams, with a ROUGE-2 score of 0.12. Even while bigram overlap is typically more difficult, the model is nevertheless able to accurately represent a significant fraction of these linguistic structures.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 01,2024 at 10:48:25 UTC from IEEE Xplore. Restrictions apply.

ROUGE-L (Longest Common Subsequence): The model's generated summaries and the reference summaries based on the longest common subsequence are aligned, as indicated by the ROUGE-L score of 0.23. This metric produces a thorough evaluation by considering word sequences as well as the structure and ordering of words.

V. CONCLUSION AND FUTURE WORK

Using the CNN/Daily Mail dataset, the assessment metrics Rouge-1, Rouge-2, and Rouge-L highlight the transformer model's efficacy in news summarising. The model performs well in unigram alignment (Rouge-1) and preserves structural coherence (Rouge-L); however, it has difficulty capturing exact bigram overlap (Rouge-2). The foundation for upcoming advancements and research paths is laid by these results. Subsequent research endeavours must concentrate on enhancing the model by means of fine-tuning its parameters, expanding its dataset, and integrating domain-specific expertise. Furthermore, human assessment studies will offer a subjective perspective on the level of quality and information contained in the summaries that are generated.

Future research endeavours to improve the performance of the transformer model by tackling these factors comprehensively. This will enable the model to be more versatile and applicable to a wider range of news genres, as well as real-world information management scenarios. This iterative process of extension and improvement ensures that news summary approaches continue to evolve in line with the dynamic nature of natural language processing research.

REFERENCES

[1] T. T, M. Bohra, P. Dadure, and P. Pakray, “Comparative analysis of T5 model for abstractive text summarization on different datasets,” SSRN Electronic Journal, 2022 doi: https://doi.org/10.2139/ssrn.4096413.

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, December 5). Attention Is All You Need. ArXiv.org. doi: https://doi.org/10.48550/arXiv.1706.03762

[3] Dean, Jeffrey, and Sanjay Ghemawat. “MapReduce: Simplified Data Processing on Large Clusters.” Communications of the ACM, vol. 51, no. 1, 1 Jan. 2008, pp. 107–113, dl.acm.org/citation.cfm?id=1327492, https://doi.org/10.1145/1327452.1327492.

[4] Lin, C.-Y. (2004, July 1). ROUGE: A Package for Automatic Evaluation of Summaries. ACLWeb; Association for Computational Linguistics. doi: https://aclanthology.org/W04-1013.

[5] Madhuri, J. N., & Ganesh Kumar, R. (2019, March 1). Extractive Text Summarization Using Sentence Ranking. IEEE Xplore. doi: https://doi.org/10.1109/IconDSC.2019.8817040

[6] Liao, P., Zhang, C., Chen, X., & Zhou, X. (2019). Improving Abstractive Text Summarization with History Aggregation. ArXiv (Cornell University). doi: https://doi.org/10.48550/arxiv.1912.11046

[7] Liu, X., & Liutong Xv. (2019). Abstract Summarization Based on the Combination of Transformer and LSTM. doi: https://doi.org/10.1109/icicas48597.2019.00199

[8] C. Bhatt, Akanksha, R. Chauhan, D. Kumar, T. Singh and S. Sharma, "Deep Learning for Handwritten Character Recognition," 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), Ghaziabad, India, 2023, pp. 379-385, doi: 10.1109/ICSEIET58677.2023.10303610.

[9] Gupta, A., Chugh, D., Anjum, A., & Rahul Katarya. (2021). Automated News Summarization Using Transformers. ArXiv (Cornell University).doi : https://doi.org/10.48550/arxiv.2108.01064

[10] Saeed, M. Y., Awais, M., Talib, R., & Younas, M. (2020). Unstructured Text Documents Summarization With Multi-Stage Clustering. IEEE Access, 8, 212838–212854.doi: https://doi.org/10.1109/access.2020.3040506

[11] Krushnadeo Tanaji Belerao and S. B. Chaudhari, “Summarization using mapreduce framework based big data and hybrid algorithm (HMM and DBSCAN),” Sep. 2017, doi: https://doi.org/10.1109/icpcsi.2017.8392320.

[12] R. Chauhan, G. S. Wadhwa, D. K. Verma, K. N. Vaghela and C. Bhatt, "Exploring the Intersection of Language Processing and Image Generation in a Conversational AI Application," 2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC), Rajkot, India, 2023, pp. 468-473, doi: 10.1109/R10-HTC57504.2023.10461849.

[13] T. Chowdhury, A. Mukherjee, and Susanta Kumar Chakraborty, “An Efficient MapReduce-Based Adaptive K-Means Clustering for Large Dataset,” Dec. 2017, doi: https://doi.org/10.1109/inis.2017.40.

[14] Tanvir Habib Sardar, Ahmed Rimaz Faizabadi, and Z. Ansari, “An evaluation of MapReduce framework in cluster analysis,” Jul. 2017, doi: https://doi.org/10.1109/icicict1.2017.8342543.

[15] M.-H. Su, C.-H. Wu, and H.-T. Cheng, “A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2061–2072, 2020, doi: https://doi.org/10.1109/taslp.2020.3006731.

[16] M. V. P. T. Lakshika, H. A. Caldera, and W. V. Welgama, “Abstractive Web News Summarization Using Knowledge Graphs,” Nov. 2020, doi: https://doi.org/10.1109/icter51097.2020.9325453.

[17] Hafeez, R., Khan, S., Muhammad Azeem Abbas, & Maqbool, F. (2018). Topic based Summarization of Multiple Documents using Semantic Analysis and Clustering. doi: https://doi.org/10.1109/honet.2018.8551325

[18] Akter, S., Aysa Siddika, Md. Palash Uddin, Hossain, D., Shikhor Kumer Roy, & Masud Ibn Afjal. (2017). An extractive text summarization technique for Bengali document(s) using K-means clustering algorithm. doi: https://doi.org/10.1109/icivpr.2017.7890883

[19] X. Mao, H. Yang, S. Huang, Y. Liu, and R. Li, “Extractive summarization using supervised and unsupervised learning,” Expert Systems with Applications, vol. 133, pp. 173–181, Nov. 2019, doi: https://doi.org/10.1016/j.eswa.2019.05.011.

[20] Rahul, S. Rauniyar, and Monika, “A Survey on Deep Learning based Various Methods Analysis of Text Summarization,” IEEE Xplore, Feb. 01, 2020. https://ieeexplore.ieee.org/document/9112474 (accessed May 13, 2023).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 01,2024 at 10:48:25 UTC from IEEE Xplore. Restrictions apply.

Analysing the anomalous relationship between precipitation and PM_{10} concentrations in Peninsular India

Manvi Nilaya[1], *Vaibhavi*[1], *Swayam* Agrawal[1], and *Lovleen* Gupta[1*]

[1]Department of Environmental Engineering, Delhi Technological University, New Delhi, India

Abstract. Global air quality trends suggest decreased PM_{10} levels with increased precipitation, however, Peninsular India exhibits a distinct pattern. This research investigates the correlation between precipitation and PM_{10} concentrations in Peninsular India, with special focus on the anomalous relationship between the two variables, where rainfall does not consistently lead to lower PM_{10} concentration levels. The study examines data from 11 air quality monitoring stations across four states, considering PM_{10} concentrations, precipitation intensities and meteorological data of wind. Statistical analyses were employed to study the distribution patterns of PM_{10} mass concentrations corresponding to the type of rainfall during the monsoon season in every city of research. To investigate further, the correlation between PM_{10} levels before the rainfall and the average precipitation intensity during the rainfall duration was established. Results have revealed that the impact of precipitation on PM_{10} concentration levels varies depending upon the pollutant's concentration before the rainfall duration and precipitation intensities. Light rainfall (precipitation intensity between 2.5 mm/day to 7.6 mm/day) leads to a decrease in PM_{10} concentration levels in particular meteorological conditions, whereas very light rainfall (precipitation intensity less than 2.5 mm/day) increases the PM_{10} concentration levels post rainfall. The scavenging effect of precipitation is investigated by measuring the difference in PM_{10} concentration levels before and after a certain precipitation intensity. When initial PM_{10} concentration is lower than the seasonal average, results show a little increase in PM_{10} concentration levels after a light rainfall lasting less than 10 hours. This research also explores the fluctuations in the PM_{10} concentration levels under different wind directions and speeds. It is observed that increasing wind speed in one direction improves the clearance effect on PM_{10} concentrations. The objective of this research is to close the knowledge gap between regional meteorological conditions (primarily related to precipitation and wind dynamics) and regional air quality by offering data-backed observations. This facilitates effective air pollution management for such disparate air quality patterns.

Keywords. Air Quality, Anomalous Relationship, Peninsular India. PM_{10}, Precipitation

*Corresponding author : lgupta@dce.ac.in

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

1 Introduction

Particulate Matter10 (PM_{10}) is one of the major air pollutants which has negative effects on human health [1]. Previous research suggests a correlation between exposure to high PM_{10} concentration levels and adverse health effects like asthma and other allergic conditions affecting the immune system [2]. Spatial and temporal variations are observed in PM_{10} concentrations due to meteorological factors, topography, and local emission sources [3]. Anthropological sources like power plants, industrial burning, and household biomass burning are primary contributors to rising PM_{10} concentration levels [4]. Central Pollution Control Board (CPCB) under the Ministry of Environment, Forests, and Climate Change has set National Ambient Air Quality Standards (NAAQS) to regulate and monitor PM10 concentrations in India. The annual average PM_{10} concentration should not exceed 60 micrograms per cubic meter (ugm-3) and the 24 hours average should not exceed 100 micrograms per cubic meter more than once per year on an average over a three-year period [5]. A nationwide network of air quality monitoring stations is operated under state pollution control boards and CPCB to track compliance with these standards.

2 Literature Review

Previous research works have explored the influence of meteorological factors such as wind speed, wind direction, precipitation, temperature, and humidity on PM_{10} concentration levels. The distribution of PM_{10} particles is impacted by precipitation in two different ways, subject to other meteorological conditions. Generally, the first way involves the removal of PM_{10} particles from the atmosphere through the process of wet deposition. The other way leads to an increase in PM_{10} levels due to particle formation and growth under specific conditions. Wind speed and direction also play a pivotal role in the dispersion and transportation of PM_{10} particles [6]. Low wind speeds cause poor dispersion of PM_{10}, resulting in higher concentration levels, whereas high wind speeds improve the spreading of PM_{10} and weakens its concentrations. Seasonal fluctuations in rainfall, humidity, temperature, and wind dynamics also affect the concentration and distribution of PM_{10} in the atmosphere [7].

Many previous research studies suggest that high rainfall intensities typically attribute to lower PM_{10} concentration levels in the air through the process of wet scavenging [8]. This result is supported by the observations of the rain composition and the PM_{10} levels demonstrated the efficiency of rainfall in the clearance of PM_{10} from the air [9]. Characteristics of ions in both PM_{10} and rainwater were influenced by factors such as rainfall amount, duration, humidity levels, and wind direction. These findings support the role of rainfall in reducing PM_{10} pollution and associated contamination in regions. Similarly, research in Northern India consistently shows a decrease in PM_{10} levels following rainfall events [10].

In contrast, the states of Kerala, Karnataka, Tamil Nadu, and Andhra Pradesh in Peninsular India exhibit a trend where higher rainfall does not always lead to lower PM_{10} levels. This unique relationship between rainfall and PM_{10} concentrations is the central focus of our research. Previous studies on air quality in South India lack an exploration of how rainfall impacts PM_{10} levels. Despite the understanding of the impact of rainfall on air quality dynamics, limited research has been done to examine the regional factors contributing to this anomaly. Addressing this research gap underscores the necessity for a targeted study using statistical analyses to comprehend the interaction between rainfall, wind patterns, and PM_{10} concentrations.

3 Data And Methodology

3.1 Climate Sources

The region under consideration is Peninsular India encompassing the states of Kerala, Karnataka, Tamil Nadu, and Andhra Pradesh. This area experiences a climate which relies heavily on monsoons for precipitation. With moderate to high year-round temperatures and occasionally substantial rainfall over 2,000 mm annually, the tropical monsoon climate is the most humid. Lengthy periods of above 18 °C (64 °F) during the winter and early summer; summers are extremely hot in low-lying areas, with highs of above 50 °C (122 °F). Rainfall averages in the region range from 750 to 1,500 mm per year during the rainy season, which runs from June to September. Rainfall in the region is mostly caused by the southwest monsoon, which occurs from June to September [11].

3.2 Air Quality Concentration and Meteorological Data

Average hourly PM_{10} concentration is analysed over a 4-month period spanning from June to September for the year 2022. Since, our major focus is to see how PM_{10} concentration behaves in response to varying precipitation, wind velocity as well as direction in the southern peninsula of India, we have focused upon 11 ambient air quality monitoring sites as per Appendix A belonging to 4 different states of Peninsular India: Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu. The coordinates of these sites are encircled in Figure 1. We have limited the analysis only to the monsoon period of these states in order to better capture the relationship between PM_{10} concentrations and precipitation. Even though air pollution has been monitored since 2015 for most of the considered sites, improvement is observed in the quality, quantity and consistency of data within the last five or six years. It is important to note that meteorological data, including wind direction, wind velocity, and precipitation, has been obtained on an hourly basis for the months of June to September, mirroring the PM_{10} data collection frequency. PM_{10}, wind direction, and wind velocity data were obtained from the CPCB website. Precipitation data is sourced from the NASA POWER data access viewer available at the link referred in [12].

4 Results

4.1 Effects of Precipitation of Different Intensities on PM_{10} Concentrations

Statistical analyses were conducted to examine the patterns of PM_{10} mass concentration distribution that correspond to the various types of rainfall that occur each month of the monsoon season for the selected cities of Indian peninsular states. (Figure 2). In accordance with the Indian Meteorological Department (IMD), Pune, rainfall types are categorized based on intensity as follows: Very light rain, where the rate of rainfall ranges from 0 to 2.5 millimetres per day. Light rain, with a rainfall rate ranging from 2.6 to 7.6 millimetres per day. Moderate to heavy rain, characterized by a rainfall rate exceeding 7.6 millimetres per day. In Amravati, Andhra Pradesh, the highest average PM_{10} concentration occurs on non-rainy days while the lowest is observed in moderate rainfall in the months of June and July. The concentration averages of PM_{10} during light and non-rainy days

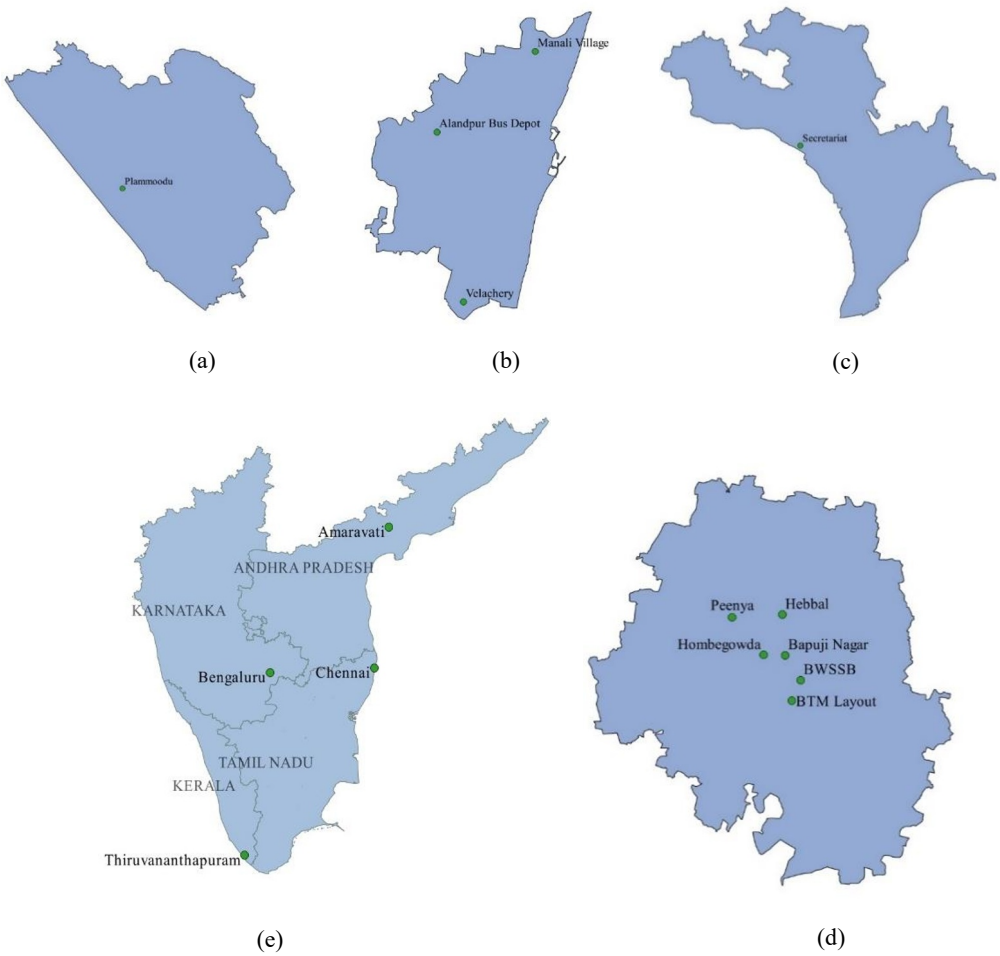


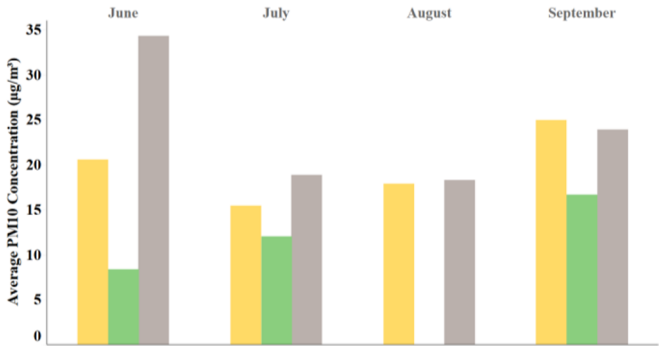
Fig.1. Stations of (a) Thiruvananthapuram, Kerala (b) Chennai, Tamil Nadu (c) Amravati, Andhra Pradesh (d) Bengaluru, Karnataka (e) Peninsular India

are comparable in August while it is higher in light rain days in September. In Bengaluru, Karnataka, peak average PM_{10} concentration is recorded on days with moderate to light rainfall throughout the season whereas the lowest is noted during heavy rainfall in August. Chennai, Tamil Nadu records higher average PM_{10} concentrations in moderate to light rain days than on non-rainy days, except in the month of July. From the results, Thiruvananthapuram in Kerala experiences highest concentrations of PM_{10} in non-rainy days and lowest during moderate rainfall.

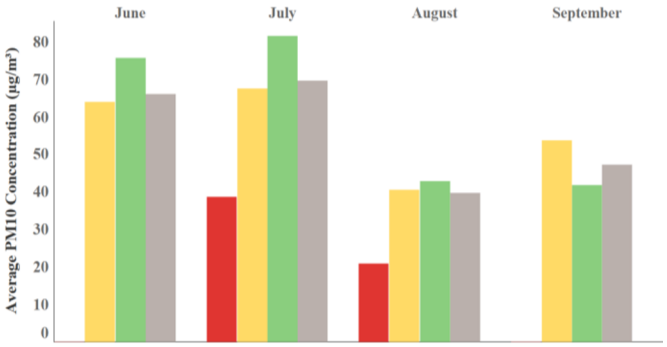
Therefore, a moderate rainfall has noticeable scavenging influence on PM_{10} concentrations in the cities Amravati and Thiruvananthapuram. Furthermore, it is evident that light to moderate rainfall causes the concentration of the pollutant to rise in Chennai. In Bengaluru, light rain days tend to increase PM_{10} concentrations in August and September, while moderate rainfall has a clearance effect on the pollutant. However, an anomalous effect on PM_{10} concentration is observed in Bengaluru during the moderate rainfall days of June.

4.2 Influence of Precipitation on Different Initial Concentrations of PM_{10}

To learn more about how precipitation affects PM_{10} concentrations, 508 rainfall instances were identified in all cities with a precipitation duration of 2 hours or more during the time frame of research. The concentration of PM_{10} before and after precipitation was measured and the difference between the concentration levels was calculated. This value was then plotted against different intensities of rainfall to analyse the impact of precipitation on concentration of the pollutant (Figure 3). In the figure, the blue scatter points represent decreases in PM_{10} concentration during rainfall, whereas red scatter points indicate increases. We divided our analyses into four scopes or ranges of investigation based on the properties of distribution of the scattered points in Figure 3.



(a)



(b)

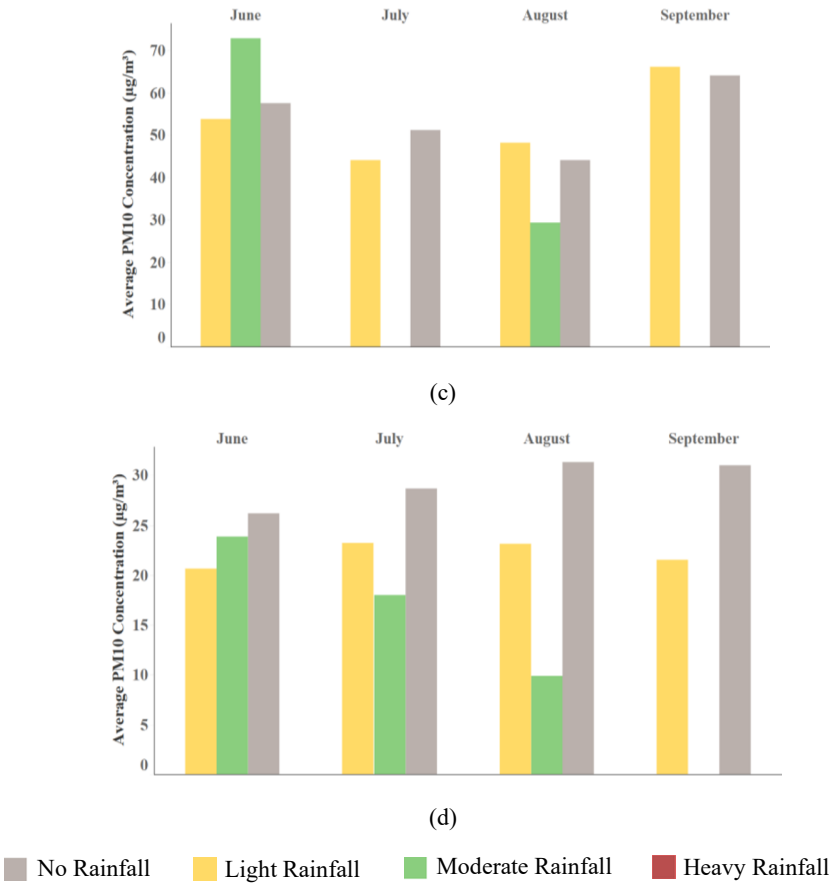


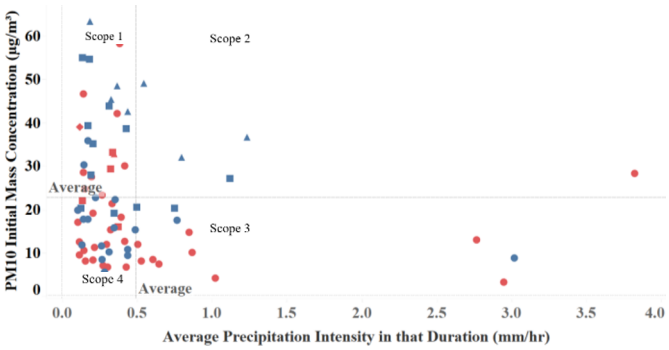
Fig.2. PM_{10} mass concentration between precipitation intensities between June 2022 to September 2022. (a) Amravati, Andhra Pradesh, (b) Bengaluru, Karnataka, (c) Chennai, Tamil Nadu, (d) Thiruvananthapuram, Kerala

Scope 1 is defined when the initial PM_{10} concentration is greater than the season's average concentration and the precipitation intensity is lesser than the average intensity of the season. In Amravati, a total of 26 rainfall instances were identified out of which 14 cases, accounting for 54% displayed a decrease in concentration levels of the pollutant after the rainfall duration. Chennai observes 55 (77.5%) such cases out of 71 total rainfall instances and Bengaluru observes 38 (73%) such cases out of 52 total rainfall instances. In Thiruvananthapuram, 19 cases accounting for 70% of the total observed rainfall instances (27) experience a decrease in PM_{10} concentration levels. This indicates that during period of relatively poor air quality, light precipitation will have an obvious scavenging influence on the pollutant concentration. The higher is the pollutant concentration, the greater is the clearance effect of rainfall.

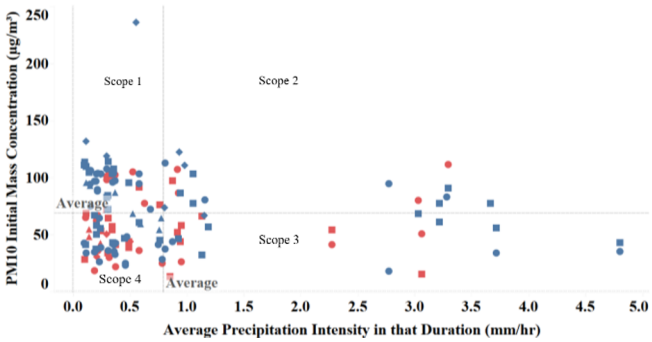
Scope 2 is characterised by an initial PM_{10} concentration and precipitation intensity more than the average of the season. All cities show notable decrease in PM_{10} concentration levels in relatively higher precipitation intensities. Like scope 1, the results indicate the removal of pollutants by precipitation in all cities.

Scope 3 is defined when the initial PM_{10} concentration is less than the average but the precipitation intensity is higher than the average of the season. Amravati observes 13 such rainfall instances, where 9 cases (69%) show an increase or no effect of PM_{10} concentration levels. Among 17 rainfall instances identified in Thiruvananthapuram, PM_{10} concentration increases or stays the same in 11 cases, constituting 65% of the total. Only 10 cases of the total 17 rainfall instances (59%) in Bengaluru result in a small decrease in PM_{10} concentration after rainfall, while the other instances accounting for 41% record an increase in pollutant concentration level after rainfall. In Chennai, 20 cases out of 39 (51%) total rainfall instances display an increase in PM_{10} concentration levels after the rainfall. This suggest that during periods of fine air quality and low initial PM_{10} concentrations, relatively stronger precipitation leads to an increase in PM_{10} concentration levels.

Scope 4 is characterised by an initial PM_{10} concentration and precipitation intensity lesser than the average of the season. In Amravati, 21 cases out of total 36 (58%) rainfall instances in this scope, indicates an increase in PM_{10} concentration level. Bengaluru records 33 out 72 total rainfall cases (46%) and Thiruvananthapuram experiences 16 out of 29 total rainfall instances (55%) where PM_{10} concentration level rises after the rainfall duration. Chennai observes 41 out of 77 (53%) total rainfall instances where either no effect or an increase in the pollutant concentration level is observed. This indicates that during



(a)



(b)

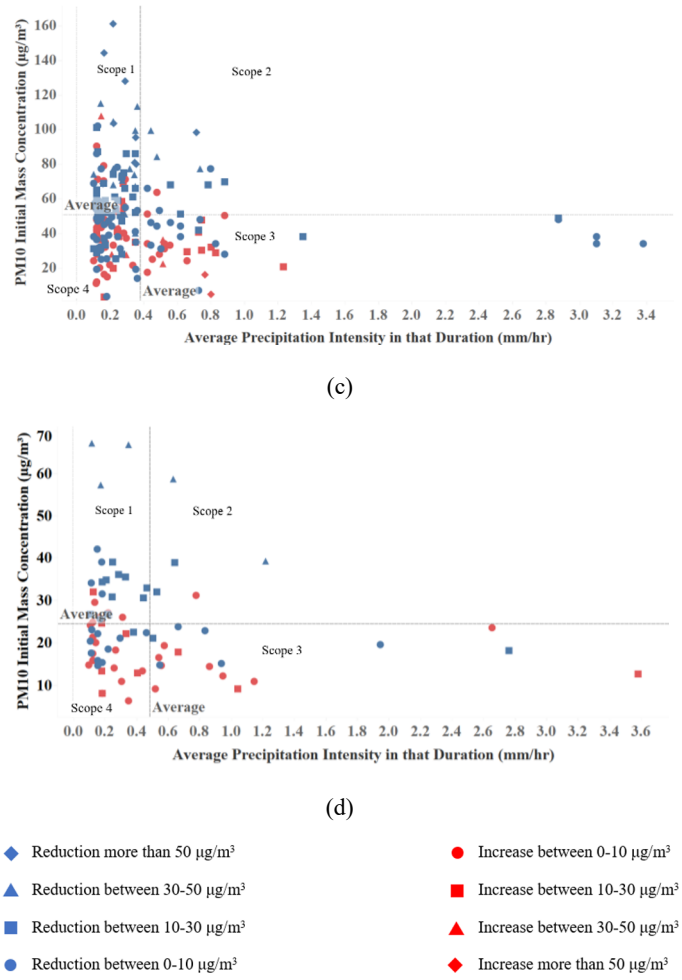


Fig.3. The correlation between the initial PM_{10} concentration and precipitation intensity over the course of rainfall. (a) Amravati, Andhra Pradesh, (b) Bengaluru, Karnataka, (c) Chennai, Tamil Nadu, (d) Thiruvananthapuram, Kerala

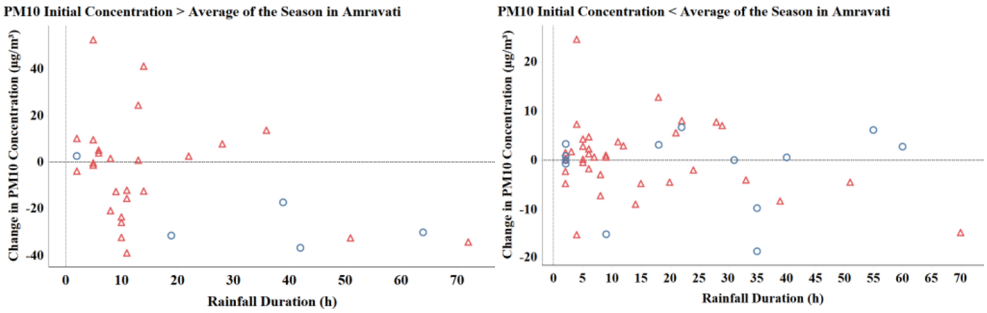
periods of relatively good air quality, light precipitation might lead to an increase in pollutant concentration levels. Consequently, a higher initial PM_{10} concentration corresponds to a greater decrease in the pollutant concentration level during periods of low rainfall.

To investigate precipitation's scavenging effect on PM_{10} concentration in more detail, we examined the correlation between the fluctuation in PM_{10} concentration levels before and after the rainfall with the duration of precipitation with varying intensities (Figure 4). For initial PM_{10} concentration less than the average of the season, precipitation of varying intensities leads to a moderate increase in PM_{10} concentration, with most pronounced spike observed when the rainfall intensity is less than average and the duration is less than 10 hours in all cities. Conversely, when the initial PM_{10} concentration exceeds average, precipitation of different intensities and durations result in a reduction in PM_{10} concentration to varying degrees.

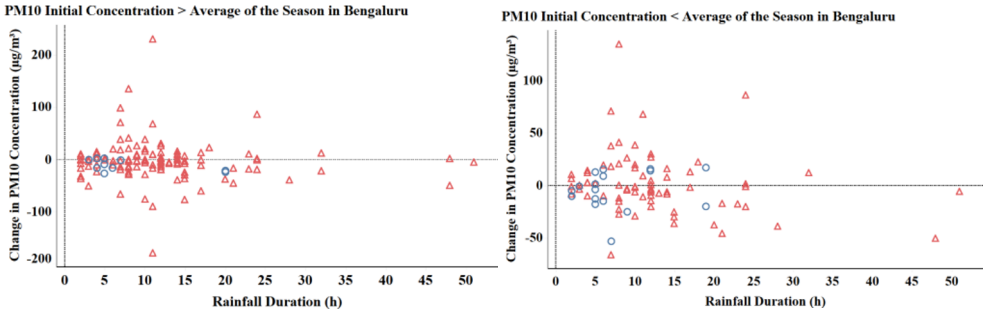
4.3 Removal of PM_{10} by Winds of Varying Velocities

This section utilises CBPF plots for the four cities, which were generated using PM_{10} concentrations above the 90th percentile for the monsoon season, to determine how varying wind speeds in all directions scavenge PM10 concentrations (Figure 5). Due to the urban nature of the cities that were selected and the possibility of major emission sources in the vicinity of the monitoring stations, a high percentile was selected. This will assist to better understand the contribution of point sources to the PM_{10} concentration at these sites.

In Amravati, Andhra Pradesh, higher wind speeds are mostly found in the west and southwest directions, where PM_{10} concentrations are relatively lower. On the other hand, PM_{10} concentrations are higher in the east and southeast, and the peak concentration occurs at mean wind speeds of 4 to 6 m/s. In Bengaluru, Karnataka, the prevailing winds mainly flow from the west. Lower concentrations of wind are detected in the westward direction across all wind speeds, while higher concentrations are found in the south direction within the wind speed range of 3 to 5m/s.



(a)



(b)

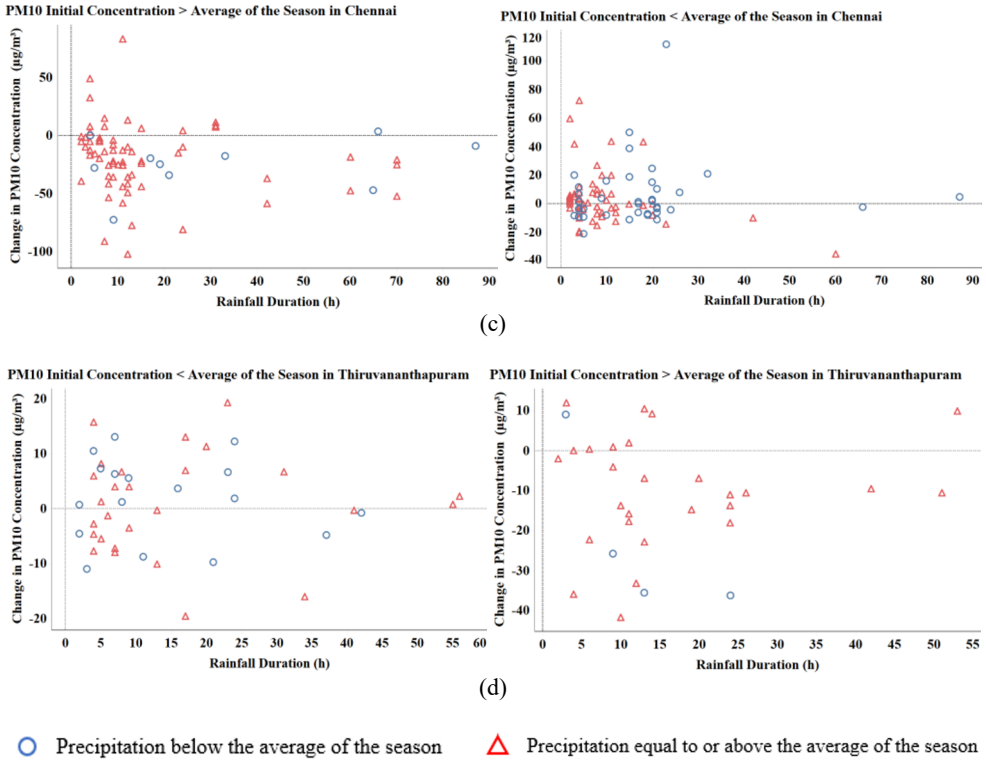
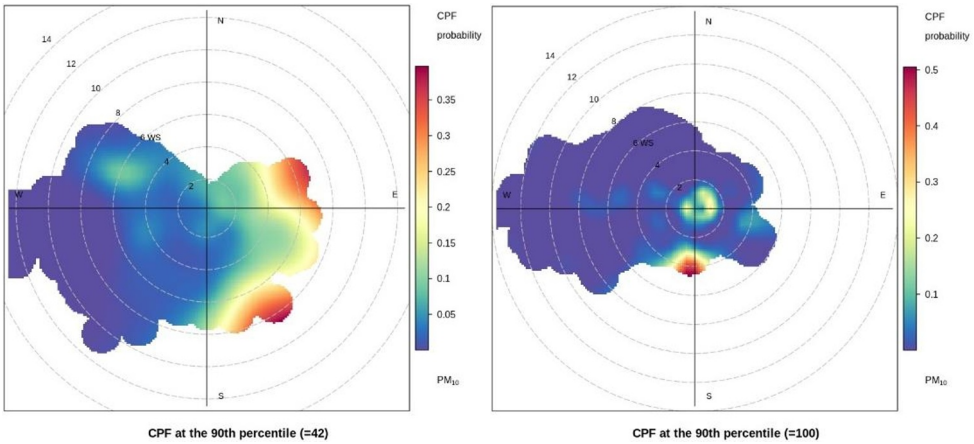
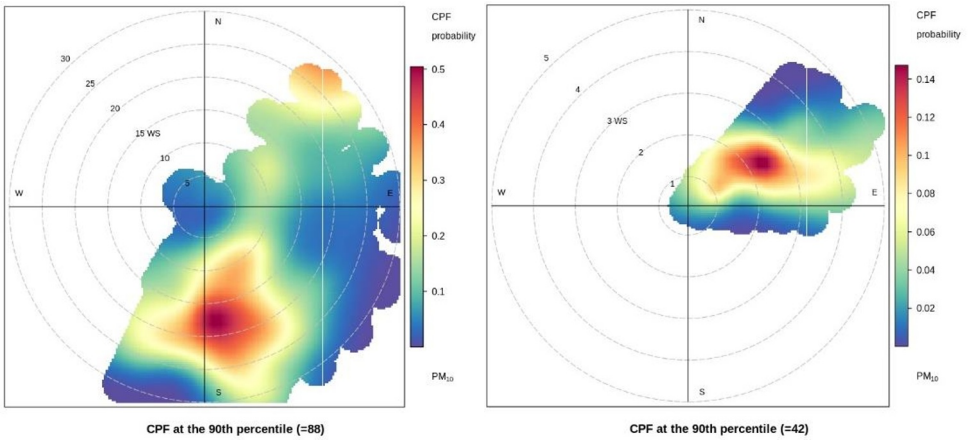


Fig.4. Relationship between duration of precipitation and variation in PM_{10} concentration in at (a) Amravati, Andhra Pradesh, (b) Bengaluru, Karnataka, (c) Chennai, Tamil Nadu, (d) Thiruvananthapuram, Kerala.



(a) (b)



(c) (d)

Fig. 5. Mass concentration variation under different wind speed and wind direction in (a) Amravati, Andhra Pradesh, (b) Bengaluru, Karnataka, (c) Chennai, Tamil Nadu, (d) Thiruvananthapuram, Kerala

The plots also indicate that in Chennai, Tamil Nadu, winds predominantly blow from the northeast to the southwest. Lower concentrations occur in the east, southeast, and south directions at high wind speeds above 25 m/s. Higher concentrations are present in the south direction when wind speeds range from 10 to 25 m/s. In Thiruvananthapuram, Kerela, winds are primarily blowing from the northeast to east directions. Lower PM_{10} concentrations are observed in the northeast when wind speeds exceed 4 m/s. Higher concentrations exist toward the northeast when wind speeds range from 1 to 3 m/s. This section underscores the impact of changing wind patterns on the spread of PM_{10} concentration in Chennai, Amravati, Bengaluru, and Thiruvananthapuram. More rapid wind velocities link with lower pollutant levels, especially evident in Chennai and Amravati, where concentrations drop over 25m/s and top at 4 to 6 m/s separately. Bengaluru shows lower concentrations with west-heading winds (3 to 5m/s) and higher levels in the south, while Thiruvanthapuram sees the same inclination at varying wind paces.

5 Conclusion

Based on the study's findings, several conclusions can be made. First, different cities have distinct effects from rainfall intensity on PM_{10} concentrations. PM_{10} concentrations in Amravati and Thiruvananthapuram are significantly scavenged by moderate rainfall. On the other hand, Chennai's PM_{10} concentrations often rise with mild to moderate rainfall. Rainfall in Bengaluru has a more mixed effect; light rains raise PM_{10} concentrations in August and September, while moderate rains generally clear the air, except for an unusual spike in June.

Second, the initial levels of the pollutant determine how precipitation affects PM_{10} concentrations. When the air quality is relatively poor, precipitation lowers PM_{10} concentration levels. Furthermore, a higher PM_{10} level before the rainfall leads to greater declines in the pollutant's concentration levels, particularly during periods of low rainfall. However, if the air quality is relatively clean, high rainfall results in rising PM_{10} concentrations.

Thirdly, wind speed and wind direction play a crucial role in the dispersion of PM_{10}. Different cities experience different directions predominant directions of wind, and hence this produces a varying impact on PM_{10} concentrations also. For example, Amravati experiences lowest PM_{10} concentrations with the occurrence of west and southwest winds with a peak speed between 3 m/s and 5 m/s. However, Bengaluru observes highest PM_{10} concentrations during the presence of west and southwest winds, whereas winds from east and northeast lead to lowest PM_{10} concentrations. Similar impact of wind direction and speed on precipitation is observed in Bengaluru and Chennai. Furthermore, regular monsoonal winds from east and northeast result in effective scavenging impact as they gain velocity as they reach Thiruvananthapuram.

The study aims to showcase the varying impact of meteorological factors, wind speed, wind direction, and precipitation on PM_{10} concentration levels in the four Indian cities of the Peninsular India throughout the monsoon season. This research thrives to serve as a critical insight for fruitful air pollution management efforts and its implementation.

References

1. Pantavou, K., Psiloglou, B., Lykoudis, S., Mavrakis, A., and Nikolopoulos, G. K. (2018). "Perceived air quality and particulate matter pollution based on field survey data during a winter period." *International Journal of Biometeorology*, **62(12)**, 2139–2150.
2. Unal, Y. S., Toros, H., Deniz, A., and Incecik, S. (2011). "Influence of meteorological factors and emission sources on spatial and temporal variations of PM10 concentrations in Istanbul Metropolitan Area." *Atmospheric Environment*, **45(31)**, 5504–5513.
3. Sumesh, R. K., Rajeevan, K., Resmi, E. A., and Unnikrishnan, C. K. (2017). "Particulate matter concentrations in the southern tip of India: Temporal variation, meteorological influences, and source identification." *Earth Systems and Environment*, **1(2)**.
4. Venkataraman, C., Brauer, M., Tibrewal, K., Sadavarte, P., Ma, Q., Cohen, A., Chaliyakunnel, S., Frostad, J., Klimont, Z., Martin, R. V., Millet, D. B., Philip, S., Walker, K., and Wang, S. (2018). "Source influence on emission pathways and ambient PM2.5 pollution over India (2015–2050)." *Atmospheric Chemistry and Physics*, **18(11)**, 8017–8039.
5. Central Pollution Control Board, Ministry of Environment, Forest, and Climate Change, Government of India.(2019). National Ambient Air Quality Status and Trends 2019. https://cpcb.nic.in/upload/NAAQS_2019.pdf
6. Bodor, K., Bodor, Z., Keresztesi, Á., and Szép, R. (2020). "PM10 concentration reduction due to the wet scavenging in the Ciuc Basin, Romania." *Acta Universitatis Sapientiae, Agriculture and Environment*, **12(1)**, 1–8.
7. Fortelli, A., Scafetta, N., and Mazzarella, A. (2016). "Influence of synoptic and local atmospheric patterns on PM10 air pollution levels: A model application to Naples (Italy)." *Atmospheric Environment*, **143**, 218–228.
8. Kim, S., Hong, K.-H., Hwandon, J., Young-Jae, P., Moojong, P., and Young, S. (2014). "Effect of precipitation on air pollutant concentration in Seoul, Korea." *Asian Journal of Atmospheric Environment*, **8(4)**, 202–211.
9. Gao, B., Ouyang, W., Cheng, H., Xu, Y., Lin, C., and Chen, J. (2019). "Interactions between rainfall and fine particulate matter investigated by simultaneous chemical composition measurements in downtown Beijing." *Atmospheric Environment*, **218**, 117000.

10. Singh, B., Gahlaut, V., and Dutta, H. N. (2022). "A study of air pollution episode over Delhi National Capital Region, India during October 29, 2016- November 03, 2016." *Ecology, Environment and Conservation*.
11. Chouhan, T. S. (1992). Desertification in the World and Its Control. Scientific Publishers. ISBN 978-8-1723-3043-9.
12. NASA POWER DATA Access Viewer: https://power.larc.nasa.gov/data-access-viewer

APPENDIX A. List of Air Quality Monitoring Stations

State	Station Name		Latitude	Longitude
Andhra Pradesh	**1**	**Secretariat, Amaravati - APPCB**	16.5062	80.648
Kerala	**2**	**Plammoodu, Thiruvananthapuram - Kerala PCB**	8.5492	76.9214
Karnataka	**3**	**BTM Layout, Bengaluru - CPCB**	12.9152	77.6104
	4	**BWSSB Kadabesanahalli, Bengaluru - CPCB**	12.9441	77.6227
	5	**Bapuji Nagar, Bengaluru - KSPCB**	12.9791	77.601
	6	**Hebbal, Bengaluru - KSPCB**	13.0359	77.597
	7	**Hombegowda Nagar, Bengaluru - KSPCB**	12.9793	77.5711
	8	**Peenya, Bengaluru - CPCB**	13.0324	77.527
Tamil Nadu	**9**	**Alandur Bus Depot, Chennai - CPCB**	13.1284	80.1981
	10	**Manali Village, Chennai - TNPCB**	13.1992	80.2846
	11	**Velachery Res. Area, Chennai - CPCB**	12.9791	80.2214

Journal of Scientific & Industrial Research
Vol. 83, August 2024, pp. 830-835
DOI: 10.56042/jsir.v83i8.2358

Analysing the Bottleneck in Crankcase Cover Manufacturing using Simulation and Modelling

Sumit Chawla[1]* & Ranganath M Singari[2]
[1]Applied Science Department (Mechanical), Bharati Vidyapeeth's College of Engineering, Delhi 110 063, India
[2]Mechanical, Production, and Industrial Engineering Department, Delhi Technological University, Delhi 110 042, India

Received 04 June 2023; revised 05 June 2024; accepted 17 July 2024

Before constructing the actual production plant architecture, the simulation builds a virtual production model that is identical to the real environment and offers insights into the future. By using simulation, we can test the behaviour of the expensive and sophisticated manufacturing system without physically putting any money in it. In this study, Flexsim is used for the first time to model and simulate the production of two-wheeler crankcase covers. In this work, a simulation model for crankcase cover production processes in the automotive sector is developed. Flexsim simulation tool is used as an optimization tool for analysing the bottleneck or performing the line balancing. The results indicate that by eliminating the bottleneck in the production line, it increases the line efficiency as well as the production throughput. Initially production line efficiency is just 51.8% when bottleneck is present. After bottleneck is removed, it increases not only throughput but also line efficiency to 97.29%. All manufacturing sectors will profit from these findings.

Keywords: Bottleneck, Line balancing, Optimization, Plant layout, Production model

Introduction

Simulation is a tool for modelling and analysing complex systems. By creating a virtual representation of a real system, simulations allow researchers, engineers, and analysts to study the behaviour of the system and make predictions about how it will perform under different conditions. The use of computer simulation can help to reduce the cost and time required to develop, test, and improve real-world systems, and can also provide valuable insights into the underlying processes and mechanisms of the system being studied.[1]

Analysis of the various production systems is aided by simulation. It resolved issues with resource management, planning, shop floor operations, material handling, storage, and manufacturing. Different companies now provide various simulation software. This software lowers the cost of manufacturing and offers efficient resource management.[2] The development of computer simulations can be traced back to the 1950s when early computer systems were used to simulate complex systems and processes. The creation of programming languages in the following decade made it easier to implement these simulations. Over time, ready-made simulation software emerged, allowing for more complex and accurate simulations to be created, including those that utilized 3D modelling. These advancements have played a crucial role in shaping the fields of science, engineering, and technology, as well as numerous other industries. With the increasing demand for efficient and cost-effective solutions, many companies have started developing advanced simulation software. These programs allow for the creation of virtual models of various processes, providing valuable insights and enabling informed decision-making. The advancements in this technology have made it easier for businesses to optimize their operations, reduce costs, and make more informed decisions. This, in turn, has led to the growing popularity of simulation software across various industries.[3]

Flexsim is a software tool for modelling, simulating, and optimizing complex systems and processes. It is used in various industries such as manufacturing, healthcare, supply chain, and transportation. Flexsim provides a 3D environment for building virtual models and simulating real-world situations, allowing users to experiment with different scenarios and optimize their systems. It can be used for various purposes such as capacity planning, process improvement, resource optimization, and risk analysis. Flexsim offers a range

*Author for Correspondence
E-mail: chawlasumit4@gmail.com

of features such as process visualization, data analysis, and reporting, making it a valuable tool for organizations looking to improve efficiency and reduce costs. Flexsim, Arena, Simul8, Anylogic, Witness, and Promodel are some of the simulation software available in the marketplace. Software for discrete event simulation called Flexsim is employed for challenging optimization issues. Flexsim is an extremely potent modelling and simulation tool. This tool may be used to analyse appointment scheduling issues and manufacturing issues. Additionally, it aids management in adjusting work schedules and hiring staff for upcoming projects. The simulation also reveals any system bottlenecks or inefficiencies.[4]

Scheduling, production planning, and design issues may all be successfully solved using the simulation technique. These methods aid in solving decision-making issues and producing more accurate outcomes based on actual dynamic circumstances.[5] The experimenter tab in Flexsim is used to construct experiments that are used to optimise the performance of the manufacturing plant and the set of solutions.[6] The Flexsim tool can construct a hybrid simulation model that optimises a variety of parameters, including makespan, due date, tardiness, work-in-progress inventory, machine utilisation, etc.[7] This tool offers a comprehensive output report that may be examined to gauge the system's performance.[8]

Yaser *et al.*[9] discussed the use of modelling and simulation as a tool for making decisions in the context of improving the performance of a production cell in a manufacturing plant. The use of the Flexsim simulation tool is mentioned as a way to analyse different scenarios and make recommendations for improvement. The results of the simulation indicate that an improved scenario resulted in an increase in production level and the elimination of the need for overtime labour costs. Wang *et al.*[10] established a thorough framework for the logistics distribution centre layout design problem. A small logistics distribution center's structure and operational procedures were examined and planned using logistics simulation software to create the simulation model and the distribution centre model of logistics is optimised using Flexsim simulation software.

Hassan *et al.*[11] used Flexsim to analyse the behaviour of a 500 Kg capacity shea nut processing plant and found that the initial plant design had bottlenecks and low efficiency (35.7%). After improving the layout, the plant efficiency increased to 83.3% with shorter lead times, improved machine utilization, and increased throughput capacity. The results showed that effective layout design and simulation can improve traditional shea nut processing practices. Ljaskovska *et al.*[12] presented the results of analysing and applying the study of mechanical systems in various industries, such as electronics, processing, and food. The importance of simulating and visualizing the processes of creating technological systems and analysing their relationships is highlighted to improve the efficiency of resource and time use in production processes. Luscinski *et al.*[13] focused on the development of a simulation model of a flexible manufacturing system using Flexsim 3D software. The aim is to achieve and maintain both strategic and operational goals in a global competitive market. The model is designed to match both production and market requirements in the face of variable demand for products made in many variants and short lead times.

Watnakornbuncha *et al.*[14] reduced production costs of Eureka lemon shampoo products by using Flexsim software simulations. Data was collected from Pasutara Farm in Thailand. Results showed a 39.58% reduction in time cost by implementing three stages of production improvement on the nine-workstation production line. The research suggests improvements to the production processes based on the results. Rao *et al.*[15] improved the productivity of the spool casing assembly process in the horn armature core. The study uses method study, time study and Value Stream Mapping (VSM) to identify the value-added and non-value-added activities and transportation wastes in the current state process. The future state VSM is developed to reduce these wastes and the results of both current and future state time study are simulated using Flexsim simulation software.

Flexsim is a simulation software that can also be used to model and analyse queues in public healthcare systems to reduce waiting times. The software can be used to create a virtual representation of a healthcare system and simulate the flow of patients through various stages of the process. By analysing the data generated through simulation, healthcare managers can identify bottlenecks and areas for improvement, and test different strategies for reducing waiting times before implementing them in real-life settings.[16] Flexsim is also used to simulate a production line and visualize it in real-time, allowing you to test different scenarios, analyse results, and make informed

decisions about optimizing the production process. Flexsim graphical user interface makes it easy to build and modify simulation models, and its advanced simulation engine provides fast and accurate results. The software can also be used to analyse data and present results in customizable reports and charts.[1] Flexsim simulation environment can also be used for personnel management on the production line by creating a virtual simulation of the production process. This can help to analyse and optimize various aspects of personnel management such as staffing levels, work scheduling, training, and motivation. Flexsim can also help to identify bottlenecks in the production process, evaluate the impact of changes in personnel management policies, and estimate the costs and benefits of different personnel management strategies.[17]

Modern techniques, such as Computer-Aided Design (CAD) and Computer-Aided Manufacturing (CAM) software, can help improve machine shop layout design by providing faster and more efficient methods of modelling and simulation. These tools allow designers to create virtual models of the shop floor, test different layout configurations, and optimize the use of space and resources. By using these modern techniques, machine shops can reduce the time it takes to design layouts, minimize the risk of design errors, and improve overall efficiency and productivity.[18] According to previous study, most researchers have used the Flexsim tool to solve modelling and simulation issues. The literature analysis also reveals that there is little research in the domain of modelling and simulation of crankcase cover manufacture in the automotive sector. The primary goal of this work is to create a 3D simulation model based on a general process model. Later, bottleneck will be identified, and removed for increasing the production line efficiency and production throughput.

Materials and Methods

Analysing the process model is the initial stage in simulating any industry's layout. The next step is to turn this simulation model into a process model. Flexsim is an effective tool for turning any theoretical process model into a realistic three-dimensional simulation model.

Process Model

A crankcase cover is made by going through a number of processes, including melting raw materials in a furnace, die casting, fettling, drilling, machining, polishing, surface treatment, and painting. In Fig. 1, the overall process model for the crankcase cover manufacturing process is displayed. Our main aim of the study is to find out the bottleneck present in the production line and to eliminate this bottleneck to increase the line efficiency.

Flexsim Simulation Approach

A simulation model of the crankcase cover manufacturing process is created using the Flexsim

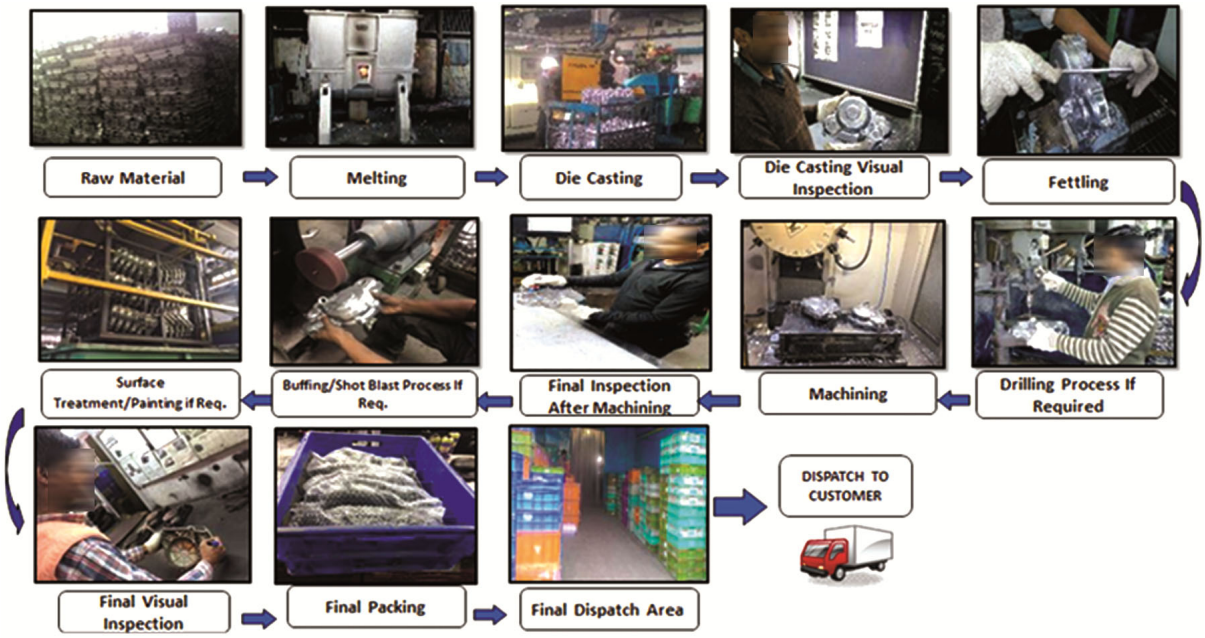


Fig. 1 — General process model for crankcase cover manufacturing

simulation tool. Additionally, this programme creates infinitely extensible models that may be freely expanded.[19] With the use of this instrument, a 3-D model that cannot be established in its actual form may be examined.[20]

Assumptions

When creating a simulation model, the following assumptions are taken into account.

(a) Every machine has sufficient capability to function.

(b) It is expected that the raw material arrival rate will follow an exponential distribution with a location value of zero and a scale value of 4 minutes.

(c) A triangular distribution with a minimum setup time of 0.05 minutes, a maximum setup time of 0.15 minutes, and a typical setup time of 0.01 minutes is considered.

(d) The processing time is also assumed to be triangular distribution.

The four fundamental elements are present in Flexsim. A source, sink, processor, and queues are among these elements.[20]

Simulation Model

The arrival rate, processing time, and resource number are the usual three inputs of the simulation model. Throughput, line efficiency, and states are typically used to gauge the model output. Crankcase covers are the flow components in this simulation model. The source, queue, processor, and sink are the fixed resources employed in this simulation model. Additionally, loading and unloading flow items are done by operators. The operators are coordinated by a dispatcher. The real database system of the organisation must occasionally be used to update the virtual model. The crankcase cover manufacturing 3D simulation model is shown in Fig. 2. To execute this simulation model, the model's parameters must be defined. The various operations performed for the manufacturing of crankcase cover are assumed to be triangular distribution. Initially raw material (aluminium bricks of 5 kg) is melted in the furnace. Then, molten metal is carried to die casting machine with the help of a ladle. The time taken by cold chamber Die Casting Machine (DCM) is approximately 1.5 minutes. After that, Manual Fettling (MF) is done for 0.18 minutes for cleaning some extra material. Then drilling machine is done for 0.35 minutes if required some holes for particular places in the part. Then, main finishing operation is done on Vertical Milling Centre (VMC) for 4 minutes. After that, Buffing Machine (BM) is used to remove dirt and polish the part. Last, part is processed in Painting Shop (PS) for various coatings on the part for 21.3 minutes for removing any surface defect. Data used in this study is collected from the Super Auto Limited company, Faridabad, India.

Flexsim reset and run options are used to run the model during simulations, and the dashboard option presents all system analysis in the form of numerous charts. Any uncertainty that was present in the production system is eliminated through analysis of the virtual model.

Elimination of Bottleneck in Production Line

The run speed during simulation is taken as 2401 sec. On running the simulation model for 8-hour (28800 sec) shift, it has been observed that the bottleneck is created on the queue 4 as shown in Fig. 2. The production line efficiency of this model is calculated as 51.8%. The throughput during an 8-hour shift is shown in Fig. 3. It can also be seen from this diagram that bottleneck is present after drilling machine due to which throughput value of VMC is 115 as compared to 218 value of drilling machine. Since the bottleneck is created before VMC machine, So, add one more VMC machine in production line as shown in Fig. 4. After removing this bottleneck, the production line efficiency becomes 97.29%. The increase in throughput value after eliminating bottleneck is also shown in Fig. 5.

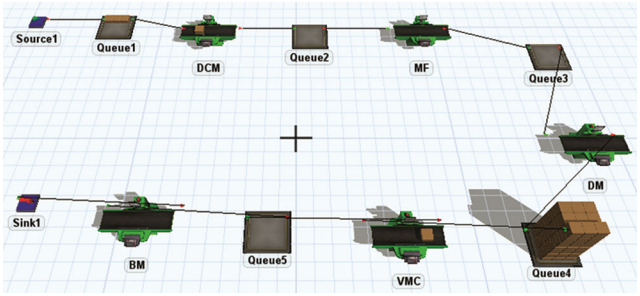


Fig. 2 — Simulation model of crankcase cover manufacturing

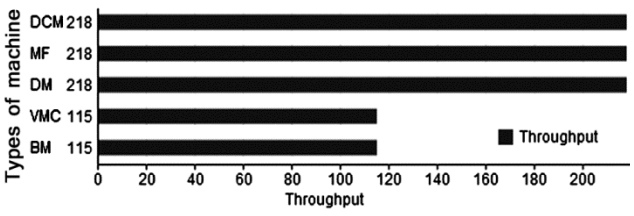


Fig. 3 — Throughput during 8-hour shift

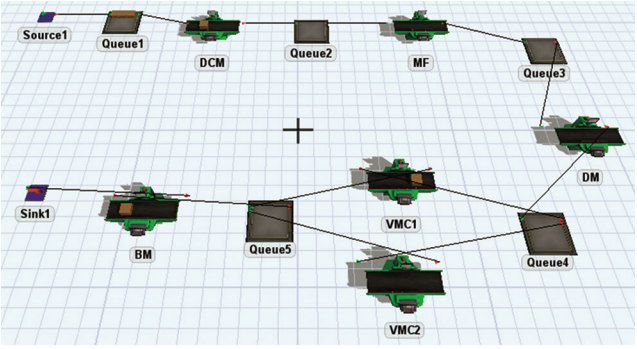


Fig. 4 — Simulation model after removing bottleneck

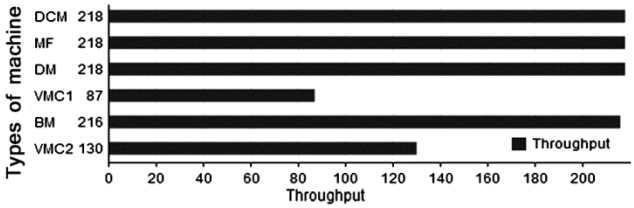


Fig. 5 — Throughput after removing bottleneck

Results and Discussion

There are various line balancing methods to eliminate bottlenecks in a production line discussed as follows:[21]

a) Increase capacity: This can be done by adding more resources, such as employees, machines, or raw materials, to the production process.
b) Reallocate resources: Moving resources from low-demand areas to high-demand areas can help balance production and eliminate bottlenecks.
c) Improve efficiency: Streamlining processes, reducing waste, and increasing productivity can help reduce bottlenecks.
d) Prioritize tasks: Focusing on the most important tasks first can help ensure that resources are used efficiently, and bottlenecks are minimized.
e) Automation: Automating certain tasks can help increase productivity and reduce the likelihood of bottlenecks occurring.
f) Improved scheduling: Scheduling production in a way that considers potential bottlenecks can help prevent them from occurring.
g) Collaboration and communication: Encouraging collaboration and communication among different departments can help identify and resolve bottlenecks more effectively.

There are some specific ways to eliminate bottleneck in any simulation model such as increasing the inter-arrival time of source1, decreasing the setup time/processing time or adding some more machines in production line. The most preferable method for line balancing is to increase the capacity i.e. adding more machines in production lines. Since the bottleneck is created before VMC machine, so, add one more VMC machine in production line as shown in Fig. 4. After removing this bottleneck, the production line efficiency becomes 97.29%. The increase in throughput value after eliminating bottleneck is also shown in Fig. 5.

By simulating the behaviour of the production system, it can determine whether the production targets can be met within the specified time and evaluate the performance of each machine in the production line. The simulation results can provide valuable information about the capacity, efficiency, and bottlenecks in the production process, which can be used to identify areas for improvement and optimize the production line for maximum performance.

Effective operations and machinery are crucial for the automotive industry to sustain its growth and meet the demands of the market. Efficient and reliable processes and equipment help improve production capacity and product quality, leading to increased customer satisfaction and competitiveness in the market.

Modelling and simulation have certainly had a significant impact on the process of launching new businesses. They provide a powerful tool for evaluating and testing new ideas, concepts, and designs in a virtual environment before they are implemented in the real world. This allows entrepreneurs to identify and address potential challenges and make informed decisions, leading to more effective and efficient launch processes. The results indicate that by eliminating the bottleneck in the production line, it increases the line efficiency as well as the production throughput.

Conclusions

By creating a virtual model of the manufacturing process, it's possible to evaluate various scenarios and test different solutions to identify the root cause of any bottlenecks and determine the most effective way to remove them. Initially production line efficiency is 51.8% when bottleneck is present at queue 4. By adding one more vertical milling centre in the production line, queue 4 bottleneck is removed, and it increases not only throughput but also line efficiency to 97.29%. All manufacturing sectors will profit from these findings. The development of new, effective manufacturing facilities for a variety of sectors can

also benefit from the use of this simulation model. By employing a global table with various part sets and automatically updating the database with corporate databases, further research may also be expanded. Future research might also examine how adding buffers affects system performance within the design of the processing stations.

References

1 Ishak A, Zubair F & Cendani A S, Production line simulation in vise using the Flexsim application, in *IOP Conf Series: Mater Sci Eng* (IOP Publishing) **1003(1)** (2020) 1–6, doi: 10.1088/1757-899X/1003/1/012103.

2 Bjorklund S, Yamani N & Lloyd T, Simulation software for real time forecasting as an operational support, *Proc of the Second World Conf POM 15th Annual POM Conf* (Cancun, Mexico) 2004, 1–30.

3 Kliment M, Trojan J, Pekarcikova M & Kronova J, Creation of simulation models using the Flexsim software module, *Advanced Logistic Systems - Theory and Practice*, **16(1)** (2022) 41–50, doi: 10.32971/als.2022.004.

4 Krishna L S R, Saketh T S S & Kumar M V S, A simulation based approach for studying the effect of buffers on the performance of an FMS, *Int J Eng Sci Invent*, **4(1)** (2015) 5–9.

5 Fauadi M, Azimi N Z, Anuar N I, Ali M M, Sued M K & Ramlan S, Simulation approach for logistical planning in a warehouse : A review, *Orient J Comput Sci Technol*, **11(4)** (2018) 201–208.

6 Jarernram J & Samattapapong N, Parallel machine scheduling using simulation software, *Proc* of the *ICMEIT* (Pattaya, Thailand) 2018, 150–155.

7 Rodrigues R P, Pinho A F D & Sena D C, Application of hybrid simulation in production scheduling in job shop systems, *Simul Trans of the Soc Model Simul Int*, **96(3)** (2019) 1–16, doi: 10.1177/0037549719861724.

8 Nie X & Wang L, Simulation process design for scheduling mode of railway container terminals based on Flexsim, *Proc J Phys : Conf Ser*, **1176(5)** (2019) 1–6, doi: 10.1088/1742-6596/1176/5/052012.

9 Yaser S, Abdelatif N, Fahim I, Emad Y, Saleh A, & Kassem S S, FlexSim simulation to enhance productivity of a production cell: A case study, *Proc* of the *3rd NILES* (Giza, Egypt) 2021, 312–317. doi: 10.1109/NILES53778.2021.9600545.

10 Wang S, Wang S M & Zhang N, Flexsim-based simulation and optimization of green logistics distribution center, *Proc of the 14th ICCMS* (Chongqing, China) 2022, 76–82, doi: 10.1145/3547578.3547590.

11 Hassan N A, Arogundade A I, Iyenagbe U B & Musa D I, Simulation and analyses of shea nuts (vitallaria paradoxa) processing plant using FlexSim, *J Fut Sustain*, **3(2)** (2023) 67–74, doi: 10.5267/j.jfs.2022.11.006.

12 Ljaskovska S, Martyn Y, Malets I & Velyka O, Optimization of parameters of technological processes means of the flexsim simulation simulation program, *Proc of the IEEE 3rd ICDSMP* (Lviv, Ukraine) 2020, 391–397, doi: 10.1109/DSMP47368.2020.9204029.

13 Luscinski S & Ivanov V, A simulation study of industry 4.0 factories based on the ontology on flexibility with using flexsim software, *Manag Prod Eng Rev*, **11(3)** (2020) 74–83, doi: 10.24425/mper.2020.134934.

14 Watnakornbuncha D, Muangwai A, Wichitphongsa W, Amdee N & Inthawongse C, Processing cost reduction of lemon products in community enterprises using Flexsim simulation, *Proc of the IEEM* (Kuala Lumpur, Malaysia) 2022, 670–674, doi: 10.1109/IEEM55944.2022.9989661.

15 Rao G V P, Nallusamy S & Raman P, Enhancement of production in subassembly line of a medium scale industry using different lean tools and FlexSim simulation software, *Int J Eng Res Africa*, **44** (2019) 229–239, doi: 10.4028/www.scientific.net/jera.44.229.

16 Amalia P & Cahyati N, Queue analysis of public healthcare system to reduce waiting time using flexsim 6.0 software, *Int J Ind Optim*, **1(2)** (2020) 101–110, doi: 10.12928/ijio.v1i2.2428.

17 Krynke M, Personnel management on the production line using the Flexsim simulation environment, *Manuf Technol*, **21(5)** (2021) 657–667, doi: 10.21062/mft.2021.073.

18 Patil R J, Kubade P R & Kulkarni H B, Optimization of machine shop layout by using Flexsim software, *Proc of the AIP Conf* (AIP Publishing) **2200(1)** 2019, doi: 10.1063/1.5141203.

19 Kliment M, Trojan J, Pekarcíková M & Kronova J, Creation of simulation models using the Flexsim software module, *ALS-Theory and Practice*, **16(1)** (2022) 41–50, doi: 10.32971/als.2022.004.

20 Kessel D, Jeon J, Jung J, Oh E & Kim C, Logistical simulation modeling for planning a soil remediation process, *Sci Technol Nucl Install*, **2019** (2019) 1–13, doi: 10.1155/2019/6789506.

21 Kikolski M, Identification of production bottlenecks with the use of plant simulation software, *Eng Manag Prod Serv*, **8(4)** (2016) 103–112, doi: 10.1515/emj-2016-0038.

Journal of Scientific & Industrial Research
Vol. 83, August 2024, pp. 830-835
DOI: 10.56042/jsir.v83i8.2358

Analysing the Bottleneck in Crankcase Cover Manufacturing using Simulation and Modelling

Sumit Chawla[1]* & Ranganath M Singari[2]
[1]Applied Science Department (Mechanical), Bharati Vidyapeeth's College of Engineering, Delhi 110 063, India
[2]Mechanical, Production, and Industrial Engineering Department, Delhi Technological University, Delhi 110 042, India

Received 04 June 2023; revised 05 June 2024; accepted 17 July 2024

Before constructing the actual production plant architecture, the simulation builds a virtual production model that is identical to the real environment and offers insights into the future. By using simulation, we can test the behaviour of the expensive and sophisticated manufacturing system without physically putting any money in it. In this study, Flexsim is used for the first time to model and simulate the production of two-wheeler crankcase covers. In this work, a simulation model for crankcase cover production processes in the automotive sector is developed. Flexsim simulation tool is used as an optimization tool for analysing the bottleneck or performing the line balancing. The results indicate that by eliminating the bottleneck in the production line, it increases the line efficiency as well as the production throughput. Initially production line efficiency is just 51.8% when bottleneck is present. After bottleneck is removed, it increases not only throughput but also line efficiency to 97.29%. All manufacturing sectors will profit from these findings.

Keywords: Bottleneck, Line balancing, Optimization, Plant layout, Production model

Introduction

Simulation is a tool for modelling and analysing complex systems. By creating a virtual representation of a real system, simulations allow researchers, engineers, and analysts to study the behaviour of the system and make predictions about how it will perform under different conditions. The use of computer simulation can help to reduce the cost and time required to develop, test, and improve real-world systems, and can also provide valuable insights into the underlying processes and mechanisms of the system being studied.[1]

Analysis of the various production systems is aided by simulation. It resolved issues with resource management, planning, shop floor operations, material handling, storage, and manufacturing. Different companies now provide various simulation software. This software lowers the cost of manufacturing and offers efficient resource management.[2] The development of computer simulations can be traced back to the 1950s when early computer systems were used to simulate complex systems and processes. The creation of programming languages in the following decade made it easier to implement these simulations. Over time, ready-made simulation software emerged, allowing for more complex and accurate simulations to be created, including those that utilized 3D modelling. These advancements have played a crucial role in shaping the fields of science, engineering, and technology, as well as numerous other industries. With the increasing demand for efficient and cost-effective solutions, many companies have started developing advanced simulation software. These programs allow for the creation of virtual models of various processes, providing valuable insights and enabling informed decision-making. The advancements in this technology have made it easier for businesses to optimize their operations, reduce costs, and make more informed decisions. This, in turn, has led to the growing popularity of simulation software across various industries.[3]

Flexsim is a software tool for modelling, simulating, and optimizing complex systems and processes. It is used in various industries such as manufacturing, healthcare, supply chain, and transportation. Flexsim provides a 3D environment for building virtual models and simulating real-world situations, allowing users to experiment with different scenarios and optimize their systems. It can be used for various purposes such as capacity planning, process improvement, resource optimization, and risk analysis. Flexsim offers a range

*Author for Correspondence
E-mail: chawlasumit4@gmail.com

of features such as process visualization, data analysis, and reporting, making it a valuable tool for organizations looking to improve efficiency and reduce costs. Flexsim, Arena, Simul8, Anylogic, Witness, and Promodel are some of the simulation software available in the marketplace. Software for discrete event simulation called Flexsim is employed for challenging optimization issues. Flexsim is an extremely potent modelling and simulation tool. This tool may be used to analyse appointment scheduling issues and manufacturing issues. Additionally, it aids management in adjusting work schedules and hiring staff for upcoming projects. The simulation also reveals any system bottlenecks or inefficiencies.[4]

Scheduling, production planning, and design issues may all be successfully solved using the simulation technique. These methods aid in solving decision-making issues and producing more accurate outcomes based on actual dynamic circumstances.[5] The experimenter tab in Flexsim is used to construct experiments that are used to optimise the performance of the manufacturing plant and the set of solutions.[6] The Flexsim tool can construct a hybrid simulation model that optimises a variety of parameters, including makespan, due date, tardiness, work-in-progress inventory, machine utilisation, etc.[7] This tool offers a comprehensive output report that may be examined to gauge the system's performance.[8]

Yaser *et al.*[9] discussed the use of modelling and simulation as a tool for making decisions in the context of improving the performance of a production cell in a manufacturing plant. The use of the Flexsim simulation tool is mentioned as a way to analyse different scenarios and make recommendations for improvement. The results of the simulation indicate that an improved scenario resulted in an increase in production level and the elimination of the need for overtime labour costs. Wang *et al.*[10] established a thorough framework for the logistics distribution centre layout design problem. A small logistics distribution center's structure and operational procedures were examined and planned using logistics simulation software to create the simulation model and the distribution centre model of logistics is optimised using Flexsim simulation software.

Hassan *et al.*[11] used Flexsim to analyse the behaviour of a 500 Kg capacity shea nut processing plant and found that the initial plant design had bottlenecks and low efficiency (35.7%). After improving the layout, the plant efficiency increased to 83.3% with shorter lead times, improved machine utilization, and increased throughput capacity. The results showed that effective layout design and simulation can improve traditional shea nut processing practices. Ljaskovska *et al.*[12] presented the results of analysing and applying the study of mechanical systems in various industries, such as electronics, processing, and food. The importance of simulating and visualizing the processes of creating technological systems and analysing their relationships is highlighted to improve the efficiency of resource and time use in production processes. Luscinski *et al.*[13] focused on the development of a simulation model of a flexible manufacturing system using Flexsim 3D software. The aim is to achieve and maintain both strategic and operational goals in a global competitive market. The model is designed to match both production and market requirements in the face of variable demand for products made in many variants and short lead times.

Watnakornbuncha *et al.*[14] reduced production costs of Eureka lemon shampoo products by using Flexsim software simulations. Data was collected from Pasutara Farm in Thailand. Results showed a 39.58% reduction in time cost by implementing three stages of production improvement on the nine-workstation production line. The research suggests improvements to the production processes based on the results. Rao *et al.*[15] improved the productivity of the spool casing assembly process in the horn armature core. The study uses method study, time study and Value Stream Mapping (VSM) to identify the value-added and non-value-added activities and transportation wastes in the current state process. The future state VSM is developed to reduce these wastes and the results of both current and future state time study are simulated using Flexsim simulation software.

Flexsim is a simulation software that can also be used to model and analyse queues in public healthcare systems to reduce waiting times. The software can be used to create a virtual representation of a healthcare system and simulate the flow of patients through various stages of the process. By analysing the data generated through simulation, healthcare managers can identify bottlenecks and areas for improvement, and test different strategies for reducing waiting times before implementing them in real-life settings.[16] Flexsim is also used to simulate a production line and visualize it in real-time, allowing you to test different scenarios, analyse results, and make informed

decisions about optimizing the production process. Flexsim graphical user interface makes it easy to build and modify simulation models, and its advanced simulation engine provides fast and accurate results. The software can also be used to analyse data and present results in customizable reports and charts.[1] Flexsim simulation environment can also be used for personnel management on the production line by creating a virtual simulation of the production process. This can help to analyse and optimize various aspects of personnel management such as staffing levels, work scheduling, training, and motivation. Flexsim can also help to identify bottlenecks in the production process, evaluate the impact of changes in personnel management policies, and estimate the costs and benefits of different personnel management strategies.[17]

Modern techniques, such as Computer-Aided Design (CAD) and Computer-Aided Manufacturing (CAM) software, can help improve machine shop layout design by providing faster and more efficient methods of modelling and simulation. These tools allow designers to create virtual models of the shop floor, test different layout configurations, and optimize the use of space and resources. By using these modern techniques, machine shops can reduce the time it takes to design layouts, minimize the risk of design errors, and improve overall efficiency and productivity.[18] According to previous study, most researchers have used the Flexsim tool to solve modelling and simulation issues. The literature analysis also reveals that there is little research in the domain of modelling and simulation of crankcase cover manufacture in the automotive sector. The primary goal of this work is to create a 3D simulation model based on a general process model. Later, bottleneck will be identified, and removed for increasing the production line efficiency and production throughput.

Materials and Methods

Analysing the process model is the initial stage in simulating any industry's layout. The next step is to turn this simulation model into a process model. Flexsim is an effective tool for turning any theoretical process model into a realistic three-dimensional simulation model.

Process Model

A crankcase cover is made by going through a number of processes, including melting raw materials in a furnace, die casting, fettling, drilling, machining, polishing, surface treatment, and painting. In Fig. 1, the overall process model for the crankcase cover manufacturing process is displayed. Our main aim of the study is to find out the bottleneck present in the production line and to eliminate this bottleneck to increase the line efficiency.

Flexsim Simulation Approach

A simulation model of the crankcase cover manufacturing process is created using the Flexsim

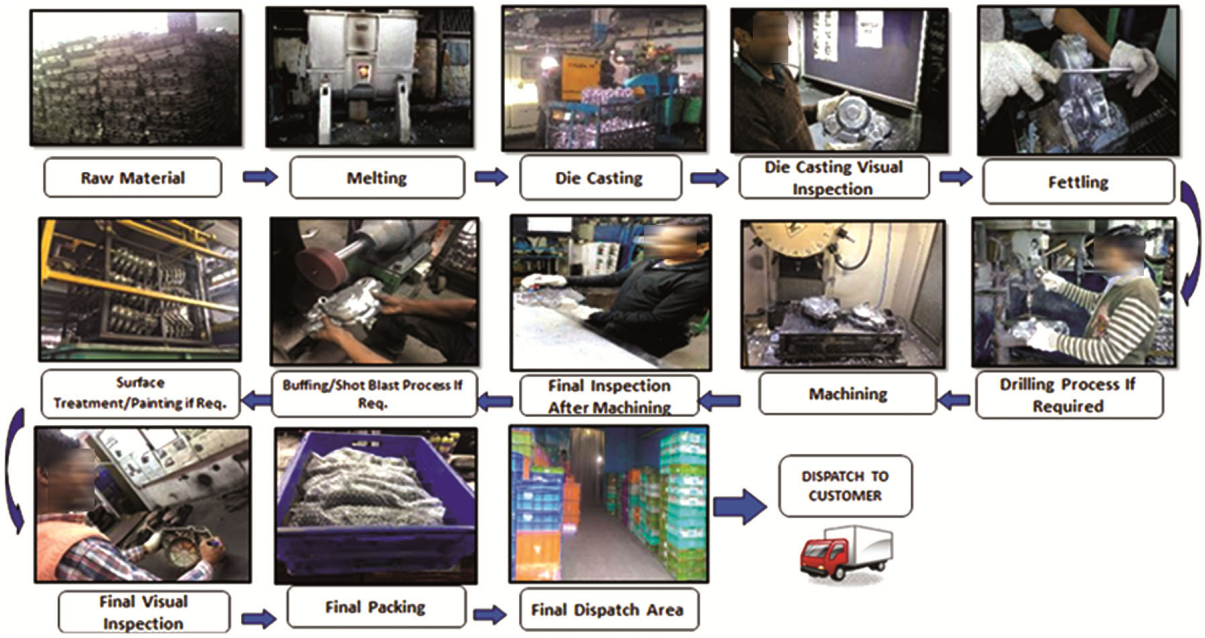


Fig. 1 — General process model for crankcase cover manufacturing

simulation tool. Additionally, this programme creates infinitely extensible models that may be freely expanded.[19] With the use of this instrument, a 3-D model that cannot be established in its actual form may be examined.[20]

Assumptions

When creating a simulation model, the following assumptions are taken into account.

(a) Every machine has sufficient capability to function.

(b) It is expected that the raw material arrival rate will follow an exponential distribution with a location value of zero and a scale value of 4 minutes.

(c) A triangular distribution with a minimum setup time of 0.05 minutes, a maximum setup time of 0.15 minutes, and a typical setup time of 0.01 minutes is considered.

(d) The processing time is also assumed to be triangular distribution.

The four fundamental elements are present in Flexsim. A source, sink, processor, and queues are among these elements.[20]

Simulation Model

The arrival rate, processing time, and resource number are the usual three inputs of the simulation model. Throughput, line efficiency, and states are typically used to gauge the model output. Crankcase covers are the flow components in this simulation model. The source, queue, processor, and sink are the fixed resources employed in this simulation model. Additionally, loading and unloading flow items are done by operators. The operators are coordinated by a dispatcher. The real database system of the organisation must occasionally be used to update the virtual model. The crankcase cover manufacturing 3D simulation model is shown in Fig. 2. To execute this simulation model, the model's parameters must be defined. The various operations performed for the manufacturing of crankcase cover are assumed to be triangular distribution. Initially raw material (aluminium bricks of 5 kg) is melted in the furnace. Then, molten metal is carried to die casting machine with the help of a ladle. The time taken by cold chamber Die Casting Machine (DCM) is approximately 1.5 minutes. After that, Manual Fettling (MF) is done for 0.18 minutes for cleaning some extra material. Then drilling machine is done for 0.35 minutes if required some holes for particular places in the part. Then, main finishing operation is done on Vertical Milling Centre (VMC) for 4 minutes. After that, Buffing Machine (BM) is used to remove dirt and polish the part. Last, part is processed in Painting Shop (PS) for various coatings on the part for 21.3 minutes for removing any surface defect. Data used in this study is collected from the Super Auto Limited company, Faridabad, India.

Flexsim reset and run options are used to run the model during simulations, and the dashboard option presents all system analysis in the form of numerous charts. Any uncertainty that was present in the production system is eliminated through analysis of the virtual model.

Elimination of Bottleneck in Production Line

The run speed during simulation is taken as 2401 sec. On running the simulation model for 8-hour (28800 sec) shift, it has been observed that the bottleneck is created on the queue 4 as shown in Fig. 2. The production line efficiency of this model is calculated as 51.8%. The throughput during an 8-hour shift is shown in Fig. 3. It can also be seen from this diagram that bottleneck is present after drilling machine due to which throughput value of VMC is 115 as compared to 218 value of drilling machine. Since the bottleneck is created before VMC machine, So, add one more VMC machine in production line as shown in Fig. 4. After removing this bottleneck, the production line efficiency becomes 97.29%. The increase in throughput value after eliminating bottleneck is also shown in Fig. 5.

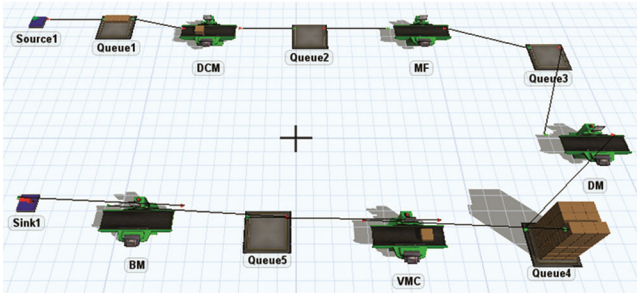


Fig. 2 — Simulation model of crankcase cover manufacturing

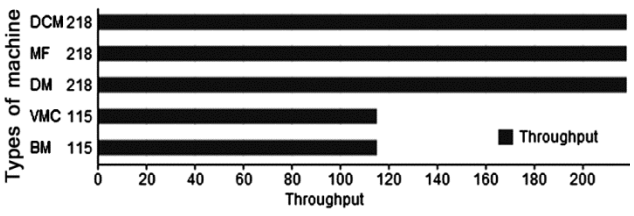


Fig. 3 — Throughput during 8-hour shift

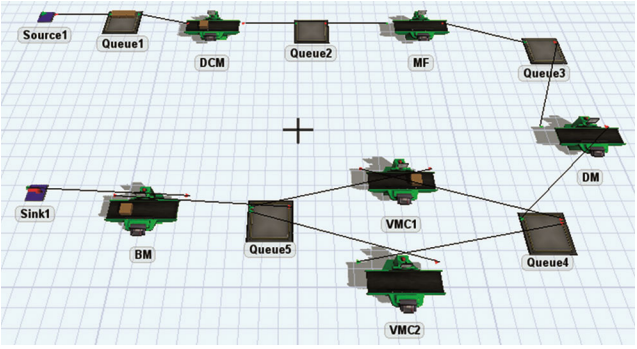


Fig. 4 — Simulation model after removing bottleneck

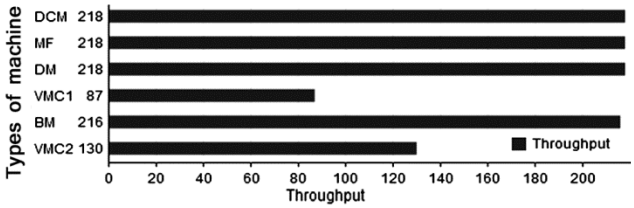


Fig. 5 — Throughput after removing bottleneck

Results and Discussion

There are various line balancing methods to eliminate bottlenecks in a production line discussed as follows:[21]

a) Increase capacity: This can be done by adding more resources, such as employees, machines, or raw materials, to the production process.
b) Reallocate resources: Moving resources from low-demand areas to high-demand areas can help balance production and eliminate bottlenecks.
c) Improve efficiency: Streamlining processes, reducing waste, and increasing productivity can help reduce bottlenecks.
d) Prioritize tasks: Focusing on the most important tasks first can help ensure that resources are used efficiently, and bottlenecks are minimized.
e) Automation: Automating certain tasks can help increase productivity and reduce the likelihood of bottlenecks occurring.
f) Improved scheduling: Scheduling production in a way that considers potential bottlenecks can help prevent them from occurring.
g) Collaboration and communication: Encouraging collaboration and communication among different departments can help identify and resolve bottlenecks more effectively.

There are some specific ways to eliminate bottleneck in any simulation model such as increasing the inter-arrival time of source1, decreasing the setup time/processing time or adding some more machines in production line. The most preferable method for line balancing is to increase the capacity i.e. adding more machines in production lines. Since the bottleneck is created before VMC machine, so, add one more VMC machine in production line as shown in Fig. 4. After removing this bottleneck, the production line efficiency becomes 97.29%. The increase in throughput value after eliminating bottleneck is also shown in Fig. 5.

By simulating the behaviour of the production system, it can determine whether the production targets can be met within the specified time and evaluate the performance of each machine in the production line. The simulation results can provide valuable information about the capacity, efficiency, and bottlenecks in the production process, which can be used to identify areas for improvement and optimize the production line for maximum performance.

Effective operations and machinery are crucial for the automotive industry to sustain its growth and meet the demands of the market. Efficient and reliable processes and equipment help improve production capacity and product quality, leading to increased customer satisfaction and competitiveness in the market.

Modelling and simulation have certainly had a significant impact on the process of launching new businesses. They provide a powerful tool for evaluating and testing new ideas, concepts, and designs in a virtual environment before they are implemented in the real world. This allows entrepreneurs to identify and address potential challenges and make informed decisions, leading to more effective and efficient launch processes. The results indicate that by eliminating the bottleneck in the production line, it increases the line efficiency as well as the production throughput.

Conclusions

By creating a virtual model of the manufacturing process, it's possible to evaluate various scenarios and test different solutions to identify the root cause of any bottlenecks and determine the most effective way to remove them. Initially production line efficiency is 51.8% when bottleneck is present at queue 4. By adding one more vertical milling centre in the production line, queue 4 bottleneck is removed, and it increases not only throughput but also line efficiency to 97.29%. All manufacturing sectors will profit from these findings. The development of new, effective manufacturing facilities for a variety of sectors can

also benefit from the use of this simulation model. By employing a global table with various part sets and automatically updating the database with corporate databases, further research may also be expanded. Future research might also examine how adding buffers affects system performance within the design of the processing stations.

References

1 Ishak A, Zubair F & Cendani A S, Production line simulation in vise using the Flexsim application, in *IOP Conf Series: Mater Sci Eng* (IOP Publishing) **1003(1)** (2020) 1–6, doi: 10.1088/1757-899X/1003/1/012103.

2 Bjorklund S, Yamani N & Lloyd T, Simulation software for real time forecasting as an operational support, *Proc of the Second World Conf POM 15th Annual POM Conf* (Cancun, Mexico) 2004, 1–30.

3 Kliment M, Trojan J, Pekarcikova M & Kronova J, Creation of simulation models using the Flexsim software module, *Advanced Logistic Systems - Theory and Practice*, **16(1)** (2022) 41–50, doi: 10.32971/als.2022.004.

4 Krishna L S R, Saketh T S S & Kumar M V S, A simulation based approach for studying the effect of buffers on the performance of an FMS, *Int J Eng Sci Invent*, **4(1)** (2015) 5–9.

5 Fauadi M, Azimi N Z, Anuar N I, Ali M M, Sued M K & Ramlan S, Simulation approach for logistical planning in a warehouse : A review, *Orient J Comput Sci Technol*, **11(4)** (2018) 201–208.

6 Jarernram J & Samattapapong N, Parallel machine scheduling using simulation software, *Proc* of the *ICMEIT* (Pattaya, Thailand) 2018, 150–155.

7 Rodrigues R P, Pinho A F D & Sena D C, Application of hybrid simulation in production scheduling in job shop systems, *Simul Trans of the Soc Model Simul Int*, **96(3)** (2019) 1–16, doi: 10.1177/0037549719861724.

8 Nie X & Wang L, Simulation process design for scheduling mode of railway container terminals based on Flexsim, *Proc J Phys : Conf Ser*, **1176(5)** (2019) 1–6, doi: 10.1088/1742-6596/1176/5/052012.

9 Yaser S, Abdelatif N, Fahim I, Emad Y, Saleh A, & Kassem S S, FlexSim simulation to enhance productivity of a production cell: A case study, *Proc* of the *3rd NILES* (Giza, Egypt) 2021, 312–317. doi: 10.1109/NILES53778.2021.9600545.

10 Wang S, Wang S M & Zhang N, Flexsim-based simulation and optimization of green logistics distribution center, *Proc of the 14th ICCMS* (Chongqing, China) 2022, 76–82, doi: 10.1145/3547578.3547590.

11 Hassan N A, Arogundade A I, Iyenagbe U B & Musa D I, Simulation and analyses of shea nuts (vitallaria paradoxa) processing plant using FlexSim, *J Fut Sustain*, **3(2)** (2023) 67–74, doi: 10.5267/j.jfs.2022.11.006.

12 Ljaskovska S, Martyn Y, Malets I & Velyka O, Optimization of parameters of technological processes means of the flexsim simulation simulation program, *Proc of the IEEE 3rd ICDSMP* (Lviv, Ukraine) 2020, 391–397, doi: 10.1109/DSMP47368.2020.9204029.

13 Luscinski S & Ivanov V, A simulation study of industry 4.0 factories based on the ontology on flexibility with using flexsim software, *Manag Prod Eng Rev*, **11(3)** (2020) 74–83, doi: 10.24425/mper.2020.134934.

14 Watnakornbuncha D, Muangwai A, Wichitphongsa W, Amdee N & Inthawongse C, Processing cost reduction of lemon products in community enterprises using Flexsim simulation, *Proc of the IEEM* (Kuala Lumpur, Malaysia) 2022, 670–674, doi: 10.1109/IEEM55944.2022.9989661.

15 Rao G V P, Nallusamy S & Raman P, Enhancement of production in subassembly line of a medium scale industry using different lean tools and FlexSim simulation software, *Int J Eng Res Africa*, **44** (2019) 229–239, doi: 10.4028/www.scientific.net/jera.44.229.

16 Amalia P & Cahyati N, Queue analysis of public healthcare system to reduce waiting time using flexsim 6.0 software, *Int J Ind Optim*, **1(2)** (2020) 101–110, doi: 10.12928/ijio.v1i2.2428.

17 Krynke M, Personnel management on the production line using the Flexsim simulation environment, *Manuf Technol*, **21(5)** (2021) 657–667, doi: 10.21062/mft.2021.073.

18 Patil R J, Kubade P R & Kulkarni H B, Optimization of machine shop layout by using Flexsim software, *Proc of the AIP Conf* (AIP Publishing) **2200(1)** 2019, doi: 10.1063/1.5141203.

19 Kliment M, Trojan J, Pekarcíková M & Kronova J, Creation of simulation models using the Flexsim software module, *ALS-Theory and Practice*, **16(1)** (2022) 41–50, doi: 10.32971/als.2022.004.

20 Kessel D, Jeon J, Jung J, Oh E & Kim C, Logistical simulation modeling for planning a soil remediation process, *Sci Technol Nucl Install*, **2019** (2019) 1–13, doi: 10.1155/2019/6789506.

21 Kikolski M, Identification of production bottlenecks with the use of plant simulation software, *Eng Manag Prod Serv*, **8(4)** (2016) 103–112, doi: 10.1515/emj-2016-0038.

Analysis and mapping of the indoor acoustic environment in university study spaces

Shreya Jha[1], *Anoushka* Gupta[1], *Akshansh* Gupta[1], *Rajeev* Kumar Mishra[11], and *Ravi* Pratap Singh Jadon[1]

Department of Environmental Engineering, Delhi Technological University, Delhi, India-110042

Abstract. The university setting significantly impacts the student's learning and academic achievement. A vital aspect shaping their productivity and well-being within this environment is the indoor soundscape. Different organizations, nationally and internationally, have emphasized the importance of addressing noise-related issues and have established guidelines. This study investigates how indoor soundscapes affect student learning and achievement in university classrooms. The present study analyses noise levels at nine locations during morning, afternoon, and evening study hours using a class 1 sound level meter. Results show noise levels ranging from 49.8 to 77.6 dBA, averaging 66.1 dBA. Over 96% exceeded the ANSI, CPCB noise limit of 50 dBA, with 62.96% surpassing WHO's 55 dBA standards. Seven of nine locations maintained acceptable noise levels 95% of the time, considering a ±10 dBA fluctuation allowed by CPCB. Libraries and computer labs generally had lower noise, while engineering labs had higher levels due to noisy equipment. This research holds significance for educational institutions, policymakers, and stakeholders in creating conducive learning environments. Understanding prevailing soundscapes informs the design and management of study areas, and helps optimize learning spaces for improved productivity.

Keywords. University Campus, Learning Environment, Soundscapes, Noise Pollution Level (NPL), Noise Mapping, GIS.

1 Introduction

Constantly, we are surrounded by sound, from the tranquil buzz of nature to the cacophony of city settings. While certain sounds are enjoyable and essential for communication, unwelcome or excessive noise can heavily affect our health, overall wellness, and cognitive abilities [1,2]. Excessive or disruptive undesired sounds in a specific location are categorized as noise. When such sounds persistently occur in the surrounding environment, it is termed noise pollution [3]. Noise pollution occurs when the sound volume exceeds permissible limits for a particular environment [4]. Elevated levels of noise often stem from diverse origins, broadly categorized as natural and manmade sources [5]. Natural noise consists of sounds produced by humans, insects, birds, barking dogs, thunderstorms, wind, etc. Conversely, man-made sources encompass traffic, industrial activity, construction work, loudspeakers during festivals, railways, and firecrackers

* Corresponding author : rajeevkumarmishra@dtu.ac.in

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

[6], collectively contributing significantly to overall environmental noise pollution. Noise pollution has been increasingly impacting the worldwide populace, particularly in large urban areas [7]. Long-term exposure to loud noises can lead to both mental and physical health issues in individuals [8]. As per a recent study published by the World Health Organisation (WHO), approximately 100 million individuals in Europe are subjected to roadway noise, resulting in a yearly minimum loss of 1.6 million years of healthy life [9].

Noise in educational institutions such as schools and colleges is a major source of stress and irritability [10]. These institutions often experience high levels of external noise sources such as roadside traffic, construction activities, and other internal factors such as student conversations or sound-generating devices [11]. Factors such as teaching methodologies, student and educator behaviour, student demographics, and the nature of the learning activities can impact noise generation [12,13,14].

The operation of noisy equipment like laboratory machines, tools and audio-visual systems in a room can contribute to increased overall sound levels, leading to disturbances throughout the site space [15]. The specific type and intensity of the equipment, as well as its proximity to individuals, can impact the extent of noise disruption [16]. Old equipment that is malfunctioning or not maintained properly may produce disruptive sounds, which influences the acoustic environment [12]. Heating, ventilation, and air conditioning (HVAC) systems present in the space under study may also generate noise due to the operation of various components such as fans, compressors, air handlers, etc. Different types of HVAC systems produce different levels of noise [12,17].

Excessive external noise levels can increase indoor noise levels, making it challenging for occupants to concentrate and communicate effectively [18]. Studies have shown that on average, the noise level recorded in schools is around 57 dBA, with roadside traffic being the primary source of noise [19]. Noise from other sources like construction work or loud music can pass through windows, doors, or walls of a room[20]. The status of doors or windows in a room, whether they are open or closed, can have a substantial impact on noise levels. Closed ones provide a barrier that helps block the transmission of sound waves between rooms or from external sources, thereby reducing noise infiltration. On the other hand, open ones allow sound to travel more freely, leading to increased noise levels within the room [21,22]. Properly sealed pathways with soundproofing features can further enhance noise reduction capabilities [23]. Factors such as poor-quality acoustic, ceiling tiles, lack of acoustically treated furniture, lack of carpet, and single pane windows can enhance the background noise levels in classrooms, disrupt learning activities, and impact the performance of both students' and teachers' [24].

The combination of these noises can create disturbances in the classroom environment, thereby creating an unhealthy environment for students and teachers [11]. Various researchers have investigated the impact of various meteorological variables such as temperature, relative humidity, and wind speed on exacerbating or mitigating urban noise pollution [25]. Regression analysis revealed that wind speed plays a positive role in reducing noise levels, while temperature tends to increase ambient noise levels. Additionally, lower relative humidity was found to significantly decrease noise levels [26,27].

According to studies by the WHO, exposure to noise levels above 53 dBA can lead to increased blood pressure, heart rate, and muscle weakness [2]. While students may become accustomed to these noise levels over time because this is what they hear daily, the potential dangers associated with it should not be overlooked [1]. Studies by leading scientists suggest that intermittent exposure to 100 dBA may have little effect on human health, but continuous

exposure throughout the day can lead to psychological ailments like anxiety, depression, stress, and fatigue [9]. It may also cause cardiac, auditory, or cognitive problems in individuals of all ages from foetuses to the elderly [10,28]. It is therefore crucial for individuals to be conscious of their daily exposure to elevated levels of noise to protect their overall health. Generally, verbal communication, teaching, learning, mental tasks, and sleep are typically the most impacted [29]. Exposure to disturbing noises can lead to a delay of up to one month in the reading and comprehension abilities of young individuals. Many people may not fully comprehend the negative effects that noise can have, particularly those who are consistently exposed to lower levels of noise over extended periods [9].

Researchers and the public, are increasingly raising their concerns about this issue and there has been a surge in the development of laws and policies to reduce and mitigate the impact of these unacceptable noises on the social environment [30]. International standards for classroom listening environments have been laid down by the World Health Organization (WHO) [9], American National Standard Association (ANSI) [31], American Speech Language Hearing Association (ASHA) [32], and Indian national standards by the Central Pollution Control Board (CPCB), India [33]. Various organizations have also established standards for building design and acceptable acoustic levels to manage background noise and reverberation in classrooms.

Larger classrooms have longer reverberation times, leading to increased background noise levels. The research found that increasing classroom size is co-related with worse speech intelligibility [34]. Additionally, larger spaces may require louder speaking volumes from teachers or students to be heard, contributing to overall noise levels [12,35,36,37]. Researchers have also discovered that classrooms with tall ceilings reduced noise levels and facilitated higher comprehensibility of speech than classrooms with small height ceilings [38,39]. Ceiling shape can also affect noise levels in a variety of ways. Flat ceilings can contribute to sound reflection and reverberation, thereby increasing noise levels. In some cases, a slatted, vaulted, or pitched roof can reduce the acoustic energy input compared to a flat roof, thus reducing the overall noise level [40].

Material of construction of interior and exterior walls, ceilings, and floors can significantly affect reflection and noise excitation. Thick walls, partitions, and hard finishes provide excellent acoustic insulation, thus reducing noise transmission and reducing indoor noise levels [36,41]. Buildings made of hard materials such as brick, concrete, and stone provide better sound insulation compared to lighter materials such as wood or vinyl siding [36]. However, wear and tear of buildings over time leads to gaps, cracks, or structural weaknesses in the building envelope, allowing more external noise to infiltrate the walls of the room. Acoustic ceiling tiles or panels with sound-absorbing nature help reduce noise levels by absorbing excess sound energy. Conversely, hard roof materials such as concrete or metal can reflect sound, resulting in increased noise levels and poor sound quality [42]. The composition of furniture affects noise levels through its acoustic properties. Soft, upholstered surfaces absorb sound waves, helping to muffle noise and reduce reverberation. In contrast, dense furniture reflects sound and can contribute to increased noise levels [36,43]. Similarly, soft flooring materials such as carpets or mats absorb sound waves and reduce noise whereas, solid materials such as tile or hardwood floors reflect sound, which amplifies the noise [36]. Factors such as substandard acoustic ceiling tiles, absence of acoustically treated furniture, lack of carpeting, and single-pane windows can amplify background noise levels in classrooms. This can disrupt learning activities and impact the performance of both students and teachers [24].

In 2001, R. Murray Schafer, a renowned Canadian musician, and composer, introduced a concept known as "soundscape," which refers to any specific auditory snapshot of the

surrounding sound environment that is recognized as a distinct field of study. Schafer's definition of "soundscape analysis" encompasses the comprehensive examination of all kinds of sounds present in a particular area or region [44]. A Sound level meter is a tool that empowers soundscape
design by enabling precise measurement of ambient noise levels [45] and identification of specific
noise sources. This data informs the creation of soundscapes that are both aesthetically pleasing and functionally suited to the environment. Noise mapping is a contemporary method that illustrates noise level metrics of various areas on a unified map [5], offering a visual representation of noise level within a specific region. This tool is valuable for acoustic urban planning as it aids in pinpointing noise pollution hotspots within an area [46].

The significance of a favourable listening environment is well acknowledged; however, there remains a noticeable dearth of data on sound environments within higher education institutions in India. This study delves into a previously uncharted territory: the acoustic ecology of study areas within university settings. While past research has outlined the sound environments of university campuses, this research goes further by concentrating on the distinctive sonic ambiance within specific indoor study spaces. The purpose of this research is to quantify the acoustic environment of various study spaces within Delhi Technological University (DTU). This involves measuring noise metrics like L_{eq}, L_{10}, $L_{50,}$ $L_{95,}$ and NLP using a sound level meter, mapping their distribution across these study areas on QGIS, and comparing them to the standards prescribed by regulatory bodies like CPCB [33], ANSI [31] and WHO [9]. The authors anticipate that this study will serve as an asset for educational authorities by proposing potential improvements to existing study areas and offering guidance for future university construction projects, thereby striving to create more optimal listening conditions.

2 Materials and Methods

2.1 Study Area

The research investigation was conducted at the primary campus of Delhi Technological University (DTU) situated in New Delhi, India. Spanning an expansive total area of 163.9 acres (equivalent to 66.3 hectares), DTU's campus comprises 10% dedicated to building structures, with the remaining portion allocated to open grounds, lush gardens, and interconnected pathways. Noteworthy is the campus's sizable population, accommodating a student body of 11,000 individuals enrolled in various undergraduate and postgraduate programs. For the research endeavour, a meticulous selection process identified a total of nine indoor classrooms and study areas (referenced in Table 1 and visually depicted in

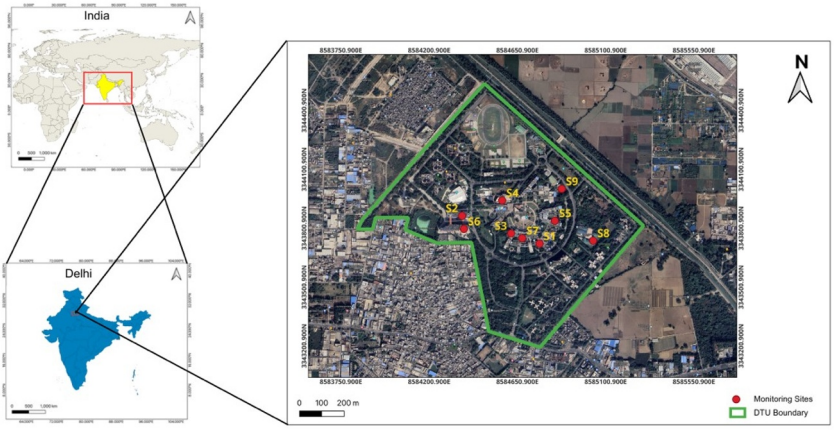


Fig**. 1**) dispersed strategically within the campus premises. The rationale behind choosing these specific locations was rooted in their distinct positional diversity, unique classroom attributes, and the widespread representation of students across the campus landscape. The objective was to capture an exemplary sample of the campus environment to analyse spatial and temporal variations in noise levels.

Table 1. List of selected sites on the campus along with their location codes, latitude, and longitude

Code	Site Name	Latitude (° N)	Longitude (° E)
S1	Water Engineering Laboratory	28.748876	77.118243
S2	Machine Learning Laboratory	28.749993	77.114696
S3	Power Electronics Laboratory	28.749286	77.116941
S4	Library Reading Classroom	28.750609	77.116529
S5	Mechanical Engineering Workshop	28.749793	77.118948
S6	Academic Block Classroom	28.749468	77.114776
S7	Old Department Classroom	28.749096	77.117449
S8	Pragya Bhawan Classroom	28.748994	77.120699
S9	SPS Classroom	28.751074	77.119260

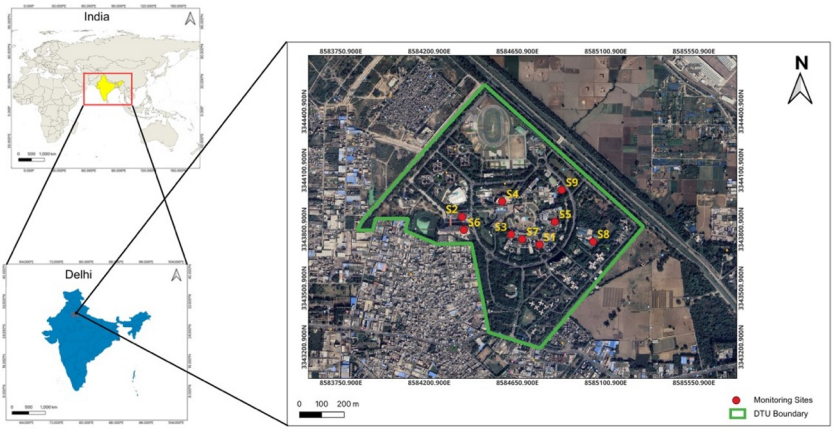


Fig. 1. Map of locations within the DTU campus selected for noise level monitoring

2.2 Methodology

2.2.1 Tools and Techniques

The study utilized a CESVA SC 310 sound level meter (or SLM) to measure noise level. This Class 1 Integrated SLM offers a frequency weighting of A, C, and Z, with a measurement range from 30 dBA to 137 dBA and an accuracy of ±0.6 dBA. Measurements were taken using the dBA scale, chosen for its suitability in assessing noise sources typical in educational settings and aligning with perceived speech levels. The SLM was calibrated according to IEC 61672-1 standard to ensure accuracy. The characteristics of the SLM [47] are provided in Table 2. The A-weighting scale was chosen due to its equal emphasis on all noise sources except for lower frequencies, aligning closely with speech levels perceived by listeners. Measurements were taken at each site as shown in Fig. **2**. During active classroom events such as discussions and presentations, as well as quiet activities like reading or writing, measurements of noise levels were taken. Furthermore, external sources of noise such as sounds from the corridor, construction work, and vehicle traffic were also observed and documented [11]. As a result, an

occupied state was defined when the classroom contained students, a teacher, furniture, and any functioning ventilation systems (e.g., heat blowers).

Table 2. Characteristics of CESVA SC 310

Product	CESVA SC 310
Type	Class 1 Integrating Sound Level Meter
Frequency Weighting	A, C, Z
Measurement Range	30 dBA to 137 dBA
Accuracy	± 0.6 dBA
Display	LCD screen
Communication	USB and RS-232 ports
Microphone	½ "condenser microphone



Fig. 2. Noise measurements taken at the monitoring site

2.2.2 Study Design

Noise measurements were conducted three times per day during the morning, noon, and evening periods. Morning measurements were taken from 09:00 to 11:00 hrs, afternoon measurements from 12:00 to 14:00 hrs, and evening measurements from 16:00 to 18:00 hrs. These measurements were carried out in classrooms S1 to S9 from Monday to Friday during study hours.

The measuring procedure was conducted according to the protocols given by the CPCB, in 2015 for ambient noise level monitoring [48]**.** The SLM was calibrated to concerning IEC 61672-1 standard [49]. Noise levels were measured in dBA scales at each location for 15:00 minutes [50,51].

The sound level meter provided the L_{eq}, L_{10}, $L_{50,}$ and L_{95} values for the entire monitoring. L_{eq} represents the equivalent continuous sound level over time. The L_{10} value indicates the peak noise component (noise levels equal to or exceeding 10% of the study duration), while the L_{50} value denotes the median noise (noise levels equal to or exceeding 50% of the study duration). Additionally, L_{95} signifies the background noise components (noise levels equalling or exceeding 95% of the study duration) [52].

LA_{eq}, the measurement used to describe the average sound level over a defined period was calculated for the period of 15 minutes and denoted by LA_{eq},15min. The formula here used is given in (1) [53].

$$LA_{eq} = 10 \log \sum f_i * 10^{Leq/10} \quad (1)$$

Where,
f_i = fraction of the total duration at a consistent level
L_{eq} = sound level in dBA.

The standard deviation (σ) was calculated for the 15-minute noise values. In statistics, the standard deviation measures the extent of dispersion or variation within a dataset. A small standard deviation suggests that the data points are clustered around the mean (or expected value) of the set, while a large standard deviation signifies that the data points cover a wide range of values.

Several research studies have indicated that a wider variation in sound distribution contributes to increased annoyance. The metric, noise pollution level (LNP) incorporates the diversity of sounds and is utilized for evaluating community noise impacts [53, 54]. Here LNP,15min was calculated and it is defined as

$$LNP = LA_{eq} + k\sigma \quad (2)$$

Where,
LNP = level of noise pollution in decibels (dBA)
LA_{eq} = average equivalent sound level (dBA)
σ = standard deviation of the sound level
k = constant with a value determined to be 2.56

To comprehend the overall noise exposure experienced by students at specific sites throughout the day, an average of the morning, afternoon, and evening LA_{eq},15min values were computed using the formula in equation 1 and labelled as LA_{eq}. Then similarly, LNP, LA_{10}, $LA_{50,}$ and LA_{95} were also calculated.

All the above-stated metrics in all nine locations were tabulated in **Error! Reference source not found.** and **Error! Reference source not found.**. Bar charts were prepared to represent the temporal and spatial noise variations. Fig**. 3** 3 depicts the spatial and temporal variation of the noise level in all 9 locations. A clustered chart was also made to represent the temporal variation in the LA_{eq},15min value, for all the nine sites during morning, afternoon, and evening periods. The values of noise standards prescribed by the ANSI [31], CPCB [33], and WHO [9] were also plotted alongside the LA_{eq}, 15-minute values.

Spatial maps of LA_{eq}, LNP, LA_{10}, $LA_{50,}$ and LA_{95} were developed using techniques like inverse distance weighted (IDW) interpolation on Quantum GIS. A version of QGIS 3.36.0-Maidenhead [7,55] was used to visualize the distribution of noise levels across different study spaces within the university. Alongside noise measurements, ambient conditions and factors that could influence the noise level in an indoor space were observed. Both quantitative and qualitative factors, which were both constant and varying in nature were identified and are listed in Table 3 and

Table **4**.

Table 3. Quantitative factors observed during monitoring

Quantitative Factors		
S.no	**Variable Name**	**Description**
1	Wind Speed	Wind speed in the classroom
2	Number of Students	Number of students in the classroom

Quantitative Factors		
S.no	**Variable Name**	**Description**
1	Wind Speed	Wind speed in the classroom
2	Number of Students	Number of students in the classroom
3	Classroom Area	Size of the classroom
4	Ceiling Height	Height of the classroom to ceiling

Table 4. Factors noticed during monitoring that are qualitative in nature

Qualitative Factors		
S.no	**Variable Name**	**Description**
1	Internal Noise Origin/Sources	Noises within the classroom
2	External Noise Origin/Sources	Noises outside the classroom
3	Day Time	Time of day when the noise level was recorded
4	Acoustic Treatment	Implementation of acoustic treatment
5	Ceiling Shape	The shape of the classroom ceiling

The factors – classroom area, ceiling height, acoustic treatment, and ceiling shape are constant, while the remaining variables – wind speed, number of students, internal and external noise sources, and day time are the varying ones that were observed during the noise level monitoring at the nine sites. The data for the constant factors was obtained from the DTU Civil Office. Each factor has some effect on the noise levels. While some factors like classroom characteristics & acoustics, the number of students, and their behaviour, have major and noticeable effects, factors like time of the day and meteorological factors have a minor.

3 Results and Discussions

Noise levels at the nine chosen study areas were monitored and documented. The study outcomes are showcased through numerical tables, charts, and maps. Specifically, Table 5 displays the recorded fifteen-minute equivalent noise levels, the standard deviation in the readings over the fifteen minutes, and the computed noise pollution levels at these nine locations during morning, afternoon, and evening shifts.

Table 5. Spatial and Temporal Variation of Equivalent noise level (LA_{eq},15min), Standard Deviation (σ), and Noise pollution level (LNP,15min)

Site	Time Of The Day	LAeq,15min (dBA)	Standard Deviation (σ)	LNP,15min (dBA)
S1	Morning	51.0	2.5	57.4
	Afternoon	72.2	2.2	77.8
	Evening	51.2	2.1	64.0
S2	Morning	54.3	8.1	75.1
	Afternoon	52.6	3.9	62.5
	Evening	54.2	2.6	60.8
S3	Morning	61.1	4.6	72.8
	Afternoon	66.3	1.6	70.5
	Evening	65.6	4.7	77.6
	Morning	52.4	3.3	60.9

S4	Afternoon	49.8	1.8	54.5
	Evening	55.3	3.3	63.7
S5	Morning	57.6	3.6	66.9
	Afternoon	77.6	11.6	107.3
	Evening	49.7	6.0	67.1
S6	Morning	62.3	4.2	73.1
	Afternoon	66.0	1.3	69.3
	Evening	62.3	3.0	69.5
S7	Morning	61.0	4.1	71.5
	Afternoon	66.8	2.0	71.9
	Evening	61.0	2.0	71.5
S8	Morning	65.0	0.9	67.2
	Afternoon	57.9	2.0	63.0
	Evening	55.2	4.7	67.4
S9	Morning	61.0	2.7	68.0
	Afternoon	60.7	4.3	71.8
	Evening	65.6	2.0	70.6

The highest decibel levels (LAeq,15min**)** recorded during the morning, afternoon, and evening shifts were 65 dBA in S8, 77.6 dBA in S5, and 65.6 dBA in S3, respectively. Conversely, the lowest levels for the three shifts were 51 dBA in S1, 49.7 dBA in S4, and 51.2 dBA in S1, respectively. Overall, the noise levels fluctuated between 49.7 dBA and 77.6 dBA throughout the day, with the highest noise levels typically observed in the afternoon, followed by the evening and morning shifts.

As mentioned earlier in the paper, numerous external and internal factors contribute to the noise levels within a given space. This variation of the varying characteristics often results in standard deviations in the measured values. For instance, abnormal standard deviations of 8.1in S2 during the morning and 11.6 in S5 during the afternoon were attributed to the utilization of noisy laboratory equipment and student discussions.

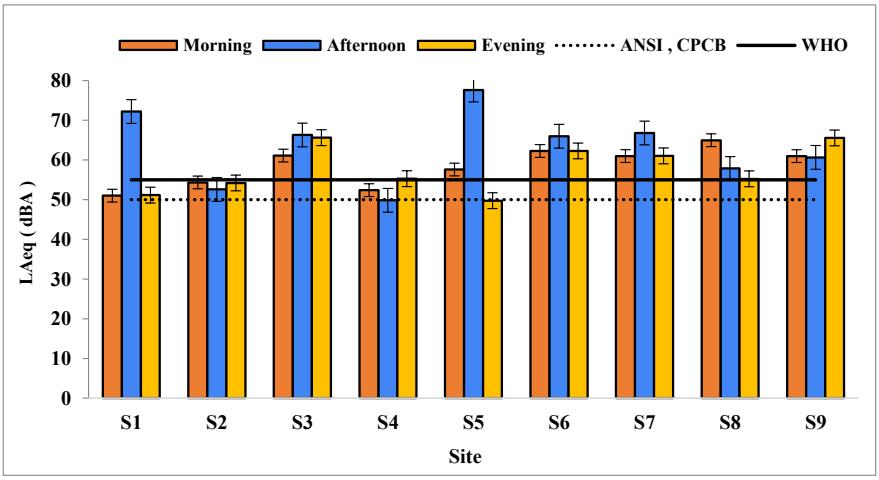


Fig. 3. Graphical representation of the spatial and temporal variation of LAeq,15min at the nine locations, and their comparison with noise standards

The findings from the objective measurements were juxtaposed against the noise limits recommended by ANSI, CPCB, and WHO for urban educational settings, which include schools and universities (Fig. **3**). The noise level assessments revealed that 96.29% of the recorded points surpassed the noise limit set by ANSI and CPCB (LA_{eq} = 50 dBA), while 62.96% of the measured points exceeded the standards outlined by the WHO (LA_{eq} = 55 dBA).

Table 6. Spatial variation of LA_{eq}, LNP, LA_{10}, LA_{50} and LA_{95}

	Noise Metrics (dBA)				
Site	**LA_{eq}**	**LNP**	**LA_{10}**	**LA_{50}**	**LA_{95}**
S1	67.5	73.2	70.3	67.4	63.7
S2	53.8	70.7	57.2	53.4	49.0
S3	64.9	74.6	68.4	64.4	60.1
S4	53.1	61.1	57.1	52.2	48.4
S5	72.9	102.5	74.1	73.0	72.3
S6	63.9	71.0	68.1	62.4	57.3
S7	63.9	71.6	68.4	64.8	57.8
S8	61.3	66.3	65.0	61.0	55.6
S9	63.0	70.4	66.5	62.7	57.9

The average noise level in the indoor spaces (average of 9 sites) was 66.1 dBA. To comprehend the overall noise exposure experienced by students at specific sites, an average of the morning, afternoon, and evening Leq values was computed and labelled as LA_{eq}. Table 6 presents the LA_{eq} values, ranging from 53.1 to 72.9 dBA. The most noise-polluted area was the mechanical engineering laboratory (S5), primarily attributed to the utilization of heavy and noisy instruments like the metal cutter. Following closely was the water engineering laboratory (S1), situated near a road, correlating the high noise levels to vehicular traffic movements and the operation of noisy lab equipment. Additionally, the high LA_{eq} in the Power Electronics

Laboratory (S3) was linked to the presence of electrical machines. Conversely, the Library and Computer Engineering laboratory were identified as the least noisy spaces. Despite this, the overall results exceeded permissible limits, as the university campus is situated in a silent zone, contrasting the recorded noise levels. Fig. **4** illustrates the map of LA_{eq} values across the sites, facilitating the identification of noise level hotspots and enabling the implementation of necessary measures to alleviate the issue.

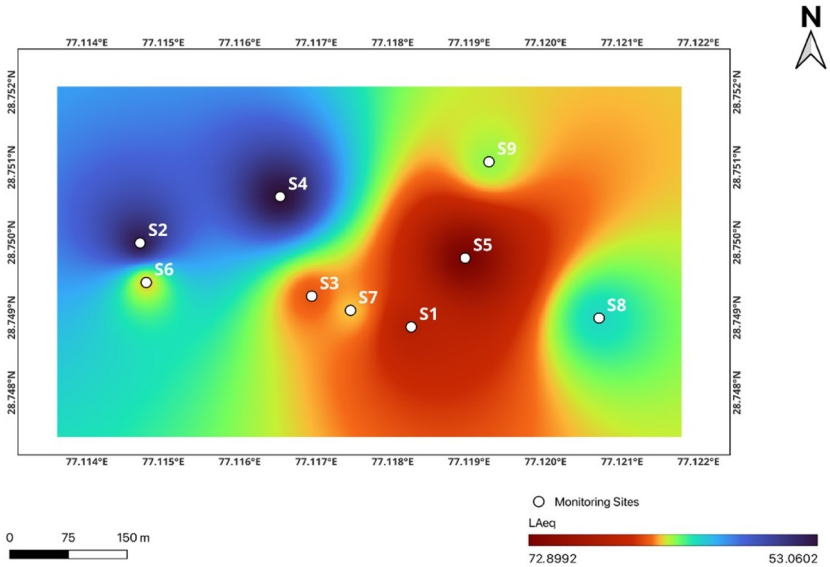


Fig. 4. Mapping of Equivalent Noise Level (LA_{eq}) in the nine sites across the DTU campus

The evaluation incorporated the Noise Pollution Level (LNP), which integrates LA_{eq} with a component signifying the noise variability, namely the standard deviation of instantaneous levels. This metric can be beneficial, particularly in scenarios where noise annoyance is a focal point. LNP is valuable for comparing noise levels in diverse environments with varying noise characteristics, as it provides a more comprehensive assessment than relying solely on LA_{eq}. Fig. **5** depicts the Mapping of Noise Pollution Level (LNP) within the nine sites across the DTU campus, visually representing the noise pollution levels present in each location. Here, we can infer that S5 is the most polluted due to the usage and stoppage of machines.

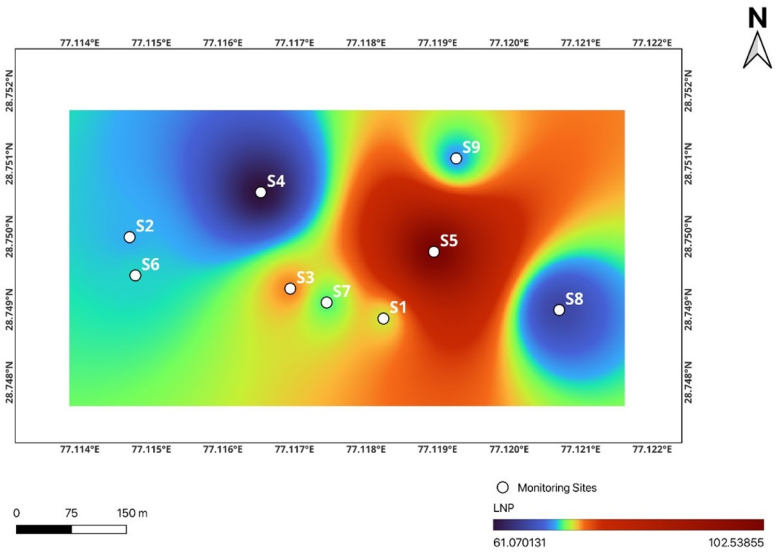


Fig. 5. Mapping of Noise Pollution Level (LNP) in the nine sites across the DTU campus

The noise metrics, i.e. the percentiles of noise level - LA_{10}, LA_{50}, and LA_{95}, were also discussed in this study. Values of these noise indicators are provided in Table 6. For example, looking at Fig. **6**, which shows the representation of LA_{10} (the ten percentile time exceeding noise level), it is observed that peak levels of intrusive noise exceed 10% of the total observation time with values ranging from 57.1 to 74.1 dBA across nine sites throughout the day. Additionally, based on Fig. **7**, showing the variation of LA_{50} (representing average noise level values), a span from 52.2 to 73 dBA indicates that only two classrooms adhere to ambient standards for about half of the time. Fig. **8** displays the background noise level (LA_{95}) and illustrates intense levels surpassing established guidelines or permissible values in three out of nine sites. Considering that the CPCB allows a maximum deviation of ±10 dBA from the noise standards, after adjusting for this range, it can be inferred that seven out of the nine sites maintain their noise levels within an acceptable range for 95% of the time.

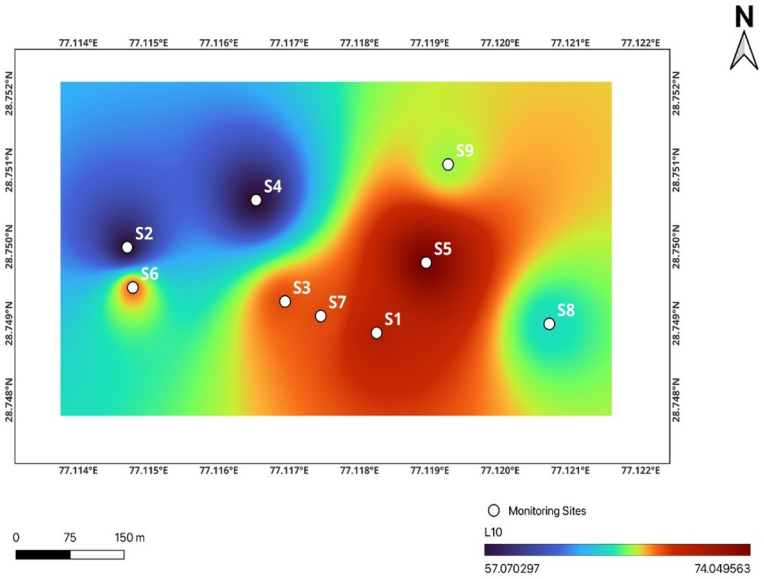


Fig. 6. Mapping of Peak Noise Level (LA_{10}) in the nine sites across the DTU campus

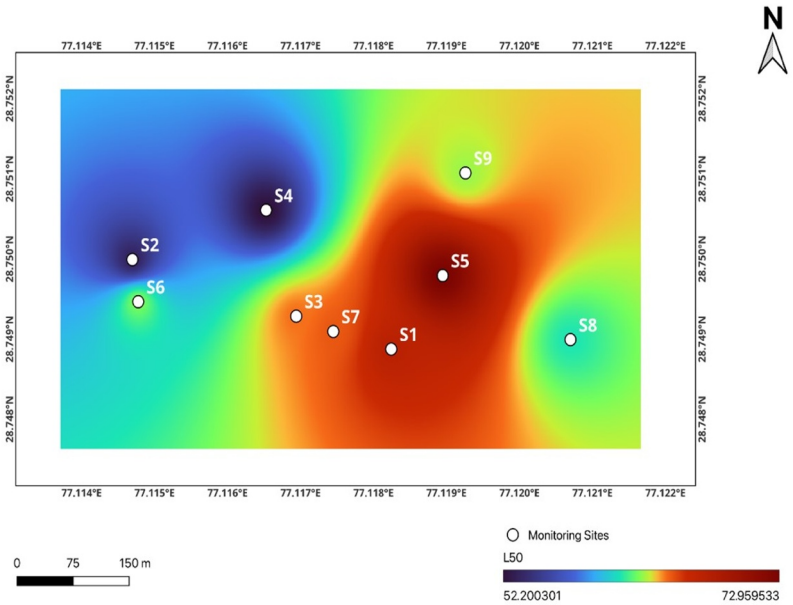


Fig. 7. Mapping of Mean Noise Level (LA_{50}) in the nine sites across the DTU campus

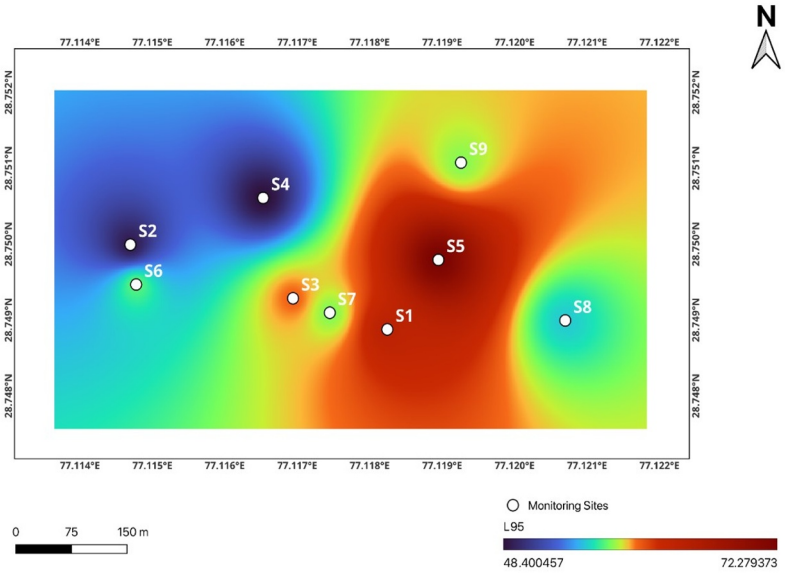


Fig. 8. Mapping of Background Noise Levels(LA95) at nine locations on the DTU campus4 Conclusions and recommendations

The research into the acoustic environment of university learning spaces has provided important insights into the impact of noise levels on student well-being and academic achievement. Various factors and their influence on classroom acoustics were also described thoroughly in the study. Objective measurements of sound pressure showed that most of the recorded points exceeded established university noise standards, highlighting the urgent need for active noise control measures in educational settings.

To effectively address noise concerns, it is recommended that educational institutions prioritize adhering to building design considerations laid down by the certified bodies, use soundproofing materials in classrooms and laboratories, as well as manually try to keep the decibels as low as possible. Student awareness programs and flexible classroom procedures should be followed to support a calmer and more productive learning environment. Keeping indoor plants is another way to reduce noise levels as they help absorb sounds. Research and development in noise mathematical modelling and acoustic simulation are encouraged to increase the effectiveness of noise abatement techniques and enhance student well-being. By integrating and applying these suggestions, universities can provide healthy learning environments to the students, boosting their academic performance and success.

5 References

[1] M. M. Damián-Chávez, P. E. Ledesma-Coronado, M. Drexel-Romo, D. I. Ibarra-Zárate, and L. M. Alonso-Valerdi, "Environmental noise at library learning commons affects student performance and electrophysiological functioning," *Physiol Behav*, vol. 241, Nov. 2021, doi: 10.1016/j.physbeh.2021.113563.

[2] K. Mealings, "A Scoping Review of the Effects of Classroom Acoustic Conditions on Primary School Children's Physical Health," *Acoust Aust*, vol. 50, no. 3, pp. 373–381, Sep. 2022, doi: 10.1007/s40857-022-00271-8.

[3] C. El Yamlahi Chahdi, B. El Wahbi, Y. El Madhi, and A. Soulaymani, "Noise pollution in school institutions: Study case at Sidi Yahya Zaer middle school in Temara, Morocco," *Social Sciences and Humanities Open*, vol. 9, no. March, 2024, doi: 10.1016/j.ssaho.2024.100877.

[4] M. H. Masum, S. K. Pal, A. A. Akhie, I. J. Ruva, N. Akter, and S. Nath, "Spatiotemporal monitoring and assessment of noise pollution in an urban setting," *Environmental Challenges*, vol. 5, Dec. 2021, doi: 10.1016/j.envc.2021.100218.

[5] M. Cai, J. Zou, J. Xie, and X. Ma, "Road traffic noise mapping in Guangzhou using GIS and GPS," *Applied Acoustics*, vol. 87, pp. 94–102, 2015, doi: 10.1016/j.apacoust.2014.06.005.

[6] L. S. Deconto, M. B. S. Novelli, D. J. da S. Lima, and E. L. A. Monteiro-Filho, "Influence of natural and anthropogenic sound sources on the soundscape of the Cananéia estuary, southeastern Brazil," *Mar Pollut Bull*, vol. 169, Aug. 2021, doi: 10.1016/j.marpolbul.2021.112502.

[7] P. Das, S. Talukdar, S. Ziaul, S. Das, and S. Pal, "Noise mapping and assessing vulnerability in meso level urban environment of Eastern India," *Sustain Cities Soc*, vol. 46, Apr. 2019, doi: 10.1016/j.scs.2019.01.001.

[8] S. Tet Leong and P. Laortanakul, "MONITORING AND ASSESSMENT OF DAILY EXPOSURE OF ROADSIDE WORKERS TO TRAFFIC NOISE LEVELS IN AN ASIAN CITY: A CASE STUDY OF BANGKOK STREETS," 2003.

[9] Organización Mundial de la Salud., *Environmental noise guidelines for European Region*. WHO Regional Office for Europe, 2018.

[10] I. Enmarker and E. Boman, "Noise annoyance responses of middle school pupils and teachers," *J Environ Psychol*, vol. 24, no. 4, pp. 527–536, Dec. 2004, doi: 10.1016/j.jenvp.2004.09.005.

[11] G. Sundaravadhanan, H. Selvarajan, and B. McPherson, "Classroom Listening Conditions in Indian Primary Schools: A Survey of Four Schools," *Noise Health*, vol. 19, no. 86, pp. 31–38, Jan. 2017, doi: 10.4103/1463-1741.199240.

[12] A. Y. Alqahtani, A. A. Makki, and H. M. Alidrisi, "Revealing factors influencing classroom noise in the universities teaching and learning environment: A design of experiments approach," *Journal of Engineering Research (Kuwait)*, vol. 11, no. 1, p. 100009, 2023, doi: 10.1016/j.jer.2023.100009.

[13] C. Visentin, M. Pellegatti, M. Garraffa, A. Di Domenico, and N. Prodi, "Individual characteristics moderate listening effort in noisy classrooms," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-40660-1.

[14] R. F. Guidini, F. Bertoncello, S. Zanchetta, M. Lúcia, and S. Dragone, "Correlations between classroom environmental noise and teachers' voice Correlações entre ruído ambiental em sala de aula e voz do professor," 2012.

[15] A. Jaramillo, M. Ermann, and P. Miller, "The teachers perspective on noise in the classroom," *Proceedings of Meetings on Acoustics*, vol. 19, no. 2013, 2013, doi: 10.1121/1.4799985.

[16] E. Sala and L. Rantala, "Acoustics and activity noise in school classrooms in Finland," *Applied Acoustics*, vol. 114, pp. 252–259, Dec. 2016, doi: 10.1016/j.apacoust.2016.08.009.
[17] C. Burke and I. Grosvenor, "School buildings," *The School I'd Like*, pp. 31–44, 2020, doi: 10.4324/9780203439074-10.
[18] B. Shield, R. Conetta, J. Dockrell, D. Connolly, T. Cox, and C. Mydlarz, "A survey of acoustic conditions and noise levels in secondary school classrooms in England," *J Acoust Soc Am*, vol. 137, no. 1, pp. 177–188, 2015, doi: 10.1121/1.4904528.
[19] B. Shield and J. E. Dockrell, "External and internal noise surveys of London primary schools," *J Acoust Soc Am*, vol. 115, no. 2, pp. 730–738, Feb. 2004, doi: 10.1121/1.1635837.
[20] A. Kaarlela-Tuomaala, R. Helenius, E. Keskinen, and V. Hongisto, "Effects of acoustic environment on work in private office rooms and open-plan offices - Longitudinal study during relocation," *Ergonomics*, vol. 52, no. 11, pp. 1423–1444, 2009, doi: 10.1080/00140130903154579.
[21] C. E. Mediastika, L. Kristanto, J. Anggono, F. Suhedi, and H. Purwaningsih, "Open windows for natural airflow and environmental noise reduction," *Archit Sci Rev*, vol. 61, no. 5, pp. 338–348, Sep. 2018, doi: 10.1080/00038628.2018.1502151.
[22] M. K. Kim, C. Barber, and J. Srebric, "Traffic noise level predictions for buildings with windows opened for natural ventilation in urban environments," *Sci Technol Built Environ*, vol. 23, no. 5, pp. 726–735, Jul. 2017, doi: 10.1080/23744731.2016.1262708.
[23] A. Kaarlela-Tuomaala, R. Helenius, E. Keskinen, and V. Hongisto, "Effects of acoustic environment on work in private office rooms and open-plan offices - Longitudinal study during relocation," *Ergonomics*, vol. 52, no. 11, pp. 1423–1444, 2009, doi: 10.1080/00140130903154579.
[24] C. Y. Choi and B. McPherson, "Noise levels in Hong Kong Primary Schools: Implications for classroom listening," *Intl J Disabil Dev Educ*, vol. 52, no. 4, pp. 345–360, Dec. 2005, doi: 10.1080/10349120500348714.
[25] P. Cech, "Smart Classroom Study Design for Analysing the Effect of Environmental Conditions on Students' Comfort," *Intelligent Environments*, vol. 21, pp. 14–23, 2016, doi: 10.3233/978-1-61499-690-3-14.
[26] D. Singh, A. Prakash, A. K. Srivastava, K. Kumar, and V. K. Jain, "The effects of meteorological parameters in ambient noise modelling studies in Delhi," *Environ Monit Assess*, vol. 185, no. 2, pp. 1873–1882, Feb. 2013, doi: 10.1007/s10661-012-2674-1.
[27] G. Clausen and D. P. Wyon, "The combined effects of many different indoor environmental factors on acceptability and office work performance," *HVAC and R Research*, vol. 14, no. 1, pp. 103–113, 2008, doi: 10.1080/10789669.2008.10390996.
[28] A. Y. Alqahtani, A. A. Makki, and H. M. Alidrisi, "Revealing factors influencing classroom noise in the universities teaching and learning environment: A design of experiments approach," *Journal of Engineering Research*, vol. 11, no. 1, p. 100009, Mar. 2023, doi: 10.1016/j.jer.2023.100009.
[29] T. B. de Souza, K. C. Alberto, and S. A. Barbosa, "Evaluation of noise pollution related to human perception in a university campus in Brazil," *Applied Acoustics*, vol. 157, Jan. 2020, doi: 10.1016/j.apacoust.2019.107023.
[30] B. Szeremeta and P. H. T. Zannin, "Analysis and evaluation of soundscapes in public parks through interviews and measurement of noise," *Science of the Total Environment*, vol. 407, no. 24, pp. 6143–6149, 2009, doi: 10.1016/j.scitotenv.2009.08.039.
[31] "ANSI-ASA_S12.60-2010_PART_1_with_2011_sponsor_page".
[32] "Acoustics in Educational Settings: Acoustics in Educational Settings: Technical Report ASHA Working Group on Classroom Acoustics." [Online]. Available: http://www.asha.org/members/deskref-journals/deskref/default

[33] "THE NOISE POLLUTION (REGULATION AND CONTROL) RULES, 2000."
[34] R. R. Mcneer, C. L. Bennett, D. B. Horn, and R. Dudaryk, "Factors affecting acoustics and speech intelligibility in the operating room: Size matters," *Anesthesia and Analgesia*, vol. 124, no. 6. Lippincott Williams and Wilkins, pp. 1978–1985, Jun. 01, 2017. doi: 10.1213/ANE.0000000000002118.
[35] S. Lapp, "Noise in Classrooms: Its Impact on Students and How to Cost Effectively Manage It."
[36] Y. Liew, "The effectiveness of using furnishings to improve the housing acoustic of low-cost flats in Malaysia," *Journal of Building Performance*, vol. 11, no. 1, pp. 107–116, 2020.
[37] C. M. Mak and Y. P. Lui, "The effect of sound on office productivity," *Building Services Engineering Research and Technology*, vol. 33, no. 3, pp. 339–345, 2012, doi: 10.1177/0143624411412253.
[38] J. F. Culling, R. Gocheva, Y. Li, and N. Kamaludin, "The effects of ceiling height and absorber placement on speech intelligibility in simulated restaurants," *Acoust Sci Technol*, vol. 41, no. 1, pp. 223–228, 2020, doi: 10.1250/ast.41.223.
[39] J. F. Culling, R. Gocheva, Y. Li, and N. Kamaludin, "The effects of ceiling height and absorber placement on speech intelligibility in simulated restaurants," *Acoust Sci Technol*, vol. 41, no. 1, pp. 223–228, 2020, doi: 10.1250/ast.41.223.
[40] E. Arvidsson, E. Nilsson, D. B. Hagberg, and O. J. I. Karlsson, "The Effect on Room Acoustical Parameters Using a Combination of Absorbers and Diffusers—An Experimental Study in a Classroom," *Acoustics*, vol. 2, no. 3, pp. 505–523, 2020, doi: 10.3390/acoustics2030027.
[41] P. J. Lee, Y. H. Kim, J. Y. Jeon, and K. D. Song, "Effects of apartment building façade and balcony design on the reduction of exterior noise," *Build Environ*, vol. 42, no. 10, pp. 3517–3528, Oct. 2007, doi: 10.1016/j.buildenv.2006.10.044.
[42] E. P. Obot, R. C. Okoro, D. E. Oku, C. N. Nwosu, and M. U. Onuu, "Acoustical Analysis of Insertion Losses of Ceiling Materials," vol. 60, no. September, pp. 15–25, 2023.
[43] M. Kamal, R. A. El-Rahman, and S. Tawfiq, "Evaluation of noise levels affecting Schools in Cairo - Egypt," *Proceedings of ISMA 2010 - International Conference on Noise and Vibration Engineering, including USD 2010*, pp. 1811–1820, 2010.
[44] H. Boucher and T. Moisey, "An Experiential Learning of a Philosophy of Music Education Inspired by the Work of Canadian Composer R. Murray Schafer," *Creat Educ*, vol. 10, no. 10, pp. 2111–2131, 2019, doi: 10.4236/ce.2019.1010153.
[45] Q. Bian, C. Zhang, C. Wang, L. Yin, W. Han, and S. Zhang, "Evaluation of Soundscape Perception in Urban Forests Using Acoustic Indices: A Case Study in Beijing," *Forests*, vol. 14, no. 7, Jul. 2023, doi: 10.3390/f14071435.
[46] K. T. Tsai, M. Der Lin, and Y. H. Chen, "Noise mapping in urban environments: A Taiwan study," *Applied Acoustics*, vol. 70, no. 7, pp. 964–972, Jul. 2009, doi: 10.1016/j.apacoust.2008.11.001.
[47] "sc310_en".
[48] "Protocol for Ambient Level Noise Monitoring," 2015.
[49] International Electrotechnical Commission., International Electrotechnical Commission. Technical committee 29 : Electroacoustics., and International Organization of Legal Metrology., *Electroacoustics : sound level meters. Part 1, Specifications*.
[50] R. Vijay, A. Sharma, T. Chakrabarti, and R. Gupta, "Assessment of honking impact on traffic noise in urban traffic environment of Nagpur, India," *J Environ Health Sci Eng*, vol. 13, no. 1, Feb. 2015, doi: 10.1186/s40201-015-0164-4.

[51] J. Romeu, M. Genescà, T. Pmies, and S. Jiménez, "Street categorization for the estimation of day levels using short-term measurements," *Applied Acoustics*, vol. 72, no. 8, pp. 569–577, Jul. 2011, doi: 10.1016/j.apacoust.2010.09.012.

[52] A. K. Guha and S. Gokhale, "Urban workers' cardiovascular health due to exposure to traffic-originated PM2.5 and noise pollution in different microenvironments," *Science of the Total Environment*, vol. 859, Feb. 2023, doi: 10.1016/j.scitotenv.2022.160268.

[53] M. Long, "Acoustic Measurements and Noise Metrics," in *Architectural Acoustics*, Elsevier, 2014, pp. 129–174. doi: 10.1016/b978-0-12-398258-2.00004-0.

[54] "Assessment and Mapping of Environmental Noise Levels in Chandigarh City: A Case Study," *Indian Journal of Pure & Applied Physics*, 2023, doi: 10.56042/ijpap.v61i10.2588.

[55] V. K. Patil, S. P. Lunawat, B. Student, K. Wagh, and S. N. Dharankar, "Noise Mapping Using Q-GIS: A Case Study of Nasik City," *International Journal for Research in Engineering Application & Management (IJREAM)*, vol. 04, p. 3, 2018, doi: 10.18231/2454-9150.2018.0377.

Kragujevac Journal of Mathematics
Volume 48(4) (2024), Pages 629–646.

APPROXIMATION BY A COMPOSITION OF APOSTOL-GENOCCHI AND PĂLTĂNEA-DURRMEYER OPERATORS

NAV SHAKTI MISHRA[1] AND NAOKANT DEO[1]

Abstract. The present paper deals with the Durrmeyer construction of operators based on a class of orthogonal polynomials called Apostol-Genocchi polynomials. For the proposed operators, we first establish a global approximation result followed by its convergence estimate in terms of usual, r-th and weighted modulus of continuity. We further study the asymptotic type results such as the Voronovskaya theorem and quantitative Voronovskaya theorem. Moreover, we estimate the rate of pointwise convergence of the proposed operators for functions of bounded variation defined on the interval $(0,\infty)$. Finally, the results are validated through graphical representations and an absolute error table.

1. Introduction

Recently, Prakash et al. [20] proposed the following positive linear sequence of operators:

$$G_n^{\alpha,\lambda}(f;x)=\sum_{k=0}^{\infty}s_{n,k}^{\alpha,\lambda}(x)f\left(\frac{k}{n}\right),\quad x\in[0,\infty), \tag{1.1}$$

where $s_{n,k}^{\alpha,\lambda}(x)=e^{-nx}\left(\frac{1+\lambda e}{2}\right)^{\alpha}\frac{g_k^{\alpha}(nx;\lambda)}{k!}$ and $g_k^{\alpha}(x;\lambda)$ is the generalized Apostol-Genocchi polynomials of order α, which belong to the class of orthogonal polynomials. These polynomials were defined for a complex variable z, $|z|<\pi$ in [16]. However, in this study we limit ourselves to a real variable $t\in[0,\infty)$. The generalized Apostol

Key words and phrases. Apostol-Genocchi polynomials, Păltănea basis, generating functions, special functions, functions of bounded variation.

2020 *Mathematics Subject Classification.* Primary: 41A25. Secondary: 41A36, 11B83.

DOI

Received: April 02, 2021.

Accepted: August 06, 2021.

Genocchi polynomial of order α, i.e., $g_k^{\alpha}(x;\lambda)$ can be estimated with the help of following generating function:

$$\left(\frac{2t}{1+\lambda e^t}\right)^{\alpha} e^{xt} = \sum_{k=0}^{\infty} g_k^{\alpha}(x;\lambda)\frac{t^k}{k!}. \tag{1.2}$$

The more explicit form of $g_k^{\alpha}(x;\lambda)$ was proposed by Luo and Srivastava in [17]. They presented some elementary properties of these polynomials and derived explicit series representation of $g_k^{\alpha}(x;\lambda)$ in terms of hypergeometric function defined by Gauss. The series is given as follows:

$$\begin{aligned} g_k^{\alpha}(x;\lambda) =& 2^{\alpha}\alpha!\binom{k}{\alpha}\sum_{n=0}^{k-\alpha}\binom{k-\alpha}{n}\binom{\alpha+n-1}{n}\frac{\lambda^n}{(1+\lambda)^{\alpha+n}} \\ &\times\sum_{j=0}^{n}(-1)^j\binom{n}{j}j^n(x+j)^{k-n-\alpha}{}_2F_1\left(\alpha+n-k,n;n+1;\frac{j}{x+j}\right), \end{aligned}$$

where $\{k,\alpha\}\in\mathbb{N}\cup\{0\}$, $\lambda\in\mathbb{R}\setminus\{-1\}$, $x\in\mathbb{R}$ and ${}_2F_1(a,b;c;t)$ denotes the Gaussian hypergeometric function defined by

$${}_2F_1(a,b;c;t) = {}_2F_1(b,a;c;t) = \sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{(c)_k}\cdot\frac{t^k}{k!}.$$

In particular, for $\alpha=1$ and $\lambda=1$, these operators reduce to classical Genocchi polynomials which are obtained by the following generating function:

$$\frac{2te^{xt}}{e^t+1} = \sum_{k=0}^{\infty} g_k(x)\frac{t^k}{k!}, \quad |t|<2\pi,\ x\in\mathbb{R},$$

where $g_k(x) = g_k(x;1)$. It can be clearly seen that $g_k(x)$ are the k^{th}-degree polynomials, few terms of which are given as follows:

$$g_1(x)=1, \quad g_2(x)=2x-1, \quad g_3(x)=3x(x-1),$$
$$g_4(x)=4x^3-6x^2+1, \quad g_5(x)=5x^4-10x^3+5x,\ldots$$

For the case $x=0$, one can obtain the so-called Genocchi numbers g_k using the relation:

$$g_k(x) = \sum_{i=0}^{k}\binom{k}{i}g_i x^{k-i}.$$

Genocchi numbers can be defined in many ways depending on the field where they are intended to be applied. They find a wide range of application in numerical analysis, combinatorics, number theory, graph theory etc. Luo [15,16] defined Apostol-Genocchi polynomials of higher order and also introduced q-Apostol-Genocchi polynomials. He studied the relationship of these polynomials with Zeta function. In the last two decades, a surprising number of papers appeared studying Genocchi numbers, their combinatorial relations, Genocchi polynomials and their generalisations along with their various expansions and integral representations. To the readers, we suggest following articles [4, 18, 21] and references therein.

In the recent past, much work has been dedicated towards the Durrmeyer type modification of linear positive operators. For instance, Dhamija and Deo [8] introduced the Durrmeyer form of Jain operators based on inverse Pólya-Eggenberger distribution. They studied its moments with the aid of Vandermonde convolution formula and analysed other approximation properties. Heilmann and Raşa [12] studied a link between Baskakov-Durrmeyer type operators and their corresponding classical Kantorovich variants. Acu and Radu [3] introduced and studied a class of operators which link α-Bernstein operators and genuine α-Bernstein Durrmeyer operators. To see more work relevant to this area, one may refer [2, 5, 7, 10, 11, 13].

Inspired by above stated researches, we now consider a Durrmeyer type modification of Apostol-Genocchi operators based on Păltănea basis on positive real line. For $f \in C[0,\infty)$ and $\rho > 0$, the operators are defined as follows:

$$\mathcal{M}_n^{\alpha,\lambda}(f;x) = \sum_{k=0}^{\infty} s_{n,k}^{\alpha,\lambda}(x) \int_0^{\infty} l_{n,k}^{\rho}(t) f(t)\, dt, \quad x \in [0,\infty), \tag{1.3}$$

where $l_{n,k}^{\rho}(t) = n\rho e^{-n\rho t} \frac{(n\rho t)^{k\rho-1}}{\Gamma(k\rho)}$ and $s_{n,k}^{\alpha,\lambda}(x)$ is defined in (1.1).

The outline of the study is as follows. We consider a Durrmeyer type construction of Apostol-Genocchi operators based on the basis function due to Păltănea [19] with real parameters α, λ and ρ. We establish approximation estimates such as a global approximation theorem and rate of approximation in terms of usual, r-th and weighted modulus of continuity. We further study asymptotic formulae such as Voronovskaya theorem and quantitative Voronovskaya theorem. The last theorem is an application of the proposed operators for the functions whose derivatives are of bounded variation. Moreover, the approximation and the absolute error therein has been shown graphically by varying the values of various parameters using Mathematica software.

2. Preliminaries

Before proceeding to our main results, we state some general lemmas which are useful throughout this paper. In addition, we have used Mathematica software wherever necessary for complex and tedious calculations such as for moments and central moments etc.

Lemma 2.1. *For $e_s(t) = t^s$, $s \in \mathbb{N} \cup \{0\}$ and $\rho > 0$, we have*

$$\int_0^{\infty} l_{n,k}^{\rho}(t) t^s = \frac{(k\rho+s-1)!}{(n\rho)^s (k\rho-1)!} = \frac{(k\rho)_s}{(n\rho)^s},$$

where the symbol $(\beta)_n = \beta(\beta+1)(\beta+2)\cdots(\beta+n-1)$, $(\beta)_0 = 1$ denotes the rising factorial.

Lemma 2.2. *For operators (1.3), the moments are obtained as follows:*

$\mathcal{M}_n^{\alpha,\lambda}(e_0;x) = 1,$

$$\mathcal{M}_n^{\alpha,\lambda}(e_1;x) = x + \frac{\alpha}{n(1+\lambda e)},$$

$$\mathcal{M}_n^{\alpha,\lambda}(e_2;x) = x^2 + \frac{x}{n}\left[\frac{1+2\alpha+\lambda e}{(1+\lambda e)} + \frac{1}{\rho}\right] + \frac{1}{n^2}\left[\frac{\alpha^2 - 2\alpha\lambda e - \alpha e^2\lambda^2}{(1+\lambda e)^2} + \frac{\alpha}{\rho(1+\lambda e)}\right],$$

$$\begin{aligned}
&\mathcal{M}_n^{\alpha,\lambda}(e_3;x)\\
=&x^3 + \frac{x^2}{n}\left[\frac{3(\alpha+\lambda e+1)}{(1+\lambda e)} + \frac{3}{\rho}\right]\\
&+ \frac{x}{n^2}\left[\frac{3\alpha^2 - 3\alpha\lambda^2e^2 - 3\alpha\lambda e + 3\alpha + \lambda^2e^2 + 2\lambda e + 1}{(1+\lambda e)^2} + \frac{3(2\alpha+\lambda e+1)}{\rho(1+\lambda e)} + \frac{2}{\rho^2}\right]\\
&+ \frac{1}{n^3}\left[\frac{\alpha^3 - 3\alpha^2\lambda^2e^2 - 6\alpha^2\lambda e - \alpha\lambda^3e^3 - 4\alpha\lambda^2e^2 - 5\alpha\lambda e}{(1+\lambda e)^3} + \frac{2\alpha}{\rho^2(1+\lambda e)}\right.\\
&\left.+\frac{3(\alpha^2 - \alpha\lambda^2e^2 - 2\alpha\lambda e)}{\rho(1+\lambda e)^2}\right],
\end{aligned}$$

$$\begin{aligned}
&\mathcal{M}_n^{\alpha,\lambda}(e_4;x)\\
=&x^4 + \frac{x^3}{n}\left[\frac{2\alpha+3\lambda e+3}{(1+\lambda e)} + \frac{6}{\rho}\right]\\
&+ \frac{x^2}{n^2}\left[\frac{25 + 12\alpha + 6\alpha^2 + 50\mathrm{e}\lambda + 25\mathrm{e}^2\lambda^2 - 6\alpha\mathrm{e}^2\lambda^2}{(1+\lambda e)^2} + \frac{18(1+\alpha+\mathrm{e}\lambda)}{(1+\lambda e)\rho} + \frac{11}{\rho^2}\right]\\
&+ \frac{x}{n^3}\left[\frac{7 + 20\alpha + 63\alpha^2 + 2\alpha^3 - 6\alpha^2e^2\lambda^2 - 9\alpha^2\lambda e}{(1+\lambda e)^3}\right.\\
&+ \frac{-5\alpha e^3\lambda^3 + 4\alpha e^2\lambda^2 + 24\alpha\lambda e + 7e^3\lambda^3 + 21\lambda e}{(1+\lambda e)^3}\\
&\left.+\frac{6(3\alpha^2 - 3\alpha e^2\lambda^2 - 3\alpha\lambda e + 3\alpha + e^2\lambda^2 + 2\lambda e + 1)}{(1+\lambda e)^2\rho} + \frac{11(1+2\alpha+\lambda e)}{(1+\lambda e)\rho^2} + \frac{6}{\rho^3}\right]\\
&\times\frac{1}{n^4}\left[\frac{\alpha^4 - 6e^2\alpha^3\lambda - 12e\alpha^3\lambda - 16e^4\alpha\lambda^4 + 8e^3\alpha\lambda^3 - 82e^3\alpha\lambda^2 - 118e^2\alpha\lambda^2 - 66\alpha + \lambda e}{(1+\lambda e)^4}\right.\\
&+ \frac{6(\alpha^3 - 3e^2\alpha^2\lambda^2 - 6e\alpha^2\lambda - e^3\alpha\lambda^3 - 4e^2\alpha\lambda^2 - 5e\alpha\lambda)}{(1+\lambda e)^3\rho}\\
&\left.+ \frac{11(\alpha^2 - \alpha e^2\lambda^2 - 2\alpha\lambda e)}{(1+\lambda e)^2\rho^2} + \frac{6\alpha}{(1+\lambda e)\rho^3}\right].
\end{aligned}$$

Proof. In the proposed operators (1.3), for $s = 0, 1, 2$ respectively we have 1.

2. For $s = 1$, again using Lemma 2.1 we have

$$\mathcal{M}_n^{\alpha,\lambda}(e_1;x) = \frac{e^{-nx}}{n}\left(\frac{1+\lambda e}{2}\right)^{\alpha}\sum_{k=0}^{\infty}\frac{g_k^{\alpha}(nx;\lambda)}{k!}k. \tag{2.1}$$

Differentiating both sides of (1.2) with respect to t and taking limits $t \to 1$ and $x \to nx$, we have

$$\sum_{k=0}^{\infty} \frac{g_k^{\alpha}(nx;\lambda)}{k!} k = 2^{\alpha} e^{nx} \left(\frac{1}{1+\lambda e}\right)^{1+\alpha} (\alpha + nx(1+\lambda e)).$$

Making use of this value in equation (2.1), we obtain the first moment.

3. Similarly for $s = 2$, we have

$$\mathcal{M}_n^{\alpha,\lambda}(e_2;x) = \frac{e^{-nx}}{n^2} \left(\frac{1+\lambda e}{2}\right)^{\alpha} \sum_{k=0}^{\infty} \frac{g_k^{\alpha}(nx;\lambda)}{k!} k^2. \tag{2.2}$$

On differentiating both sides of (1.2) with respect to t and taking limits $t \to 1$ and $x \to nx$, we have

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{g_k^{\alpha}(nx;\lambda)}{k!} k(k-1) =& 2^{\alpha} e^{nx} \left(\frac{1}{1+\lambda e}\right)^{1+\alpha} \{2ns\alpha(1+\lambda e) \\ &+ n^2x^2(1+\lambda e)^2 - (\lambda e(3+\lambda e) - \alpha + 1)\}.\end{aligned}$$

Combining this with the first order moment and equation (2.2) we obtain the third moment.

We can obtain the higher order moments in a similar way. □

Lemma 2.3. *Let us define* $\delta_n^s(x) = \mathcal{M}_n^{\alpha,\lambda}(\Phi_s;x)$, *where* $\Phi_s(t) = (e_1 - x)^s$ *and* $s = 1, 2$. *Then, from Lemma* 2.2 *we have*

$$\begin{aligned}\delta_n^{(1)}(x) &= \frac{\alpha}{n(1+\lambda e)}, \\ \delta_n^{(2)}(x) &= \frac{x}{n}\left[1 + \frac{1}{\rho}\right] + \frac{1}{n^2}\left[\frac{\alpha^2 - 2\alpha\lambda e - \alpha e^2\lambda^2}{(1+\lambda e)^2} + \frac{\alpha}{\rho(1+\lambda e)}\right].\end{aligned}$$

Furthermore,

$$\begin{aligned}\lim_{n\to\infty} n\mathcal{M}_n^{\alpha,\lambda}(\Phi_1;x) &= \frac{\alpha}{(1+\lambda e)}, \\ \lim_{n\to\infty} n\mathcal{M}_n^{\alpha,\lambda}(\Phi_2;x) &= \left(1 + \frac{1}{\rho}\right)x, \\ \lim_{n\to\infty} n^2\mathcal{M}_n^{\alpha,\lambda}(\Phi_4;x) &= \frac{3(1 + (6+8\alpha)\rho - (3+8\alpha)\rho^2 + \lambda e(1 + 6\rho - 3\rho^2))x^2}{(1+\lambda e)\rho^2}\end{aligned}$$

and

$$\lim_{n\to\infty} n^3\mathcal{M}_n^{\alpha,\lambda}(\Phi_6;x) = \frac{(1+\rho)(2+\rho+3\alpha\rho+\lambda e(2+\rho))x^5}{(1+\lambda e)^3\rho^4}.$$

Remark 2.1. Since fourth and sixth central moments are too lengthy and unnecessarily space consuming, we are omitting their values here. Instead we choose to write their limiting values, which is useful in the proofs of our main theorems.

3. Main Theorems

Theorem 3.1. *For any $f \in C_B[0,\infty)$, where $C_B[0,\infty)$ is the class of all continuous and bounded functions, we have*

$$\mathcal{M}_n^{\alpha,\lambda}(f(t);x) = f(x)$$

uniformly on any compact subset of $[0,\infty)$.

Proof. Taking into account Lemma 2.2, we can easily see that $\mathcal{M}_n^{\alpha,\lambda}(e_r;x) \to x^r$ for each $r = 0,1,2$ and hence using the well known Korovkin's theorem due to [14], operators $\mathcal{M}_n^{\alpha,\lambda}$ converge uniformly on each compact subset of $[0,\infty)$. □

3.1. **Global approximation.** Let us denote $B_f[0,\infty)$ the space of all functions f on positive real axis that satisfy the condition $|f(x)| \le H_f(1+x^2)$ where H_f is a constant depending only on f but independent of x.

Let $C_f[0,\infty)$ be the subspace of $B_f[0,\infty)$ containing all continuous f on $[0,\infty)$. The norm in $C_f[0,\infty)$ is defined by

$$||f||_2 = \sup_{x\in[0,\infty)} \frac{|f(x)|}{1+x^2}.$$

Also, let $C_f^l[0,\infty) := \left\{ f \in C_f[0,\infty) : \lim_{x\to\infty} \frac{|f(x)|}{1+x^2} \text{ is finite} \right\}$.

Theorem 3.2. *For each $f \in C_f^l[0,\infty)$, we have*

$$\lim_{n\to\infty} \left\| \mathcal{M}_n^{\alpha,\lambda}(f;x) - f \right\|_2 = 0.$$

Proof. The proof of this theorem can be given by application of Korovkin theorem [9] on the interval $[0,\infty)$. Therefore, it would suffice if we prove that

$$\lim_{n\to\infty} \left\| \mathcal{M}_n^{\alpha,\lambda}(e_l;x) - e_l \right\|_2 = 0, \quad l = 0,1,2. \tag{3.1}$$

For $l = 0$, condition (3.1) holds as operators $\mathcal{M}_n^{\alpha,\lambda}$ preserve constant functions. Next, we can write

$$\left\| \mathcal{M}_n^{\alpha,\lambda}(e_1;x) - x \right\|_2 \le \sup_{x\in[0,\infty)} \frac{\alpha}{n(1+\lambda e)(1+x^2)} \to 0,$$

for adequately large n. Therefore, the condition (3.1) is satisfied for $l = 1$.

Finally, we write

$$\left\| \mathcal{M}_n^{\alpha,\lambda}(e_2;x) - x^2 \right\|_2 \le \sup_{x\in[0,\infty)} \frac{1}{(1+x^2)} \left\{ \frac{x}{n}\left(1+\frac{1}{\rho}\right) + \frac{1}{n^2}\left(\frac{\alpha^2 + 2\alpha\lambda e - \alpha e^2\lambda^2}{(1+\lambda e)^2} + \frac{\alpha}{\rho(1+\lambda e)} \right) \right\},$$

which suggests

$$\lim_{n\to\infty} \left\| \mathcal{M}_n^{\alpha,\lambda}(e_2;x) - e_2 \right\|_2 = 0.$$

Hence, the theorem follows. □

Let $C_B[0,\infty)$ be the class of all continuous and bounded real valued functions. We define the r-th order modulus of continuity by $\omega_r(f,\delta)$ and define it as

$$\omega_r(f,\delta) = \sup_{x\in[0,\infty)} \sup_{0\le h\le\delta} |\Delta_h^r f(x)|,$$

where Δ denotes the forward difference. In particular, the usual modulus of continuity is defined for $r=1$ and is denoted by $\omega(f,\delta)$. Moreover, we define the norm as $\|f\| = \sup_{x\in[0,\infty)} |f(x)|$.

Also, the Peetre's K-functional for the function $g \in C_B^2[0,\infty)$ is defined as:

$$K_2(f;\delta) = \inf_{g\in C_B^2[0,\infty)} \left\{ \|f-g\| + \delta \|g''\| : g \in C_B^2[0,\infty) \right\},$$

where

$$C_B^2[0,\infty) = \{g \in C_B[0,\infty) : g', g'' \in C_B[0,\infty)\}.$$

The next theorem establishes the degree of approximation of the operators $\mathcal{M}_n^{\alpha,\lambda}$ in terms of the usual and second order modulus of continuity for the functions in the space $C_B[0,\infty)$.

Theorem 3.3. *For $\hbar \in C_B^2[0,\infty)$, define the auxiliary operators $\widetilde{\mathcal{Q}}_n^{\alpha,\lambda}$ as*

$$\widetilde{\mathcal{Q}}_n^{\alpha,\lambda}(\hbar;x) = \mathcal{M}_n^{\alpha,\lambda}(\hbar;x) - \hbar\left(x + \frac{\alpha}{n(1+\lambda e)}\right) + \hbar(x). \tag{3.2}$$

Then there exists a constant $C>0$ such that

$$|\mathcal{M}_n^{\alpha,\lambda}(\hbar,x) - \hbar(x)| \le C\omega_2(\hbar,\sqrt{\delta}) + \omega(\hbar,\delta_n^{(1)}(x)),$$

where

$$\delta = \delta_n^{(2)}(x) + \left(\frac{\alpha}{n(1+\lambda e)}\right)^2.$$

Proof. Using Lemma 2.2, one can easily observe that $\widetilde{\mathcal{Q}}_n^{\alpha,\lambda}((t-x);x) = 0$.

Let $f \in C_B^2[0,\infty)$, then by Taylor's expansion we have

$$f(t) = f(x) + (t-x)f'(x) + \int_x^t (t-u)f''(u)du.$$

Moreover, we can write

$$\begin{aligned}
\left|\widetilde{\mathcal{Q}}_n^{\alpha,\lambda}(f;x) - f(x)\right| &= \left|\widetilde{\mathcal{Q}}_n^{\alpha,\lambda}\left(\int_x^t (t-u)f''(u)du, x\right)\right| \\
&\le \left|\mathcal{M}_n^{\alpha,\lambda}\left(\int_x^t (t-u)f''(u)du, x\right)\right| \\
&\quad + \left|\int_x^{\left(x+\frac{\alpha}{n(1+\lambda e)}\right)} \left(\left(x+\frac{\alpha}{n(1+\lambda e)}\right) - u\right) f''(u)du\right|
\end{aligned}$$

$$\leq \left(\mathcal{M}_n^{\alpha,\lambda} \left((t-x)^2; x \right) + \left(\frac{\alpha}{n\left(1+\lambda e\right)} \right)^2 \right) \|f''\| . \tag{3.3}$$

Since we know that

$$\left| \mathcal{M}_n^{\alpha,\lambda} \left(\hbar; x \right) \right| \leq \|\hbar\| ,$$

therefore

$$\left| \widetilde{\mathcal{Q}}_n^{\alpha,\lambda}(\hbar; x) \right| \leq \left| \mathcal{M}_n^{\alpha,\lambda} \left(\hbar; x \right) \right| + \left| \hbar \left(x + \frac{\alpha}{n\left(1+\lambda e\right)} \right) \right| + |\hbar(x)| \leq 3 \|\hbar\| . \tag{3.4}$$

Finally, combining equations (3.2), (3.3) and (3.4), we get

$$\begin{aligned}
\left| \mathcal{M}_n^{\alpha,\lambda} \left(\hbar; x \right) - \hbar(x) \right| \leq & \left| \widetilde{\mathcal{Q}}_n^{\alpha,\lambda}(\hbar - f); x) - (\hbar - f)(x) \right| + \left| \widetilde{\mathcal{Q}}_n^{\alpha,\lambda}(f; x) - f(x) \right| \\
& + \left| \hbar(x) - \hbar \left(x + \frac{\alpha}{n(1+\lambda e)} \right) \right| \\
\leq & 4 \|\hbar - f\| + \left(\mathcal{M}_n^{\alpha,\lambda} \left((t-x)^2; x \right) + \left(\frac{\alpha}{n\left(1+\lambda e\right)} \right)^2 \right) \|f''\| \\
& + \left| \hbar(x) - \hbar \left(x + \frac{\alpha}{n(1+\lambda e)} \right) \right| \\
\leq & C \left\{ \|\hbar - f\| + \left(\delta_n^{(2)}(x) + \left(\frac{\alpha}{n\left(1+\lambda e\right)} \right)^2 \right) \|f''\| \right\} \\
& + \omega \left(\hbar, \frac{\alpha}{n\left(1+\lambda e\right)} \right) .
\end{aligned}$$

Taking infimum over all $f \in C_B^2[0,\infty)$ and using the result $K_2(f,\delta) \leq \omega_2\left(f, \sqrt{\delta}\right)$ due to [6], we get the desired outcome. □

Theorem 3.4. *Let $\hbar \in C_B(0,\infty)$, then for any $r > 0$, $x \in [0,r]$ and adequately large n, we have*

$$\left| \mathcal{M}_n^{\alpha,\lambda} \left(\hbar; x \right) - \hbar(x) \right| \leqslant 4 H_\hbar \left(1 + x^2 \right) \frac{D}{n} + 2\omega_{r+1} \left(\hbar, \sqrt{\frac{D}{n}} \right) ,$$

where D is a positive constant.

Proof. If $x \in [0,r]$ and $t > r+1$, then $t - x > 1$. Therefore, we have the following inequality:

$$|\hbar(t) - \hbar(x)| \leqslant 4 H_\hbar \left(1 + x^2 \right) (t-x)^2 .$$

Again for $x \in [0,r]$ and $t \in [0, r+1]$ and using the well known inequality $\omega(f, \beta\delta) \leqslant (\beta+1)\,\omega(f,\delta)\,, \beta \in (0,\infty)$, one can obtain

$$|\hbar(t) - \hbar(x)| \leqslant \left(1 + \frac{|t-x|}{\delta} \right) \omega_{r+1}(\hbar, \delta) .$$

From (3.5) and (3.5), we can write

$$|\hbar(t)-\hbar(x)| \leqslant 4H_{\hbar}\left(1+x^2\right)(t-x)^2+\left(1+\frac{|t-x|}{\delta}\right)\omega_{r+1}(\hbar,\delta).$$

Applying operator $\mathcal{M}_n^{\alpha,\lambda}$ in the above relation and making use of Cauchy-Schwarz inequality, we get

$$\begin{aligned}\left|\mathcal{M}_n^{\alpha,\lambda}(\hbar;x)-\hbar(x)\right| \leqslant & 4H_{\hbar}\left(1+x^2\right)\mathcal{M}_n^{\alpha,\lambda}\left((t-x)^2;x\right)\\ &+\left(1+\frac{1}{\delta}\mathcal{M}_n^{\alpha,\lambda}(|t-x|;x)\right)\omega_{r+1}(\hbar,\delta)\\ \leqslant & 4H_{\hbar}\left(1+x^2\right)\mathcal{M}_n^{\alpha,\lambda}\left((t-x)^2;x\right)\\ &+2\omega_{r+1}\left(\hbar,\sqrt{\mathcal{M}_n^{\alpha,\lambda}\left((t-x)^2;x\right)}\right).\end{aligned}$$

Since $\mathcal{M}_n^{\alpha,\lambda}\left((t-x)^2;x\right) \leqslant \frac{D}{n}$, where D is a positive constant, it follows that for adequately large n, we have

$$\left|\mathcal{M}_n^{\alpha,\lambda}(\hbar;x)-\hbar(x)\right| \leqslant 4H_{\hbar}\left(1+x^2\right)\frac{D}{n}+2\omega_{r+1}\left(\hbar,\sqrt{\frac{D}{n}}\right),$$

which is the required result. □

3.2. **Quantitative Voronovskaya theorem and Voronovoskaya theorem.** Let $C_B[0,\infty)$ be the subspace of $B_f[0,\infty)$ containing all continuous and bounded functions f for which $\lim\limits_{x\to\infty}|f(x)|(1+x^2)^{-1}$ is finite.

The weighted modulus of continuity $\Omega(f,\delta)$ due to [1] for each $f \in C_B[0,\infty)$ is defined as

$$\Omega(f,\delta)=\sup_{x\in[0,\infty),|h|<\delta}\frac{|f(x+h)-f(x)|}{(1+h^2+x^2+h^2x^2)}.$$

In the next theorem, we discuss the quantitative Voronovskaya theorem for the proposed operators (1.3) and derive a Voronovoskaya asymptotic result as a resulting corollary, making use of the following properties of weighted modulus of continuity. For every $f \in C_f^l[0,\infty)$,

(a) $\Omega(f,\delta)\to 0$ for $\delta\to 0$;
(b) $|f(t)-f(x)| \leq \left(1+(t-x)^2\right)(1+x^2)\,\Omega(f,|t-x|)$.

Theorem 3.5. *Let $\hbar'' \in C_2^l[0,\infty)$ and $x\in[0,\infty)$. Then we have*

$$\begin{aligned}&\left|\mathcal{M}_n^{\alpha,\lambda}(\hbar;x)-\hbar(x)-\frac{\alpha}{n(1+\lambda e)}\hbar'(x)\right.\\ &\left.-\frac{1}{2}\left(\frac{x}{n}\left(1+\frac{1}{\rho}\right)+\frac{1}{n^2}\left(\frac{\alpha^2+2\alpha\lambda e-\alpha e^2\lambda^2}{(1+\lambda e)^2}+\frac{\alpha}{\rho(1+\lambda e)}\right)\right)\hbar''(x)\right|\end{aligned}$$

$$\leq \frac{8\left(1+x^2\right)}{n}\Omega\left(\hbar'', \frac{1}{\sqrt{n}}\right).$$

Proof. By Taylor's expansion, we may write

$$\begin{aligned}\mathcal{M}_n^{\alpha,\lambda}(\hbar;x)-\hbar(x) &= \mathcal{M}_n^{\alpha,\lambda}\left(\left(\hbar(t)-\hbar(x)\right);x\right)\\ &= \mathcal{M}_n^{\alpha,\lambda}\left(\left(t-x\right)\hbar'\left(x\right)+\frac{\left(t-x\right)^2}{2}\hbar''\left(x\right)+\lambda\left(t,x\right)\left(t-x\right)^2;x\right),\end{aligned}$$

where $\lambda\left(t,x\right)=\left(\hbar''\left(\varsigma\right)-\hbar''\left(x\right)\right)/2$ is a continuous function which tends to zero at 0 and ς lies between x and t. Using Lemma 2.3, we get

$$\begin{aligned}&\left|\mathcal{M}_n^{\alpha,\lambda}(\hbar;x)-\hbar(x)-\frac{\alpha}{n\left(1+\lambda e\right)}\hbar'\left(x\right)\right.\\ &\left.-\frac{1}{2}\left(\frac{x}{n}\left(1+\frac{1}{\rho}\right)+\frac{1}{n^2}\left(\frac{\alpha^2+2\alpha\lambda e-\alpha e^2\lambda^2}{\left(1+\lambda e\right)^2}+\frac{\alpha}{\rho\left(1+\lambda e\right)}\right)\right)\hbar''\left(x\right)\right|\\ \leq&\mathcal{M}_n^{\alpha,\lambda}\left(\left|\lambda\left(t,x\right)\right|\left(t-x\right)^2;x\right).\end{aligned}$$

With simple manipulations in property (b) of weighted modulus of continuity and using $|\varsigma-x|\leq|t-x|$, we can write

$$\left|\lambda\left(t,x\right)\right|\leq 8\left(1+x^2\right)\left(1+\frac{\left(t-x\right)^4}{\delta^4}\right)\Omega\left(\hbar'',\delta\right),$$

which implies that

$$\left|\lambda\left(t,x\right)\right|\left(t-x\right)^2\leq 8\left(1+x^2\right)\left(\left(t-x\right)^2+\frac{\left(t-x\right)^6}{\delta^4}\right)\Omega\left(\hbar'',\delta\right).$$

Therefore, in view of Lemma 2.3, we can write

$$\mathcal{M}_n^{\alpha,\lambda}\left(\left|\lambda\left(t,x\right)\right|\left(t-x\right)^2;x\right)\leq 8\left(1+x^2\right)\Omega\left(\hbar'',\delta\right)\left\{\delta_n^{(2)}\left(x\right)+\frac{1}{\delta^4}\delta_n^{(6)}\left(x\right)\right\},$$

as $n\to\infty$. Choosing $\delta=\frac{1}{\sqrt{n}}$, we get the desired outcome. □

Corollary 3.1. *Let f be a bounded and integrable function on the interval $[0,\infty)$ such that the second derivative of f exists at a fixed point $x\in[0,\infty)$. Then*

$$\lim_{n\to\infty} n(\mathcal{M}_n^{\alpha,\lambda}(f;x)-f(x))=\frac{\alpha}{(1+\lambda e)}f'(x)+x\left(1+\frac{1}{\rho}\right)f''(x).$$

3.3. **Functions of derivatives of bounded variation.** Next we estimate the rate of convergence of the operators (1.3) for functions with derivatives of bounded variation defined on $[0,\infty)$.

Let $f\in \mathrm{DBV}_\tau[0,\infty)$ be the class of functions whose derivatives are of bounded variation on any finite subinterval of $[0,\infty)$ and satisfy the growth condition $|f(t)|\leq$

Kt^{τ}, $\tau > 0$ for all $t > 0$ and constant $K > 0$. For such functions, let us represent our proposed operators (1.3) in the following form

$$\mathcal{M}_n^{\alpha,\lambda}(f;x) = \int_0^{\infty} \mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t) f(t) dt, \tag{3.5}$$

where $\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t) = \sum_{k=0}^{\infty} s_{n,k}^{\alpha,\lambda}(x) l_{n,k}^{\rho}(t)$.

Lemma 3.1. *For $x \in [0,\infty)$ and adequately large n, we have*

(i) *if $0 \le y < x$, then*

$$\vartheta_n(x,y) = \int_0^y \mathfrak{q}_n^{\alpha,\rho}(x;t)\, dt \le \frac{K\delta_n^2(x)}{n(x-y)^2};$$

(ii) *if $x < z \le \infty$, then*

$$1 - \vartheta_n(x,z) = \int_z^{\infty} \mathfrak{q}_n^{\alpha,\rho}(x;t)\, dt \le \frac{K\delta_n^2(x)}{n(z-x)^2}.$$

Proof. (i) Taking into account Lemma 2.2 and proposed operators (1.3), we have

$$\begin{aligned}
\vartheta_n(x,z) &= \int_0^y \mathfrak{q}_n^{\alpha,\rho}(x;t)\, dt \\
&\le \int_0^y \mathfrak{q}_n^{\alpha,\rho}(x;t) \left(\frac{x-t}{x-y}\right)^2 dt = \frac{1}{(x-y)^2} \int_0^y (t-x)^2 \mathfrak{q}_n^{\alpha,\rho}(x;t)\, dt \\
&\le \frac{1}{(x-y)^2} \mathcal{M}_n^{\alpha,\lambda}\left((t-x)^2;x\right) \le \frac{K\delta_n^2(x)}{n(x-y)^2}.
\end{aligned}$$

Proof of (ii) is similar to (i). □

Theorem 3.6. *Consider a function f of bounded variation on every sub-interval of $[0,\infty)$ that satisfies the growth condition $|f(t)| \le Kt^{\tau}$ for some absolute constant K and $\tau > 0$. If there exists an integer γ, $(2\gamma \ge \tau)$ such that $f(t) \le O(t^{\gamma})$ for every $t > 0$, then for $\gamma > 0$, $x \in [0,\infty)$ and sufficiently large n, we have*

$$\begin{aligned}
\left|\mathcal{M}_n^{\alpha,\lambda}(f;x) - f(x)\right| \le & \frac{1}{2}\left(f'(x+) + f'(x-)\right)\delta_n^1(x) \\
& + \sqrt{\frac{K\delta_n^2(x)}{4n}}\left|f'(x+) - f'(x-)\right| + \frac{x}{\sqrt{n}} \bigvee_{x-\frac{x}{\sqrt{n}}}^{x+\frac{x}{\sqrt{n}}} (f'_x) \\
& + \frac{K\delta_n^2(x)}{nx} \sum_{k=1}^{[\sqrt{n}]} \bigvee_{x-\frac{x}{k}}^{x+\frac{x}{\sqrt{k}}} (f'_x) + \frac{K\delta_n^2(x)}{nx^2}\left|f(2x) - f(x) - xf'(x)\right| \\
& + \wp(\gamma,\tau,x) + \frac{K\delta_n^2(x)}{nx^2}\left|f(x)\right| + \sqrt{\frac{K\delta_n^2(x)}{n}} f'(x+),
\end{aligned}$$

where $\bigvee_{b}^{a}(f)$ *denotes the total variation of* f *on any finite subinterval* $[a,b]$ *of* $[0,\infty)$ *and* $\wp(\gamma,\tau,x) := 2^{\gamma}\text{Ç}\left(\int_0^{\infty}(t-x)^{2\tau}\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\right)^{\frac{\gamma}{2\tau}}$.

Proof. For $x\in[0,\infty)$, we can write for our proposed operators (3.5) that

$$\begin{aligned}\mathcal{M}_n^{\alpha,\lambda}(f;x)-f(x) &= \int_0^{\infty}\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,(f(t)-f(x))\,dt\\ &= \int_0^{\infty}\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\left(\int_x^t f'(x)\,du\right)dt. \qquad (3.6)\end{aligned}$$

Also for any $f\in DBV_{\gamma}[0,\infty)$, equality (3.7) holds true, i.e.,

$$\begin{aligned}f'(u) =& \frac{1}{2}\left(f'(x+)+f'(x-)\right)+f'_x(u)+\frac{1}{2}\left(f'(x+)-f'(x-)\right)\operatorname{sgn}(u-x)\\ &+\delta_x(u)\left(f'(u)-\frac{1}{2}\left(f'(x+)+f'(x-)\right)\right), \qquad (3.7)\end{aligned}$$

where

$$\delta_x(u)=\begin{cases}1, & u=x,\\ 0, & u\neq x.\end{cases}$$

It can be easily verified that:

$$\int_0^{\infty}\left(\int_x^t\left(f'(u)-\frac{1}{2}\left(f'(x+)+f'(x-)\right)\right)\delta_x(u)du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt=0.$$

Now in view of our proposed operators (3.5), we may write

$$\begin{aligned}&\int_0^{\infty}\left(\int_x^t\left(\frac{1}{2}\left(f'(x+)+f'(x-)\right)\right)du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\\ =&\frac{1}{2}\left(f'(x+)+f'(x-)\right)\mathcal{M}_n^{\alpha,\lambda}((t-x);x).\end{aligned}$$

Moreover

$$\begin{aligned}&\int_0^{\infty}\left(\int_x^t\left(\frac{1}{2}\left(f'(x+)-f'(x-)\right)\operatorname{sgn}(u-x)\right)du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\\ \leq&\frac{1}{2}\left|f'(x+)-f'(x-)\right|\int_0^{\infty}|t-x|\,\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\\ \leq&\frac{1}{2}\left|f'(x+)-f'(x-)\right|\left(\mathcal{M}_n^{\alpha,\lambda}\left((t-x)^2;x\right)\right)^{1/2}. \qquad (3.8)\end{aligned}$$

Making use of equations (3.7)–(3.8) and Lemma 2.2 in equation (3.6), we get

$$\mathcal{M}_n^{\alpha,\lambda}(f;x)-f(x)\leq\frac{1}{2}\left(f'(x+)+f'(x-)\right)\mathcal{M}_n^{\alpha,\lambda}((t-x);x)$$

$$
\begin{aligned}
&+\frac{1}{2}\left|f'(x+)-f'(x-)\right|\left(\mathcal{M}_n^{\alpha,\lambda}\left((t-x)^2;x\right)\right)^{1/2}\\
&+\int_0^\infty\left(\int_x^t f'_x(x)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\\
\leq&\frac{1}{2}\left(f'(x+)+f'(x-)\right)\mathcal{M}_n^{\alpha,\lambda}\left((t-x);x\right)\\
&+\sqrt{\frac{K\delta_n^2(x)}{4n}}\left|f'(x+)-f'(x-)\right|\\
&+\int_0^\infty\left(\int_x^t f'_x(x)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt.
\end{aligned}
$$

Taking absolute values on both sides and rewriting equation we have:

$$
\begin{aligned}
\left|\mathcal{M}_n^{\alpha,\lambda}(f;x)-f(x)\right|\leq&\frac{1}{2}\left(f'(x+)+f'(x-)\right)\mathcal{M}_n^{\alpha,\lambda}\left((t-x);x\right)\\
&+\sqrt{\frac{K\delta_n^2(x)}{4n}}\left|f'(x+)-f'(x-)\right|+P_{n_1}(x)+P_{n_2}(x),
\end{aligned}
\tag{3.9}
$$

where

$$
P_{n_1}(x)=\left|\int_0^x\left(\int_x^t f'_x(x)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\right|
$$

and

$$
P_{n_2}(x)=\left|\int_x^\infty\left(\int_x^t f'_x(x)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\right|.
$$

Integrating by parts after applying Lemma 3.1, and taking $y=x-\frac{x}{\sqrt{n}}$, we obtain

$$
P_{n_1}(x)\leq\int_0^{x-\frac{x}{\sqrt{n}}}\vartheta_n(x;t)\left|f'_x(t)\right|dt+\int_{x-\frac{x}{\sqrt{n}}}^x\vartheta_n(x;t)\left|f'_x(t)\right|dt.
$$

Since $f'_x(x)=0$ and $\vartheta_n(x;t)\leq 1$, it implies

$$
\begin{aligned}
\int_{x-\frac{x}{\sqrt{n}}}^x\vartheta_n(x;t)\left|f'_x(t)\right|dt&=\int_{x-\frac{x}{\sqrt{n}}}^x\vartheta_n(x;t)\left|f'_x(t)-f'_x(x)\right|dt\\
&\leq\int_{x-\frac{x}{\sqrt{n}}}^x\bigvee_t^x(f'_x)\,dt\leq\frac{x}{\sqrt{n}}\bigvee_{x-\frac{x}{\sqrt{n}}}^x(f'_x).
\end{aligned}
$$

Again using Lemma 3.1 and substituting $y=x-\frac{x}{u}$ we obtain

$$
\int_0^{x-\frac{x}{\sqrt{n}}}\vartheta_n(x;t)\left|f'_x(t)\right|dt\leq\frac{K\delta_n^2(x)}{n}\int_0^{x-\frac{x}{\sqrt{n}}}\frac{\left|f'_x(t)\right|}{(x-t)^2}dt\leq\frac{K\delta_n^2(x)}{nx}\int_1^{\sqrt{n}}\bigvee_{x-\frac{x}{u}}^x(f'_x)\,du
$$

$$\leq \frac{K\delta_n^2(x)}{nx}\sum_{k=1}^{\left[\sqrt{n}\right]}\bigvee_{x-\frac{x}{k}}^{x}(f'_x).$$

Thus, we can write $P_{n_1}(x)$ as

$$P_{n_1}(x) \leq \frac{x}{\sqrt{n}}\bigvee_{x-\frac{x}{\sqrt{n}}}^{x}(f'_x) + \frac{K\delta_n^2(x)}{nx}\sum_{k=1}^{\left[\sqrt{n}\right]}\bigvee_{x-\frac{x}{k}}^{x}(f'_x). \tag{3.10}$$

Next, to estimate $P_{n_2}(x)$, we have

$$\begin{aligned} P_{n_2}(x) &\leq \left|\int_x^{2x}\left(\int_x^t f'_x(u)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\right| + \left|\int_{2x}^{\infty}\left(\int_x^t f'_x(u)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\right| \\ &\leq A_n(x) + B_n(x), \end{aligned}$$

where

$$A_n(x) = \left|\int_x^{2x}\left(\int_x^t f'_x(u)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\right|$$

and

$$B_n(x) = \left|\int_{2x}^{\infty}\left(\int_x^t f'_x(u)\,du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt\right|.$$

Since $1-\vartheta_n(x,t)\leq 1$, by putting $t = x+\frac{x}{u}$ successively, we have

$$\begin{aligned} A_n(x) =& \left|\int_x^{2x} f'_x(u)\,(1-\vartheta_n(x,2x))\,du - \int_x^{2x} f'_x(t)\,(1-\vartheta_n(x,t)\,dt\right| \\ \leq& \frac{K\delta_n^2(x)}{nx^2}\left|f(2x)-f(x)-xf'(x)\right| + \int_x^{x+\frac{x}{\sqrt{n}}}\left|f'_x(t)\right|\left|1-\vartheta_n(x,t)\right|dt \\ &+ \int_{x+\frac{x}{\sqrt{n}}}^{2x}\left|f'_x(t)\right|\left|1-\vartheta_n(x,t)\right|dt \\ \leq& \frac{K\delta_n^2(x)}{nx^2}\left|f(2x)-f(x)-xf'(x)\right| + \frac{K\delta_n^2(x)}{n}\int_{x+\frac{x}{\sqrt{n}}}^{2x}\frac{V_x^t(f'_x)}{(t-x)^2}dt + \int_x^{x+\frac{x}{\sqrt{n}}}\bigvee_x^t(f'_x)dt \\ \leq& \frac{K\delta_n^2(x)}{nx^2}\left|f(2x)-f(x)-xf'(x)\right| + \frac{K\delta_n^2(x)}{n}\sum_{k=1}^{\left[\sqrt{n}\right]}\bigvee_x^{x+\frac{x}{\sqrt{k}}}(f'_x) + \frac{x}{\sqrt{n}}\bigvee_x^{x+\frac{x}{\sqrt{n}}}(f'_x). \end{aligned}$$

Further we estimate the value of $B_n(x)$ as follows:

$$B_n(x) = \left|\int_{2x}^{\infty}\left(\int_x^t f'_x(u)du\right)\mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)dt\right| \tag{3.11}$$

$$\leq \text{Ç} \int_{2x}^{\infty} t^{\gamma} \mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt + |f(x)| \int_{2x}^{\infty} \mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt + \sqrt{\frac{K\delta_n^2(x)}{n}} f'(x+). \tag{3.12}$$

It is obvious that $t \leq 2(t-x)$ and $x \leq t-x$, when $t \geq 2x$. Now applying Hölder's inequality in the first term of equation (3.11), we get

$$B_n(x) = 2^{\gamma} \text{Ç} \left(\int_0^{\infty} (t-x)^{2\tau} \mathfrak{q}_{n,\rho}^{\alpha,\lambda}(x;t)\,dt \right)^{\frac{\gamma}{2\tau}} + \frac{K\delta_n^2(x)}{nx^2} |f(x)| + \sqrt{\frac{K\delta_n^2(x)}{n}} f'(x+)$$

$$= \wp(\gamma,\tau,x) + \frac{K\delta_n^2(x)}{nx^2} |f(x)| + \sqrt{\frac{K\delta_n^2(x)}{n}} f'(x+). \tag{3.13}$$

Finally, combining equations (3.10)–(3.13) and putting values of $P_{n_1}(x)$ and $P_{n_2}(x)$ in (3.9), we get the required result and the theorem is proved. □

Example 3.1. Let $f(x) = x^4 - 3x^3 + 2x^2 + 1$. We choose parameters $\alpha = \lambda = 2$ and $\rho = 3$. For $n = 10, 50, 100, 200$, we have the following representations.

(a) Figure 1 shows the rate of approximation of the operators $\mathcal{M}_n^{\alpha,\lambda}$ towards the function f. Clearly the proposed operators (1.3) converge to the function f for sufficiently large n.

(b) In Figure 2, the associated absolute error $\Theta_n = |\mathcal{M}_n^{\alpha,\lambda}(f;x) - f(x)|$ is represented graphically for arbitrary values of x in interval $[0,\infty)$. It can be observed that error is monotonically decreasing for increasing n.

(c) An error estimation table is provided in Table 1 which depicts that for higher value of n, the error approaches to zero.

Therefore, it can be concluded that proposed operators (1.3) provide good approximation for n adequately large.

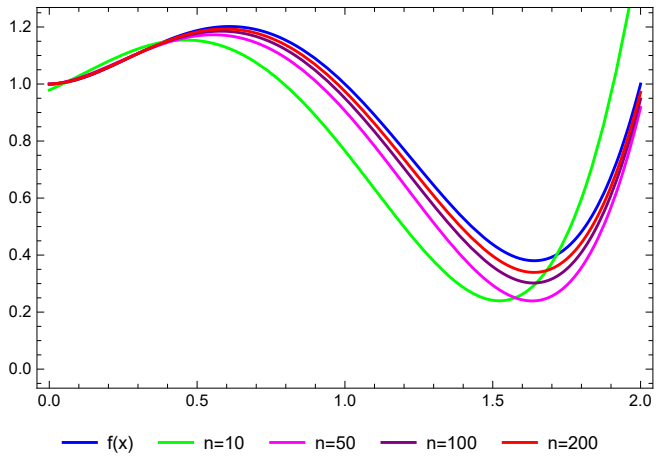


FIGURE 1. Convergence of $\mathcal{M}_n^{\alpha,\lambda}(f;x)$ for the polynomial function $f(x) = x^4 - 3x^3 + 2x^2 + 1$ with parameters $\alpha = \lambda = 2$, $\rho = 3$.

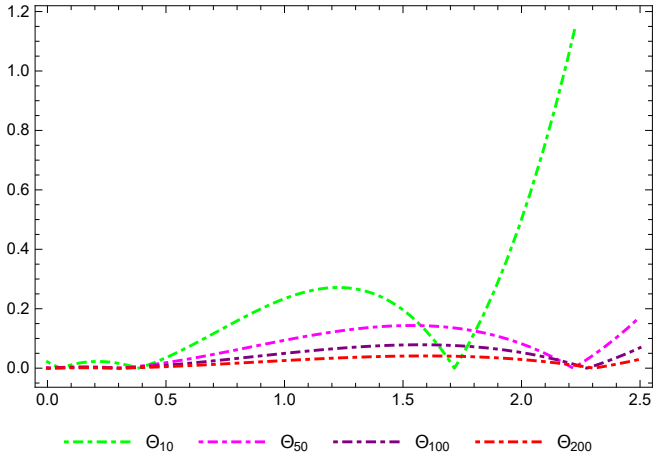


FIGURE 2. Absolute error $\Theta_n = \mathcal{M}_n^{\alpha,\lambda}(f;x) - f(x)|$ of the proposed operators for $f(x) = x^4 - 3x^3 + 2x^2 + 1$ with parameters $\alpha = \lambda = 2$, $\rho = 3$.

TABLE 1. Table for Absolute error $\Theta_n = |\mathcal{M}_n^{\alpha,\lambda}(f;x) - f(x)|$ of the proposed operators $\mathcal{M}_n^{\alpha,\lambda}$.

x	Θ_{10}	Θ_{50}	Θ_{100}	Θ_{200}
0.4	0.0100759	0.00856603	0.0045899	0.00236582
0.8	0.166213	0.0652568	0.0345534	0.0177532
1.2	0.232451	0.120425	0.0647845	0.0335316
1.6	0.0454756	0.123219	0.0698567	0.0369876
2.0	0.921829	0.022785	0.0243437	0.0154083
2.4	2.65087	0.231729	0.0971809	0.0439198
2.8	5.48687	0.691177	0.320143	0.15371
3.2	9.68409	1.40641	0.66997	0.326674
3.6	15.4968	2.42828	1.17209	0.575527
4.0	23.1792	3.80765	1.85192	0.912982

Example 3.2. Figure 3 illustrates the effect of increase in values of parameter ρ on the rate of convergence of proposed operators $\mathcal{M}_n^{\alpha,\lambda}$ for the function $f(x) = 4x(x - 1.1)(x - 1.9)$ while keeping the value of α, λ and n fixed. Here we chose $n = 10$ and $\alpha = \lambda = 2$ to show the impact of the parameter ρ clearly. One can easily deduce from the figure that as we increase the value of ρ the rate of convergence gets relatively faster.

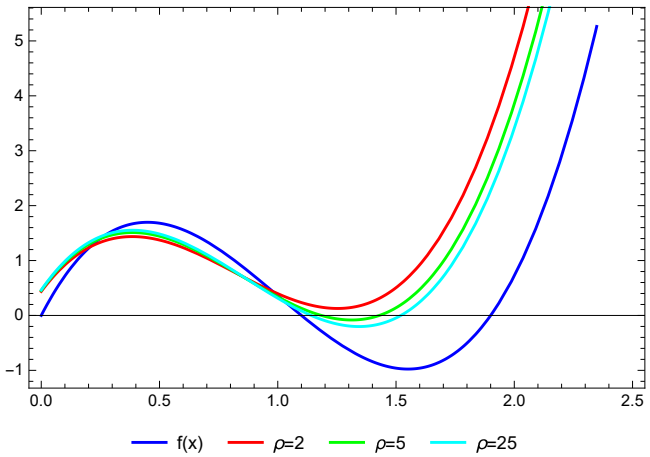


FIGURE 3. Effect of increase in parametric value of ρ for given $n = 10$, $\alpha = \lambda = 2$ on the convergence rate of proposed operators.

Acknowledgements. The authors are grateful to the referee for several valuable comments and suggestions leading to an overall improvement of this paper.

REFERENCES

[1] T. Acar, A. Aral and I. Raşa, *The new forms of Voronovskayas theorem in weighted spaces*, Positivity **20**(1) (2016), 25–40. https://doi.org/10.1007/s11117-015-0338-4

[2] A. M. Acu, T. Acar and V. A. Radu, *Approximation by modified U_n^ρ operators*, Rev. R. Acad. Cienc. Exactas Fís. Nat. (Esp.). Serie A. Matemáticas **113**(3) (2019), 2715–2729. https://doi.org/10.1007/s13398-019-00655-y

[3] A. M. Acu and V. A. Radu, *Approximation by certain operators linking the α-Bernstein and the Genuine α-Bernstein-Durrmeyer operators*, in: N.Deo et al. (Eds.), *Mathematical Analysis I: Approximation Theory*, Springer, Proceedings in Mathematics & Statistics **306**, 2020, 77–88.

[4] I. N. Cangul, H. Ozden and Y. Simsek, *A new approach to q-Genocchi numbers and their interpolation functions*, Nonlinear Analysis: Theory, Methods & Applications **71**(12) (2009), e793–e799. https://doi.org/10.1016/j.na.2008.11.040

[5] N. Deo and S. Kumar, *Durrmeyer variant of Apostol-Genocchi-Baskakov operators*, Quaest. Math. **44**(4) (2020), 1–18. https://doi.org/10.2989/16073606.2020.1834000

[6] R. A. DeVore and G. G. Lorentz, *Constructive Approximation*, Springer, Berlin, 1993.

[7] M. Dhamija, *Durrmeyer modification of Lupas type Baskakov operators based on IPED*, in: N. Deo et al. (Eds.), *Mathematical Analysis I: Approximation Theory*, Springer, Proceedings in Mathematics & Statistics **306**, 2020, 111–120.

[8] M. Dhamija and N. Deo, *Jain-Durrmeyer operators associated with the inverse Pólya–Eggenberger distribution*, Appl. Math. Comput. **286** (2016), 15–22. https://doi.org/10.1016/j.amc.2016.03.015

[9] A. D. Gadjiev, *Theorems of the type of PP Korovkins theorems*, Mat. Zametki. **20**(5) (1976), 781–786.

[10] T. Garg, A. M. Acu and P. N. Agrawal, *Further results concerning some general Durrmeyer type operators*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM **113**(3) (2019), 2373–2390. https://doi.org/10.1007/s13398-019-00628-1

[11] V. Gupta, *A note on the general family of operators preserving linear functions*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM **113**(4) (2019), 3717–3725. https://doi.org/10.1007/s13398-019-00727-z

[12] M. Heilmann and I. Raşa, *A nice representation for a link between Baskakov and Szász-Mirakjan-Durrmeyer operators and their Kantorovich variants*, Results Math. **74**(1) (2019), 1–12. https://doi.org/10.1007/s00025-018-0932-4

[13] A. Kajla, N. Ispir, P. N. Agrawal and M. Goyal, *q-Bernstein-Schurer-Durrmeyer type operators for functions of one and two variables*, Appl. Math. Comput. **275** (2016), 372–385. https://doi.org/10.1016/j.amc.2015.11.048

[14] P. P. Korovkin, *Linear Operators and Approximation Theory*, Hindustan Publication Co., Delhi 1960.

[15] Q. M. Luo, *q-Extensions for the Apostol-Genocchi polynomials*, General Mathematics **17**(2) (2009), 113–125.

[16] Q. M. Luo, *Extensions of the Genocchi polynomials and their Fourier expansions and integral representations*, Osaka J. Math. **48** (2011), 291–309.

[17] Q. M. Luo and H. M. Srivastava, *Some generalizations of the Apostol-Genocchi polynomials and the Stirling numbers of the second kind*, Appl. Math. Comput. **217**(12) (2011), 5702–5728. https://doi.org/10.1016/j.amc.2010.12.048

[18] M. A. Özarslan, *Unified Apostol-Bernoulli, Euler and Genocchi polynomials*, Comput. Math. Appl. **62**(6) (2011), 2452–2462. https://doi.org/10.1016/j.camwa.2011.07.031

[19] R. Păltănea, *Modified Szász-Mirakjan operators of integral form*, Carpathian J. Math. **24**(3) (2008), 378–385.

[20] C. Prakash, D. K. Verma and N. Deo, *Approximation by a new sequence of operators involving Apostol-Genocchi polynomials*, Math. Slovaca (to appear).

[21] H. M. Srivastava, B. Kurt and V. Kurt, *Identities and relations involving the modified degenerate hermite-based Apostol-Bernoulli and Apostol-Euler polynomials*, Rev. R. Acad. Cienc. Exactas Fís. Nat. (Esp.). Serie A. Matemáticas **113**(2) (2019), 1299–1313. https://doi.org/10.1007/s13398-018-0549-1

[1]Delhi Technological University
Department of Applied Mathematics
Bawana Road, Delhi-110042, India
Email address: navshaktimishra_phd2k18@dtu.ac.in, navshakti20@gmail.com
Email address: naokantdeo@dce.ac.in

Check for updates

Received: 12 June 2024 | Revised: 25 July 2024 | Accepted: 26 July 2024

DOI: 10.1002/pen.26915

RESEARCH ARTICLE

POLYMER ENGINEERING AND SCIENCE WILEY

Assessing rheological properties of oxidized *Moringa oleifera* gum and carboxymethyl chitosan-based self-healing hydrogel for additive manufacturing applications

Mehul Verma[1] | **Jitendra Kumar**[1] | **Abhinav Ayush Pradhan**[1] | **Nilotpal Majumder**[2] | **Sourabh Ghosh**[2] | **Roli Purwar**[1]

[1]Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University, Delhi, India

[2]Department of Textile and Fibre Engineering, Indian Institute of Technology, New Delhi, Delhi, India

Correspondence
Roli Purwar, Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University, Delhi, India.
Email: roli.purwar@dtu.ac.in

Funding information
Delhi Technological University

Abstract

Rheology plays a vital role in pneumatic three-dimensional (3D) printing of hydrogels. This study investigates the rheological behavior of a novel self-healing hydrogel (O-MOG/CMCh) formed by a Schiff base crosslinking reaction between oxidized *Moringa oleifera* gum (O-MOG), a biodegradable antimicrobial polysaccharide, and carboxymethyl chitosan (CMCh), a water-soluble biocompatible chitosan derivative. Three hydrogel formulations were designed using 5% w/v of CMCh with varied concentrations of O-MOG (3% w/v, 4% w/v, and 5% w/v) and evaluated through rheology analyses, including frequency sweeps, amplitude sweeps, oscillatory thixotropy, and gelation kinetics. These tests revealed that the material has shear thinning, self-healing properties, a high linear viscoelastic region (LVE), and gel formation times (t_{gel}) of 3.23–4.57 min. The hydrogel synthesized with 5% w/v of O-MOG composition exhibited the best characteristics for printability based on rheological assessments, and this composition was used for further printing assessment, where bi-layered 4 × 4 and 2 × 2 grids were successfully printed using 22 G (0.41 mm) and 23 G (0.34 mm) syringes. All the constructs had a printability index value of 1 ± 0.13 and spreading ratios <6.5, demonstrating the feasibility of employing the synthesized hydrogel as an acellular matrix via additive manufacturing.

Highlights

- Self-healing hydrogel was prepared by mixing the precursors through a cannula.
- Rheology was examined using standard tests for printability assessment.
- 3D printability was achieved using two different gauze syringes.
- Printability parameters were recorded and analyzed for the constructs.

KEYWORDS

3D printability, additive manufacturing, carboxymethyl chitosan, *Moringa oleifera* gum, rheology, Schiff base

 © 2024 Society of Plastics Engineers.

1 | INTRODUCTION

3D printing forms three-dimensional objects by successive adding layers of material as per the programmed model.[1] Polymeric hydrogels are generally printed using the extrusion 3D printing method. Investigating the rheological properties of hydrogel used in 3D printing is essential for understanding their deformation and flow characteristics during printing.[2] This analysis is critical for achieving desired object properties and ensuring successful printing outcomes.[3] Rheology offers valuable insights into how materials are made, whether in liquid or solid form, aiding manufacturers and researchers in optimizing printing processes, anticipating material responses, and maintaining the quality and effectiveness of 3D-printed items. This information allows us to predict whether the material can be successfully extruded through a print nozzle, how well it will hold its shape after deposition, the potential for print defects due to factors like sagging or over-extrusion, and how printing parameters might need to be adjusted to optimize results.[4] Different rheological tests include shear-thinning behavior,[5] amplitude sweep test,[6] step-strain analysis,[7] frequency sweep test,[2] yield stress analysis,[8] and time sweep experiments are described, and the results of those tests are reported before the 3D printing of polymeric materials.

Polysaccharides, natural polymers that are abundant, versatile, biocompatible, and relatively inexpensive, gained a new choice of material for 3D printing applications, especially in drug delivery, tissue engineering, and other fields.[9,10] Common polysaccharides used in 3D printing include chitosan, alginate, gelatin, and cellulose, which possess favorable biomimetic properties.[11,12] Furthermore, *Moringa oleifera* gum, a naturally occurring polysaccharide, is extracted from the Moringa tree. It has several medicinal properties, including antipyretic properties that reduce fever; antioxidant properties that guard against oxidative stress; anti-asthmatic properties that provide relief from asthmatic conditions; astringent properties that constrict and bind tissue; and rubefacient properties that, when applied, cause the skin to become red to promote blood circulation.[13] Chitosan, a naturally occurring polycationic linear polysaccharide, originates from chitin, the primary structural component found in the exoskeletons of insects and crustaceans (such as shrimps and crabs), as well as in the cell walls of fungi. It is the second most plentiful natural polysaccharide after cellulose.[14] Carboxymethyl chitosan, produced by the deacetylation of chitosan, possesses hydrophilic characteristics and has been used in 3D bioprinting applications. It has non-toxic, biocompatible, and biodegradable properties, which make it useful in biomedical utilization.[15–17]

The 3D construct is formed using different mixture ratios of alginates and carboxymethyl chitosan in a calcium chloride crosslinker solution.[18] Bider et al. synthesized a hydrogel using alginate dialdehyde and gelatin in the presence of calcium chloride and microbial transglutaminase, having 3D printability used for bone tissue engineering.[19] Another group of researchers also synthesized a 3D printable hydrogel using alginate dialdehyde and gelatin scaffolds incorporating polydopamine with silicon dioxide and copper oxide nanoparticles for bone regeneration.[20] Ettoumi et al. fabricated a nanocomposite polysaccharides hydrogel using fucoidan and chitosan with better 3D printing ability to release polyphenols.[21] Kim et al. designed a 3D printable hydrogel using oxidized and hydrazide-modified hyaluronic acids for potential use as bio-inks.[22] A biocompatible and 3D printable hydrogel was synthesized using double-quaternized chitosan with cystamine-based non-isocyanate polyurethane, tailored for biomedical applications.[23] A 3D printable hydrogel derived from carboxymethyl chitosan and oxidized hyaluronic acid via Schiff base reaction utilized for biomedical applications.[24] A 3D printable bio-ink was derived from chitosan and gallic acid for tissue engineering applications.[25] Wang et al. designed a 3D printable hydrogel in the form of the heart valve using arginine-modified polylysine and oxidized hyaluronic acid through dynamic acylhydrazone bonds.[26]

After evaluating the literature based on 3D printing, the novelty of this research includes (1) Unique biomaterial combination material combination. This specific combination has likely not been extensively studied before, especially in additive manufacturing. (2) Focus on self-healing properties: a crucial feature for advanced materials in 3D printing and other additive manufacturing processes. (3) Application in additive manufacturing: By assessing the rheological properties of this hydrogel for additive manufacturing applications, the study bridges the gap between novel biomaterials and advanced manufacturing techniques. (4) Rheological property assessment: The comprehensive analysis of rheological properties provides valuable insights into the material's behavior during the additive manufacturing process, which is critical for optimizing printing parameters and final product quality.

This research article is an enhancement in terms of application of our previously published study, which involves the synthesis of an injectable hydrogel using multialdehyde *Moringa oleifera* gum and carboxymethyl chitosan via Schiff base crosslinking mechanism used in drug delivery.[27] Based on that, the optimized concentration of both the solutions (oxidized *Moringa oleifera* gum and carboxymethyl chitosan) with some addition was selected for the current study. It is hypothesized that a

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

hydrogel derived from the combination of oxidized *Moringa oleifera* gum and carboxymethyl chitosan through Schiff base crosslinks may possess adequate rheological properties to be utilized as an ink to fabricate 3D constructs by extrusion-based printing.

This work describes a rheological analysis and extrusion-based 3D bioprinting application of a Schiff-based hydrogel synthesized by combining derivatized chitosan and oxidized *Moringa oleifera* gum. Furthermore, this research may open new avenues for developing advanced, sustainable, and self-healing materials in additive manufacturing, with applications ranging from biomedical devices to industrial components.

2 | MATERIALS AND METHODS

2.1 | Materials

Moringa oleifera gum was obtained from Nutramine Life Science, New Delhi, India (Product code 232154). Chitosan (deacetylation degree of 85%) was bought from Otto Chemical Pvt. Ltd. (C 1906). Sodium hydroxide (030167), sodium periodate (645645) ($NaIO_4$), and hydroxylamine hydrochloride (565325) were obtained from CDH chemicals. Ethylene glycol (71883), isopropyl alcohol (67-63-0), and monochloroacetic acid (79-11-8) were purchased from Thomas Baker Chemicals Pvt. Ltd. Ethanol (57101) was procured from SRL Chemicals, India. The hydrogels used in this study were synthesized in the laboratory. The oxidation % of O-MOG and the degree of substitution of CMCh were determined to be 87.8% and 67%, respectively. Furthermore, a known amount of O-MOG (3%, 4%, and 5% w/v) was dissolved in distilled water at 600 RPM for 24 h (solution 1). Similarly, CMCh (5% w/v) was dissolved in phosphate buffer saline at pH 7.4 at 800 RPM (solution 2). Then, both solutions were filled into separate syringes (10 mL) and extruded simultaneously into a test tube (20 mL) through a two-way cannula at 37°C[27] (Table 1).

TABLE 1 Concentration of precursors used for hydrogel synthesis.

Concentration of O-MOG	Concentration of CMCh	Hydrogel
3% (w/v)	5% (w/v)	O-MOG/CMCh-3
4% (w/v)	5% (w/v)	O-MOG/CMCh-4
5% (w/v)	5% (w/v)	O-MOG/CMCh-5

2.2 | Methods

2.2.1 | Rheological analysis

Rheological measurements were performed using the Anton Paar MCR 302 rheometer throughout, using 40 mm parallel-plate geometry. The temperature was set at 37°C for all tests. The sample's viscosity is measured across a range of shear rates, from low to high, to evaluate the viscosity variation. The shear rate sweep was carried out with the shear rate ranging from 0.1 to 100 s^{-1}. The amplitude sweep test was performed to check the linear viscoelastic region at a constant angular frequency of 10 rad s^{-1}, with the strain amplitude ranging from 0.01 to 100%. A frequency sweep test was conducted under a constant strain amplitude of 0.1% to evaluate the synthesized hydrogel properties over frequencies ranging from 0.1 to 100 rad s^{-1}. Loss factor (tan δ) was determined at a constant strain of 0.1% frequency ranges from 0.1 to 100 rad s^{-1}. Oscillatory step strain analysis was conducted under a low (0.01%) and high strain (100%) for three cycles over 10 min. The time sweep test was conducted at 0.1% strain and 1 Hz frequency over 60 min. Yield stress and flow point were also measured under a constant angular frequency of 10 rad s^{-1}, with the strain amplitude ranging from 0.01% to 100%.

2.2.2 | Printing

The printability tests were done using a direct write assembly (Fiber Align, Aerotech Inc., USA). Two bi-layered 16 mm × 16 mm designs were made in Robo-CAD software with 4 and 8 mm inter-filament spacing, respectively. The printing speed was kept at 1 mm/s, and the extrusion pressure was 5 ± 1.5 psi for all samples. The sizes of nozzles used were 22 G (0.41 mm), and 23 G (0.34 mm). Printability characteristics were evaluated by calculating the printability index, spreading ratio, and line width. All images were analyzed using the ImageJ software for all measurements.

The perimeter and area of the pores were calculated to determine and assess the printability of the O-MOG/CMCh hydrogels using the printability index (Pr). The Pr value evaluates how well an ink retains its shape after deposition by measuring the degree to which square pores in a grid tend to become rounded. It measures the printing resolution. The printability index was calculated using Equation (1):

$$\mathrm{Pr} = \frac{\pi}{4}.\frac{1}{C} = \frac{L^2}{16A} \quad (1)$$

where C is the circularity of the enclosed pore, L stands for the perimeter, and "A" stands for the pore area.

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Spreading ratio (SR) is another parameter used to evaluate the inks' ability to print precise structures while reducing coalescence and pore closure. SR is calculated as the ratio of the width of the printed filament to the diameter of the nozzle (Equation (2)):

$$\mathrm{SR} = \frac{\text{width of extruded strand}}{\text{diameter of the nozzle}} = \frac{t_{\text{filament}}}{d_{\text{nozzle}}} \quad (2)$$

The line widths of the printed filaments were measured by taking multiple measurements of the width of the printed lines in each construct. The measurements' averages were taken and plotted on a bar graph. Line width accuracy ensures consistent material deposition for desired part design and functionality. Ideally, the line width should be the same as the inner diameter of the nozzle.

3 | RESULTS AND DISCUSSION

This study assesses the feasibility of 3D printing for fabricated hydrogels composed of O-MOG and CMCh via the Schiff base reaction mechanism. Understanding the printability of synthesized hydrogels necessitates an examination of their rheological properties. Therefore, rheological investigations were conducted to elucidate the flow characteristics and viscoelastic attributes of all three hydrogel samples in which the concentration of O-MOG varies from 3% w/v to 5% w/v and CMCh concentration remains constant (5% w/v).

Figure 1 illustrates the flow behavior of O-MOG/CMCh hydrogels across a spectrum of shear rates ranging from 0.1 to 100 s^{-1}. All samples exhibit shear-thinning behavior, a decreasing viscosity (O-MOG/CMCh-1 = 1,293,800 mPa.s to 500 mPa.s; O-MOG/CMCh-4 = 359,020 mPa.s to 380 mPa.s; O-MOG/CMCh-5 = 2,208,800 mPa.s to 766 mPa.s) was observed with increasing shear rate. This significant reduction in viscosity is beneficial for 3D printing, enabling smooth hydrogel flow through the fine printing nozzles. Throughout the test, hydrogels with the highest amount of oxidized MOG (O-MOG/CMCh-5) exhibited the highest viscosity. In comparison, the hydrogel formulated with the lowest concentration of oxidized MOG (O-MOG/CMCh-3) demonstrated the lowest viscosity among all the samples. With a higher O-MOG concentration, more aldehyde groups become available for the Schiff base reaction. This increases Schiff base linkages, increasing the hydrogel's viscosity. These results show that the material has shear-thinning properties and can be extruded from the nozzle. Shear-thinning properties enable efficient material extrusion through fine nozzles, a crucial factor for pneumatic and extrusion-based printing techniques. The hydrogels formed by a single Schiff base linkage in Ding et al.'s work also show a drop of about three orders of magnitude over the same shear rate range[28] (15,000 Pa.s to ~0 Pa.s). Hu et al.'s work on a thermoresponsive hydrogel also shows similar outcomes and is suited for printing[29] (100,000 Pa.s to ~0).

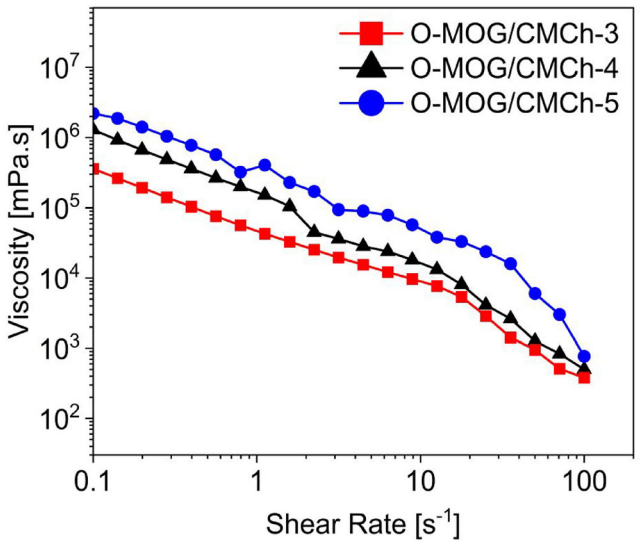


FIGURE 1 Shear rate sweep of O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5.

Dynamic strain sweeps were performed to identify the LVE region, where storage modulus (G') and loss modulus (G'') remain stable. The LVE region represents the range of applied strain where the hydrogel exhibits linear viscoelastic behavior. Within this region, G' and G'' remain constant, with G'' being lower than G' throughout. This indicates minimal energy loss and an ideal elastic response.[30] This analysis provides insights into the extent of deformation a material can endure before its internal structure deteriorates. The test identifies the LVE region where the material can be extruded without compromising its structural integrity. Amplitude sweeps conducted from 0.01% to 100% strain revealed that the hydrogels possess a large LVE region, indicating their ability to withstand a wide range of strains while maintaining linearity, as depicted in Figure 2. As the hydrogel O-MOG/CMCh-5 forms the highest number of Schiff linkages, it exhibits more solid-like behavior and consequently has the highest G' value in the LVE region. Earlier studies on dynamic glycol chitosan hydrogels formed through Schiff base linkages have demonstrated comparable results in amplitude sweep experiments.[31] The decrease in storage modulus and increase in loss modulus specifies the disruption of the inner structure of hydrogel. Therefore, the LVE region strain value (0.1%) was selected for the dynamic frequency sweep test. Yu et al. fabricated a hydrogel derived from carboxymethyl chitosan and oxidized hyaluronic acid crosslinked through the Schiff base reaction; amplitude study shows that hydrogel contains a

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

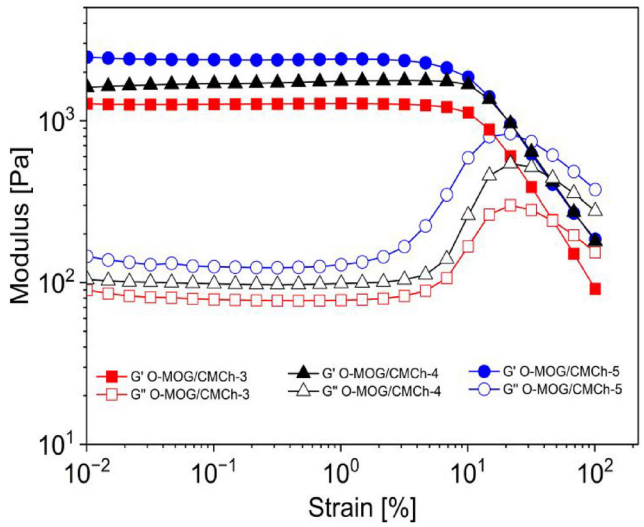


FIGURE 2 Amplitude sweep test of O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5.

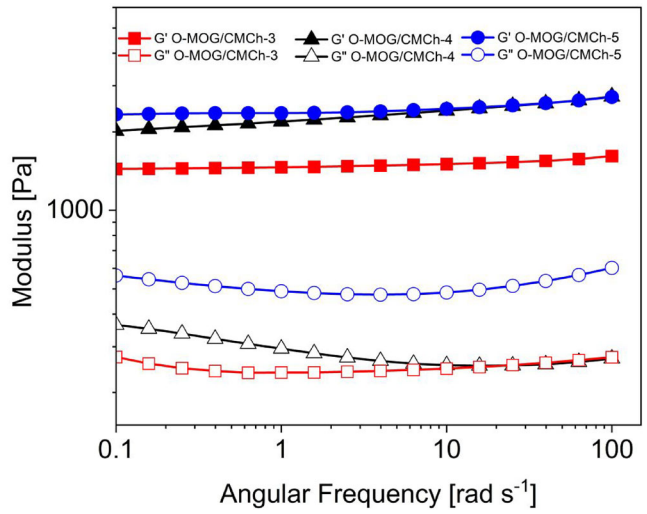


FIGURE 3 Frequency sweep test of O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5.

minimum concentration of oxidized polysaccharide shows favorable mechanical strength.[24]

By varying the applied frequency, a frequency sweep shows how easily a gel deforms under different printing speeds, impacting 3D printability. Figure 3 shows the dynamic frequency sweep analysis, and it is observed that for all concentrations, the G' consistently exceeded the G'' across the frequency range of 0.1 to 100 rad s^{-1}. Again, this indicates a predominantly elastic behavior rather than a fluid-like state in all gels. The hydrogel with the highest concentration of oxidized MOG (O-MOG/CMCh-5) also has the highest G' due to a more significant number of Schiff base linkages. Moreover, G' consistently exhibited a value five times greater than G'' for all

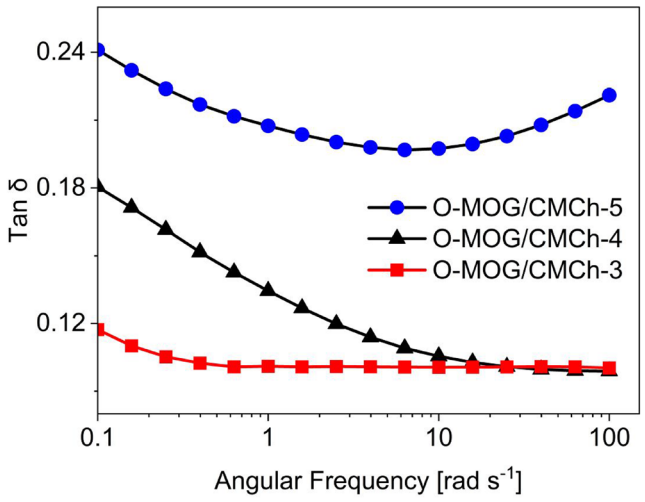


FIGURE 4 Tan δ versus angular frequency of O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5.

hydrogel concentrations, emphasizing their predominantly elastic nature. The G' and G'' values showed minimal variation across the frequency range of 0.1–100 rad s^{-1}, indicating consistent viscoelastic behavior at different frequencies. Wang et al. observed analogous findings in their study on 3D printing of a self-healing polysaccharide hydrogel, where the storage modulus exhibited frequency-independent behavior.[30,32]

In the LVE region, at 0.1% strain, the behavior of a material is characterized by its tan delta (tan δ) values. This ratio, calculated as the G''divided by the G', reflects the balance between energy dissipation (viscosity) and energy storage (elasticity). When tan δ exceeds 1, the material behaves predominantly as a viscous liquid. Conversely, tan δ below 1 values signify that the material exhibits predominantly elastic, solid-like behavior.[2] A loss factor of less than one indicates that the material is more elastic than viscous, which is desirable for maintaining structural integrity during printing and preventing excessive flow or spreading after deposition. Figure 4 shows that tan δ consistently remains below 1 across the tested frequency range for all hydrogels, indicating that elastic effects rather than viscous effects mainly govern stresses. Cheng et al. highlighted that the value tan δ ranging from 0.2 to 0.7 is ideal for printing material.[33] Our investigation revealed that among the studied samples, the O-MOG 5% gel was the sole composition exhibiting a loss factor surpassing 0.2, rendering it the most suitable candidate for printing applications.

The material needs to recover its viscosity almost instantly after extrusion. This rapid 'rebuilding' allows the extruded filament to hold its shape, preventing collapse or unwanted spreading before it solidifies. Thus, the oscillatory step strain test shows the self-healing

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

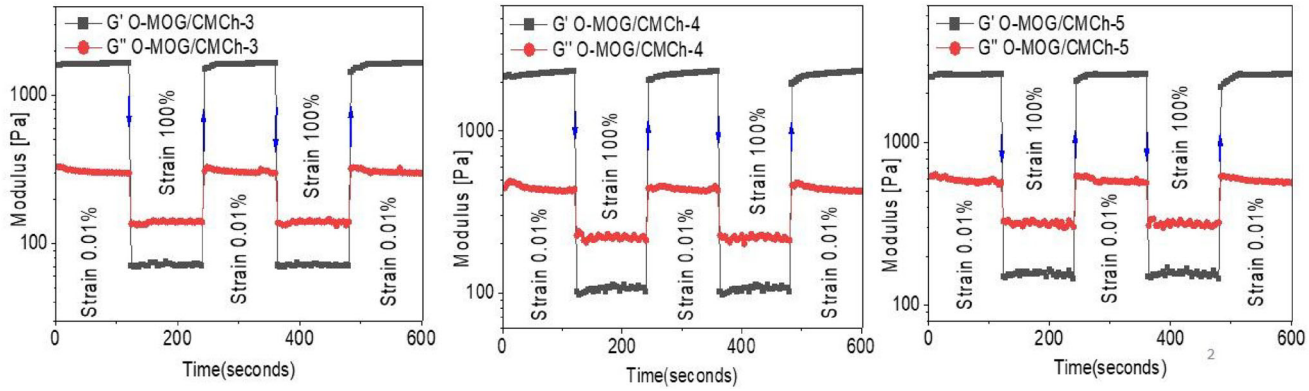


FIGURE 5 Oscillatory thixotropy test of O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5.

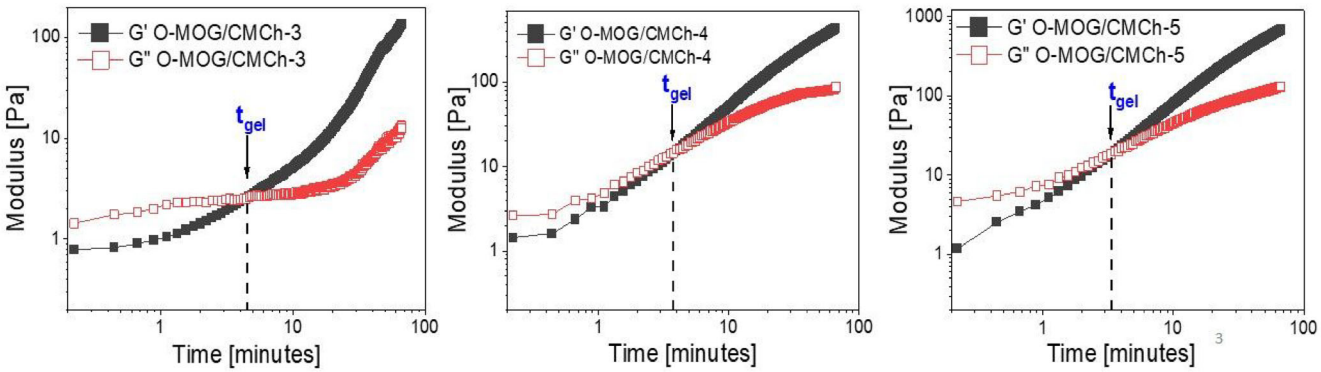


FIGURE 6 Time sweep test O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5.

property of the material. Evaluation of thixotropic properties includes testing at high and low strain amplitudes. The material's ability to rapidly crossover, demonstrating reversible liquid–solid transitions, is crucial for printed filament shape retention. Figure 5 shows the oscillatory strain experiment, in which all the hydrogels were subjected to low (0.01%) and high (100%) strain values, giving them enough time to undergo gel-to-sol and sol-to-gel transitions.

The self-recovery process was repeated for three cycles, and it was observed that under a low strain of 0.01%, all O-MOG/CMCh hydrogel quickly regained its mechanical strength. G' of all three hydrogels exhibited almost complete recovery within 20 s, highlighting the rapid and full instantaneous recovery of viscoelastic properties and the self-healing capability. Maity et al. found similar results with a chitosan-based hydrogel for strains up to 100% and showed self-healing properties.[34] Hu et al. also performed a similar study of 100% strain for a biomaterial ink and found complete recovery necessary for 3D printability.[29] This ensures that extruded filaments experience minimal to no deformation or distortion upon deposition on the printing platform, demonstrating the material's suitability for maintaining structural integrity throughout the printing process.[35]

Figure 6 shows the time sweep assessment for all hydrogels. The gelation time (t_{gel}) was determined by the crossover point, where the G' exceeded the G'', marking the onset of gelation. This transition indicated the initiation of crosslinking facilitated by forming Schiff base linkages. The rate of increase in G' outpaced that of G'', signaling a pronounced shift toward a more solid-like state in the viscoelastic behavior of the system.[36] Across the investigated formulations, gelation time ranged from 3.23 ± 0.05 to 4.59 ± 0.05 min, with the O-MOG/CMCh-3 hydrogel having the highest gelation time and O-MOG/CMCh-5 possessing the least. This is in line with theory as O-MOG/CMCh-5 gel has the highest number of Schiff linkages out of all variants due to a more significant number of aldehyde groups present so that it can undergo gelation at an increased rate. Puertas-Bartolomé et al. found similar gelation times in their study on 3D printing of a reactive hydrogel system formed by Schiff base reaction of carboxymethyl chitosan with oxidized hyaluronic acid.[37]

This test is an extension of the amplitude sweep test, and it measures the G' and G'' against shear stress to identify the yield and flow points. The yield point, where G' drops significantly, indicates the transition from solid-like to liquid-like behavior and flow. The crossover point, where

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

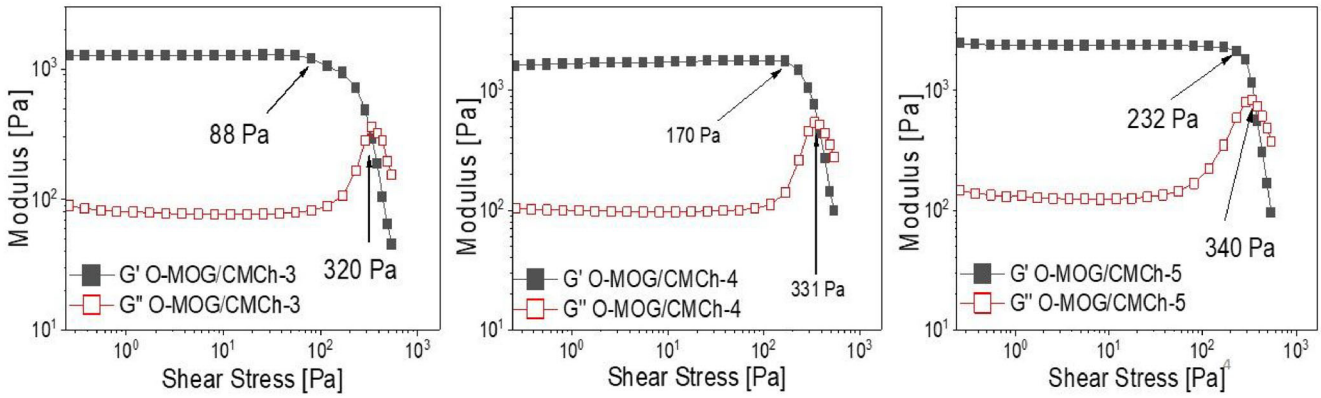


FIGURE 7 Yield stress and flow point O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5.

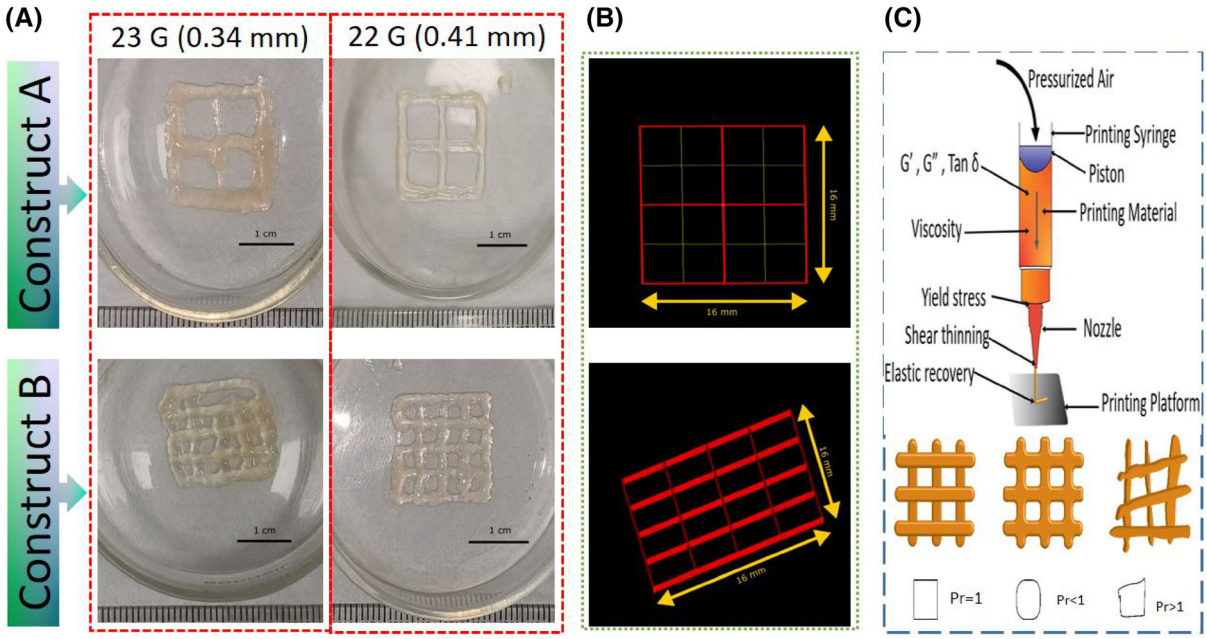


FIGURE 8 (A) Constructs A and B were built using nozzles of 0.41 mm (22 G) and 0.34 mm (23 G) inner diameters, (B) designs of constructs on RoboCAD with inter-filament distances of 8 mm and 16 mm showing two layers stacked on each other vertically, (C) influence of rheological properties on the pneumatic-based printing process and application of the printability formula for different pore shapes.

G'' surpasses G', represents the further transition into liquid-like behavior, approximating the flow point. A well-defined yield point is crucial for shape retention after extrusion, as values that are too low or too high can lead to structural defects or excessive extrusion forces, respectively. The amplitude sweep test obtained all hydrogels' yield stress and flow point values. Figure 7 shows that increasing the O-MOG concentration (3% w/v to 5% w/v) also increased yield stress values from 88 to 232 Pa. This suggests that the hydrogels become more rigid as the O-MOG concentration increases. The flow point values of hydrogels O-MOG/CMCh-3, O-MOG/CMCh-4, and O-MOG/CMCh-5 are 320, 331, and 340 Pa, respectively, suggesting an increasing resistance to flow or deformation. Haider et al. conducted a similar test on a thermosensitive hydrogel, and their reported findings align closely with the present study.[38]

3.1 | 3D printing of hydrogels

After analysis of all the rheological parameters, it was demonstrated that the hydrogel O-MOG/CMCh-5 found the best rheological properties in most tests. Therefore, the hydrogel O-MOG/CMCh-5 was chosen for printability

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

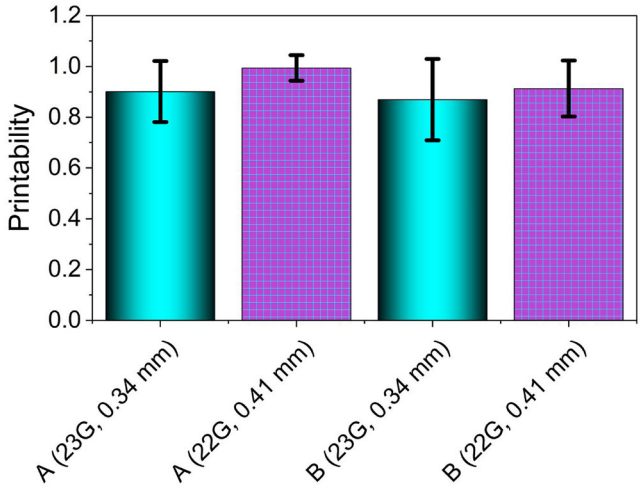


FIGURE 9 Printability indices of all constructs.

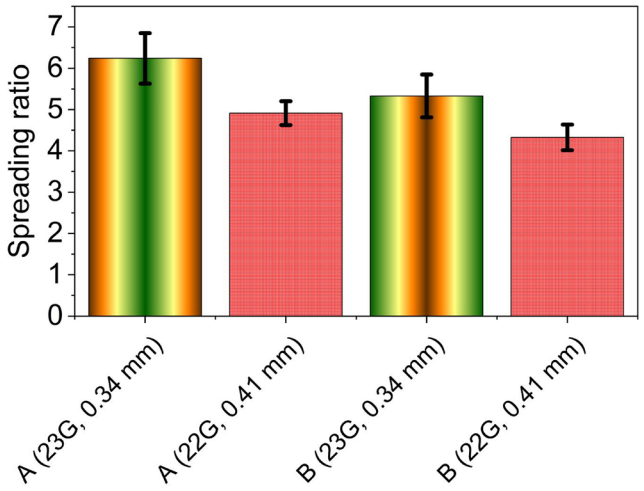


FIGURE 10 Spreading ratios of all constructs.

analysis. Two grid patterns, 4 × 4 and 2 × 2, with two layers of extruded hydrogel stacked vertically and inter-filament distances of 8 and 4 mm, respectively, were printed using nozzles with inner diameters of 0.34 mm (23 G) and 0.41 mm (22 G), resulting in a total of four constructs (Figure 8A). In the case of the 24 G nozzle, the printer could not provide the required higher pressure for extrusion. The Pr, SR, and line width were calculated for all four constructs.

Figure 8C shows the application of the printability formula for different pore shapes in which a Pr value of 1 signifies perfectly square holes. In contrast, Pr values less than 1 suggest more circular holes and merged filaments, and Pr values greater than 1 indicate irregular shapes and bumpy extrusion. Larger Pr values correlate with greater gelation and vice versa.[39,40] The mean Pr values were 0.9187878, 1.0136814, 0.8768435, and 0.9387235 for the four constructs A (23 G), A (22 G), B (23 G), and B (22 G), respectively (Figure 9). These results correspond to another work on a self-cross linkable chitosan-based hydrogel made using Schiff base linkages.[25] The constructs fabricated using the 22 G (0.41 mm) nozzle demonstrated superior Pr values, indicating that for this particular composition of O-MOG (5% w/v) and CMCh (5% w/v), 22 G represents the minimum feasible nozzle diameter.

Lower spreading ratios are preferred to produce hydrogel structures with enhanced precision.[41] The smallest spreading ratio among the provided samples was identified for construct B, printed using a 0.41 mm (22 G) inner diameter nozzle, measuring approximately 4.9. All spreading ratios recorded are below 6.5 (Figure 10). The constructs printed using nozzles with a 0.41 mm (22 G) inner diameter also showed the lowest SR values among all four grids. This observation further reinforces that 22 G nozzles are well-suited for printing for the given

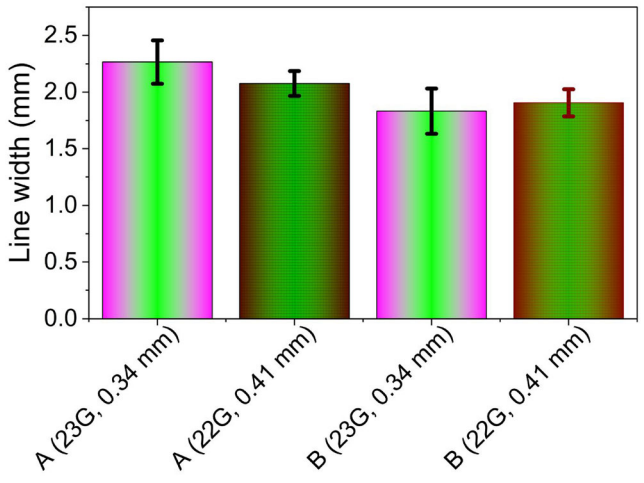


FIGURE 11 Line widths of all constructs.

composition of gels. Ideally, a spreading ratio close to 1 is desired. But Ioannidis et al.'s work on an alginate-gelatin mixture as bio-ink, spreading ratios higher than 10 have also been reported.[42]

The line width for all printed samples was determined. Figure 11 shows construct A, printed with a 0.34 mm (23 G) inner diameter nozzle, exhibited the widest line width among all samples, measuring approximately 2.23 mm. This finding corresponds with the spreading ratio of construct A (23 G), which is the highest among all sample constructs. Interestingly, this width surpassed that of constructs printed with a 0.41 mm (22 G) inner diameter nozzle. This again suggests that employing nozzles with inner diameters smaller than 22 G compromises the structural integrity of the extruded material, leading to unsatisfactory outcomes. Once again, citing the work by Ioannidis et al., line widths close to 3 mm from 25 G diameter nozzles have been reported.[42]

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 | CONCLUSION

The most common rheological assessments were carried out for three O-MOG/CMCh hydrogels containing varied concentrations of O-MOG (3% w/v to 5% w/v) with constant CMCh concentration (5% w/v) formed by Schiff base reaction. Time sweep test showed that hydrogel containing 5% w/v of O-MOG, undergoing sol–gel transition at 3.23 ± 0.05 min, and amplitude sweep test demonstrated a relatively high mechanical strength of $G' \approx 2.4$ kPa. Thus, O-MOG/CMCh-5 exhibited the best properties for printing purposes. At the same time, all O-MOG/CMCh hydrogels are highly shear thinning with good structure recovery properties. Oscillatory thixotropy results show that the material has excellent self-healing capabilities. The rheological properties suggest that the O-MOG/CMCh hydrogels, formed by Schiff base linkage, are fit for 3D printing applications. The O-MOG/CMCh-5 hydrogel was then employed to successfully print 4×4 and 2×2 grids using 0.41 mm (22 G) and 0.34 mm (23 G) inner diameter nozzles. On visual analyses of the constructs, it is observed that the extruded filament is more uniform when the gel is extruded from the 22 G nozzle. The uniformity begins to deteriorate as the nozzle size decreases to 23 G. The printability index of all constructs is within the acceptable range, with the constructs printed via a 0.41 mm (22 G) inner diameter nozzle showing the best results as their Pr values are closest to one. The lowest Pr value is ≈0.87 for construct B, printed using a 23 G nozzle. Spreading ratio and line width data also show better results for grids printed using 22 G nozzles. This suggests that the material cannot recover completely after it is extruded through the 0.34 mm (23 G) inner diameter nozzle. A possible reason for this is that extruding the gel through nozzles with diameters smaller than the 22 G nozzle disrupts the internal structure of the gel to a degree where it begins to clog the nozzle. Hence, a 22 G inner diameter nozzle is best suited for this composition. Further material optimization is needed to use nozzles of lower diameters and get better results for Pr and SR values. Different compositions, different preparation methods, or the addition of crosslinkers may lead to better shape fidelity, mechanical properties, and uniform extrudability with lower-diameter nozzles.

ACKNOWLEDGMENTS

The author greatly acknowledges Delhi Technological University, Delhi, for providing financial assistance to conduct the research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data will be made available on request.

ORCID

Roli Purwar https://orcid.org/0000-0001-8003-999X

REFERENCES

1. He X, Lu Q. A review of high internal phase Pickering emulsions: stabilization, rheology, and 3D printing application. *Adv Colloid Interface Sci*. 2024;324:103086. doi:10.1016/j.cis.2024.103086
2. Herrada-Manchón H, Fernández MA, Aguilar E. Essential guide to hydrogel rheology in extrusion 3D printing: how to measure it and why it matters? *Gels*. 2023;9(7):517. doi:10.3390/gels9070517
3. Tejada-Ortigoza V, Cuan-Urquizo E. Towards the development of 3D-printed food: a rheological and mechanical approach. *Foods*. 2022;11(9):1191. doi:10.3390/foods11091191
4. Siacor FDC, Chen Q, Zhao JY, et al. On the additive manufacturing (3D printing) of viscoelastic materials and flow behavior: from composites to food manufacturing. *Addit Manuf*. 2021;45:102043. doi:10.1016/j.addma.2021.102043
5. Torabi H, McGreal H, Zarrin H, Behzadfar E. Effects of rheological properties on 3D printing of poly(lactic acid) (PLA) and poly(hydroxy alkenoate) (PHA) hybrid materials. *ACS Appl Polym Mater*. 2023;5(6):4034-4044. doi:10.1021/acsapm.3c00271
6. Liu Y, Hildner M, Roy O, et al. On the selection of rheological tests for the prediction of 3D printability. *J Rheol*. 2023;67(4):791. doi:10.1122/8.0000612
7. Azad MA, Olawuni D, Kimbell G, Badruddoza AZM, Hossain MS, Sultana T. Polymers for extrusion-based 3D printing of pharmaceuticals: a holistic materials-process perspective. *Pharmaceutics*. 2020;12(2):124. doi:10.3390/pharmaceutics12020124
8. Jayathilakage R, Rajeev P, Sanjayan J. Rheometry for concrete 3D printing: a review and an experimental comparison. *Buildings*. 2022;12(8):1190. doi:10.3390/buildings12081190
9. Zamboulis A, Michailidou G, Koumentakou I, Bikiaris DN. Polysaccharide 3D printing for drug delivery applications. *Pharmaceutics*. 2022;14(1):145. doi:10.3390/pharmaceutics14010145
10. Gebeyehu A, Surapaneni SK, Huang J, et al. Polysaccharide hydrogel based 3D printed tumor models for chemotherapeutic drug screening. *Sci Rep*. 2021;11(1):372. doi:10.1038/s41598-020-79325-8
11. Wu X, Cheng X, Kang M, Dong R, Zhao J, Qu Y. Natural polysaccharide-based hydrogel bioprinting for articular cartilage repair. *Front Mater*. 2023;10:10. doi:10.3389/fmats.2023.1204318
12. Teixeira MC, Lameirinhas NS, Carvalho JPF, Silvestre AJD, Vilela C, Freire CSR. A guide to polysaccharide-based hydrogel bioinks for 3D bioprinting applications. *Int J Mol Sci*. 2022;23(12):6564. doi:10.3390/ijms23126564
13. Tamboli QY, Patange SM, Mohanta YK, et al. Moringa oleifera gum-assisted synthesis and characterization of CoAg(x)Fe(2-x)O(4): insight into structural, magnetic, optical, and biomedical properties. *ACS Omega*. 2024;9(3):3835-3845. doi:10.1021/acsomega.3c06578
14. Aranaz I, Alcántara AR, Civera MC, et al. Chitosan: an overview of its properties and applications. *Polymers (Basel)*. 2021;13(19):3256. doi:10.3390/polym13193256
15. El-Araby A, Janati W, Ullah R, Ercisli S, Errachidi F. Chitosan, chitosan derivatives, and chitosan-based nanocomposites: eco-

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

friendly materials for advanced applications. *Front Chem*. 2024; 11:1327426. doi:10.3389/fchem.2023.1327426

16. Kumar J, Purwar R. Biodegradable, biocompatible, and self-healing, injectable hydrogel based on oxidized Azadirachta indica gum and carboxymethyl chitosan via dynamic imine-linkage for biomedical application. *Iran Polym J*. 2024.
17. Kumar J, Purwar R. Injectable Mesquite gum and carboxymethyl chitosan hydrogel using Schiff Base crosslinks: a versatile platform for drug delivery in wound care. 2024.
18. Mohabatpour F, Duan X, Yazdanpanah Z, et al. Bioprinting of alginate-carboxymethyl chitosan scaffolds for enamel tissue engineering in vitro. *Biofabrication*. 2023;15(1):15022. doi:10.1088/1758-5090/acab35
19. Bider F, Miola M, Clejanu CE, et al. 3D bioprinting of multifunctional alginate dialdehyde (ADA)–gelatin (GEL) (ADA-GEL) hydrogels incorporating ferulic acid. *Int J Biol Macromol*. 2024; 257:128449. doi:10.1016/j.ijbiomac.2023.128449
20. Kim M, Schöbel L, Geske M, Boccaccini AR, Ghorbani F. Bovine serum albumin-modified 3D printed alginate dialdehyde-gelatin scaffolds incorporating polydopamine/SiO2-CaO nanoparticles for bone regeneration. *Int J Biol Macromol*. 2024;264:130666. doi:10.1016/j.ijbiomac.2024.130666
21. Ettoumi F-Z, Huang H, Xu Y, et al. Supramolecular assembly of dual crosslinked nanocomposite polysaccharides hydrogel: integration of injectable, self-healing, and pH-responsive platform for sustained delivery of polyphenols. *Food Hydrocoll*. 2024;154:110108. doi:10.1016/j.foodhyd.2024.110108
22. Kim HS, Kim JS, Hwang J, Lee IY, Lee KY. Decoupling stiffness and toughness of self-healing hydrogels for complex tissue regeneration via 3D bioprinting. *Chem Eng J*. 2024;487:150551. doi:10.1016/j.cej.2024.150551
23. Laurén I, Farzan A, Teotia A, Lindfors NC, Seppälä J. Direct ink writing of biocompatible chitosan/non-isocyanate polyurethane/cellulose nanofiber hydrogels for wound-healing applications. *Int J Biol Macromol*. 2024;259:129321. doi:10.1016/j.ijbiomac.2024.129321
24. Yu Z, Li Q, He X, et al. A multifunctional hydrogel based on nature polysaccharide fabricated by Schiff base reaction. *Eur Polym J*. 2023;197:112330. doi:10.1016/j.eurpolymj.2023.112330
25. Gwak MA, Lee SJ, Lee D, Park SA, Park WH. Highly gallol-substituted, rapidly self-crosslinkable, and robust chitosan hydrogel for 3D bioprinting. *Int J Biol Macromol*. 2023;227:493-504. doi:10.1016/j.ijbiomac.2022.12.124
26. Wang P, Li J, Zhang W, et al. 3D printed heart valve mediated nitric oxide sustained release reduced potential for calcification and inflammatory capacity. *Chem Eng J*. 2023;469:143892. doi:10.1016/j.cej.2023.143892
27. Kumar J, Purwar R. Self-healing, biocompatible injectable hydrogel based on multialdehyde Moringa oleifera gum and carboxymethyl chitosan: a suitable platform for drug delivery in wound healing application. *ChemistrySelect*. 2024;9(9): e202400309. doi:10.1002/slct.202400309
28. Ding X, Li G, Xiao C, Chen X. Enhancing the stability of hydrogels by doubling the Schiff Base linkages. *Macromol Chem Phys*. 2019;220(3):1800484. doi:10.1002/macp.201800484
29. Hu C, Hahn L, Yang M, et al. Improving printability of a thermoresponsive hydrogel biomaterial ink by nanoclay addition. *J Mater Sci*. 2021;56(1):691-705. doi:10.1007/s10853-020-05190-5
30. Barrulas RV, Corvo MC. Rheology in product development: an insight into 3D printing of hydrogels and aerogels. *Gels*. 2023; 9(12):986. doi:10.3390/gels9120986
31. Efstathiou S, Wemyss AM, Patias G, et al. Self-healing and mechanical performance of dynamic glycol chitosan hydrogel nanocomposites. *J Mater Chem B*. 2021;9(3):809-823. doi:10.1039/d0tb02390f
32. Wang Y, Han L, Zhang X, et al. 3D bioprinting of an electroactive and self-healing polysaccharide hydrogels. *J Tissue Eng Regen Med*. 2022;16(1):76-85. doi:10.1002/term.3238
33. Cheng Y, Qin H, Acevedo NC, Jiang X, Shi X. 3D printing of extended-release tablets of theophylline using hydroxypropyl methylcellulose (HPMC) hydrogels. *Int J Pharm*. 2020;591: 119983. doi:10.1016/j.ijpharm.2020.119983
34. Maity S, Datta A, Lahiri S, Ganguly J. A dynamic chitosan-based self-healing hydrogel with tunable morphology and its application as an isolating agent. *RSC Adv*. 2016;6(84):81060-81068. doi:10.1039/C6RA15138H
35. Bercea M. Rheology as a tool for fine-tuning the properties of printable bioinspired gels. *Molecules*. 2023;28(6):2766. doi:10.3390/molecules28062766
36. Gering C, Rasheed A, Koivisto JT, Párraga J, Tuukkanen S, Kellomäki M. Chemical modification strategies for viscosity-dependent processing of gellan gum. *Carbohydr Polym*. 2021; 269:118335. doi:10.1016/j.carbpol.2021.118335
37. Puertas-Bartolomé M, Włodarczyk-Biegun MK, del Campo A, Vázquez-Lasa B, San RJ. 3D printing of a reactive hydrogel bio-ink using a static mixing tool. *Polymers (Basel)*. 2020;12(9): 1986. doi:10.3390/polym12091986
38. Haider MS, Ahmad T, Yang M, et al. Tuning the thermogelation and rheology of poly(2-oxazoline)/poly(2-oxazine)s based thermosensitive hydrogels for 3D bioprinting. *Gels*. 2021;7(3): 78. doi:10.3390/gels7030078
39. Lim W, Shin SY, Cha JM, Bae H. Optimization of polysaccharide hydrocolloid for the development of bioink with high printability/biocompatibility for coextrusion 3D bioprinting. *Polymers (Basel)*. 2021;13(11):1773. doi:10.3390/polym13111773
40. Majumder N, Roy C, Doenges L, Martin I, Barbero A, Ghosh S. Covalent conjugation of small molecule inhibitors and growth factors to a silk fibroin-derived bioink to develop phenotypically stable 3D bioprinted cartilage. *ACS Appl Mater Interfaces*. 2024;16(8):9925-9943. doi:10.1021/acsami.3c18903
41. Freeman FE, Kelly DJ. Tuning alginate bioink stiffness and composition for controlled growth factor delivery and to spatially direct MSC fate within bioprinted tissues. *Sci Rep*. 2017; 7(1):17042. doi:10.1038/s41598-017-17286-1
42. Ioannidis K, Danalatos RI, Champeris Tsaniras S, et al. A custom ultra-low-cost 3D bioprinter supports cell growth and differentiation. *Front Bioeng Biotechnol*. 2020;8:580889. doi:10.3389/fbioe.2020.580889

How to cite this article: Verma M, Kumar J, Pradhan AA, Majumder N, Ghosh S, Purwar R. Assessing rheological properties of oxidized *Moringa oleifera* gum and carboxymethyl chitosan-based self-healing hydrogel for additive manufacturing applications. *Polym Eng Sci*. 2024; 1-10. doi:10.1002/pen.26915

15482634, 0, Downloaded from https://4spepublications.onlinelibrary.wiley.com/doi/10.1002/pen.26915 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Assessment of PM_{10} in Colombo, Sri Lanka and Its Potential Source Regions

Thrividya Rathnayake, *Lovleen* Gupta

Department of Environmental Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, 110042 Delhi, India

rathnayakemudiyanselag_2k22ene06@dtu.ac.in, lgupta@dce.ac.in

Abstract. Air pollution, specifically the presence of particulate matter (PM_{10}) is a growing concern in many parts of the world, including Sri Lanka because of their ability to cause series health risks, destructive environmental damage and their negative effects on the nation's economy. Since Colombo is the commercial capital of Sri Lanka, this study analyzed the PM_{10} concentrations in Colombo over one-year period; impact of meteorological parameters on the PM_{10} levels and potential source regions. To accomplish this, data on 24-hour average PM_{10} concentrations and meteorological parameters were collected at Colombo from Jan to Dec 2022. Colombo experienced 40.5 ± 26.6μgm^{-3} average PM_{10} concentrations in 2022 which is about 3 times higher than WHO guidelines (15μgm^{-3}). Seasonally, Northeast monsoon season experienced the highest average PM_{10} concentration (~56.8 μgm^{-3}) in Colombo that may be attributed to low rainfall (~3.8mm) and low wind speed (~12.1mph). Conversely, minimum average PM_{10} level (~24.2 μgm^{-3}) occur during Southwest monsoon which may be caused by higher rainfall (~8.4mm) and wind speed (~17.9mph). CBPF was used to evaluate the potential source regions for high PM_{10} levels in Colombo, indicating potential sources could be vehicular emission, thermal power plant, waste combustion and industrial emissions. The results of this study assist policy makers in charting the control strategies in the region.

Keywords: PM_{10}, Health Risks, Colombo, Potential Source Regions, Policy Makers

1 INTRODUCTION

Air pollution is one of the most complicated environmental issues that must be addressed as it has been linked to various health issues leading to mortality [1]. Of all the air pollutants, Particulate Matter (PM) pollution, which consists of particles floating in the air is responsible for ~8.7 million deaths globally [2]. PM_{10} leads to reduced visibility and have the ability to cause bronchial irritation, inflammation, and an increase in respiratory system

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

reactivity, wheezing, asthmatic flare-ups, respiratory infections, chronic bronchitis, and chronic obstructive pulmonary disease may all be triggered by these effects [3].
Sri Lanka lies in the Indian Ocean, in the southwest of the Bay of Bengal and is categorized as a tropical country [4]. Colombo, Sri Lanka's capital city serves as the country's economic, cultural, and political hub and is located on the island's western coast. While statistics from 2018 indicate that 13% of Sri Lanka's population lives in Colombo, air pollution problem is getting worse due to the city's fast urbanization, accumulation of businesses, commercial sectors, and rising living standards, etc.[5] (S.Nandasena & Premasiri, 2021). Vehicular emissions, burning of organic waste, agricultural byproducts, industrial emissions (particularly from thermal power plants), petroleum refining, and trans-boundary movement of air pollutants from the Asian continent all have an impact on air quality of Colombo [6]. Sri Lanka has historically made significant policy interventions in reducing air pollution through the establishment of necessary legal and regulatory mechanisms, institutional strengthening, and carrying out necessary capacity development activities [7]. The World Health Organization (WHO) air quality guidelines were published in 2005, updated in 2021 and Sri Lanka's air quality standards, including the standards for PM_{10}, were amended and published in August 2008 [8].
In Sri Lanka, regular monitoring of air pollution with automatic air quality stations began in 1997 with two monitoring stations in Colombo [9]. Since then there have been number of air quality studies with limited datasets in different parts of Sri Lanka. A study in Colombo found that the ratio of suspended particulate matter ($TSP:PM_{10}:PM_{2.5}$) is 4:3:2, with PM_{10} accounting for 70% of total particles [10]. Cooking with biomass, car exhaust, and road dust contribute to particle chemical composition [11]. In 2020, a literature review has been conducted and this showed the wellsprings of air contamination can be classified into three essential groups: transport, industry (counting power generation) and homegrown use [9]. There was another study in 2021 which investigated the potential impacts of ambient particulate matter (PM) on respiratory disease hospitalization in Kandy in Sri Lanka for 2019 [12].
This study presents the variation in PM_{10} over Colombo throughout the year from January – December 2022, analysis the impact of meteorological parameters on PM_{10} and explores the potential source regions responsible for high PM10 concentration.

2 METHODOLOGY

2.1 Study Area

The Palk Strait divides Sri Lanka, an island in the Indian Ocean, from peninsular India. Its latitude and longitude are 5°55′ and 9°51′ N and 79°41′ and 81°53′ E, respectively. Its maximum length and width are 268 miles (432 km) and 139 miles (224 km), respectively. Colombo is the commercial capital city of Sri Lanka which is located on the western coast of Sri Lanka, along the Indian Ocean. The coastline extends from north to south, providing access to important ports like the Colombo Port and the newer Port City development. Particularly during the monsoon season, the topographical features in this area have a significant effect on the spatial patterns of winds, seasonal rainfall, temperature, relative humidity, and other climatic aspects. First Inter-monsoon Season (March – April), Southwest Monsoon Season (May – September), Second Inter-monsoon Season (October – November), and Northeast Monsoon Season (December – February) are the four climate seasons that can be used to describe the climate in Sri Lanka over a 12-month period [4].



Fig. 1. Study Area [13]

2.2 Data Collection

24-hour average PM_{10} data of Colombo were monitored and gathered by National Building Research Organization (NBRO) which is a government organization in Sri Lanka for one-year period (2022). There are some blank days for monitoring PM_{10} concentrations due to the power failures and other unavoidable situations. If continuously more than 75% 24-hr data was not available, monthly average value is not calculated. Data was screened any negative values or $PM_{2.5}/PM_{10}$ is more than 1, those data have been eliminated. For this study, annual average Air Quality standards were considered as $50 \mu gm^{-3}$ for PM_{10} which were amended and gazette in 2008 [8].
CSV file of 24 hours daily mean values of meteorological parameters such as rainfall, mean temperature and wind speed in Colombo are freely available at https://www.kaggle.com/datasets/rasulmah/sri-lanka-weather-dataset?resource=download.

3 RESULTS AND DISCUSSION

3.1 Daily, Monthly and Annual PM_{10} Variation

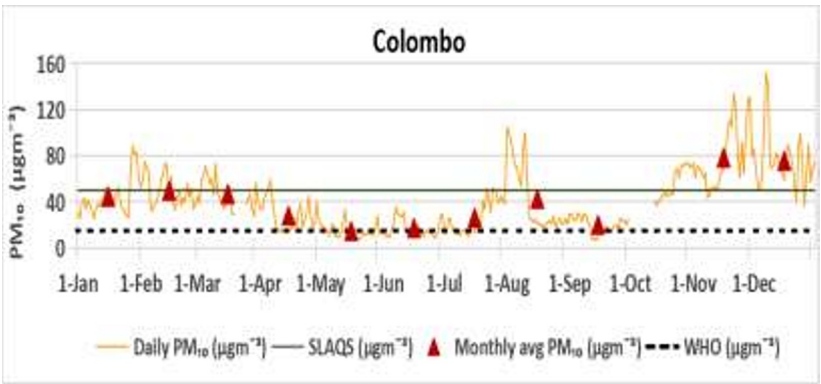


Figure 2. Daily and monthly average variation of PM_{10} concentrations in Colombo

Table 1. Annual summary of PM_{10} data in Colombo (N stands for the number of available data per year and Min (Minimum), Max (Maximum), Median, Mean, Q1 (first quartile), Q3, and SD (Standard Deviation) values are mentioned in μgm^{-3}).

	N	Min	Max	Median	Mean	Q1	Q2	Q3	SD
Colombo	360	6.5	152.0	33.9	40.5	19.3	33.9	56.4	26.6
SLAQS = 50 μgm^{-3} WHO Guidelines = 15 μgm^{-3}									

Colombo, Sri Lanka's capital city serves as the country's economic, cultural, and political hub and is located on the island's western coast. During Jan PM10 concentrations are varying below the SLAQS ($50\mu gm^{-3}$), while pollution concentration during some days of both Feb and March exceeding the permissible limit. From Apr to Jul pollutant level is less than $50\mu gm^{-3}$ and again during Aug there is a sudden increase. PM_{10} amount during Sept and Oct are not crossing the standard limits. Nov and Dec months are experiencing the highest PM_{10} concentrations throughout the study period. Colombo marks the maximum PM_{10} concentration as $152\mu gm^{-3}$ during Dec. According to the summary of PM_{10} data in Colombo the average concentration in 2022 is $40.5\mu gm^{-3}$ which is not higher than SLAQ standards.

3.2 Seasonal Variation

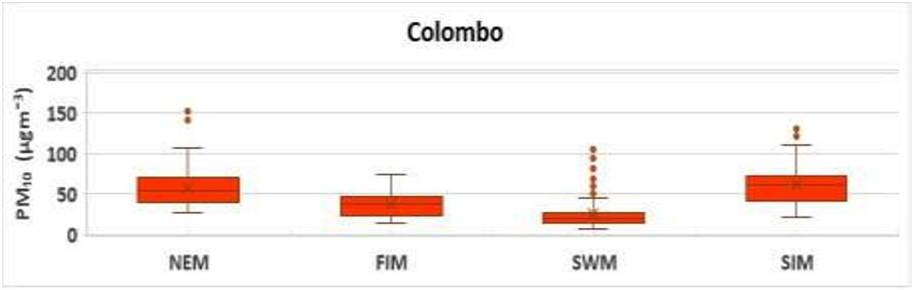


Fig. 4. Seasonal variation of PM_{10} concentrations in Colombo in 2022 (NEM – Northeast monsoon, FIM – First inter-monsoon, SWM – Southwest monsoon, SIM – Second inter-monsoon)

Fig. 4 provides a clear picture of variation of PM_{10} during 4 different climate seasons in Sri Lanka. Northeast monsoon (Dec – Feb) is experiencing the maximum pollutant concentrations while Southwest monsoon (May – Sep) is having the minimum pollution level.

3.3 Effects of Meteorological Parameters

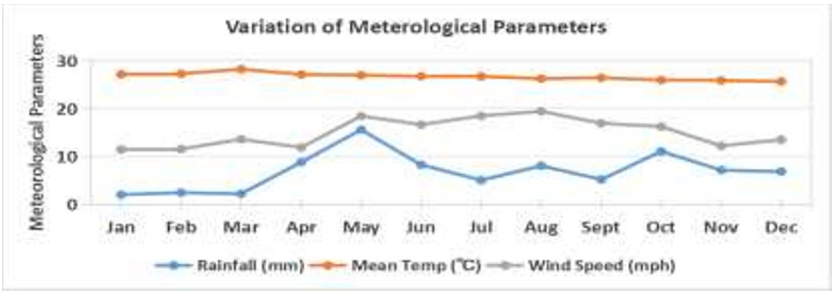


Figure 5. Monthly average rainfall, temperature and wind speed in Colombo in 2022

Behavior of PM_{10} concentrations in Colombo over 2022 can be explained by analyzing meteorological parameters such as rainfall, temperature and wind speed. From Jan to March, Colombo is experiencing low rainfall and low wind speed. From Jan to March mean temperature is slightly increasing and again from March to Dec temperature is decreasing. During the whole year mean temperature is fluctuating between 18 - 30ºC. May is having highest rainfall and highest wind speed which is resulting lower pollutant concentrations. Nov and Dec are facing low rainfall, low winds peed and low temperature that leads to maximum PM_{10} levels in Colombo.

3.4 Role of Local Wind Speed and Wind Direction

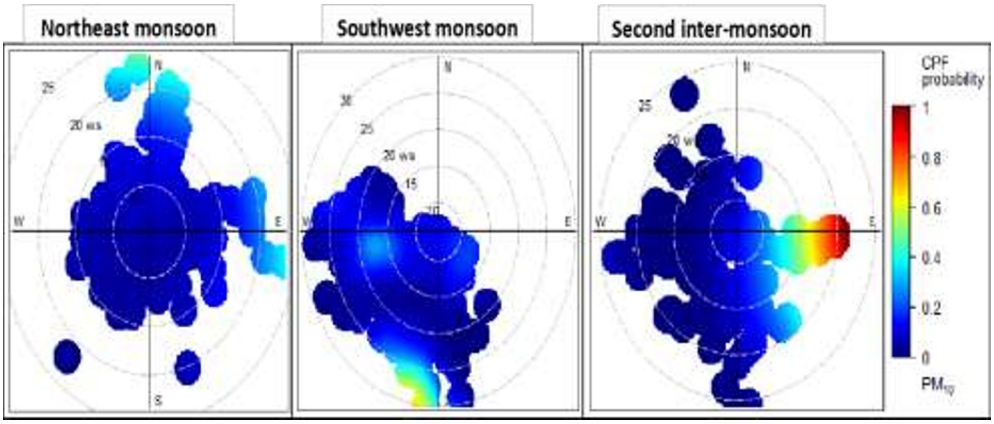


Figure 6: PM_{10} CBPF plots at Colombo for three seasons

During northeast monsoon, high PM_{10} concentrations occur at high wind speed (20-25 mph) from N and SE directions. Likewise, during southwest monsoon high PM_{10} level occur at 30 mph high wind speed from S and SW, while in second inter-monsoon high pollution level is experiencing at 20mph wind speed from E. These plots reveal that cross boundary pollution or stack emissions are present rather than the non-buoyant ground-level sources in Colombo. Potential sources in this area

could be vehicular emission, thermal power plant, waste combustion and industrial emissions.

4 Conclusion

Behavior of PM_{10} over Colombo throughout the year from January –December 2022, impact of meteorological parameters on PM_{10} and the potential source regions responsible for high PM_{10} concentration are analyzed in this study. Daily, monthly and seasonal variations reveal that high PM_{10} level is experiencing during Jan, Nov and Dec months (Northeast monsoon). From May – July (Southwest monsoon), Colombo is experiencing the lowest PM_{10} concentrations. This seasonal trend could be caused by the meteorological parameters as during northeast monsoon Colombo is facing low rainfall, low temperature and low wind speed while Colombo is experiencing highest rainfall, wind speed and temperature during southwest monsoon. CBPF plots indicate that cross boundary or stack emissions are happening in Colombo rather than ground level emissions. Vehicular emission, thermal power plant, waste combustion and industrial emissions could be the potential sources in the study area. This evidence from the study can be utilized for further air quality and exposure measurement for serving as a platform for further researches, eventually leading to environmental policy, health plan review, and population welfare plans not only in Colombo but also in all over Sri Lanka.

References

[1] J. Rovira, J. L. Domingo, and M. Schuhmacher, “Air quality, health impacts and burden of disease due to air pollution (PM10, PM2.5, NO2 and O3): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain),” *Sci. Total Environ.*, 2019, doi: https://doi.org/10.1016/j.scitotenv.2019.135538.

[2] World Health Organization, “Air Pollution.” Accessed: Jul. 20, 2018. [Online]. Available: http://www.who.int/airpollution/en/

[3] S. Nandasena and S. Premasiri, *Air Quality Assessment for Health and Environment Policies in Sri Lanka*. Air Resource Management and National Ozone Unit, Ministry of Environment, 2021.

[4] S. L. Department of Meteorology, “Climate of Sri Lanka.” [Online]. Available: https://www.meteo.gov.lk/index.php?option=com_content&view=article&id=94&Itemid=310&lang=en&lang=en#2-southwest-monsoon-season-may-september

[5] International Water Management Institute, “Assessing and Planning City Region Food System - Colombo, Sri Lanka,” 2008. [Online]. Available: https://ruaf.org/assets/2019/12/CRFS-factsheet-Colombo.pdf

[6] IQAir, “Air Quality in Sri Lanka.” [Online]. Available: https://www.iqair.com/in-en/sri-lanka

[7] Ministry of Mahaweli Development and Environment, “Clean Air 2025: An Action Plan for Air Quality Mnagement.” [Online]. Available: https://www.ccacoalition.org/policy-database/clean-air-2025-action-plan-air-quality-management

[8] Y. L. S. Nandasena, A. R. Wickremasinghe, and N. Sathiakumar, “Air pollution and health in Sri Lanka: a review of epidemiologic studies,” *BMC Public Health*, vol. 10, no. 1, p. 300, Dec. 2010, doi: 10.1186/1471-2458-10-300.

[9] O. A. Ileperuma, “Review of air pollution studies in Sri Lanka,” *Ceylon J. Sci.*, pp.

225–238, 2020.

[10] C. . and Premasiri, H.D.S., Samarakkody, R.P., Annakkage and G. B. M. A. Basnayake, "Particulate Pollution and Ratio of SPM:PM10:PM2.5 in Colombo Atmosphere," *Proc. Sri Lanka Assoc. Adv. Sci.*, 2003.

[11] H. L. Seneviratne, M.C.S., Waduge, V.A., Handagiripathira and K. G. S. Mahakumbura, P.D., Attanayake T.N. and Jayawardena, "Air pollution monitoring using Nuclear related Analytical Technique, Energy Dispersive X-ray fluorescence in Colombo," *Proceeedings First Natl. Symp. Air Resour. Manag.*, 2004.

[12] S. . S. Priyankara, L. . M.; Jayaratne, R.; Morawska, R. . Abeysundara, S.; Weerasooriya, S. C. . Knibbs, L.D.; Dharmage, and G. Yasaratne, D.; Bowatte, "Ambient PM2.5 and PM10 Exposure and Respiratory Disease Hospitalization in Kandy, Sri Lanka.," *Int. J. Environ. Res. Public Heal.*, vol. 18, 2021, doi: https://doi.org/10.3390/ijerph18189617.

[13] Wikipedia, "Sri Lanka - Colombo." [Online]. Available: https://en.m.wikipedia.org/wiki/File:Sri_Lanka_-_Colombo_%28original_size%29.png

2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT) | 979-8-3315-4068-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICONSCEPT61884.2024.10627829

Augmented GAN: Advancing Generative Adversarial Networks through Enhanced Discriminator Training and Generator Stability

Aman Prakash[1], Ryan Binu Varghese[2], and Lavi Tanwar[3]

[1,2,3]Department of Electronics and Communications Engineering, Delhi Technological University

Email: [1]amanbahrain@gmail.com, [2]ryan.binu@gmail.com, [3]lavi.tanwar@dtu.ac.in

***Abstract*—General Adversarial Networks (GANs) have emerged as a powerful framework for generating reliable and transformative synthetic data in areas such as image generation, image and text synthesis, and data augmentation. This paper presents a comprehensive guide to building a Generative Adversarial Neural Network using TensorFlow and Python. Through our research on this topic, we delve into the implementation details, providing step-by-step instructions for constructing both the generator and discriminator networks using TensorFlow, a popular deep-learning framework. Furthermore, we explore techniques for optimizing GAN performance, including architectural modifications, loss function selection, and training strategies. We demonstrate how to train a GAN on real-world datasets through practical examples, showcasing its capability to generate high-quality synthetic samples that closely resemble the training data distribution. In this paper, we aim to match fashion products such as clothing and accessories using GAN. This paper also focuses on improving the robustness and generalization of Convolutional Neural Network (CNN) models, such as image classification and product recommendation. By combining image results with GANs and incorporating these artificial data into the training process, we aim to improve the performance of CNN models in real-world applications.**

***Index Terms*—Generative Adversarial Networks (GANs), Generator, Discriminator, Tensorflow datasets, Deep-Learning, Data augmentation, Batch Normalization, Convolution Neural Networks, Fashion-mnist,**

I. Introduction

The ability of generative adversarial networks (GANs) to generate realistic generated data in a variety of disciplines has drawn a lot of attention. GANs were first described by Ian Goodfellow and colleagues in 2014. They consist of two neural networks, a generator and a discriminator, that compete within a framework based on game theory. To trick the discriminator, which learns to discern between real and fake data, the generator produces patterns in the data. Effectiveness in tasks including data augmentation, picture creation, and transfer has been demonstrated by this adversarial approach.

Our research advances fashion image synthesis by integrating GANs with Convolutional Neural Networks (CNNs) to enhance image generation and classification. Unlike the FashionGAN by Yirui Cui et al. (2018) [1], which focuses on image synthesis with conditional GANs, our approach incorporates advanced data augmentation and batch normalization techniques. This leads to improved robustness, generalization, and model stability. Additionally, our framework optimizes CNN performance in tasks such as image classification and product recommendation, areas not addressed by previous works.

Key contributions in this papers research include:

1) **Integration of GANs and CNNs:** Combining GANs for image generation with CNNs for classification and recommendation.
2) **Enhanced Techniques:** Using advanced data augmentation and batch normalization for better model performance.
3) **Comprehensive Framework:** Addressing both image synthesis and classification within a unified framework.

Section II provides an overview of previous studies conducted on GANs and image synthesis. Section III details the methodology, while Section IV presents experimental results and analysis.

II. Literature Review

Generative Adversarial Networks have emerged as a groundbreaking framework in the field of generative modeling, revolutionizing various domains such as image synthesis, data augmentation, and style transfer. The use of Generative Adversarial Networks is a critical topic to delve into, for the intricacies of the GAN architecture, exploring its fundamental components, the adversarial training process, and techniques for optimizing performance using TensorFlow.

Researchers have worked extensively to address the issue of generating realistic and diverse fashion images, leading to the development of specialized architectures like FashionGAN. FashionGAN tackles the challenge of synthesizing high-resolution fashion model images by incorporating conditional GANs and innovative discriminator designs. Through multi-scale generators and pairwise discriminators, FashionGAN has shown remarkable advancements in capturing intricate fashion details and diverse styles. Moreover, its ability to generate fashion images conditioned on specific attributes contributes significantly to enhancing the realism and variability of the generated outputs. As a result, FashionGAN stands as a pivotal

979-8-3315-4068-5/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:47 UTC from IEEE Xplore. Restrictions apply.

contribution in the field of fashion image synthesis, facilitating applications ranging from virtual try-on systems to fashion design automation.

Riska Dhenabayu(2023) [2] proposed the idea of Fashion-GAN based on thematic maps and evolution analysis, GANs emerge as promising in fashion, trending alongside deep learning. Future research should explore GANs in accessory and textile design for creative expression, sustainability, and circular fashion. Focus on fairness-aware GAN frameworks addresses ethical concerns, promoting diversity and responsible usage. GANs have immense potential to revolutionize fashion by inspiring innovation and exploration in the field.

In their study, *A. Jain et al.(2019)* [3] proposed a web service for designing customized clothing images based on user preferences, potentially extending to personalized outfits tailored to individual body dimensions. However, limitations include dependence on the training dataset, requiring constant updates, and potential inaccuracies in replicating real-world outfits. Model implementation involved STACKGAN and the FASHION-GEN dataset, with cloud deployment via Google's Cloud ML platform and final deployment through a user-friendly web interface. This approach enables the creation of virtual fashion models from text inputs and offers interactivity through various color palette options, marking a significant step in democratizing fashion design.

In order to provide a definitive response to the question of whether GANs are scalable, *Minguk Kang et al(2023)* [4] presented a novel strategy. The suggested architecture is scalable to model sizes that allow text-to-image synthesis. Although it was orders of magnitude faster and allowed latent interpolation and stylization, it produced competitive results with auto-regressive and diffusion models trained with comparable resources. Large-scale generative models have a whole new design space thanks to the GigaGAN architecture, which also restores important editing features that were difficult to use when diffusion and auto-regressive models were used.

H.Yang et al(2020) [5] In order to mitigate the problem of imbalanced datasets, he discusses the difficulties involved and suggests IDA-GAN, or Imbalanced Data Augmentation Generative Adversarial Networks (GAN), as an augmentation technique. It is very challenging to overcome this obstacle when training GAN models. The problem is solved by using GAN training in conjunction with a variational autoencoder. Experimental results on benchmark datasets such as Fashion-MNIST, GTSRB, CIFAR-10, SVHN, and MNIST compare IDA-GAN against state-of-the-art GAN models. IDA-GAN produces more diverse minority samples of higher quality and improves imbalanced classification tasks across various evaluation metrics.

The study presented by *Wei Sun et al(2019)* [6] introduces presents a brand-new generative model that creates lifelike pictures of humans changing into desired positions. The convolutional LSTM and U-Net architectures work together to enable the model to capture appearance information and acquire semantic visual context. Discriminators are used to generate high-quality images by guaranteeing pose consistency and image authenticity. This study demonstrates how crucial it is to combine architectural elements for efficient image generation and offers directions for improving semantic comprehension and visual realism even more.

Karan Jain et al(2020) [7] explores the shortcomings of GAN. CapsGAN outperformed DCGAN in the quality and precision aspect, according to the qualitative analysis. On the quantitative front, CapsGAN emerges victorious as it has experienced less loss than DCGAN. Even though DCGAN was quicker, CapsGAN outperformed it. Therefore, it can be concluded that CapsGAN performed better on the classic model when compared to the other model.

Han Yan et al(2023) [8]proposed the unsupervised disentanglement of texture and shape features of fashion items, offering potential benefits such as reduced annotation burden and increased flexibility in training. By decomposing these features into hierarchical representations, the TSD-GAN enables the generation of mixed-style fashion items, leveraging both coarse and fine styles hidden within the texture and shape features.

Paolo Andreini et al(2020) [9] recommended building large datasets of fully annotated images of Petri plates fast and cheaply using a synthetic picture generator. With the artificially generated data, a convolutional neural network may be trained for semantic image segmentation.

Han Yan et al(2021) [10] proposed FadGAN, a novel network for fashion design. It accurately maps and preserves the identity and texture of different fashion items. FadGAN disentangles input features into structure and texture codes, then generates new items by combining these codes.

X. Wu et al.(2021) [11] fused computer science with Dunhuang clothing design using GANs and style transfer algorithms. ClothGAN, trained on Fashion MNIST and Dunhuang photos, outperformed other models in creating visually appealing clothing designs, validated by measures like Inception Score and Human Preference Score. This innovative approach combines traditional cultural elements with AI to drive breakthroughs in fashion design.

Mehmet Günel et al(2018) [12] introduced a novel approach to outfit image editing using natural language instructions. It achieves realistic image manipulation while preserving untouched areas with experiments demonstrating superior performance over baseline methods.

III. Proposed Methodology

This study proposes a new approach to increase the generalizing ability and robustness of GANs in the context of fashion image synthesis and optimization. This method integrates GAN and Convolutional Neural Networks to advance image classification and improve CNN performance in image-related tasks.

A. Data Collection and Preprocessing

Data collection involves collecting various datasets of fashion product images from various online sources, including fashion websites, e-commerce platforms, fashion blogs, and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:47 UTC from IEEE Xplore. Restrictions apply.

social networks. The collected images are pre-processed to ensure consistency and facilitate model training [13]. Preprocessing steps include resizing all images to a standard size (e.g., 128 x 128 pixels), then cropping to remove extra backgrounds or margins, and pixel values normalization to facilitate model training, calibration, and data cleaning. It is a technique to remove duplicate, atypical, or low-quality photos.

Normalization Formula:

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

B. FashionGAN Architecture

FashionGAN's architecture is designed to efficiently generate synthetic fashion images that reflect realistic traits. Architectural choices, including network depth, layer configurations, and activation functions, are explored and optimized to enhance performance and stability [14].

C. Training FashionGAN

FashionGAN trains fashion collections to learn the distribution of fashion product images. To increase the accuracy of FashionGAN, we use adversarial training in which the discriminator and the generator compete against each other to improve the fashion quality. In addition, corrective measures such as reduction or weight loss are used to prevent strengthening and improve flexibility.

Adversarial Loss Formula:

$$L_{\text{adv}} = -\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] - \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

D. Data Augmentation

For model generalization purposes, the original dataset is supplemented by incorporating FashionGAN artificial images, which increases diversity. Improved data quality is achieved through the use of current data augmentation techniques, including random rotation, rotation, and color distortion. The augmented dataset is a more complete training set for further training of the CNN model.

E. CNN Model Training

Pre-trained CNN models are fine-tuned using transfer learning on the augmented dataset. This involves retraining the models while selectively updating parameters to adapt to fashion image features [15]. The CNN models are optimized to effectively classify and recognize fashion products, leveraging the enriched dataset generated by FashionGAN and traditional data augmentation techniques.

F. Performance Evaluation

Accuracy, precision, recall, and F1 score are among the metrics used to assess how well the CNN model performed both with and without data augmentation. Statistical tests are used to assess the significance of observed performance variations. Furthermore, we also performed a qualitative analysis to visually examine the quality of the generated images and the effectiveness of the CNN model in real-world imaging applications.

Algorithm 1: Training FashionGAN for Fashion Design

Input: Training dataset $\mathcal{D}$ with images of fashion items, Learning rates α_D, α_G, Number of training iterations N, Mini-batch size m **Output:** Trained generator network G for generating fashion designs Initialize discriminator network D with weights θ_D Initialize generator network G with weights θ_G **for** *number of training iterations* N **do**-_

each mini-batch of size m Sample a mini-batch of m real images $\{x^{(1)}, \ldots, x^{(m)}\}$ from dataset $\mathcal{D}$ Sample a mini-batch of m noise vectors $\{z^{(1)}, \ldots, z^{(m)}\}$ from prior $p_z(z)$

Update Discriminator: Compute discriminator loss:

$$L_D = -\frac{1}{m}\sum_{i=1}^{m}[\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$$

Update discriminator by ascending its stochastic gradient:

$$\theta_D \leftarrow \theta_D + \alpha_D \nabla_{\theta_D} L_D$$

Update Generator: Compute generator loss:

$$L_G = -\frac{1}{m}\sum_{i=1}^{m}\log D(G(z^{(i)}))$$

Update generator by descending its stochastic gradient:

$$\theta_G \leftarrow \theta_G - \alpha_G \nabla_{\theta_G} L_G$$

Return Trained generator network G

F1 Score Formula:

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

The proposed methodology provides a comprehensive framework for advancing fashion image synthesis and augmentation, leveraging the synergistic capabilities of GANs and CNNs [16]. Combining fresh approaches in data collection, preprocessing, model architecture design, training strategy, and performance evaluation makes this approach an efficient means of improving the performance of AI-based fashion applications.

TABLE I: Hyperparameters of FashionGAN

Hyperparameter	Values
Rate of Learning (Generator)	0.0001
Rate of Learning (Discriminator)	0.00001
Size of Batch	128
Optimized Using	Adam
Rate of Dropout (Discriminator)	0.4
Activation Function	LeakyReLU & Sigmoid
No. of Epochs	2000

G. Hyperparameter Optimization

Hyperparameters play a crucial role in optimizing the performance of FashionGAN. The selected hyperparameters,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:47 UTC from IEEE Xplore. Restrictions apply.

including learning rates for the generator and discriminator, batch size, optimizer choice, dropout rate, activation functions, and the number of epochs, are meticulously tuned to enhance FashionGAN's training efficiency, stability, and overall performance. For example, the dropout regularization strategy minimizes overfitting by randomly deactivating neurons during training, and the Adam optimizer option enables variable learning rates and effective gradient-based optimization. Furthermore, varying the learning rates and batch size affects the training process's memory efficiency and convergence speed, which maximizes FashionGAN's capacity to produce superior synthetic fashion images.

Hyperparameter tuning is a critical step in optimizing the performance of the FashionGAN model. Several hyperparameters were fine-tuned through extensive experimentation:

1) **Rate of Learning**: To guarantee stable training, distinct learning rates were assigned to the discriminator and generator. In order to keep the discriminator from overwhelming the generator, its learning rate should be lower.
2) **Batch Size**: The computing power at hand was taken into consideration when choosing the batch size. While smaller batches add more noise and may result in better generalization, larger batches produce more stable gradient estimates.
3) **Optimizer Selection**: Because of its momentum and adjustable learning rate characteristics, which aid in accelerating convergence and escaping local minima, the Adam optimizer was selected.
4) **Dropout Rate**: The dropout rate was fine-tuned to balance between retaining enough information and preventing overfitting. A dropout rate of 0.4 was found to be optimal for the discriminator.
5) **Number of Epochs**: The number of training epochs was determined based on the convergence of the loss functions. Training for too many epochs can lead to overfitting, while too few epochs result in underfitting.

Through meticulous hyperparameter tuning, the FashionGAN model achieved optimal performance, generating high-quality synthetic fashion images with realistic features.

IV. Experimentation and Results

A. Dataset

We used the "Fashion MNIST dataset" from Kaggle, which is a dataset of grayscale pictures of clothing items, in this study to illustrate the imbalance dataset. The Fashion MNIST dataset, a well-known benchmark in the field of computer vision, provides a robust platform for evaluating image classification and generation models. Every picture in the dataset is 28 by 28 pixels and falls into one of ten distinct categories, providing a wide variety of fashion items for comprehensive analysis. The 10 classes in the dataset are 1. Tops, 2. Trouser, 3. Pullover, 4. Dress, 5. Coat, 6. Sandal, 7. Shirt, 8. Sneaker, 9. Bag, 10. Ankle boots.

Before model training, the Fashion MNIST dataset underwent several standard preprocessing steps to ensure optimal performance and compatibility with the chosen machine learning framework. These preprocessing steps are crucial for enhancing the quality of the input data and facilitating efficient model training. The steps included:

1) **Normalization**: During training, the grayscale image pixel values were standardized to the interval [0, 1] in order to enhance numerical precision and encourage convergence. Normalizing the pixel values guarantees a consistent and efficient learning process for the model while also assisting in lowering computational complexity. To do this, divide the value of each pixel by 255.
2) **Shuffling**: In order to guarantee that the model receives random data input and to stop the model from unintentionally discovering order-related patterns in the dataset, we shuffled the training dataset. The purpose of this shuffling step is to ensure that the model does not memorize the order of the samples and that it generalizes well to new data.
3) **Batching**: The dataset was divided into fixed-size batches, which allowed us to use batch gradient descent for training. Batching not only improves the efficiency of the training process by reducing memory overhead but also speeds up convergence by enabling parallel processing. Batch gradient descent ensures that the model updates are more stable and less noisy compared to stochastic gradient descent.

B. Performance Evaluation

The performance evaluation of our proposed FashionGAN model encompasses a comprehensive analysis of various aspects, including dataset description, training procedure, evaluation metrics, results, generated images, and conclusion.

1) Dataset Description: The Fashion MNIST dataset is the basis for evaluating the performance of FashionGAN. Fashion MNIST, a well-known reference dataset in computer vision, provides grayscale representations of various fashion items such as t-shirts and pants [17]. Additionally, it contains images of dresses and shoes with gray scale values. Each image in the dataset has a dimension of 28×28 pixels, totaling 60,000 training samples along with 10,000 testing samples. The utilization of Fashion MNIST facilitates standardized evaluation and comparison with existing methods.

2) Training Procedure: The training procedure of FashionGAN revolves around adversarial training, a paradigm introduced by generative adversarial networks.

The training process involves several key steps:

1) **Data Preprocessing**: The Fashion MNIST dataset is preprocessed to normalize pixel values between 0 and 1, ensuring numerical stability during training.
2) **Model Architecture**: Deep CNNs are used in FashionGAN to build a generator and a discriminator.
3) **Loss Function**: Reducing the binary cross-entropy loss between the discriminator's predicted labels and the ground truth labels is the aim of training.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:47 UTC from IEEE Xplore. Restrictions apply.

4) **Optimization**: The Adam optimizer is employed to minimize the generator and discriminator losses separately. Adaptive learning rates provided by Adam facilitate efficient convergence during training.

TABLE II: Summary of Discriminator and Generator Models

Model	Description
Discriminator	• Input shape: (28, 28, 1) • Architecture: – RandomFlip("horizontal") – Conv2D(32, 5) – LeakyReLU(0.2) – Dropout(0.4) – Conv2D(64, 5) – LeakyReLU(0.2) – Dropout(0.4) – Conv2D(128, 5) – LeakyReLU(0.2) – Dropout(0.4) – Conv2D(256, 5) – LeakyReLU(0.2) – Dropout(0.4) – Flatten() – Dropout(0.4) – Dense(1, activation='sigmoid')
Generator	• Input shape: (128,) • Output shape: (28, 28, 1) • Architecture: – Dense(7*7*128, input_dim=128) – LeakyReLU(0.2) – Reshape((7,7,128)) – UpSampling2D() – Conv2D(128, 5, padding='same') – BatchNormalization() – LeakyReLU(0.2) – UpSampling2D() – Conv2D(128, 5, padding='same') – BatchNormalization() – LeakyReLU(0.2) – Conv2D(128, 4, padding='same') – BatchNormalization() – LeakyReLU(0.2) – Conv2D(128, 4, padding='same') – BatchNormalization() – LeakyReLU(0.2) – Conv2D(1, 4, padding='same', activation='sigmoid')

3) Evaluation Metrics: To quantify the performance of FashionGAN, we employ the following evaluation metrics:

- **Generator Loss (G Loss)**: The discrepancy between the created image and the actual data is measured by the generator loss. More realistic visuals are produced by the generator when there are less generator losses [18].
- **Discriminator Loss (D Loss)**: The discriminator loss quantifies a discriminator's capacity to discern between authentic and counterfeit photographs. High discriminative capabilities are correlated with less discriminator loss.

4) Results: FashionGAN undergoes training for 2000 epochs, with the loss curves monitored throughout the training process.

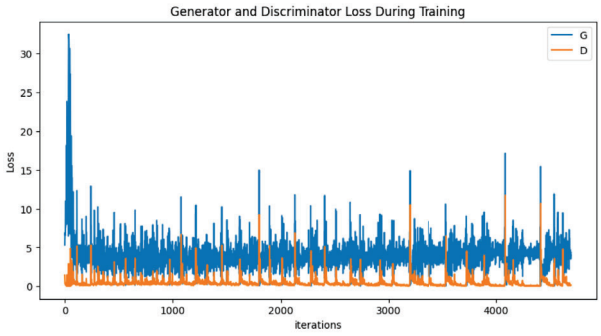


Fig. 1: Loss Curves of FashionGAN

The Figure 1, which shows the loss curves for the generator and discriminator during training, depicts the training dynamics of FashionGAN. The generator's loss is steadily declining, suggesting that it is getting better at producing realistic fashion photos each time. This is encouraging since it shows that the generator is picking up on the Fashion MNIST dataset's underlying data distribution.

On the other hand, the discriminator loss exhibits a more fluctuating pattern. Initially, there are significant oscillations, reflecting the adversarial nature of GAN training where the discriminator and generator are constantly trying to outsmart each other. However, as training progresses, the discriminator loss converges to a stable value. This stabilization is a strong indicator of successful training convergence, where both the generator and discriminator have reached a state of equilibrium.

5) Images Generated: Visual inspection of generated images provides qualitative insights into the performance of FashionGAN, highlighting its effectiveness in synthesizing high-quality fashion images. Figure 2 showcases samples of generated fashion images at various epochs.

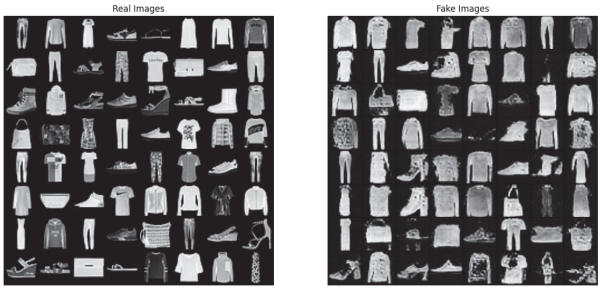


Fig. 2: Generated Fashion Images by FashionGAN vs Fake Images from fashionmnist dataset

In conclusion, FashionGAN demonstrates remarkable performance in generating realistic fashion images. Through adversarial training, it effectively learns to capture the underlying distribution of fashion items, thereby producing visually appealing and diverse images. The combination of quantitative metrics and qualitative evaluation reinforces the efficacy of FashionGAN in the domain of fashion image synthesis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:47 UTC from IEEE Xplore. Restrictions apply.

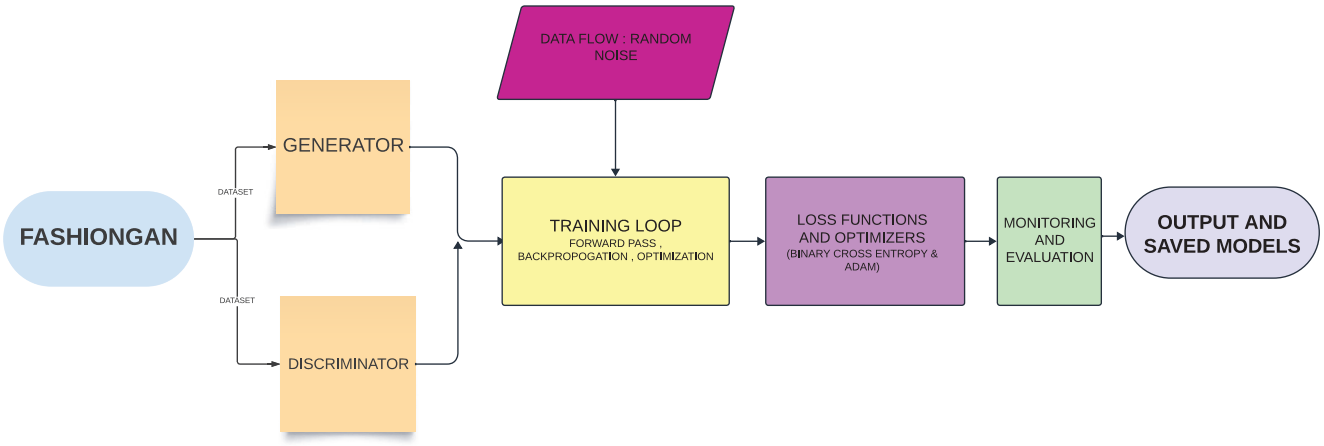


Fig. 3: FashionGAN Block Diagram

V. Conclusion and Future Work

In this study, we introduced a Generative Adversarial Network architecture tailored for fashion image generation. Through the use of TensorFlow and its vast tool library, we created a discriminator and generator model especially for the fashion MNIST dataset. Regarding the quality and variety of the photographs, our performance review shows good results.

The discriminator model, equipped with a multi-layer convolutional neural network, demonstrated robust performance in differentiating the real from synthesized images. Its architecture, consisting of many convolutional layers with batch normalization, proved effective in capturing intricate features of fashion items.

On the other hand, the generator model, employing an intricate network of dense and convolutional layers, showcased its capability in generating realistic fashion images. By utilizing techniques such as leaky ReLU activation, upsampling, and batch normalization , the generator could efficiently capture the underlying patterns and textures.

Overall, our work contributes to the advancement of generative modeling in the fashion domain and sets the stage for future exploration and innovation in this field.

References

[1] Y. R. Cui, Q. Liu, C. Y. Gao, and Z. Su, "Fashiongan: Display your fashion design using conditional generative adversarial nets," *Computer Graphics Forum*, vol. 37, no. 7, pp. 109–119, 2018.

[2] R. Dhenabayu and H. Fazlurrahman, "Potential researches of gan in fashion areas," in *2023 6th International Conference of Computer and Informatics Engineering (IC2IE)*, IEEE, 2023.

[3] D. Jain, R. Modi, Jikadra, and S. Chachra, "Text to image generation of fashion clothing," *6th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 355–358, 2019.

[4] M. Kang, J. Zhu, R. Zhang, J. Park, E. Shechtman, S. Paris, and T. Park, "Scaling up gans for text-to-image synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10124–10134, 2023.

[5] H. Yang and Y. Zhou, "Ida-gan: A novel imbalanced data augmentation gan," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 8299–8305, 2021.

[6] W. Sun *et al.*, "Pose guided fashion image synthesis using deep generative model," *arXiv preprint arXiv:1906.07251*, 2019.

[7] K. Jain, *Fashion Outfit Design Image Synthesis Using Comparative Study of Generative Adversarial Networks*. PhD thesis, National College of Ireland, Dublin, 2020.

[8] H. Yan, H. Zhang, J. Shi, J. Ma, and X. Xu, "Toward intelligent fashion design: A texture and shape disentangled generative adversarial network," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, pp. 1–23, Feb 2023.

[9] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, and F. Scarselli, "Image generation by gan and style transfer for agar plate image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 184, p. 105268, 2020.

[10] H. Yan, H. Zhang, J. Shi, J. Ma, and X. Xu, "Inspiration transfer for intelligent design: A generative adversarial network with fashion attributes disentanglement," *IEEE Transactions on Consumer Electronics*, vol. 69, pp. 1152–1163, Nov. 2023.

[11] Q. Wu, B. Zhu, B. Yong, Y. Wei, X. Jiang, R. Zhou, and Q. Zhou, "Clothgan: Generation of fashionable dunhuang clothes using generative adversarial networks," *Connection Science*, vol. 33, no. 2, pp. 341–358, 2021.

[12] M. Günel, A. Erdem, and E. Erdem, "Language guided fashion image manipulation with feature-wise transformations," in *arXiv preprint arXiv:1808.04000*, 2018.

[13] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, pp. 6999–7019, Dec. 2022.

[14] H. Yan, H. Zhang, J. Shi, J. Ma, and X. Xu, "Toward intelligent fashion design: A texture and shape disentangled generative adversarial network," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 3, pp. 1–23, 2023.

[15] A. García-Pérez, A. Ziegenbein, E. Schmidt, F. Shamsafar, A. Fernández-Valdivielso, R. Llorente-Rodríguez, and M. Weigold, "Cnn-based in situ tool wear detection: A study on model training and data augmentation in turning inserts," *Journal of Manufacturing Systems*, vol. 68, pp. 85–98, Jun 2023.

[16] J. Xu, Z. Zhang, and X. Hu, "Extracting semantic knowledge from gans with unsupervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[17] J. Xin, T. J. Yi, V. P. Yi, P. J. Yu, and Z. A. A. Salam, "Convolutional neural network for fashion images classification (Fashion-MNIST)," *Journal of Applied Technology and Innovation*, vol. 7, no. 4, p. 11, 2023.

[18] G. Iglesias, E. Talavera, and A. Díaz-Álvarez, "A survey on gans for computer vision: Recent research, analysis and taxonomy," *Computer Science Review*, vol. 48, p. 100553, 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:47 UTC from IEEE Xplore. Restrictions apply.

International Journal of Food Science +Technology

Check for updates

Original article

Bioactive profiling of two varieties of Indian legumes: adzuki and mung beans

Ritu Sharma, Devendra Kumar & Rajinder K. Gupta*

Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, 110042 Delhi, India

(Received 5 March 2024; Accepted in revised form 4 July 2024)

Summary Nutritional benefits of legumes make them an important part of sustainable diets. They are rich not only in dietary fibres, proteins, and micronutrients but also in the bioactive compounds that can exhibit various biological activities. In the present study, we aimed to investigate the nutritional analysis, phytochemical composition, and *in vitro* anti-obesity activity of two varieties of legumes indigenous to India namely, adzuki beans (*Vigna angularis*) and mung beans (*Vigna radiata*). GC–MS, UHPLC-QTOF-MS, ICP-MS analyses were done for bioactive compounds, fatty acids, macro-elements, micro-elements, protein, and non-proteinogenic amino acids. GC–MS analysis showed that the major constituents identified were hydrocarbons, fatty acids, and terpenes, while the presence of various secondary metabolites, macro- and micro-nutrients were detected using UHPLC-QTOF-MS and ICP-MS, respectively. FAME analysis revealed a higher concentration of linoleic acid in adzuki beans, which is nutritionally beneficial for human consumption. Amino acid profiling revealed the presence of essential, non-essential, and non-proteinogenic amino acids. Further, adzuki beans and mung beans were screened for cytotoxicity on 3T3-cell line and it was observed that adzuki beans were less cytotoxic as compared to mung beans. Adzuki beans showed higher HMG-Co-A reductase enzyme inhibition activity (60%) as compared to mung beans (2.79%). The results indicate that these two legumes are a very good source of bioactive metabolites and can be developed into novel cholesterol-lowering and anti-obesity functional foods.

Keywords Adzuki bean, amino acids (protein and non-protein), bioactive compounds, HMG-Co-A reductase, mung bean, obesity.

Introduction

Obesity is a metabolic disorder caused by an abnormal build-up of body fat. It can cause diabetes, cardiovascular disease, metabolic syndrome, and other comorbidities, affecting global health (Nagao & Yanagita, 2008). When an adult's body mass index (BMI) is greater than 30 kg m^{-2}, they are classified as obese or overweight. According to the reports, obesity will affect approximately 20% of the adult population by 2030 (Smith & Smith, 2016). Obesity aetiology is highly intricate and it includes economic, social, genetic, and environmental factors that interact to avoid varied degrees to induce obesity development. Until now, researchers have actively explored effective techniques to treat obesity. The anti-obesity drugs like Setmelanotide, Metreleptin, Orlistat, Phentermine/Topiramate, Naltrexone/Bupropion, Liraglutide, and Semaglutide control body weight by reducing energy absorption, resulting in weight loss (Müller *et al.*, 2022; Chakhtoura *et al.*, 2023). According to WHO, some of these anti-obesity drugs are banned due to adverse negative side effects like diarrhoea, abdominal pain, flatulence, bloating, and a decrease in the absorption of fat-soluble vitamins. Therefore, it is necessary to develop novel anti-obesity drugs with minimal side effects in a lucrative manner that would benefit from natural sources.

In the present work, we studied two varieties of beans/legumes of India, i.e., adzuki beans (*Vigna angularis*) and mung beans (*Vigna radiata*). Adzuki beans (Family: Fabaceae), an underutilised legume, are grown in over 30 countries, and in India, the cultivation is limited to areas like the north-eastern and northern hill zones. It is widely consumed in China and Japan as a dessert (sekihan) (Hirota & Takahama, 2017). In recent years, adzuki beans have received a lot of attention due to their high nutritional content and distinctive phytochemicals, which can aid in treating various metabolic disorders like diabetes, obesity, fatty liver diseases, and hyperlipidaemia. According to Lee *et al.* (2019), the Chinese variant of adzuki bean water extract can aid in

*Correspondent: E-mail: rkg67ap@dtu.ac.in, rkg67ap@yahoo.com

doi:10.1111/ijfs.17358

© 2024 Institute of Food Science & Technology (IFST).

weight control by inhibiting fat accumulation in 3T3-L1 adipocytes. It may be hypothesised that since the Chinese variant exhibits anti-obesity properties, it may be of interest to explore the anti-obesity properties of the Indian variant as well. Similarly, mung beans (Family: Fabaceae), also known as green gram, are grown in India, Sri Lanka, China, and Africa. It is an excellent source of protein, vitamins, and amino acids. It is a traditional soy food that has been utilised for more than 2000 years as a dietary supplement and an herbal remedy. But although several reports briefly discuss the anti-obesity properties of mung beans, in-depth studies of the same are not widely available.

On the basis of an extensive literature review, it was inferred that several members of the legume bean family exhibit anti-obesity properties. It has been found that regular consumption of legumes can help prevent some chronic diseases (Tor-Roca *et al.*, 2020). Further, legumes are widely available globally and can be locally sourced easily, but their potential for the development of nutraceuticals has not been fully explored. Hence, the current study aims to provide new insights into the phytochemical composition of the selected legumes and their anti-obesity activity. Firstly, we evaluated the bioactive compounds in the selected legumes, namely adzuki beans and mung beans, grown in India, by GC–MS. Then, the metabolomic technique, Ultra-high-performance liquid chromatography quadrupole time-of-flight mass spectrometry (UHPLC-QTOF-MS), was used to determine and quantify the secondary metabolites present in the selected legumes with high sensitivity and resolution. Elemental analysis (ICP-MS), fatty-acid composition (FAME), and amino acid analysis were also done to examine the nutritional profile of these selected legumes. Additionally, we assessed the *in-vitro* anti-obesity potential of adzuki beans and mung beans.

Materials and methods

Plant materials and reagents

Adzuki beans (Ab) were purchased from Himjoli Products, Delhi, India. Mung beans (Mg) were purchased from Patanjali store, Delhi, India. The selected legumes were authenticated by CSIR-NIScPR, Raw Materials Herbarium, and Museum, Delhi (RHMD), India. All the reagents and chemicals were of analytical grade and purchased from Sigma-Aldrich, St Louis, MO, USA (Table 1).

Preparation of extracts

Adzuki beans and mung beans were coarsely grounded for extraction, and 50 g of each sample was extracted using different polarity solvents in a Soxhlet apparatus.

Table 1 Nutritional composition analysis of the selected legumes on dry basis

Parameter	Adzuki beans	Mung beans
Nutritional profile		
Crude protein (g/100 g)	18.54	23.67
Dietary fibre (g/100 g)	16.41	29.44
Carbohydrate (g/100 g)	61.91	60.95
Fat (g/100 g)	2.56	0.56
Saturated fat (%)	0.91	BLQ
MUFA (%)	BLQ	BLQ
PUFA (%)	1.46	0.24
Trans fat (%)	BLQ	BLQ
Mineral profile (ppm)		
Na	19.6204 ± 0.09	24.93173 ± 0.04
Mg	3306.419 ± 0.02	2286.16 ± 0.09
Al	11.9918 ± 0.021	13.2238 ± 0.07
K	15883.36 ± 0.03	16063.99 ± 0.12
^{43}Ca	448.0509 ± 0.04	190.9574 ± 0.09
^{44}Ca	750.3771 ± 0.01	315.3264 ± 0.06
Cr	0.1021 ± 0.007	0.1211 ± 0.08
Mn	28.5288 ± 0.1	15.5899 ± 0.07
Fe	38.2998 ± 0.09	14.3856 ± 0.01
Co	0.0807 ± 0.08	0.0623 ± 0.05
Ni	1.2351 ± 0.01	2.6272 ± 0.07
Cu	7.7803 ± 0.03	11.9612 ± 0.16
Zn	33.0152 ± 0.08	25.4955 ± 0.21
As	0.0053 ± 0.01	0.0138 ± 0.15
Se	0.1797 ± 0.03	0.4138 ± 0.06
Rb	26.1909 ± 0.06	29.1123 ± 0.05s
Sr	3.8409 ± 0.03	6.0270 ± 0.03
Mo	4832.165 ± 0.01	12240.99 ± 0.01
Cd	0.0818 ± 0.02	0.1900 ± 0.09
Pb	0.0169 ± 0.12	0.0180 ± 0.09
U	0.0115 ± 0.21	0.0123 ± 0.17
Fatty acid composition (Area %, g/100 g oil sample)		
Lauric acid	0.08	0.11
Caproic acid	0.00	0.02
Myristic acid	0.33	0.61
Pentadecanoic acid	0.19	0.18
Palmitic acid	27.20	27.19
Palmitoleic acid	0.18	0.13
Heptadecanoic acid	0.32	0.27
Cis-10-pentadecanoic acid	0.09	0.00
Stearic acid	4.98	6.20
Oleic acid	7.01	17.88
Linoleic acid	36.80	27.00
Arachidic acid	0.80	0.00
ϒ- linolenic acid	0.10	12.73
Linolenic acid	19.50	3.28
Cis-11-Eicosenoic acid	0.18	1.60
Heniecosanoic acid	0.11	0.00
Cis-11-14-Eicosadienoic	0.00	0.08
Cis-8,11,14-Eicosatrienoic	0.06	0.00
Behenic acid	1.06	0.00
Erucic acid	0.00	1.42
Methyl cis-5,8,11,14	0.50	0.00
Cis-5,8,11,14,17-eicosanoic	0.17	0.00
Lignoceric acid	0.34	0.36
Nervonic acid	0.00	0.94

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1 (*Continued*)

Parameter	Adzuki beans	Mung beans
Amino acid profile (g/100 g dry sample)		
Essential amino acids (g/100 g dry sample)		
Arginine	1.432 ± 0.012	0.863 ± 0.008
Histidine	0.414 ± 0.001	0.371 ± 0.002
Isoleucine (Ile)	0.876 ± 0.002	0.920 ± 0.019
Leucine (Lue)	1.420 ± 0.006	1.519 ± 0.021
Lysine (Lys)	1.033 ± 0.011	1.025 ± 0.015
Methionine (Met)	0.197 ± 0.001	0.110 ± 0.001
Phenylalanine (Phe)	1.061 ± 0.013	1.182 ± 0.012
Threonine (Thr)	0.774 ± 0.001	0.715 ± 0.004
Tryptophan (Trp)	2.210 ± 0.022	0.069 ± 0.00
Valine (Val)	1.009 ± 0.01	1.070 ± 0.009
Non-essential amino acids (g/100 g dry sample)		
Alanine (Ala)	0.785 ± 0.002	0.799 ± 0.005
Aspartate (Asp)	2.210 ± 0.027	2.266 ± 0.028
Cysteine (Cys)	1.140 ± 0.018	0.084 ± 0.00
Glutamic Acid (Glu)	2.928 ± 0.025	3.342 ± 0.106
Glycine (Gly)	0.797 ± 0.005	0.754 ± 0.014
Proline (Pro)	0.943 ± 0.008	0.875 ± 0.013
Serine (Ser)	0.977 ± 0.007	1.022 ± 0.010
Tyrosine (Tyr)	0.512 ± 0.003	0.375 ± 0.001
Non-proteinogenic amino acids (g/100 g dry sample)		
Phosphoserine (p- Ser)	0.069 ± 0.001	0.021 ± 0.002
Taurine (Tau)	0.047 ± 0.002	0.026 ± 0.003
Phospho ethanol amine (PEA)	0.002 ± 0.00	ND
α Amino adipic acid (a- AAA)	ND	0.054 ± 0.006
β-Alanine (b-Ala)	0.064 ± 0.001	0.045 ± 0.002
Ethanol amine	0.013 ± 0.001	0.008 ± 0.00
Ornithine (Orn)	0.003 ± 0.00	0.004 ± 0.00
Urea	0.145 ± 0.009	ND
1-Methylhistidine (1 Mehis)	ND	0.089 ± 0.002
Hydroxyproline (Hypro)	0.839 ± 0.009	0.0790 ± 0.001

BLQ, below detection limit; ND, not detected.

To ensure complete extraction process, exhaustive extraction was applied with each solvent for 10–12 h (Abdel-Aal *et al*., 2015). The extracts were concentrated at 40 °C under reduced pressure using a rotary evaporator. The concentrated extracts were labelled and stored at 4 °C until further use (Table 2).

Proximate analysis

Proximate analysis was carried out according to the procedure of Association of Official Analytical Chemist (AOAC) in order to determine dietary fibre, protein, fat, and carbohydrate content in the selected legumes (Baur & Ensminger, 1977; Table 3).

Phytochemical analysis

The qualitative analysis of different extracts of both legumes was done following the standard method (Bankole *et al*., 2016). Various chemical tests such as Dragendorff's test and Mayer's test (for alkaloids), Shinoda's test (for flavonoids), Ferric chloride test (for tannins), Lead acetate test (for phenols), Killer killiani test (for glycosides), terpenoids test, and Liebermann-Burchard test (for steroids) were performed (Table 4).

Determination of elemental composition by ICP-MS

0.25 g of each sample was diluted to 10 mL nitric acid and then analysed by ICP-MS. The same procedure was used to analyse NIST standards. The blank sample solution was also prepared following the above procedure without adding the sample. The instrument calibration was assessed by analysing three certified samples from NIST (Ibrahim *et al*., 2019; Table 5).

Fatty acid profiling

Fatty acid methyl ester (FAME) analysis was done to evaluate the fatty acid content as the amount of methyl esters present in the samples using the method described by Ryan *et al*. (2007), with slight modifications. 0.15 g of the extracted oil sample was accurately weighed in a 5 mL glass vial. 200 µL of 2 N methanolic KOH solution was then added and vortexed for 2 min. The mixture was heated in a water bath shaker at 55 °C for 10 min. After the mixture was cooled, 1.0 mL of 5% HCl solution was added and then heated for 10 min in a water bath shaker at 70 °C for 10 min. 2 mL of petroleum ether was added, and the mixture was vortexed for 2 min. The upper layer was separated and analysed by GC–MS (Shimadzu, GC-2010, Kyoto, Japan) (Table 6).

Gas-chromatography-mass Spectrum analysis

GC–MS (Shimadzu GCMS-QP2010, Kyoto, Japan) was used to determine the volatile compounds in the different crude extracts of both legumes. Samples were injected in GC–MS under the following conditions – helium gas was used as carrier gas. The samples were introduced in split mode at 260 °C. The column flow rate was 1.23 mL min^{-1} in electron ionisation (EI) mode. The ion source temperature was 220 °C, and the interface temperature was 270 °C. The oven temperature was planned for 60 °C (2 min) to 280 °C (26 min), and the solvent delay time was 3.50 min. NIST (National Institute of Standards and Technology, version 1.10 beta, Shimadzu) mass spectral database was used to identify the separated peaks (Grover & Patni, 2013; Table 7).

UHPLC-QTOF-MS identification

UHPLC-QTOF-MS was used to identify secondary metabolites in different crude extracts of the selected legumes, using a previously described method with some modifications (Hong *et al*., 2021). The mass spectrometric

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2 Qualitative phytochemical analysis of the selected underutilised legumes

	Phytochemicals					
Sample Code	**Alkaloid**	**Reducing sugar**	**Flavonoid**	**Tannin**	**Phenol**	**Saponin**
Ab-H	+	–	–	–	–	–
Ab-D	+	–	–	–	–	–
Ab-M	–	–	–	–	–	–
Mg-H	+	–	–	–	–	–
Mg-D	+	–	–	–	–	–
Mg-M	–	+	–	–	–	–

+: presence of phytochemical; –: absence of phytochemical.
D, DCM extract; H, n-hexane extract; M, Methanol extract.

analysis was conducted in positive mode (ESI+). ESI source conditions were as follows: source temperature, 120 °C; desolvation gas flow, 950 L h^{-1}; cone gas flow, 50 L h^{-1}; capillary voltage, 3.22 keV. HPLC (SYNAPT-XS HDMS, Waters, Manchester, UK) fitted with a controller, AD pump, degasser, AD auto sampler, and AD column, coupled with a quadrupole time-of-flight mass spectrometer (QTOF-MS), was used for UHPLC and mass spectrometric analysis. Briefly, the extracts were mixed with 1% formic acid (10 mL) in water and kept for 10 min. 10 mL methanol and 10 mL acetonitrile were added and vortex for 1 min and centrifuged at 2516 ***g*** for 5 min. The supernatant was diluted with acidified water. Then the extract was injected into the instrument for the analysis. The chromatographic separation of the samples was done by 100 mm × 2.1 mm column C18 (Acquity BEH 2.1, Waters). The injection volume was 5 µL. Secondary metabolites were eluted using a binary mobile phase at a flow rate of 0.2 mL min^{-1}, where solvent A was LC–MS grade water containing 1% formic acid and solvent B was 1% formic acid with acetonitrile. The data acquisition and processing were performed using ChemSpider software (RSC, Washington DC, USA).

Amino acid analysis

The amino acid composition of the selected samples was measured using Automatic Amino Acid Analyser L-8900 (Hitachi Co. Ltd., Tokyo, Japan) as per the method described by Chakrabarti *et al.* (2018). Dried and finely powdered samples were hydrolysed with 6 N HCl at 110 °C for 22 h. The hydrolysed samples were dried in Nitrogen Evaporator (PCi Analytic Private Limited, Maharashtra, India). 0.02 N HCl was added in each sample and then kept in the Auto sampler. 20 µL sample injection was used. Tryptophan, cysteine, and methionine were destroyed by acid hydrolysis; these amino acids were treated with specific reagents. Cysteine and methionine were oxidised with performic acid and treated with 48% hydrobromic acid. Both the samples were hydrolysed with 4 N methane sulfonic acid and 3-(2-aminoethyl) indole for the estimation of tryptophan. The rest of the methods were same for all the amino acids. The ninhydrin derivative of hydroxyproline and proline was observed at 440 nm, and all other amino acids were observed at 570 nm. For the quantification of the contents of amino acids in the legumes, the peak areas of the detected amino acids were compared with those of authentic standards provided with the equipment. Amino Acids Mixture Standard Solutions, Type B and Type AN-2 (Wako Pure Chemical Industries, Limited, Osaka, Japan) were used. Standard solutions for tryptophan and glutamine (Sigma-Aldrich) were prepared before analysis.

Anti-obesity activity

Neutral red uptake assay (NRU assay) for cytotoxicity

Cytotoxicity of methanol extract of Ab and Mg sample on 3T3-L1 cell line (Procured from NCCS Pune) was determined by NRU Assay. The 3T3-L1 cells (5000–8000 cells per well) were cultured in 96 well plates for 24 h in Dulbecco's modified eagle medium (DMEM-AT149-1L) supplemented with 10% fetal bovine serum (FBS- HIMEDIA-RM 10432) and 1% antibiotic solution at 37 °C with 5% CO_2. After 24 h, the medium was removed, and fresh culture medium was added to each well of the plate. 5 µL of treatment dilutions (1, 10, 50, 100, 250, and 500 µg mL^{-1}) were added to the defined wells and incubated for 24 h. 100 µL of NRU (SRL Chem-36 248) (40 µg mL^{-1} in PBS – phosphate-buffered saline) was added to the wells and incubated (Heal Force-Smartcell CO_2 Incubator-Hf-90) for 1 h. After that medium was removed, NRU was dissolved in 100 µL of NRU Destain solution. Plates were read at 550/660 nm using an Elisa Plate Reader (iMark BioRad, Hercules, CA, USA). IC_{50} values of both samples were calculated by using the software Graph Pad Prism-6 (Vajrabhaya & Korsuwannawong, 2016).

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3 Bioactive compounds identified by GC–MS in the different solvent extracts of Adzuki beans

RT	Compound name	%Area	MW	MF	n-Hexane	DCM	Methanol
5.341	2,4-Dihydroxy-2,5-dimethyl-3(2h)-furan-3-one	1.46	144	$C_6H_8O_4$	–	–	+
5.400	Decane	0.19	142	$C_{10}H_{22}$	+	–	–
6.345	N-alpha, n-omega-di-cbz-l-arginine	1.05	442	$C_{22}H_{26}N_4O_6$	+	–	–
6.508	Butyl acetoxyacetate	0.69	174	$C_8H_{14}O_4$	–	–	+
6.625	Tricyclo(5.2.1.0(2,6))dec-4-ene	0.14	134	$C_{10}H_{14}$	+	–	–
7.058	Endo-tricyclo[5.2.1.0(2.6)]decane	0.10	136	$C_{10}H_{16}$	+	–	–
7.422	3-(2-Aminoethyl)-4-hydroxy-4-methyl-5- oxazolidin-2-one	2.26	228	$C_{11}H_{20}N_2O_3$	–	–	+
7.630	Maltol	0.20	126	$C_6H_6O_3$	–	+	–
7.859	2-Butyl-1-octanol	0.03	186	$C_{12}H_{26}O$	+	–	–
7.871	Benzyl alcohol, tms derivative	0.02	180	$C_{10}H_{16}OSi$	–	+	–
7.933	5-Butylnonane	0.02	184	$C_{13}H_{28}$	+	–	–
8.013	N-Hexadecane	0.02	226	$C_{16}H_{34}$	+	–	–
8.113	3-Methylundecane	0.04	170	$C_{12}H_{26}$	+	–	–
8.159	3-Hydroxy-2,3-dihydromaltol	1.28	144	$C_6H_8O_4$	–	–	+
8.570	Dodecane	0.43	170	$C_{12}H_{26}$	+	+	–
8.700	Phloroglucinol, trimethylsilyl ether	3.61	198	$C_9H_{14}O_3Si$	–	–	+
8.748	6-Methyltridecane	0.05	198	$C_{14}H_{30}$	+	–	–
9.348	1-Chloroctadecan	0.05	288	$C_{18}H_{37}Cl$	+	–	–
9.608	2,7,10-Trimethyldodecane	0.07	212	$C_{15}H_{32}$	+	–	–
9.674	5-Hydroxymethylfurfural	3.96	126	$C_6H_6O_3$	–	–	+
10.032	Tridecane	0.06	184	$C_{13}H_{28}$	+	–	–
10.245	(E)-1-Trimethylsilyl-2-nonene	0.78	198	$C_{12}H_{26}Si$	–	–	+
10.301	Trimethylsilyl 3-methyl-2-furoate	13.85	198	$C_9H_{14}O_3Si$	–	–	+
11.001	2-Bromo dodecane	0.04	248	$C_{12}H_{25}Br$	+	–	–
11.053	Docosane	0.06	310	$C_{22}H_{46}$	+	–	–
11.087	Trimethylsilyl 2-acetoxyacetate	0.09	190	$C_7H_{14}O_4Si$	–	–	+
11.406	Tetradecane	0.68	198	$C_{14}H_{30}$	+	–	–
11.669	3-Methyl-2-ketobutyric acid tbdms	0.54	230	$C_{11}H_{22}O_3Si$	–	–	+
12.104	Tetradecamethylcycloheptasiloxane	0.33	518	$C_{14}H_{42}O_7Si_7$	–	–	+
12.534	1-[(trimethylsilyl)oxy]- cyclohexene	0.09	170	$C_9H_{18}OSi$	–	–	+
12.698	Pentadecane	0.10	212	$C_{15}H_{32}$	+	+	–
12.772	3 tms derivative of 2-deoxyuridine	0.17	444	$C_{18}H_{36}N_2O_5Si_3$	–	–	+
13.916	N-Hexadecane	0.49	226	$C_{16}H_{34}$	+	–	–
15.068	Heptadecane	0.03	240	$C_{17}H_{36}$	+	–	–
16.164	Octadecane	0.35	254	$C_{18}H_{38}$	+	+	–
16.410	Isopropyl myristate	0.08	270	$C_{17}H_{34}O_2$	–	–	+
16.445	Tetradecamethylhexasiloxane	0.10	458	$C_{14}H_{42}O_5Si_6$	+	–	–
16.552	Neophytadiene	0.06	278	$C_{20}H_{38}$	–	+	–
16.641	6,10,14-Trimethylpentade can-2-one	0.16	268	$C_{18}H_{36}O$	+	–	–
16.644	Trimethylsilyl myristate	0.27	300	$C_{17}H_{36}O_2Si$	–	+	–
17.490	Methyl hexadecanoate	3.65	270	$C_{17}H_{34}O_2$	+	+	+
17.649	Trimethylsilyl pentadecanoate	0.21	314	$C_{18}H_{38}O_2Si$	–	+	–
17.878	Dibutyl phthalate	0.56	278	$C_{16}H_{22}O_4$	–	+	–
18.187	Palmitic acid	14.53	256	$C_{16}H_{32}O_2$	+	+	+
18.528	Palmitic acid, tms derivative	0.18	328	$C_{19}H_{40}O_2Si$	–	–	+
18.946	Allyl stearate	0.17	324	$C_{21}H_{40}O_2$	–	–	+
19.135	Linoleic acid, methyl ester	1.68	294	$C_{19}H_{34}O_2$	+	+	+
19.180	Methyl linolenate	0.73	292	$C_{19}H_{32}O_2$	+	–	+
19.247	Methyl 9,10-dibromostearate	0.10	454	$C_{19}H_{36}Br_2O_2$	+	–	–
19.426	Methyl stearate	0.18	298	$C_{19}H_{38}O_2$	+	–	+
19.596	Palmidrol	1.25	299	$C_{18}H_{37}NO_2$	–	+	+
19.778	Z-7-tetradecenal	21.55	210	$C_{14}H_{26}O$	–	–	+
19.868	2- Azidocholestan-3-ol	23.15	429	$C_{27}H_{47}N_3O$	+	+	–
20.038	Stearic acid	1.79	284	$C_{18}H_{36}O_2$	+	+	+
20.520	Tributyl acetylcitrate	0.36	402	$C_{20}H_{34}O_8$	+	–	–
20.755	Allyl stearate	0.13	324	$C_{21}H_{40}O_2$	–	–	+
20.852	3-Cyclopentylpropionic acid, 2-dimethylaminoethyl ester	0.25	213	$C_{12}H_{23}NO_2$	–	–	+

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3 (*Continued*)

RT	Compound name	%Area	MW	MF	n-Hexane	DCM	Methanol
20.954	Glycidyl palmitate	0.26	312	$C_{19}H_{36}O_3$	–	–	+
21.159	Caprylic acid monoethanol amide	0.17	187	$C_{10}H_{21}NO_2$	+	+	+
21.215	2-Dodecylcyclobutanone	0.10	238	$C_{16}H_{30}O$	+	–	–
21.220	1-(1-Adamantyl)-3-(1-piperidinyl)-1-propanone	0.87	275	$C_{18}H_{29}NO$	–	+	–
21.394	Lauric acid ethanolamide	0.24	243	$C_{14}H_{29}NO_2$	–	+	+
21.417	Palmidrol	0.06	299	$C_{18}H_{37}NO_2$	+	–	–
21.636	8-Methyl-6-nonenamide	0.23	169	$C_{10}H_{19}NO$	–	–	+
21.780	Heneicosane	0.07	296	$C_{21}H_{44}$	+	–	–
21.846	Octadecanamide	0.07	283	$C_{18}H_{37}NO$	+	–	+
21.967	Triphenyl phosphate	0.39	326	$C_{18}H_{15}O_4P$	+	+	+
22.262	2-Formylhexadecane	0.18	254	$C_{17}H_{34}O$	–	–	+
22.409	Methyl 2-cyclohexyl-2-methylpentanoate	0.73	212	$C_{13}H_{24}O_2$	–	–	+
22.412	Monopalmitin tms	0.14	474	$C_{25}H_{54}O_4Si_2$	+	–	–
22.415	Ethyl 3-oxooctadecanoate	0.44	326	$C_{20}H_{38}O_3$	–	+	–
22.587	2-Ethylbutyric acid, decyl ester	0.14	256	$C_{16}H_{32}O_2$	–	+	–
22.588	Pentacosane	0.07	352	$C_{25}H_{52}$	+	–	–
22.628	1,2-Dipalmitin	0.76	568	$C_{35}H_{68}O_5$	–	–	+
22.903	Bis(2-ethylhexyl) phthalate	0.39	390	$C_{24}H_{38}O_4$	+	–	–
23.288	Hexadecanoic acid, phenylmethyl ester	0.69	346	$C_{23}H_{38}O_2$	+	–	–
23.361	Tetratetracontane	0.23	618	$C_{44}H_{90}$	+	–	–
23.366	Terephthalic acid, but-3-enyl heptadecyl ester	0.21	458	$C_{29}H_{46}O_4$	–	–	+
23.369	1,22-Dibromodocosane	0.07	466	$C_{22}H_{44}Br_2$	–	+	–
23.407	Docosyl docosanoate	0.19	648	$C_{44}H_{88}O_2$	–	+	–
23.990	4-Cyanobenzoic acid, undec-10-enyl ester	0.37	299	$C_{19}H_{25}NO_2$	–	–	+
24.196	1-Glyceryl stearate	0.34	358	$C_{21}H_{42}O_4$	–	–	+
24.429	Trimethylsilyl derivative of 2-monoolein	0.24	500	$C_{27}H_{56}O_4Si_2$	–	–	+
24.520	Dioctyl terephthalate	0.11	390	$C_{24}H_{38}O_4$	+	+	–
24.625	2-[(Trimethylsilyl)oxy] ethyl palmitate	0.42	372	$C_{21}H_{44}O_3Si$	–	–	+
24.665	Benzyl linoleate	0.99	370	$C_{25}H_{38}O_2$	+	–	–
24.741	Benzyl linolenate	0.44	368	$C_{25}H_{36}O_2$	+	–	–
24.786	Linalol oxide, trimethylsilyl ether	0.55	242	$C_{13}H_{26}O_2Si$	–	–	+
24.893	1-Bromotetracosane	0.15	416	$C_{24}H_{49}Br$	–	–	+
24.911	9-Octadecenamide	0.36	281	$C_{18}H_{35}NO$	+	–	–
24.933	Dioctyl sebacate	1.47	426	$C_{26}H_{50}O_4$	–	+	–
25.813	Tocopherol-delta-tms-derivative	0.41	474	$C_{30}H_{54}O_2Si$	–	+	+
25.845	2-Phenyl-1,3-dioxan-5-yl icosanoate	0.20	474	$C_{30}H_{50}O_4$	+	–	–
26.045	3,5-Cholestadiene	0.88	368	$C_{27}H_{44}$	–	–	+
26.057	Stigmasteryl acetate	3.12	454	$C_{31}H_{50}O_2$	+	+	+
26.575	8-Methyltocol	5.27	402	$C_{27}H_{46}O_2$	+	+	+
26.680	(3-beta)-Stigmast-5-en-3-ol	0.35	414	$C_{29}H_{50}O$	+	+	–
26.805	3,5-Dedihydro-stigmastan-6,22-dien	0.08	394	$C_{29}H_{46}$	+	+	–
26.898	Tocopherol-beta-tms-derivative	0.41	488	$C_{31}H_{56}O_2Si$	–	+	–
27.142	Stigmasteryl acetate	3.12	454	$C_{31}H_{50}O_2$	+	–	+
27.277	Campesterol, methyl ether	0.12	414	$C_{29}H_{50}O$	–	+	+
27.851	O-Xylotocopherol	4.87	416	$C_{28}H_{48}O_2$	+	+	+
28.310	Tetracontane	0.59	562	$C_{40}H_{82}$	+	–	–
28.445	Stigmast-5-en-3-ol, oleate	2.21	678	$C_{47}H_{82}O_2$	+	+	+
28.566	Cerotic acid	0.09	396	$C_{26}H_{52}O_2$	+	–	–
30.052	1-Nonyl-1,2,3,4-tetrahydronaphthalene	0.23	258	$C_{19}H_{30}$	+	–	–
30.848	Campestrol	0.61	400	$C_{28}H_{48}O$	+	+	–
31.430	Stigmasterol	10.09	412	$C_{29}H_{48}O$	+	+	+
32.301	Tetracosamethyl cyclododecasiloxane	0.74	888	$C_{24}H_{72}O_{12}Si_{12}$	+	–	–
32.779	Clionasterol	7.95	414	$C_{29}H_{50}O$	+	+	+
33.090	Fucosterol	0.55	412	$C_{29}H_{48}O$	+	–	–
33.184	Alpha-amyrin, trimethylsilyl ether	1.78	498	$C_{33}H_{58}OSi$	–	+	–
33.584	Olean-12-en-3- one	4.82	424	$C_{30}H_{48}O$	–	–	+
33.656	Olean-12-en-3-ol	7.93	426	$C_{30}H_{50}O$	+	+	–

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3 (*Continued*)

RT	Compound name	%Area	MW	MF	n-Hexane	DCM	Methanol
34.179	Stigmasta-4,22-dien-3-one	1.01	410	$C_{29}H_{46}O$	+	+	+
34.529	Handianol	0.15	426	$C_{30}H_{50}O$	+	–	–
34.741	Cycloartane-3beta.,25-diol	0.28	444	$C_{30}H_{52}O_2$	+	–	–
34.831	Stigmasta-3,5-dien-7-one	0.22	410	$C_{29}H_{46}O$	–	+	–
35.846	(24 s)-Stigmast-4-en-3-one	0.73	412	$C_{29}H_{48}O$	+	+	+
35.876	Stigmast-4-en-3-one	0.20	412	$C_{29}H_{48}O$	–	+	+

Inhibition of HMG-Co-A reductase enzyme activity

The enzyme inhibitory activity of HMG-Co-A reductase in methanol extract of Ab and Mg was evaluated using the HMG-Co-A reductase assay kit (Sigma-Aldrich). The enzyme experiments were performed at 37 °C as per the manufacturer's instructions. The reaction mixture (200 µL) for both samples was prepared by adding the reagents in the following order: reaction buffer; sample (10 µL); NADPH (10 µL); HMG-Co-A reductase enzyme substrate (20 µL). The samples were incubated for 1 h to complete the reaction. Finally, 100 µL of 4-(Dimethylamino) benzaldehyde reagent (DMAB) was added to each well and incubated for 5 min. After 5 min, the absorbance was recorded at 490 nm with a microplate reader (iMark Biorad). Enzyme activity was calculated against the slope value obtained from the NADPH standard curve. All assays were performed in triplicate, and all the data were presented as mean ± standard deviation (Kumar *et al.*, 2019).

Results and discussion

Proximate analysis

The nutritional values of the selected legumes are presented in Table 1. From the data, it can be observed that both the legumes are an excellent source of carbohydrate. The carbohydrate content in adzuki beans was 61.91 g/100 g, which was higher as compared to mung beans (60.95 g/100 g). However, the trend was the opposite with respect to the dietary fibre of the legumes. Mung beans contained the highest dietary fibre concentration as compared to adzuki beans. The increase in the carbohydrate content in both the legumes is probably due to the decrease in dietary fibre (Archana *et al.*, 2013). Also, the dietary fibre in food is an indication of the level of non-digestible carbohydrate. Mung beans have the highest protein content (23.67 g/100 g) than adzuki beans (18.54 g/100 g). The highest fat content was found in Adzuki beans. The fat content in both the legumes was relatively low compared to cowpea and chickpea (4.8 g/100 g, 5.2 g/100 g, respectively), as reported by Azmah *et al.* (2023). Generally, fats are relatively minor constituents in legumes and range from 0.5 to 5.5 g (Hymavathi *et al.*, 2020). The results of the nutritional composition of the legumes were in agreement with the previous reports (Jide *et al.*, 2023; Oguuniyi *et al.*, 2023; Zafar *et al.*, 2023). The previously reported data slightly differed from the present data, probably due to the difference in the cultivars of the legumes.

Phytochemical analysis

All three different extracts (n-hexane, DCM, and methanol) of the selected legumes were screened for the presence of different phytoconstituents (Table 2). The data revealed the presence of reducing sugars and alkaloids. The alkaloids were identified in the n-hexane and DCM extracts of both legumes, while reducing sugars were identified in the methanol extract of mung beans. Flavonoids, tannins, phenols, and saponins were not identified in any of the extracts. All the phytoconstituents found in legumes are responsible for various biological activities.

ICP-MS analysis

Elemental analysis was carried out for the selected legumes. A total of 21 macro-elements and micro-elements were present in these legumes at different concentrations (Table 1). The major elements found were potassium (K), magnesium (Mg), and molybdenum (Mo). Among all the analysed elements, K was found to have the highest concentration. K is an essential mineral for the efficient functioning of tissues, the body's cells, and organs. Increased potassium intake in diets can help maintain healthy blood pressure levels. The concentration of heavy metals like cadmium (Cd), arsenic (As), and lead (Pb) were below the standard limit values set by FSSAI.

FAME analysis

The percentage of fatty acid methyl esters (FAMEs) varied between the selected legumes (Table 1). Twenty-four fatty acids were detected and quantified as percentage using GC–MS analysis. In adzuki beans, the predominant constituents were linoleic acid (36.80%) and palmitic acid (27.20%), followed by

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4 Bioactive compounds identified by GC–MS in the methanol extracts of mung beans

RT	Compound name	% Area	MW	MF
Methanol extract				
5.325	Ethyl benzoate	20.75	**150**	$C_9H_{10}O_2$
5.918	Benzaldehyde diethyl acetal	0.98	180	$C_{11}H_{16}O_2$
8.338	1-Chloro-2-(diethoxymethyl) benzene	0.12	214	$C_{11}H_{15}ClO_2$
8.850	Ethyl 2-ethoxybenzoate	0.09	194	$C_{11}H_{14}O_3$
9.619	2,4-Ditert-butylphenol	0.42	206	$C_{14}H_{22}O$
10.251	Guanosine	0.21	283	$C_{10}H_{13}N_5O_5$
10.616	1-Hexadecene	0.06	224	$C_{16}H_{32}$
11.529	Alpha-Methyl-d-galactopyranoside	0.33	194	$C_7H_{14}O_6$
11.865	3-Hydroxypropyl benzoate	0.19	180	$C_{10}H_{12}O_3$
12.563	Tetradecanoic acid	0.17	228	$C_{14}H_{28}O_2$
12.883	1-Heptadecene	0.03	238	$C_{17}H_{34}$
13.620	Phthalic acid, butyl undecyl ester	0.09	376	$C_{23}H_{36}O_4$
14.274	Methyl hexadecanoate	1.34	270	$C_{17}H_{34}O_2$
14.358	Methyl 3-(3,5-ditert-butyl-4-hydroxyphenyl) propanoate	0.74	292	$C_{18}H_{28}O_3$
15.045	Palmitic acid	13.63	256	$C_{16}H_{32}O_2$
15.478	Methyl 2-hydroxyhexadecanoate	0.27	286	$C_{17}H_{34}O_3$
15.699	Heptadecanoic acid	0.17	270	$C_{17}H_{34}O_2$
15.941	Cis-Linoleic acid methyl ester	1.84	294	$C_{19}H_{34}O_2$
15.995	Methyl linolenate	0.83	292	$C_{19}H_{32}O_2$
16.215	Methyl stearate	0.69	298	$C_{19}H_{38}O_2$
16.757	Linolic acid	17.67	280	$C_{18}H_{32}O_2$
16.895	Stearic acid	2.33	284	$C_{18}H_{36}O_2$
17.713	Carbonic acid, 2-dimethylaminoethyl neopentyl ester	2.03	203	$C_{10}H_{21}NO_3$
18.214	Lauric acid ethanolamide	0.26	243	$C_{14}H_{29}NO_2$
18.391	Eicosanoic acid	0.33	312	$C_{20}H_{40}O_2$
19.175	3-Cyclopentylpropionic acid, 2-dimethylaminoethyl ester	0.51	213	$C_{12}H_{23}ONO_2$
19.313	Fumaric acid, 2-dimethylaminoethyl dodecyl ester	0.33	355	$C_{20}H_{37}NO_4$
19.355	Diglycol dibenzoate	0.57	314	$C_{18}H_{18}O_5$
19.645	2-Monopalmitin	7.18	330	$C_{19}H_{38}O_4$
19.957	Docosanoic acid	0.11	340	$C_{22}H_{44}O_2$
20.387	Methyl tricosanoate	0.10	368	$C_{24}H_{48}O_2$
21.055	Glycerol 1-monolinolate	6.05	354	$C_{21}H_{38}O_4$
21.182	Glycerin 1-stearate	0.83	358	$C_{21}H_{42}O_4$
21.294	Dioctyl terephthalate	0.28	390	$C_{24}H_{38}O_4$
22.764	Delta-tocopherol	0.32	402	$C_{27}H_{46}O_2$
23.694	Gamma-tocopherol	2.55	416	$C_{28}H_{48}O_2$
24.470	Cholesterol	0.27	386	$C_{27}H_{46}O$
25.683	Ergost-5-en-3-ol	1.29	400	$C_{28}H_{48}O$
26.066	Stigmasterol	3.05	412	$C_{29}H_{48}O$
26.971	Gamma-sitosterol	5.43	414	$C_{29}H_{50}O$
27.969	Cycloartenol	0.13	426	$C_{30}H_{50}O$

oleic acid (7.01–34.09%), linolenic acid (19.50%), accompanied by a very small amount (0.09–7.01%) of others. In mung beans, the major fatty acids were palmitic acid (27.19%), linoleic acid (27.00%), oleic acid (17.88%), and ϒ- linolenic acid (12.73%). Adzuki bean had higher linoleic acid and linolenic acid as compared to mung bean but mung bean had higher oleic acid. Among all, linoleic acid was the predominant component and was present in the highest quantity in adzuki beans (36.80%). Linoleic acid (ω-6 polyunsaturated fatty acid, PUFA) is an essential fatty acid, which has various biological effects on human body cells, such as flexibility, cell membrane, lower cholesterol levels, reduced cardiovascular disease, inflammation regulation, and immune system function. It is the most abundant PUFA in the legumes, accounting for about 90% of dietary ω-6 PUFA consumption, a higher intake of which is associated with a 15% lower risk of coronary heart disease incidence (Azmah *et al*., 2023). According to Okereke & Banigo, the nutritional value of legumes can be significantly improved by the presence of a single essential fatty acid, which can be beneficial for human health (Obioma Okereke & Banigo, 2021).

GC–MS analysis

The phytochemicals present in the different extracts of adzuki beans and mung beans by GC–MS analysis are summarised in Tables 3 and 4. The data revealed that adzuki beans and mung beans extracts encompassed a total of 120 and 49 compounds, respectively. In Adzuki beans extracts, the major bioactive constituents were 2-azidocholestan-3-ol (23.15%), z-7-tetradecenal (21.55%), palmitic acid (14.53%), trimethylsilyl-3-methyl-2-furoate (13.85%), and stigmasterol (10.09%). The previous study revealed that adzuki beans contained the highest concentration of aliphatic aldehydes and alcohols (Bi *et al*., 2021). The previously reported data differed from that described here, probably because Bi *et al*. used a Chinese variant of adzuki beans. According to them, only two hydrocarbons were detected in adzuki beans, but in the present study, more than two hydrocarbons were detected. In the methanol extract of mung beans, ethyl benzoate (20.75%), palmitic acid (13.63%), and linoleic acid (17.67%) were identified as the chief phytoconstituents. The data was comparable with the previous report (Ullah Khan *et al*., 2017). It was notable that palmitic acid, dodecane, stearic acid, tetracontane, stigmasterol, δ-tocopherol, and γ-tocopherol were the common compounds detected in both the sample's extracts. Palmitic acid and stearic acid are the saturated fatty acids that are consumed most frequently in the Western diet (van Rooijen & Mensink, 2020). They aid in regulating glucose metabolism and developing insulin resistance in type 2 diabetes (Carta *et al*., 2017). The phytosterols like stigmasterol, γ-sitosterol, and β-sitosterol are plant sterols that are

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 5 Quantification of secondary metabolites using UHPLC-QTOF-MS in adzuki beans

Peak no.	Tentative metabolites	RT (min)	MF	MM	$[M-H]^+$	Error (ppm)	Compound ID
n-hexane extract							
1.	Stigmasterol	20.6780	$C_{29}H_{48}O$	412.3705	413.3789	2.7	CSID4444352
2.	Ergostan-3-ol	27.7371	$C_{28}H_{50}O$	402.696	403.3596	2.8	CSID532070
3.	(20S)-Protopanaxadiol	27.7371	$C_{30}H_{52}O_3$	460.3897	461.3975	−2.7	CSID9388412
4.	22β-Hydroxycholesterol	27.074167	$C_{27}H_{46}O_2$	402.653	403.3596	6.1	CSID146693
5.	γ-Butyrobetain	29.6280	$C_7H_{15}NO_2$	145.9765	145.9843	9.7	CSID705
6.	Lecithin	29.6009	$C_{44}H_{78}NO_8P$	783.5779	784.5857	0.7	CSID24766800
DCM extract							
7	Stigmasterol	28.024983	$C_{29}H_{48}O$	412.3709	413.3787	2.3	CSID4444352
8	1-Linoleoyl-2-α-linolenoyl-sn-glycero-3-phosphocholine	28.965117	$C_{44}H_{78}NO_8P$	779.5478	780.5556	2.3	CSID24766802
Methanol extract							
9	Nystose	1.040	$C_{24}H_{42}O_{21}$	666.0902	667.0980	5.1	CSID145907
10	2,3-Dimethoxy-5-geranyl-6-methyl-1,4-benzoquinone	11.881	$C_{19}H_{26}O_4$	318.1832	319.1910	2.2	CSID4444053
11	Dopaxanthin quinone	18.413	$C_{18}H_{16}N_2O_8$	388.0867	389.0945	−8.6	CSID30791578
12	Hexadecenal	20.195	$C_{16}H_{30}O$	238.2035	239.2113	3.4	CSID4444172
13	Poriferasta-5,7-dien-3beta-ol	26.743	$C_{29}H_{48}O$	412.3713	413.3791	3.0	CSID13183578
14	Tyramine	28.427	$C_8H_{11}NO$	137.0849	138.0927	9.7	CSID5408
15	1-Linoleoyl-2-α-linolenoyl-sn-glycero-3-phosphocholine	28.965	$C_{44}H_{78}NO_8P$	779.5470	780.5548	1.1	CSID24766802

Table 6 Quantification of secondary metabolites using UHPLC-QTOF-MS in mung beans

Peak no.	Tentative metabolites	RT (min)	MF	MM	$[M-H]^+$	Error (ppm)	Compound ID
Methanol extract							
1.	1,5-Anhydro-D-fructose	2.39435	$C_6H_{10}O$	139.1513	140.1591	2.2	CSID112421
2.	Baycovin	2.39435	$C_6H_{10}O$	139.1513	140.1591	2.2	CSID2943
3.	L-Rhamno-1,4-lactone	2.39435	$C_6H_{10}O$	139.1513	140.1591	2.2	CSID4573835
4.	3,6-Anhydro-β-D-mannopyranose	2.39435	$C_6H_{10}O$	139.1513	140.1591	2.2	CSID58830027
5.	2-Methylbutannitril	19.98815	C_5H_9N	82.1242	83.132	7.9	CSID27284
6.	(2S)-2-Methylbutannitril	19.98815	C_5H_9N	82.1242	83.132	7.9	CSID557380
7.	Piperidin	19.98815	$C_5H_{11}N$	84.1392	85.147	6.1	CSID7791
8.	2-Methyl-2,5-cyclohexadien-1,4-diol	19.98815	$C_7H_{10}O_2$	125.1452	126.153	2.3	CSID58829837
9.	(5β,8α,9β,10α)-Kaurane-17,18-dioic acid	25.54818	$C_{20}H_{30}O_4$	271.1186	272.1264	−5.3	CSID24703318
10.	13-apo-β-carotenone	25.54818	$C_{18}H_{26}O$	218.0922	219.1000	5.3	CSID4516011
11.	p-cymene	25.54818	$C_{10}H_{14}$	110.1120	111.1198	7.4	CSID7183
12.	Butyl propanoate	25.54818	$C_7H_{14}O_2$	129.1772	130.185	5.7	CSID11045
13.	9,10-DiHODE	25.54818	$C_{18}H_{32}O_4$	311.1362	312.444	2.3	CSID17220749
14.	(4Z)-4-[(3S)-3-Hydroxybutylidene]-3,5,5-trimethyl-2-cyclohexen-1-one	25.54818	$C_{13}H_{20}O_2$	207.2892	208.297	4.6	CSID30825341
15.	g-Butyrobetaine	27.38128	$C_7H_{15}NO_2$	106.1007	107.1085	2.1	CSID705
16.	L-γ-Palmitoyl-α-lysolecithin	30.17642	$C_{24}H_{50}NO_7P$	472.3427	473.3505	2.6	CSID405287
17.	Oleic acid	30.17642	$C_{18}H_{34}O_2$	281.4532	282.461	6.2	CSID393217
18.	(6Z)-6-Octadecenoic acid	30.17642	$C_{18}H_{34}O_2$	281.4532	282.461	6.2	CSID4444569
19.	cis-Vaccenic acid	30.17642	$C_{18}H_{34}O_2$	281.4532	282.461	6.2	CSID4445888

known to be beneficial for human health. According to the research, phytosterols could be a complementary treatment for obesity and diabetes (Jiang *et al.*, 2019). Based on the GC–MS results, the most common volatile compounds in both legumes were hydrocarbons, fatty acids, and terpenes, which is comparable to the results of other researchers (Mohammadi & Ostovar, 2023).

UHPLC-QTOF-MS analysis

UHPLC-QTOF-MS is an advantageous technology that, when used as a targeted and non-targeted profiling approach in either negative or positive ionisation modes, may identify analytes by accurately quantifying their ionic mass and fragmentation patterns. Herein, this technique is used to identify the secondary

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 7 Targeted molecules identified by UHPLC-QTOF-MS in the selected legumes

S.no.	Compound name	Molecular weight	Ab	Mg
1.	Catechin-7-O-glucoside	452.1319	+	+
2.	Catechin	290.0790	+	+
3.	Epicatechin	290.0790	+	+
4.	Quercetin	302.0427	+	+
5.	Gallocatechin	2037.4728	–	+
6.	Gallic acid	170.1210	+	+
7.	Caffeic acid	180.0423	+	+
8.	Para-coumaric acid	164.0473	+	+
9.	Glycitein	284.2610	+	+

metabolites in different extracts of the selected legumes and are presented in Tables 5 and 6. All the identified compounds were analysed on the basis of mass-to-charge (m/z) values of mass spectrometry in positive ionisation mode. To identify the maximum metabolites in the legumes, we analysed all the extracts using the targeted and non-targeted approach. The UHPLC-QTOF-MS chromatograms of targeted metabolites of both samples are provided in Figure S1a–i.

Non-targeted metabolites

The identified metabolites belong to several natural product classes including flavonoids, sugars, amino acids, fatty acids, fatty acid derivates, and other organic acids. In adzuki beans extracts, a total of 15 secondary metabolites were identified based on their mass fragmentation pattern. The data confirms the presence of phytosterols, phenolic compounds, and terpenoids. Mung beans extract contains 19 metabolites, which are primarily phytosterols, flavonoids, and phenolic compounds. Oligosaccharides were also present in slight amount. The present study is one of the first reports to identify secondary metabolites through a non-targeted approach for these legumes using UHPLC-QTOF-MS. Polyphenols have the ability to possess anti-inflammatory and antioxidant properties. Flavonoids can show anti-tumour, anti-inflammatory, and anti-oxidant properties and are considered as an essential component in cosmetics, nutraceutical, and medicinal applications (Shabbir *et al*., 2022). In this study, some carboxylic acids like para-coumaric acid, vernolic acid, gypsogenic acid were detected, which are beneficial for microbial growth by acting as a vitamin for microbial nutrition (Dahibhate & Kumar, 2022).

Targeted metabolites

A total of 9 metabolites were detected in the positive ion mode, as shown in Table 7. We found that catechin-7-O-glucoside, catechin, epicatechin, quercetin, gallocatechin, gallic acid, caffeic acid, para-coumaric acid, and glycitein were significantly present in both the selected legumes. This concurred with previous research, which found that (+)-catechin was the predominant phenolic compound in Korean adzuki beans (Wang *et al*., 2023). Several studies have reported that catechin and their derivatives have been examined for their possible health advantages, particularly their function in the treatment of obesity (Basu *et al*., 2023).

Amino acid profiling

The amino acid profile of the selected legumes revealed some interesting results, as shown in Table 1. The selected legumes contained essential, non-essential, and non-proteinogenic amino acids. All the essential amino acids, viz. arginine, leucine, histidine, lysine, methionine, tryptophan, phenylalanine, valine, threonine, and isoleucine, were found in significant amount. In adzuki beans, essential (47.60%), non-essential (46.99%), and non-proteinogenic amino acids (5.39%) were present in significant quantity. Arginine, leucine, and tryptophan consisted 48.55% of the essential amino acids. The data revealed that the major essential and non-essential amino acids in adzuki beans were tryptophan and glutamic acid, respectively. In mung beans, essential amino acids (44.34%), non-essential amino acids (53.80%), and non-proteinogenic amino acids (1.84%) were significantly present, in which leucine and phenylalanine consisted 34.43% of the essential amino acids. The data revealed that the major essential and non-essential amino acids in mung beans were leucine and glutamic acid. In both the legumes, the major amino acid was glutamic acid. These amino acids have distinct health advantages, e.g. leucine promotes protein biosynthesis, glutamic acid plays a significant role in brain disorders such as schizophrenia, Parkinson disease, and epilepsy (Duangmal & Sangsukiam, 2015; Duan *et al*., 2016). Moreover, several non-proteinogenic amino acids were also present, out of which hydroxyproline was major among adzuki beans, while 1-methylhistidine was the major non-proteinogenic amino acid in mung beans. According to the reports, legumes contain non-proteinogenic amino acids and are particularly abundant in seeds of the Fabaceae family (Rodrigues-Corrêa & Fett-Neto, 2019). They play a variety of functions in the human body, including potential health advantages. β-Alanine is the precursor of pantothenic acid (vitamin B5), and it has been widely used as a precursor for many significant industrial chemicals in food, medicine, and environmental applications (Wang *et al*., 2021). Recently, branched chain essential amino acids and non-proteinogenic amino acids like taurine, orinithine are known to have a positive effect on the regulation of body weight, muscle protein synthesis, and glucose homeostasis (Murakami, 2015; Oh *et al*., 2018; Siddik & Shin, 2019). Thus, the results confirmed that all these legumes are a rich source of nutritionally important essential, non-essential, and non-proteinogenic amino acids.

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Anti-obesity activity

Cell viability through neutral red uptake assay (NRU assay)

The results of cell viability through NRU assay at different concentrations of methanol extract of adzuki beans and mung beans are shown in Fig. 1. It can be observed that both the samples exhibited cell viability values of more than 80%, which indicates good biocompatibility of both the methanol extracts with the 3T3-cell line. The IC_{50} values for adzuki beans and mung beans extract were found to be 921.7 and 174.1 μg mL^{-1}, respectively. The decrease in cell viability is generally proportional to the extent of cytotoxicity. From the dose–response curve, it is clear that the extent of cytotoxicity is lower in both extracts. Also, lower the IC_{50} value of the extract, higher will be the cytotoxicity. From the IC_{50} value data, mung beans methanol extract is lower than adzuki beans methanol extract. Thus, mung beans extract showed maximum cytotoxicity on 3T3- cell line. Hence, it may be interpreted that the methanol extract of adzuki beans is less cytotoxic and more cytocompatible than mung beans methanol extract.

Inhibition of HMG-Co-A reductase enzyme activity

The inhibitory effect of methanol extract of adzuki beans and mung beans on HMG-Co-A reductase enzyme was assessed by *in vitro* assay. HMG-CoA (3-hydroxy-3-methyl-glutayl-coenzyme A) reductase is the key enzyme in the biosynthesis of cholesterol. The marked downregulation of hepatic HMG-CoA reductase contributes to hypocholesteremic effect (Alberts *et al.*, 1980; Peng *et al.*, 2011; Lestari *et al.*, 2017). Each extract sample was capable of inhibiting HMG-Co-A reductase enzyme in a dose-dependent manner. Among the methanol extract of adzuki beans and mung beans, adzuki beans showed an inhibitory effect of more than 50% on HMG-Co-A reductase activity. The inhibitory effect of adzuki beans extract was 60%, as shown in Fig. 2. The results indicate that

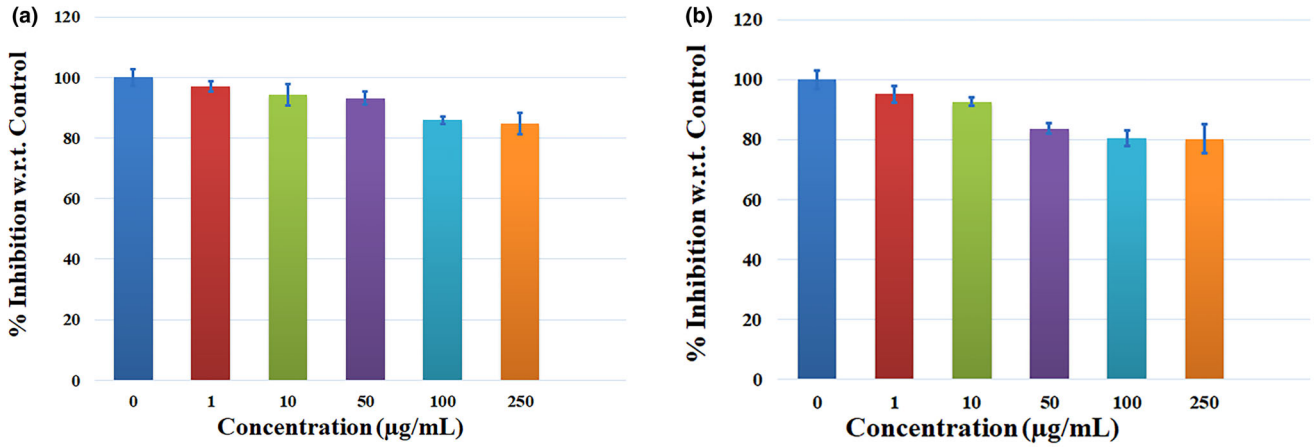


Figure 1 Cytotoxicity study of methanol extract of (a) adzuki beans (b) mung beans.

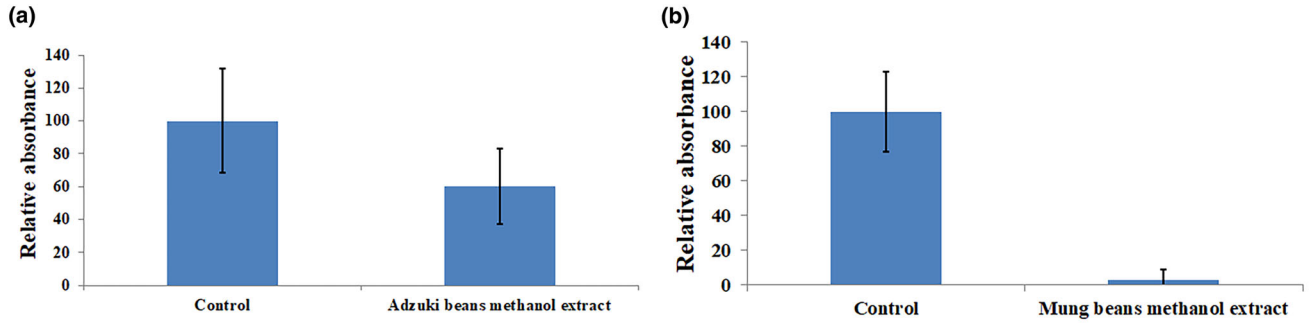


Figure 2 Inhibitory effect of (a) adzuki beans methanol extract and (b) mung beans methanol extrat against HMG-Co-A reductase enzyme.

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the inhibition of HMG-CoA reductase leads to reduction in cholesterol level and possibly to mitigate obesity.

Conclusion

This study focussed on the new insights into understanding the bioactive compounds present in the two varieties of legumes of India and evaluated their *in vitro* cholesterol-lowering and anti-obesity activity. Both the legumes are an excellent source of amino acids, carbohydrate, dietary fibre, and protein. The GC–MS data revealed the existence of significant amount of fatty acids, terpenes, and hydrocarbons. UHPLC-QTOF-MS showed the presence of metabolites like catechin, epicatechin, quercetin, caffeic acid, para-coumaric acid, and gallic acid, which may be responsible for the anti-obesity activity. Moreover, the *in vitro* cholesterol-lowering activity of adzuki and mung bean methanol extract demonstrated that methanol extract of adzuki beans can effectively inhibit the HMG-Co-A reductase enzyme activity. The results suggest that adzuki beans and mung beans can be used to mitigate the cardiovascular and obesity-related disorders. In our view, these two beans (adzuki and mung beans) can be developed into novel ready-to-eat functional foods for consumers suffering from cardiovascular disease and obesity.

Acknowledgements

We are thankful to Prof. Rina Chakraborty (University of Delhi), Dr. Manoj Pandey (Indian Pharmacopoeia Commission, India), Mr. Hitesh Shirmal (Waters, India) and Mr. Anoop Patyal (Panjab University, India), Ms. Shalini tyagi and Mr. Robin (Intertek, India) for helpful discussion.

Author contributions

Ritu Sharma: Investigation. **Devendra Kumar:** Project administration. **Rajinder K. Gupta:** Supervision; conceptualization.

Funding

This research received no external funding.

Conflict of interest

The authors declare that there is no conflict of interest.

Ethical approval

This study does not involve any human or animal testing.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ijfs.17358.

Data availability statement

The data that support the findings of this study are available from the authors upon reasonable request.

References

Abdel-Aal, E.I., Haroon, A.M. & Mofeed, J. (2015). Successive solvent extraction and GC-MS analysis for the evaluation of the phytochemical constituents of the filamentous green alga *Spirogyra longata*. *Egyptian Journal of Aquatic Research*, **41**, 233–246.

This publication was extremely good for extraction and identification of secondary metabolites, the kind of work we have done.

Alberts, A.W., Chen, J., Kuron, G. *et al.* (1980). Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme a reductase and a cholesterol-lowering agent. *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 3957–3961.

Archana, G., Sabina, K., Babuskin, S. *et al.* (2013). Preparation and characterization of mucilage polysaccharide for biomedical applications. *Carbohydrate Polymers*, **98**(1), 89–94.

Azmah, U.N., Makeri, M.U., Bagirei, S.Y. & Shehu, A.B. (2023). Compositional characterization of starch, proteins and lipids of long bean, dwarf long bean, mung bean and french bean seed flours. *Measurement: Food*, **12**, 100111.

Bankole, A.E., Adekunle, A.A., Sowemimo, A.A., Umebese, C.E., Abiodun, O. & Gbotosho, G.O. (2016). Phytochemical screening and *in vivo* antimalarial activity of extracts from three medicinal plants used in malaria treatment in Nigeria. *Parasitology Research*, **115**, 299–305.

Basu, T., Selman, A., Reddy, A.P. & Reddy, P.H. (2023). Current status of obesity: protective role of Catechins. *Antioxidants*, **12**(2), 1–21.

Baur, F.J. & Ensminger, L.G. (1977). The Association of Official Analytical Chemists (AOAC). *Journal of the American Oil Chemists' Society*, **54**(4), 171–172.

Bi, S., Wang, A., Lao, F. *et al.* (2021). Effects of frying, roasting and boiling on aroma profiles of adzuki beans (*Vigna angularis*) and potential of adzuki bean and millet flours to improve flavor and sensory characteristics of biscuits. *Food Chemistry*, **339**(17), 127878.

In this publication the chinese authors have worked on their chienese variety of adzuki beans, which was very useful for our own work.

Carta, G., Murru, E., Banni, S. & Manca, C. (2017). Palmitic acid: physiological role, metabolism and nutritional implications. *Frontiers in Physiology*, **8**, 1–14.

Chakhtoura, M., Haber, R., Ghezzawi, M., Rhayem, C., Tcheroyan, R. & Mantzoros, C.S. (2023). Pharmacotherapy of obesity: an update on the available medications and drugs under investigation. *EClinicalMedicine*, **58**, 101882.

Chakrabarti, R., Clark, W.D., Sharma, J.G., Goswami, R.K., Shrivastav, A.K. & Tocher, D.R. (2018). Mass production of *Lemna minor* and its amino acid and fatty acid profiles. *Frontiers in Chemistry*, **6**, 1–16.

We have used their methodology for amino acid analysis and non-proteinogenic amino acids.

Dahibhate, N.L. & Kumar, K. (2022). Metabolite profiling of *Bruguiera cylindrica* reveals presence of potential bioactive compounds. *PeerJ Analytical Chemistry*, **4**, e16.

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Duan, Y., Li, F., Li, Y. *et al.* (2016). The role of leucine and its metabolites in protein and energy metabolism. *Amino Acids*, **48**, 41–51.
Duangmal, K. & Sangsukiam, T. (2015). Effect of germination period on changes in mungbean and azuki bean protein and amino acid profiles. *Acta Horticulturae*, **1088**, 563–568.
Grover, N. & Patni, V. (2013). Phytochemical characterization using various solvent extracts and GC-MS analysis of methanolic extract of *Woodfordia fruticosa* (L.) Kurz. leaves. *International Journal of Pharmacy and Pharmaceutical Sciences*, **5**(4), 291–295.
Hirota, S. & Takahama, U. (2017). Inhibition of pancreatin-induced digestion of cooked rice starch by adzuki (*Vigana angularis*) bean flavonoids and the possibility of a decrease in the inhibitory effects in the stomach. *Journal of Agricultural and Food Chemistry*, **65**, 2172–2179.
Hong, Y., Wang, Z., Barrow, C.J., Dunshea, F.R. & Suleria, H.A.R. (2021). High-throughput screening and characterization of phenolic compounds in stone fruits waste by LC-ESI-QTOF-MS/MS and their potential antioxidant activities. *Antioxidants*, **10** (2), 1–22.
Hymavathi, T.V., Roberts, T.P.P., Jyothsna, E. & Sri, V.T. (2020). Proximate and mineral content of ready to use minor millets. *International Journal of Chemical Studies*, **8**, 2120–2123.
Ibrahim, S., Aslam, M. & Musthafa, M.M. (2019). Trace elemental fingerprinting of selected herbs used in Ayurveda using XRF and ICPMS. *Journal of Pharmacognosy and Phytochemistry*, **8**, 3429–3433.
Jiang, L., Zhao, X., Xu, J. *et al.* (2019). The protective effect of dietary phytosterols on cancer risk: a systematic meta-analysis. *Journal of Oncology*, **2019**, 1–11.
Jide, A.O., Alademeyin, J.O. & Ajongbolo, K.F. (2023). Adzuki (*Vigna angularis*) beans as food: chemical composition, nutrition and quality identities. *Journal of Advanced Education and Sciences*, **3**, 5–8.
Kumar, V., Sharma, P., Bairagya, H.R., Sharma, S., Singh, T.P. & Tiku, P.K. (2019). Inhibition of human 3-hydroxy-3-methylglutaryl CoA reductase by peptides leading to cholesterol homeostasis through SREBP2 pathway in HepG2 cells. *Biochimica et Biophysica Acta, Proteins and Proteomics*, **1867**, 604–615.
Lee, P.S., Teng, C.Y., Hsieh, K.F. *et al.* (2019). Adzuki bean water extract attenuates obesity by modulating M2/M1 macrophage polarization and gut microbiota composition. *Molecular Nutrition & Food Research*, **63**, 1–11.
Lestari, D., Sukandar, E.Y. & Fidrianny, I. (2017). In silico study of apigenin and apigetrin as inhibitor of 3-hydroxy-3-methyl-glutaylcoenzyme a reductase. *Asian Journal of Pharmaceutical and Clinical Research*, **10**, 284–287.
Mohammadi, N. & Ostovar, N. (2023). Chemical composition, fatty acid composition, and volatile compounds of a traditional Kurdish fermented cereal food: Tarkhineh. *Food Chemistry Advances*, **2**, 100187.
Müller, T.D., Blüher, M., Tschöp, M.H. & DiMarchi, R.D. (2022). Anti-obesity drug discovery: advances and challenges. *Nature Reviews Drug Discovery*, **21**, 201–223.
Murakami, S. (2015). Role of taurine in the pathogenesis of obesity. *Molecular Nutrition & Food Research*, **59**, 1353–1363.
Nagao, K. & Yanagita, T. (2008). Bioactive lipids in metabolic syndrome. *Progress in Lipid Research*, **47**, 127–146.
Obioma Okereke, G. & Banigo, E.O.I. (2021). Proximate composition of some selected legumes and quality attributes of oils extracted from soybean, oil bean and groundnut seeds. *Asian Food Science Journal*, **20**, 49–60.
Oguuniyi, Q.A., Ogbole, O.O., Akin-Ajani, O.D. *et al.* (2023). Carbohydrate profiling of different accessions of three underutilized legumes: a potential source of valuable nutraceuticals and pharmaceuticals. *Starch/Staerke*, **76**, 2300184.
Oh, S.-H., Kim, M., Yu, O.-K., Kim, S.-G. & Cha, Y. (2018). Attenuation of obesity by ornithine enriched blueberry liquid starter in 3T3-L1 preadipocyte and high-fat diet fed obese mouse model. *The FASEB Journal*, **31**, lb475.
Peng, C.H., Liu, L.K., Chuang, C.M., Chyau, C.C., Huang, C.N. & Wang, C.J. (2011). Mulberry water extracts possess an anti-obesity effect and ability to inhibit hepatic lipogenesis and promote lipolysis. *Journal of Agricultural and Food Chemistry*, **59**, 2663–2671.
Rodrigues-Corrêa, K.C.S. & Fett-Neto, A.G. (2019). Abiotic stresses and non-protein amino acids in plants. *Critical Reviews in Plant Sciences*, **38**, 411–430.
van Rooijen, M.A. & Mensink, R.P. (2020). Palmitic acid versus stearic acid: effects of interesterification and intakes on cardiometabolic. *Nutrient Journal*, **12**, 1–24.
Ryan, E., Galvin, K., O'Connor, T.P., Maguire, A.R. & O'Brien, N.M. (2007). Phytosterol, squalene, tocopherol content and fatty acid profile of selected seeds, grains, and legumes. *Plant Foods for Human Nutrition*, **62**, 85–91.
Shabbir, U., Tyagi, A., Ham, H.J. & Oh, D.H. (2022). Comprehensive profiling of bioactive compounds in germinated black soybeans via UHPLC-ESI-QTOF-MS/MS and their anti-Alzheimer's activity. *PLoS One*, **17**, e0263274.
Siddik, M.A.B. & Shin, A.C. (2019). Recent progress on branched-chain amino acids in obesity, diabetes, and beyond. *Endocrinology and Metabolism*, **34**, 234–246.
This publication was extremely useful to understand the importance of branched-chain amino acids in obesity and type-2 diabetes.
Smith, K.B. & Smith, M.S. (2016). Obesity statistics. *Primary Care; Clinics in Office Practice*, **43**, 121–135.
Tor-Roca, A., Garcia-Aloy, M., Mattivi, F., Llorach, R., Andres-Lacueva, C. & Urpi-Sarda, M. (2020). Phytochemicals in legumes: a qualitative reviewed analysis. *Journal of Agricultural and Food Chemistry*, **68**, 13486–13496.
Ullah Khan, I., Mahiya, S., Singh Rathore, B. & Singh, B. (2017). Volatile phyto-chemical compounds screening in alcoholic extract of western rajasthan growing mung-bean (*Vigna radiata*) by gas chromatography-mass spectroscopy. *Trends in Biosciences*, **10**, 1588–1595.
Vajrabhaya, L.O. & Korsuwannawong, S. (2016). Cytotoxicity evaluation of *Clinacanthus nutans* through dimethylthiazol diphenyltetrazolium bromide and neutral red uptake assays. *European Journal of Dentistry*, **10**, 134–138.
Wang, L., Mao, Y., Wang, Z., Ma, H. & Chen, T. (2021). Advances in biotechnological production of β-alanine. *World Journal of Microbiology and Biotechnology*, **37**, 1–11.
Wang, Y., Chen, L., Liu, H. *et al.* (2023). Characterization of the synergistic inhibitory effect of cyanidin-3-O-glucoside and catechin on pancreatic lipase. *Food Chemistry*, **404**, 134672.
Zafar, S.H., Umair, M. & Akhtar, M. (2023). Nutritional evaluation, proximate and chemical composition of mungbean varieties/cultivars pertaining to food quality characterization. *Food Chemistry Advances*, **2**, 100160.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. UHPLC-QTOF-MS chromatograms of targeted metabolites of selected legumes (a) catechin-7-O-glucoside, (b) quercetin, (c) epicatechin, (d) caffeic acid, (e) glycitein, (f) gallic acid, (g) para-coumaric acid, (h) gallocatechin, (i) catechin.

© 2024 Institute of Food Science & Technology (IFST).

13652621, 0, Downloaded from https://ifst.onlinelibrary.wiley.com/doi/10.1111/ijfs.17358 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Journal of Inorganic and Organometallic Polymers and Materials
https://doi.org/10.1007/s10904-024-03224-y

RESEARCH

Biopolymer-CMTG and m-BPDM Based Hydrogel Composite for Promising Sensing of Zinc, Cadmium, and Mercury in Aqueous Medium

Priyanka Yadav[1] · Sudhir G. Warkar[1] · Anil Kumar[1]

Received: 27 March 2024 / Accepted: 21 June 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
The detection of spectroscopically silent metal ions is challenging due to their electronic configuration (d^{10}). A practical approach to overcome this issue is the use of complex-based sensing platforms for metal ion detection. However, sensing using these ligand-based complexes occurs only in organic media, hindering their large-scale applications. Therefore, the current study aims to develop a biopolymer-based hydrogel composite for colorimetric sensing in aqueous medium. The meta-benziporphodimethene (m-BPDM)-modified carboxymethyl tamarind gum (CMTG)/polyacrylamide (PAM) hydrogel was developed via in situ incorporation of m-BPDM into hydrogel matrix. Solid-state UV-visible spectroscopy, FTIR, MXRD, and SEM characterized the m-BPDM-modified CMTG-based hydrogel composite. The as-synthesized m-BPDM-modified hydrogel was applied as a sensor for the colorimetric sensing of Zn^{+2}, Hg^{+2}, and Cd^{+2} metal ions. It demonstrated a color change from pinkish red to dark blue in the aqueous solution of metal salts. The change in color of hydrogel upon contact with the metal solution was also validated by Solid UV-visible spectroscopy. Further, the impact of temperature, the concentration of heavy metal ions, solution pH on sensing time, and sensing of zinc ions in *E. coli* cells were investigated. The sensor's performance was also assessed in industrial effluents to check its applicability in real-time applications. The quantitative determination of Zn^{+2}, Hg^{+2}, and Cd^{+2} from industrial effluents was confirmed using atomic absorption spectroscopy (AAS). This suggests that synthesized hydrogel can be utilized as a sensor for the visual on-site detection of zinc, cadmium, and mercury metal ions in an aqueous medium.

Keywords Biosensor · Biopolymers · Hydrogel · m-BPDM

1 Introduction

In recent decades, the rapid diversification of industrialization and urbanization has accelerated economic growth but has also contributed to an exponential increase in population, thereby leading to several environmental issues, such as pollution of the water, soil, and air [1]. Discharging untreated waste from agricultural activities and industries such as mining, metal rinse processes, tanning, textile, etc., is a major factor behind the increasing water pollution levels [2]. Specifically, the textile industry is a major polluter; the dyeing and finishing processes often lead to the indiscriminate release of various dyes, pigments, heavy metals, etc., into the water bodies [3, 4]. Therefore, it is imperative to manage and remediate wastewater to safeguard life within the ecosystem in the contemporary era [5, 6]. Among the numerous pollutants found in the ecological cycle, 'Heavy metal ions stand out as a particular and grave concern due to their persistence of high toxicity, excessive permeability, non-degradability, and potential for long-term environmental and public health, which causes physiological and development disorders [7, 8]. Heavy metal ions include, but are not limited to, copper (Cu), magnesium (Mg), manganese (Mn), lead (Pb), nickel (Ni), zinc (Zn), iron (Fe), mercury (Hg), plutonium (Pu), chromium (Cr), arsenic (As), cobalt (Co), cadmium (Cd), etc. [9]. Heavy metals are

✉ Sudhir G. Warkar
sudhirwarkar@gmail.com

✉ Anil Kumar
anil_kumar@dce.ac.in

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

Published online: 17 August 2024

non-biodegradable and can contaminate drinking water supplies when not effectively detected. Although living beings require trace amounts of heavy metals such as Cu, Co, Zn, Fe, Mg, Mn, etc. since they are considered to be essential for human metabolism regulation even the presence of essential heavy metal ions in inappropriate amounts could result in critical health issues for living beings [7, 10, 11].

Among all metals, Zn, Cd, and Hg are more harmful since they are spectroscopically silent in nature [12]. Due to this, detecting these metal ions cannot be achieved through techniques such as electron paramagnetic resonance (EPR) or nuclear magnetic resonance spectroscopy (NMR). Therefore, visual detection techniques are becoming more attractive in various fields because they enable qualitative and semi-quantitative analysis without complex instrumentation. Visual detection can prove to be a rapid diagnostic method as this is most useful in disaster situations due to its low cost, speed, and simplicity [12]. Therefore, researchers are highly interested in calorimetric sensing of these metals in aqueous media.

The deterioration of the environment has also increased because of the industrial revolution and human activity. The coastal areas are extremely exposed due to significant pollutant flows into the ocean and other water bodies. Heavy metals are highly affecting the environment because of their chronic toxicity, non-biodegradability, and environmental bioaccumulation. In addition, the health of humans is at risk by heavy metals, which can enter food chains and be biomagnified [13]. According to the World Health Organization (WHO), the maximum permitted amounts of heavy metal ions in drinking water should not exceed 3 ppm for Zn^{+2}, 0.003 ppm for Cd^{+2}, and 0.001 ppm for Hg^{+2} and the toxicity of heavy metals may have several health risks [14]. Especially in humans, heavy metals have adverse effects since they can accumulate in human organs via inhaling air and drinking water, which leads to illnesses like cancer, hypertension, and kidney failure, which also could be carcinogenic. Zn^{+2} levels serve a predominant role in living cells and tissues of living beings, which require regular monitoring for signal transmitters in gene expression, cell division & growth, apoptosis, and a neural signal transmitter [15]. In addition, it has been found that the dysregulation or dysfunction of zinc homeostasis strongly correlates with neurological disease and growth retardancy, such as Parkinson's disease, epilepsy, etc. [16, 17]. Some heavy metals such as Cd, Hg, Pb, and Pu are not biologically active and can cause serious major threats to the bodies of living beings due to their tendency for bioaccumulation. These toxic heavy metal ions accumulate in living bodies via the food chain, resulting in ecologically induced irreversible pollution and negative impacts on the organs [18, 19]. Specifically, Cd and Hg metals are considered to be toxic even at very minute concentrations and pose several adverse effects, including chronic physiological effects, kidney inflammation, brain disorders, and renal dysfunction. These metal ions are also utilized in refining, phosphate fertilizers, and the combustion of fossil fuels [20, 21]. Therefore, there is an immediate need for regularity in detecting heavy metal ions for the well-being of living beings and for environmental protection [22–24].

Ultimately, it is beneficial for human health and the environment to eliminate heavy metal ions effectively. Although various sophisticated techniques, such as atomic absorption spectroscopy, EDX, ICPMS, etc., are available for heavy metal detection, they require expensive equipment. They are time-intensive and complicated processes requiring high operating skills [25]. Hence, there is a high demand for a cost-effective method that is easy to operate. Sensing is one of the popular and widespread techniques due to their ease of use, low cost, sensitivity, and selectivity toward different types of analytes. Sensors that change color in response to external stimuli are called Colorimetric sensors [26]. Due to their easy fabrication, quick detection, high sensitivity and selectivity, and easy naked-eye detection, colorimetric sensors, and biosensors offer promising potential for detecting metallic cations, anions, organic dyes, drugs, and pesticides. In the available literature, various fluorescent probes were used to detect heavy metal ions [27–34]. *Zareh M. et al.* utilized naphthol and thiazole-based sensors to visualize Hg^{+2} with a quick color change from colorless to yellow [35]. Furthermore, *Kusuma S.* et al. have synthesized gold nanoparticle-based sensors to detect heavy metal ions [36]. *Wen S. et al.* reported a tricarboyanine-based chemosensor for detecting Ag^{0} from industrial waste and showed a color change from dark blue to light pink [37]. *Tong S.* and co-workers developed an optic sensor for Hg^{+2} detection and showed that the limit of detection (LOD) is 0.15 μM [38]. Most of the available literature is based on fluorescent and chromophores-based sensors, which possess some constraints, such as poor water solubility, reactivity towards other co-existing ions, delay in response time, and high backstage emission. Our previous report used meta-Benziporphodimethene (m-BPDM) to sense metal ions in an organic medium [16, 39]. m-BPDM is hydrophobic and consists of pyrrole moiety and carbocyclic units [40]. It is a great challenge for researchers to make it feasible to detect Zn^{+2}, Hg^{+2}, and Cd^{+2} ions in an aqueous medium. Since to utilize m-BPDM as a sensor in aqueous media is not possible. Therefore, to overcome this constraint, m-BPDM has been incorporated into the hydrogel matrix to sense Zn^{+2}, Hg^{+2}, and Cd^{+2} [41]. Usually, Hydrogels are made up of a polymer matrix and have a three-dimensional (3-D) porous structure that holds significant water and fluids in their 3-D network structure [42]. *Ham J.* et al. have reported using

stretchable hydrogel for calorimetric sensing of ammonia [43]. *Wang J. et al.* have reported the colorimetric detection of copper ions using chitosan/glutamic acid/agarose/Ag nanocomposite-based hydrogel [44]. *Peng Z. et al.* have synthesized a hydrogel sensor for the detection of lead metal ions [45]. *Kumar A. et al.* have synthesized guar gum-based hydrogel for sensing heavy metal ions in aqueous media [41]. Some advantages of the hydrogel-based matrix over other existing materials are comprised in Table 1. To the best of our knowledge, there is a very scarce report available on the hydrogel-based matrix for sensing heavy metals (Zn^{+2}, Hg^{+2}, and Cd^{+2}) in aqueous media. Therefore, this work aims to design a simple, cost-effective, easy method for detecting toxic heavy metal ions on-site in real-time.

In this work, we have synthesized a novel meta-Benziporphodimethene-modified Carboxymethyl tamarind gum/Polyacrylamide (m-BPDM/CMTG/PaaM) biopolymeric hydrogel, which was then employed as a colorimetric biosensor. The synthesized probe was utilized in the selective detection of Zn^{+2}, Hg^{+2}, and Cd^{+2} ions in water and further for detecting Zn^{+2} in bacteria '*E. Coli.*' and other parameters, including time, pH, and concentration, were examined using m-BPDM modified CMTG/PaaM hydrogel. In addition, it was discussed in industrial waste to check its utility in real-time application. It was concluded that the synthesized material could be used as an effective and rapid biosensor for detecting Zn^{+2}, Hg^{+2}, and Cd^{+2} ions in aqueous systems.

Table 1 Comprises some advantages of the hydrogel-based matrix over other existing materials

Advantages of synthesized hydrogel composites over existing sensing material		Ref.
Biocompatibility	Synthesized hydrogel composite is biocompatible in nature	[46]
Ease of Synthesis	The synthesis method is quite easy than existing sensing material	[41]
Solubility	Synthesized hydrogel composite can sense in aqueous media, while most of the existing sensing material required organic solvent to respond	[47]
Response Time	Most of the fluorescent sensors have low response time and back stage emission	[48]
Durability	They are more robust and stand in a wide range of environmental conditions, but some of the existing sensing materials, such as silicon-based sensors and metal-based sensors, degrade in harsh conditions	[49]
Cost-effective	They are more cost-effective since they are composed of biopolymers and polymers. Hence can be utilized for large-scale applications	[50]
Limitation of synthesized hydrogel composites over existing sensing material		
Response Time	Synthesized hydrogel composites may have a slower response time than electronic sensors	[51]

2 Methods

2.1 Materials

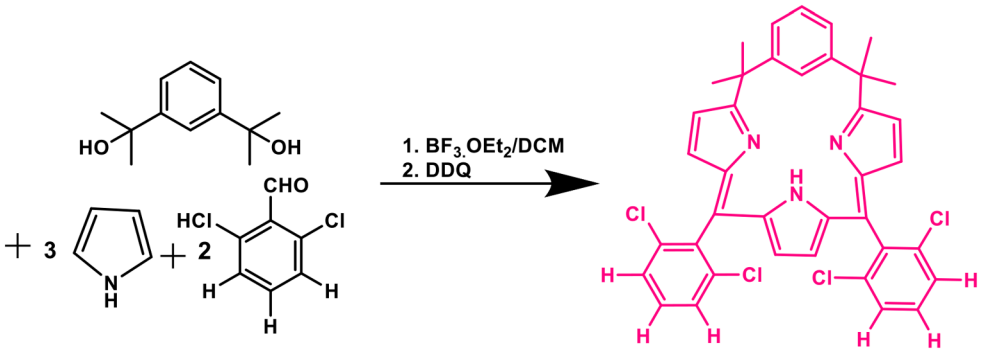
Pyrrole and potassium persulfate (KPS) were purchased from CDH (India). Functionalized aldehydes, α, α di-hydroxy-1, 3-di-isopropylbenzene, 2,3-dichloro-5,6- di-cyano-p-benzoquinone (DDQ), and Silica gel were purchased from Thermo-Fischer, India. Carboxymethyl formulation of Tamarind Gum (CMTG) was obtained from Hindustan Gum and Chemicals Ltd., Haryana (India). Sodium Dodecyl Sulphate (SDS), N, N'-methylenebisacrylamide (MBA), and Acrylamide (aam) Merck were obtained from Merck, Germany. Luria broth was obtained from Genex, India. Escherichia coli (E. Coli.), Luria broth medium (2%), phosphate buffer solution, and saline solution (0.9%) were used as obtained. Dichloromethane (DCM), Pyrrole, and functionalized aldehydes in liquid form were distilled before use, and other reagents such as Acetone and Hexane were utilized as purchased. All the solutions used were made in distilled water. The experiments were performed in triplicates to ensure the reproducibility of the synthesized sensor. The statistical analysis was done, and the error bar was inserted in the data points. Statistical differences were also analyzed for industrial effluent, and sensing of Zn^{+2} in E. *Coli.* cells (mean ± SD) using a one-way analysis of variance (ANOVA) ($p < 0.05$) was determined to be statistically significant.

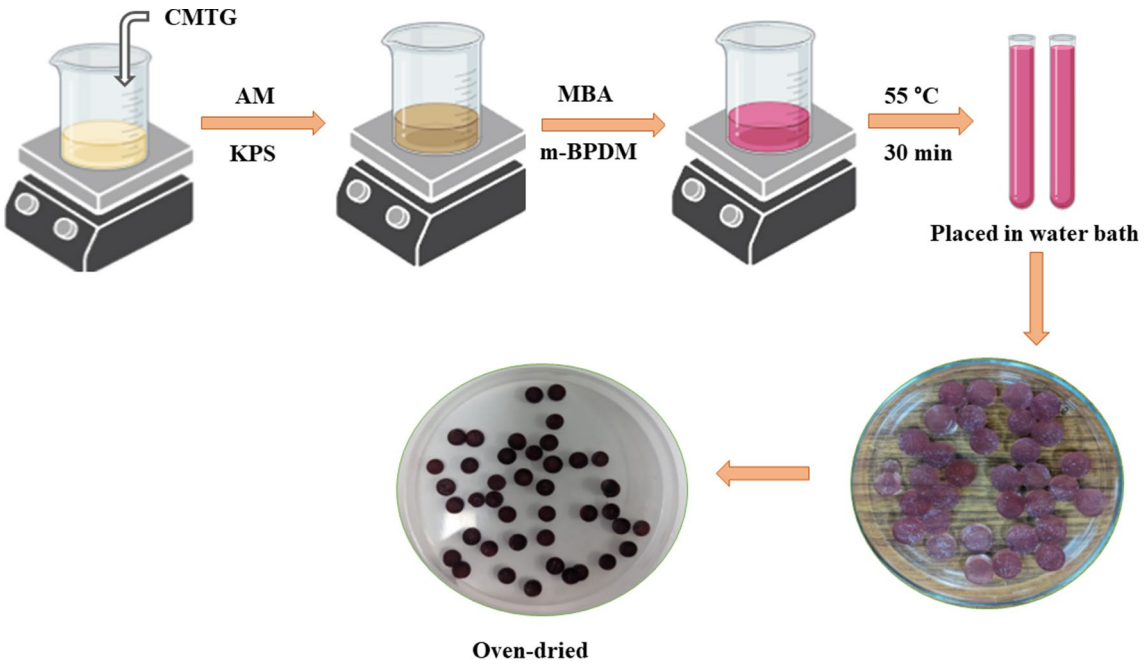
2.2 Synthesis

2.2.1 Synthesis of CMTG/Paam Hydrogel

Synthesis of CMTG/ Paam bipolymeric hydrogel was performed. In a 50 mL beaker, 0.4 g CMTG was dissolved in 15 mL distilled water, 2.0 g acrylamide (aam) was added, and stirring was maintained for 25 min. Then, adding 0.06 g KPS, initiator, and 0.030 g MBA, the crosslinker was done and stirred vigorously for 30 min. After that, the mixture was poured into test tubes and kept in a water bath at 55 °C for 30 min. Then, the product was collected and cut into circular slices, further submerged in distilled water to remove alkalinity and unreacted chemicals [52]. The obtained product was kept at room temperature overnight and placed in an oven at 40 °C. It was pale yellow in color.

Scheme 1 Synthesis of m-BPDM





Scheme 2 Synthesis of CMTG/PAM/m-BPDM hydrogel composite

2.2.2 Synthesis of m-BPDM-incorporated CMTG/Paam Hydrogel Composite

m-BPDM was synthesized, as reported earlier [16]. 0.4 g CMTG was dissolved in 12 mL distilled water and stirred until a homogeneous solution was formed. Then 2.0 g acrylamide (aam) was added and agitated for 25 min, followed by 0.08 g KPS, an initiator. In addition, 5 mg m-BPDM was added in acetone: water (1:1) ratio. In the final synthesis step, 0.035 g MBA crosslinker was added and agitated for 30 min at room temperature. Then, the reaction mixture was poured into the test tube and placed in the water bath at 55 °C for 30 min. Then, the obtained product was removed by breaking test tubes, cutting them into pieces, and immersing them in double distilled water to eliminate unreacted precursors. The product was dried in the oven at 40 °C to get the persistent weight of the hydrogel pieces. The obtained product was pinkish-red in color, and the Scheme 1 and Scheme 2 are presented.

2.3 Characterization

The UV–visible spectra were observed using a Shimadzu UV 1800 Spectrophotometer, and solid-state UV–visible absorption spectra were obtained using PerkinElmer Lambda 750 Spectrometer for solid samples The FTIR spectra were obtained using an FTIR spectrophotometer Model: Nicolet iS50 FTIR, and the spectra were recorded within a range of 4000–400 cm^{-1}. PXRD patterns of synthesized material were analyzed using a high-resolution MXRD Rigaku Ultima IV employed with Cu Kα radiation having k = 1.54 Å. The PXRD patterns were recorded with a step size of 0.02º having a scan rate of 1.0 s per step. The analysis was conducted at 298 K in the range of 2θ = 5–70º. The SEM morphology was analyzed using an SEM Model-JEOL.

2.4 Sensing Studies

To perform the sensing study, synthesized m-BPDM incorporated hydrogel was utilized to recognize Zn^{+2}, Cd^{+2}, and Hg^{+2}. The m-BPDM incorporated modified hydrogel shows abilities in sensing and scanning solutions of alkali metal salts, alkaline earth metal salts, $ZnCl_2$, $HgCl_2$, and $CdCl_2$ of different concentrations. This study involves various parameters, including time, temperature, and concentration-dependent sensitivity of synthesized hydrogel. A study was conducted to check the effect of pH and concentration of used metal ions on synthesized sensors. The sensor was used to detect the presence of Zn^{+2} in *E. Coli.* by lysing the cell in SDS detergent and was also employed to sense metal ions in industrial wastewater. All experiments involved in supporting this study were performed in triplicate.

3 Results and Discussion

3.1 Synthesis

m-BPDM was synthesized using the reported method mentioned in our previous reports [16]. m-BPDM is an inorganic moiety that is not soluble in water due to its hydrophobic nature. Therefore, a fixed ratio of 1:1 of water and acetone was the most effective composition for dissolving m-BPDM into a hydrogel matrix to overcome this constraint. Then, m-BPDM was incorporated into the biopolymer-based hydrogel matrix. The observed color of synthesized hydrogel is pinkish-red, suggesting that the m-BPDM is successfully incorporated in the CMTG/PaaM hydrogel represented in Scheme 3. Further, confirmation of m-BPDM incorporation into hydrogel was done by using UV-visible spectroscopy. Hence, m-BPDM modified CMTG/PaaM Bipolymeric hydrogel was successfully synthesized.

3.2 Characterization

3.2.1 UV-Vis Spectroscopy

The UV-visible spectra of m-BPDM, m-BPDM/CMTG/PaaM hydrogel, M^{+2} (Zn^{+2}, $Zn^{+2}/Cd^{+2}/Hg^{+2}$), m-BPDM/CMTG/PaaM hydrogels are represented in Fig. 1 The obtained spectra are similar to the available literature [16]. The recorded UV-vis spectrum of free base m-BPDM and solid-state UV- vis spectrum of the synthesized hydrogel is recorded at ambient temperature to ensure the incorporation of m-BPDM into the hydrogel matrix shown in Fig. 1a&b, respectively. The UV-vis spectrum of m-BPDM incorporated hydrogel shows a broad pattern due to discrete conjugation in the system having a soret band with high energy at 355 nm, and bands with lower energy are observed at 501 nm and 612 nm. It was observed that upon metalation, the visible color changed from pinkish red to dark blue. In The UV-vis spectrum, a shift in the lower energy band from 612 to 629 nm was observed, corresponding to the redshift upon metal insertion into m-BPDM-modified- hydrogel. Similar spectra were observed for all metalated hydrogels (Fig. 1c&d), and the band values are identical to the available literature [16, 53]. Further, the UV-vis absorption spectrum of water was also recorded to check out the leaching nature of hydrogel, and it was observed that no peak for m-BPDM was present in the aqueous medium after the sensing experiment. Hence, it may be concluded that m-BPDM interacts physically with the hydrogel matrix.

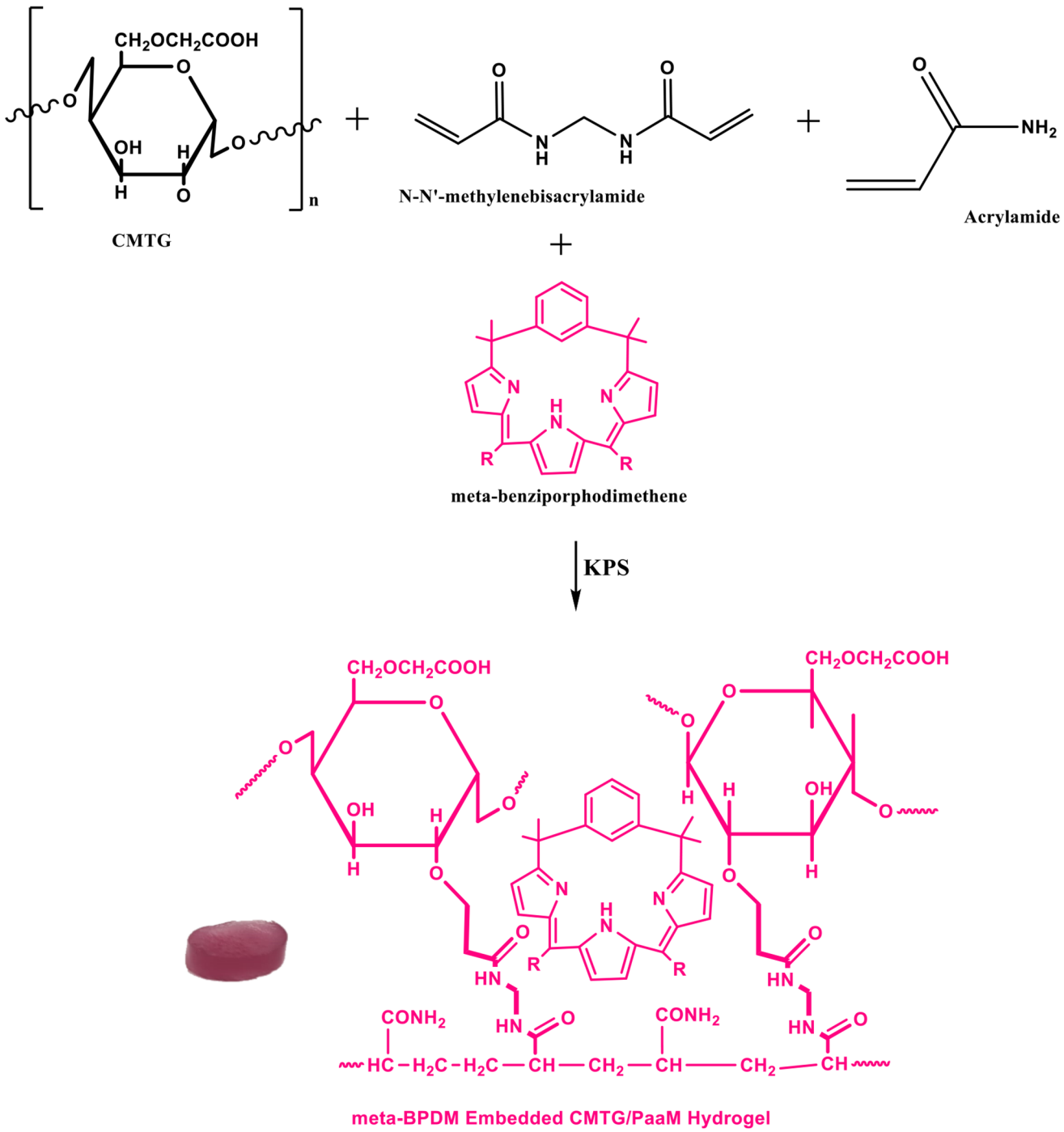
3.2.2 FTIR

The Fourier-transform infrared spectroscopy (FTIR) spectra of CMTG/PaaM hydrogel, m-BPDM/ CMTG/PaaM hydrogel, Zn^{2+}-m-BPDM/ CMTG/PaaM hydrogel, and $Zn^{2+}/Cd^{2+}/Hg^{2+}$-m-BPDM/CMTG/PaaM hydrogel are represented in Fig. 2. The broad bands were observed at 3287 and 3440 $cm^{-1,}$ attributed to the hydroxy (-OH) stretching; the bands at 2791 and 2925 cm^{-1} are associated with the saturated asymmetric C-H stretching, and the peaks at 1616 and 1624 cm^{-1} are due to C = O stretching in CMTG/PaaM hydrogel and m-BPDM/CMTG/PaaM hydrogel respectively [54]. The other bands at 1388, 1239, 964 in CMTG/PaaM hydrogel and at 1449, 1308, and 1061 in m-BPDM/ CMTG/PaaM hydrogel are attributed to C-H bending, C-N bending, and C-O-C glycosidic linkage respectively [55]. The bands in spectra of CMTG/PaaM hydrogel and m-BPDM/CMTG/PaaM hydrogel were almost similar. They showed a slight shift in the band, suggesting some interaction in the matrix, but m-BPDM is not chemically involved with the hydrogel matrix [56].

While in case of Zn^{2+}-m-BPDM/ CMTG/PaaM hydrogel, and $Zn^{2+}/Cd^{2+}/Hg^{2+}$-m-BPDM/CMTG/PaaM hydrogel the peaks were obtained at 3283, 2936, 1604, 1413, 1326, 1071, and 3321, 2941, 1646, 1427, 1322, and 1172 corresponds to -OH stretching, C-H stretching, -C = O stretching, C-H bending, C-N bending, and C-O-C glycosidic linkage respectively. The change observed in the intensity of the absorption band only may be attributed to the metalation of modified polymeric hydrogel matrices [57].

3.2.3 PXRD

The PXRD patterns of synthesized hydrogels are shown in Fig. 3, and the broad spectra were observed in the range of 21–24º, suggesting the amorphous nature of hydrogel matrices [58]. Almost similar spectra were observed in all



Scheme 3 Synthesis of m-BPDM incorporated CMTG/PaaM Biopolymeric hydrogel

hydrogels; a broad peak with low intensity was observed at 23.10º in CMTG/PaaM hydrogel. A peak at 22.29º was observed in the m-BPDM-modified incorporated CMTG/PaaM hydrogel, suggesting some interaction occurs upon incorporating the m-BPDM into the hydrogel matrix. In the case of Zn^{2+}-m-BPDM/CMTG/PaaM hydrogel, a slight enhancement in peak intensity was observed, which signifies an increase in the hydrogel's crystallinity after incorporating zinc ions into the hydrogel matrix. Further, a similar trend was observed upon sensing another metal ion, such as Cd^{+2}, Hg^{+2}, and $Zn^{+2}/Cd^{+2}/Hg^{+2}$ in m-BPDM/CMTG/PaaM hydrogel [41]. The probable cause of this trend may be the complex formation of metal ions with incorporated m-BPDM into the hydrogel matrix. Hence, it may be concluded that the change in intensity was the only difference

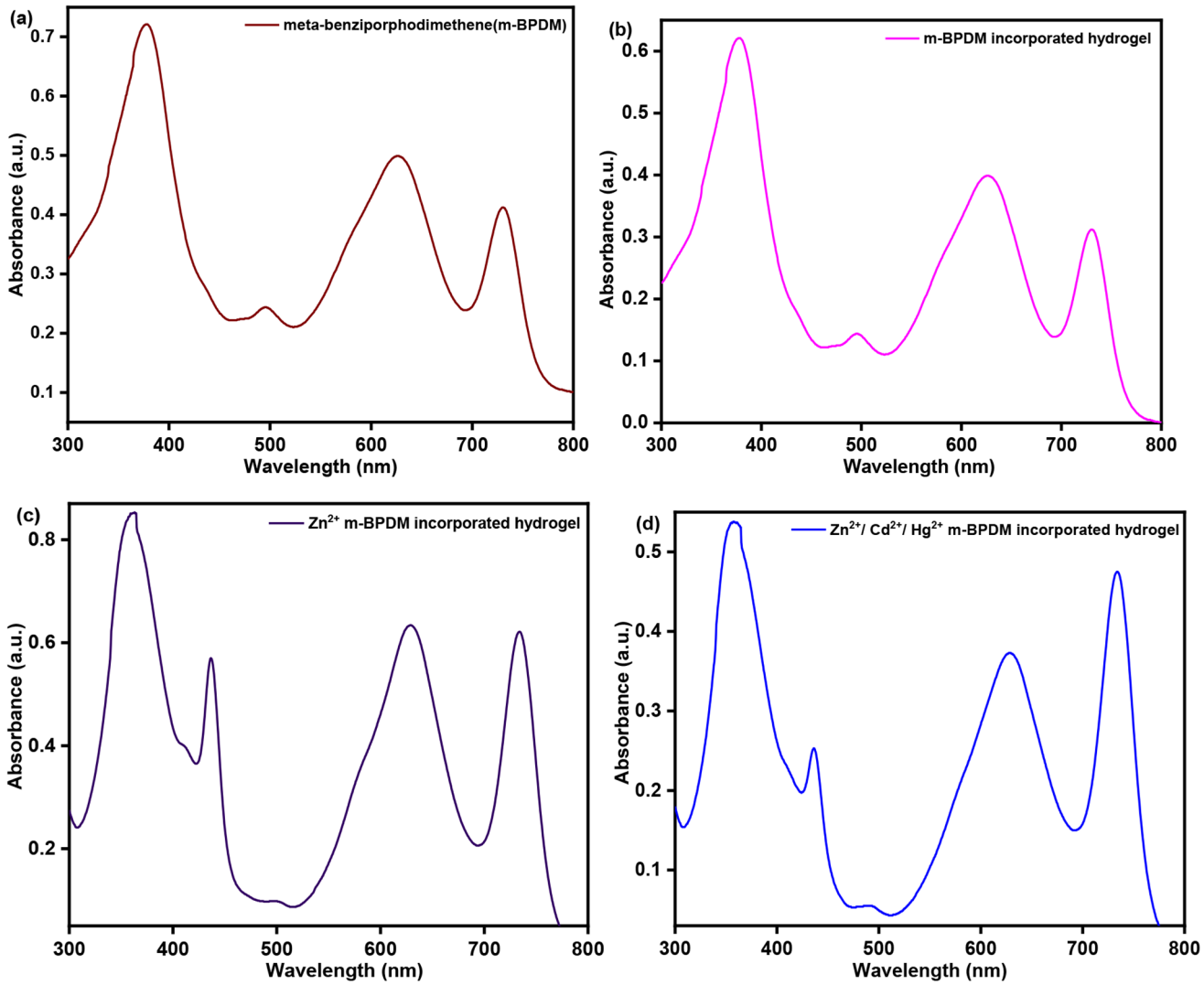


Fig. 1 Plots for (**a**) UV-vis spectrum of m-BPDM, (**b**) Solid state UV-vis spectrum of m-BPDM incorporated CMTG/PaaM hydrogel, (**c**) Zn^{2+}-m-BPDM/ CMTG/PaaM hydrogel, and (**d**) Zn^{2+}/Cd^{2+}/Hg^{2+}-m-BPDM/CMTG/PaaM hydrogel

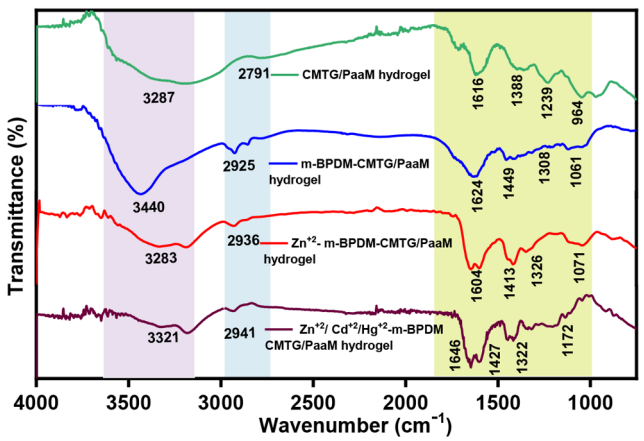


Fig. 2 FTIR plots of CMTG/PaaM hydrogel, m-BPDM incorporated CMTG/PaaM hydrogel, Zn^{2+} -m-BPDM/ CMTG/PaaM hydrogel, and $Zn^{2+//}Cd^{2+}/Hg^{2+}$ -m-BPDM/CMTG/PaaM hydrogel

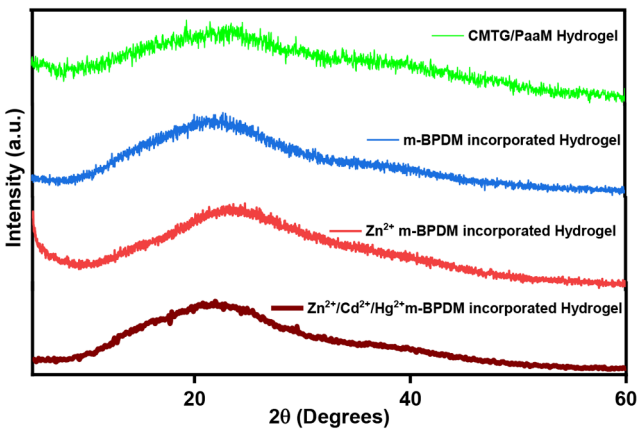


Fig. 3 PXRD Plots for CMTG/PaaM hydrogel, m-BPDM incorporated CMTG/PaaM hydrogel, Zn^{+2}-m-BPDM/ CMTG/PaaM hydrogel, and $Zn^{+2}/Cd^{+2}/Hg^{+2}$ -m-BPDM/CMTG/PaaM hydrogel

interpreted by the PXRD patterns, suggesting the m-BPDM, and metalation does not affect the hydrogel matrix.

3.2.4 SEM

The SEM micrographs of synthesized hydrogels are shown in Fig. 4. The CMTG/PaaM hydrogel shows layered morphology (Fig. 3a). While in the case of m-BPDM/CMTG/PaaM hydrogel and metalated (Zn^{+2}, Zn^{+2} /Cd^{+2}/Hg^{+2}) -m-BPDM/CMTG/PaaM hydrogel, a layered structure having uneven roughness in morphology was observed. The probable cause of this is the metal insertion into hydrogel matrices, which helps it to form a complex with embedded m-BPDM into the hydrogel matrix. The sensed metal binds itself with the m-BPDM moiety, which is physically engulfed into a hydrogel matrix [57]. Hence, no significant change in morphology was observed, suggesting that m-BPDM physically interacts with the hydrogel matrix.

3.3 Sensing Studies

Our main goal is to design a biosensor that can sense Zn, Cd, and Hg metal ions in aqueous media. It is challenging to detect all three metal ions simultaneously because they are all spectroscopically silent. We have developed a facile pathway between the inorganic moiety m-BPDM and the hydrogel matrix to overcome the hydrophobicity of m-BPDM. The synthesized hydrogel was pinkish-red in color. Then, the synthesized m-BPDM modified CMTG/Paam hydrogel was utilized to check the sensing ability in water media. To perform this, the synthesized probe was kept in the predetermined concentration of metal salt and incubated at 100 rpm. Then, the color of the modified hydrogel was changed upon metalation, as shown in Fig. 5a.

However, other alkali, alkaline metal salts, and metal ions commonly co-exist with heavy metals in water systems. Therefore, the detection of toxic heavy spectroscopically silent metals has become necessary and challenging [59]. The effect of co-existing ion experiments was performed and it was observed that the synthesized hydrogel probe does not show any significant color change due to interfering ions, as presented in Fig. 5b. Although the polymer-based hydrogel matrix consists of a polar group, synthesized hydrogel can be responsible for capturing metal ions in the hydrogel matrix. But it cannot be inspected by the naked eye on-site detection. Hence, this study attributed that the synthesized probe's color changed after binding with zinc, cadmium, and mercury ions individually and in mixtures of these three metal ions, which allows the sensor to be used for on-site inspections. m-BPDM is bound explicitly to these metal ions when surrounded by other metal ions, causing the color change may attributed that the m-BPDM/CMTG/PaaM with Zn^{+2}, $Cd^{+2,}$ and Hg^{+2} has a relatively higher binding constant to the different metal ions as our previous report [40]. Hence, the synthesized hydrogel can be used as a 'sensor' to estimate the Zn^{+2}, Cd^{+2}, and Hg^{+2} ions even in the presence of other metal ions.

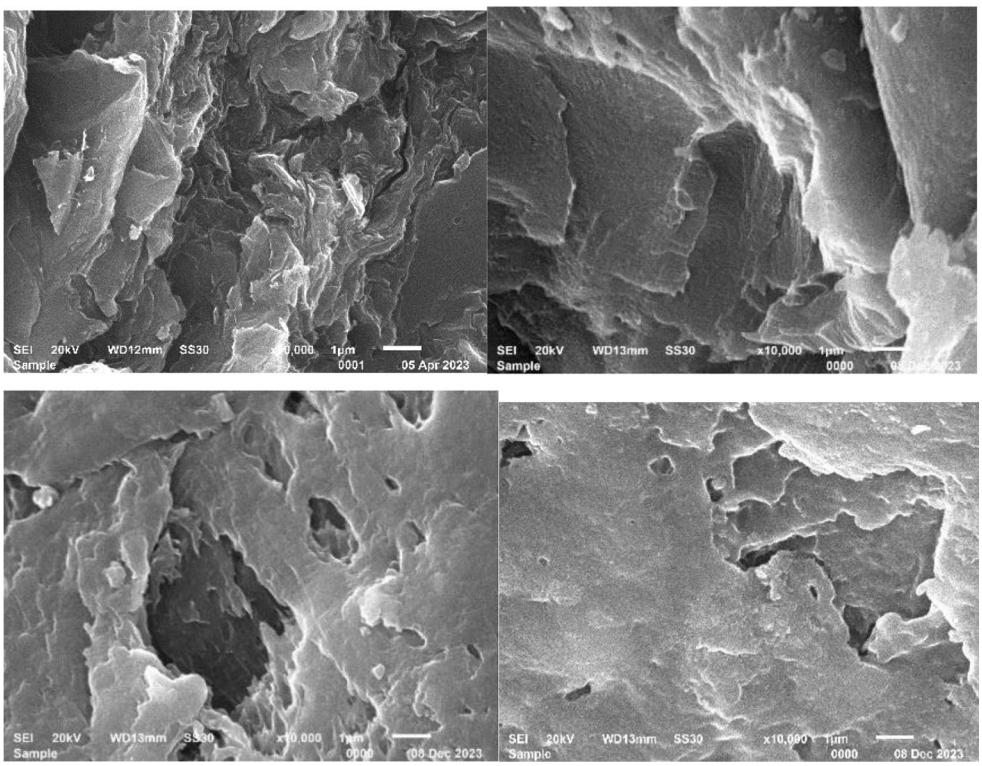


Fig. 4 SEM micrographs of (**a**) CMTG/PaaM hydrogel (Blank hydrogel), (**b**) m-BPDM/CMTG/PaaM hydrogel, (**c**) Zn^{+2} -m-BPDM/CMTG/PaaM hydrogel, and (**d**) Zn^{+2}/ Cd^{+2}/ Hg^{+2} -m-BPDM/CMTG/PaaM hydrogel

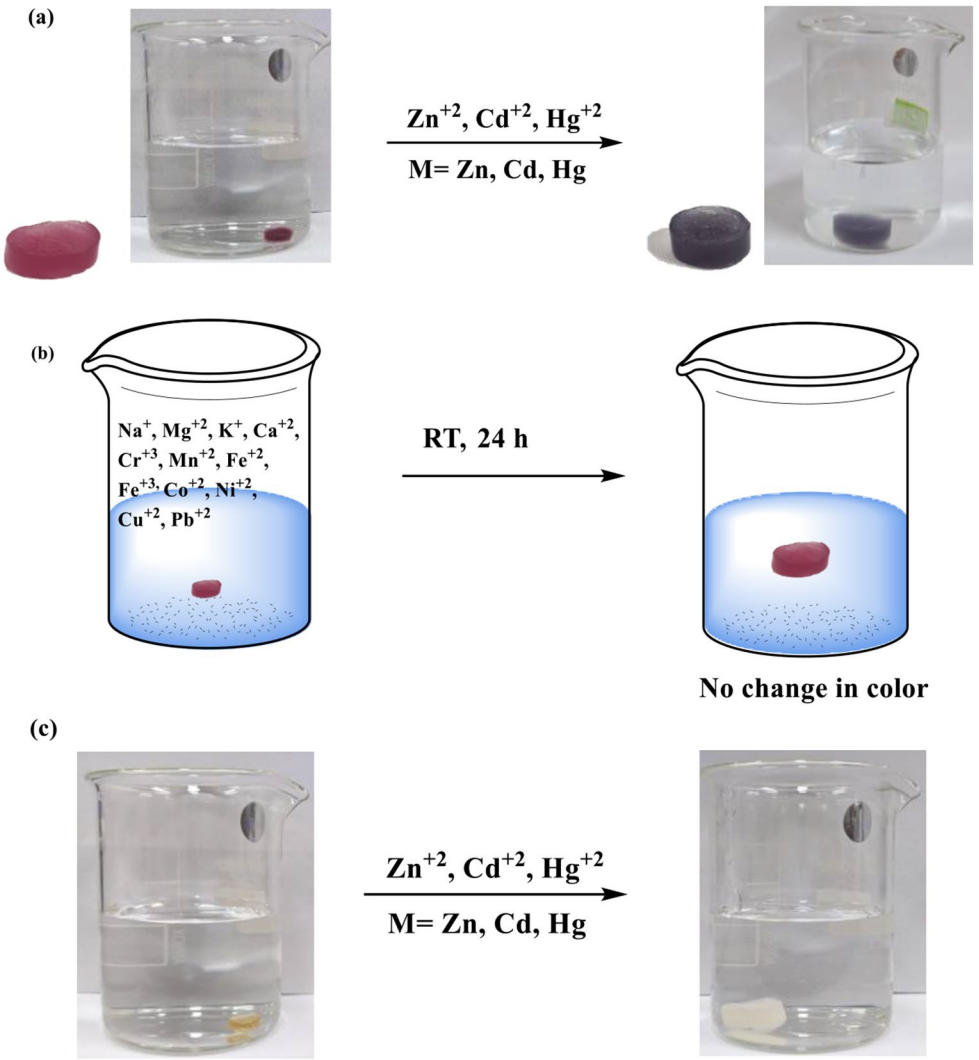


Fig. 5 Sensing studies of (**a**) Color change in the presence of Zn^{+2}, Cd^{+2}, and Hg^{+2}. (**b**) Effect of other ions, and (**c**) control experiment

Similarly, sensing experiments were performed for CMTG/PaaM hydrogel (control). In order to check the sensing of Zn^{+2}, $Cd^{+2,}$ and Hg^{+2} in aqueous media via CMTG/PaaM hydrogel, experiment was conducted with similar condition and it was observed that no color change was inspected by naked eyes as presented in Fig. 5c.

3.4 Impact of Sensing Time

We have evaluated the sensor's selectivity based on the color change with time. In acidic or basic conditions, the sensor has no significant effect and cannot be reversed once metalated. We submerged a piece of hydrogel in a test tube filled with a solution having various concentrations of metal ions. It was observed that modified hydrogel submerged with Hg^{+2} ion (0.5 mg L^{-1}) showed changes in 1 h at 33 °C and in 20 min at 60 °C. The time for sensing in Zn^{+2} and Cd^{+2} was higher at 33 °C and 60 °C, as shown in Fig. 6. Sensing studies of the minimum concentration of metal ion detected by hydrogel at 60 °C, respectively.

3.5 Temperature and Concentration-dependent Study

To elucidate the effect of temperature on a fixed concentration of metal ions, we performed a study to observe the color change of the synthesized sensor, as shown in Fig. 7a. It has been observed that at a constant temperature (30 °C) and fixed concentration (0.05 M) of individual metal ions, the color change was obtained in 40 min for Zn^{+2}, 100 min for Cd^{+2}, and 18 min for Hg^{+2}. While, at 60 °C, the colorimetric sensing was observed in 4 min for Hg^{+2}, 6 min for Zn^{+2}, and 55 min for Cd^{+2}. The possible reason for the sensing time may be attributed to binding or stability constants of particular metal ions with m-BPDM ligands [40]. The Hg^{+2} ions have different properties among other metal ions, while Zn^{+2} and Cd^{+2} have similar properties. It is concluded that

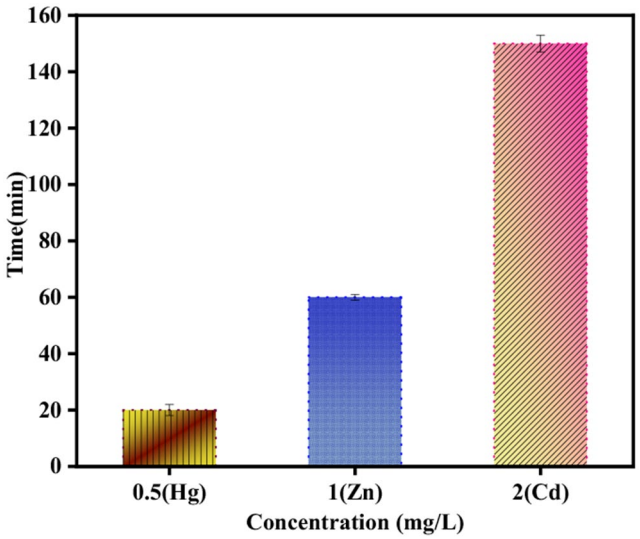


Fig. 6 Sensing studies of the minimum concentration of metal ion detected by hydrogel at 60 °C

according to the Hard–Soft-Acid-Base (HSAB) theory, a discrete conjugation pathway observed in the m-BPDM contains sp^2 and sp^3 at meso carbons, along with three pyrrolic –NH components in conjugation inside the m-BPDM binding sites. Therefore, m-BPDM has electronic conjugation with diffused electrons, so it should be treated as a soft ligand. Hence, Hg^{+2} binds comparatively better than Zn^{+2} and Cd^{+2}, considering Hg^{+2} is a soft acid, and the m-BPDM is a moderate or soft ligand.

Further, studies were conducted to observe the sensing time by the varying temperature at various concentrations of Zn^{+2}, Cd^{+2}, and Hg^{+2} respectively, as represented in Fig. 7b-d. It was concluded that as the temperature of solution increases, sensing time decreases, and the hydrogel requires a longer time to sense metal ions in low concentrations, while less time is needed for high metal ion concentrations. Hence, the detection time decreases as the temperature and metal ion concentration increase in aqueous solutions.

3.6 Selectivity and Co-existing Ions Interference Studies

To elucidate interference exhibited by alkaline, alkali, and other metal ions, a predetermined amount of metal salt solution (50 mg L^{-1}) was used to observe a color change in the hydrogel. Therefore, several heavy metal ions were used in the aqueous solution to analyze the synthesized hydrogel, including Pb^{+2}, Ni^{+2}, Co^{+2}, Cr^{+3}, Cd^{+2}, Hg^{+2}, and Zn^{+2}. The prefixed amount of modified hydrogel was dipped in salt solution individually and stirred at 100 rpm at room temperature. The observed color change is shown in Fig. 8. Among all metal ions, Zn^{+2}, Cd^{+2}, and Hg^{+2} show the colorimetric response by changing the color from pinkish red to dark blue.

3.7 Impact of pH and Concentration of Metal Ion

A study was carried out to check the effect of pH on the probe. To perform this study, different solutions of pH were made in the range of 2–10, and then synthesized hydrogel as a probe was used to check the colorimetric response of hydrogel in metal salts solution of zinc, cadmium, and mercury ions at different pH 2–10 solutions. A significant color change was observed in all pH solutions, but the results were slightly better near pH 6–7. This may be attributed to the fact that metal ions do not form hydroxides, and the protonation of acidic groups in the hydrogel matrix does not occur in this pH range, as depicted in Fig. 9a.

A study was also carried out to check the impact of metal salt concentration on synthesized sensor. To conduct this study, different concentrations (50 ppm to 500 ppm) of metal salts were prepared in an aqueous solution. As the concentration of metal ions increases from 50 ppm to 500 ppm, the time required for sensing decreases, suggesting that metal ion concentration is inversely proportional to the sensing time, as represented in Fig. 9b. Hence, it is concluded that the sensing time is in the order of $Hg^{+2} > Zn^{+2} > Cd^{+2}$.

Table 2 Impact of sensing time at different temperatures

Test Performed	Solvent (mL)	Amount of synthesized Probe (mg)	Speed (RPM)	Time(h) of the color change of hydrogel at 33°C	Time(h/min) of the color change of hydrogel at 60 °C
H_2O	50	30	100	No change	No change
$ZnCl_2$(1 mg L^{-1})	50	30	100	3 h	1 h
$HgCl_2$ (0.5 mg L^{-1})	50	30	100	1 h	20 min
$CdCl_2$ (2 mg L^{-1})	50	30	100	4.5 h	2.5 h
$ZnCl_2$(1 mg L^{-1}) + $HgCl_2$ (0.5 mg L^{-1}) + $CdCl_2$ (2 mg L^{-1})	50	30	100	1 h	20 min

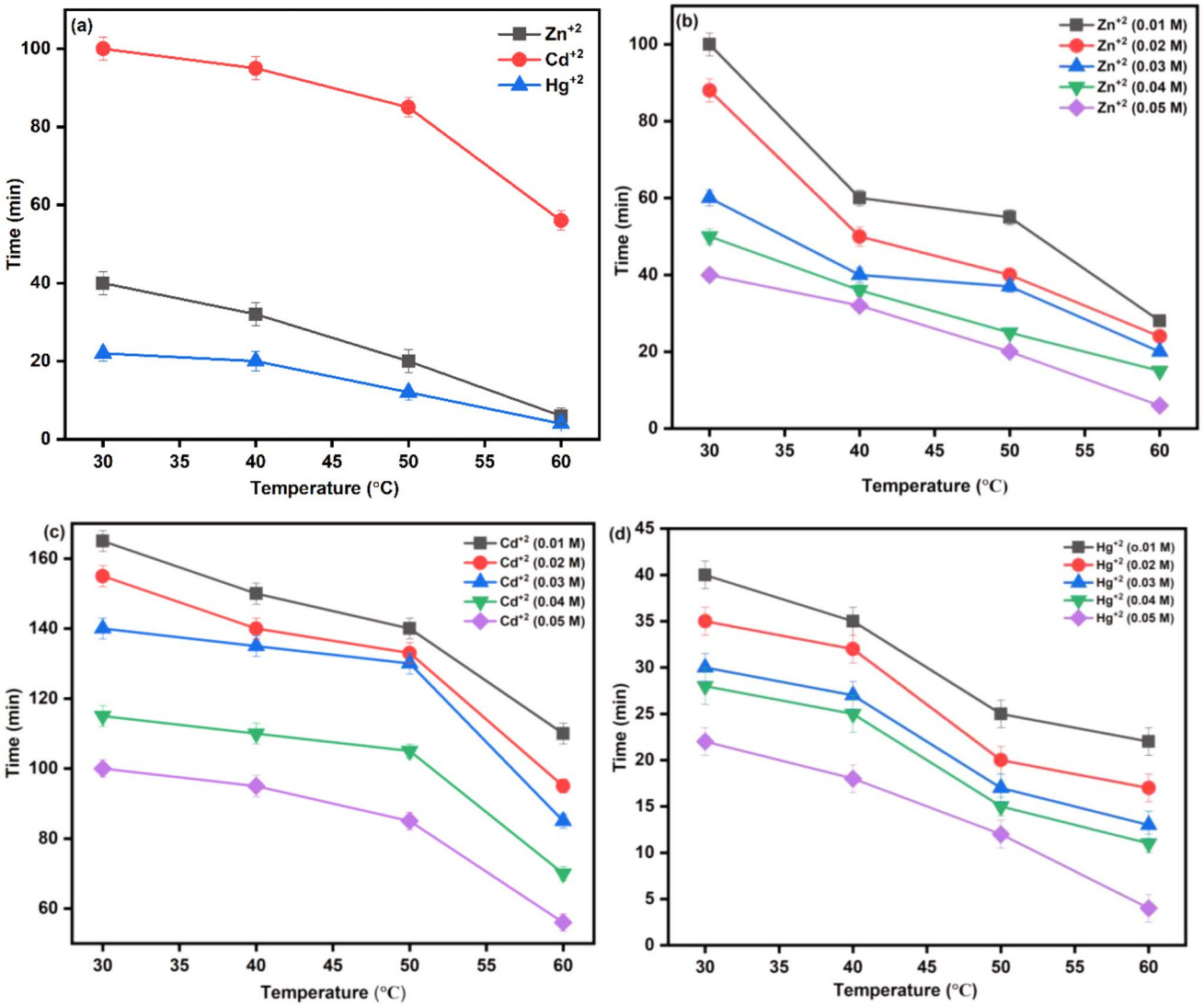


Fig. 7 (**a**) Temperature-dependent sensing at concentration 0.05 M (mol L^{-1}), (**b**), (**c**), and (**d**) concentration-dependent sensing study of Metal ion (M^{+2}) respectively

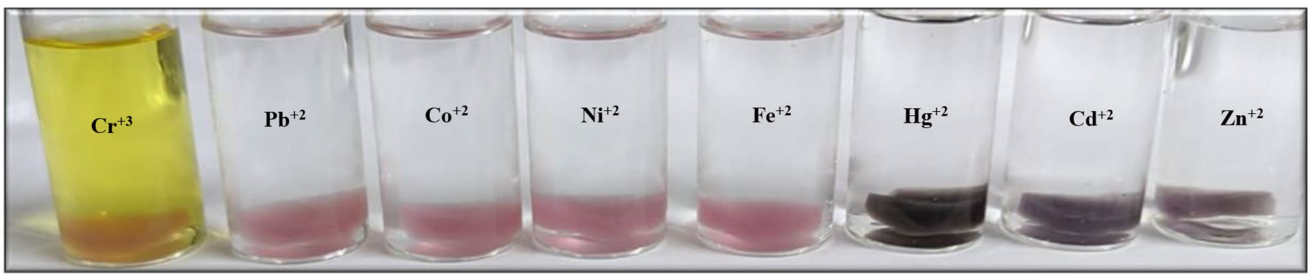


Fig. 8 The pictograph for the effect of co-existing ions on sensing

3.8 Proposed Mechanism of Sensing

In this research article, our aim is to sense Zn^{+2}, Cd^{+2}, and Hg^{+2} in aqueous media using m-BPDM based hydrogel matrix. m-BPDM is an inorganic moiety that is insoluble in water and responsible for sensing of Zn^{+2}, Cd^{+2}, and Hg^{+2}. To overcome the insolubility constraint, a CMTG/PaaM-based hydrogel matrix was used. The Biopolymer CMTG provides a hydrogel matrix that is cost-effective and biocompatible. The CMTG is composed of the D-form

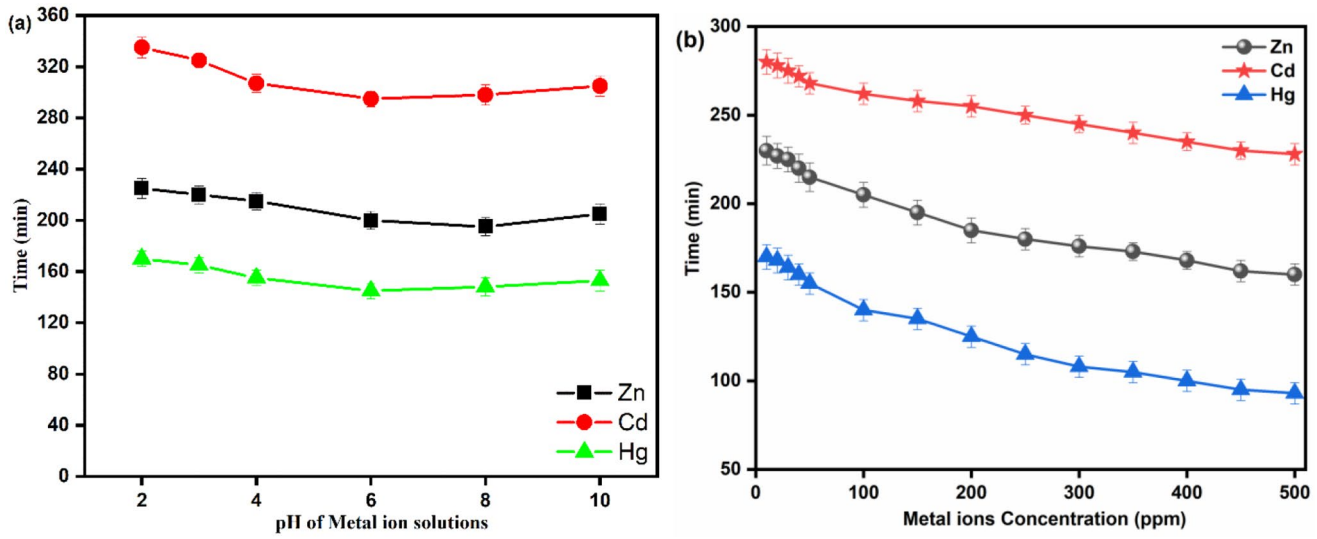


Fig. 9 (**a**) Plot of variation in pH of metal ions (1ppm) v/s time, and (**b**) metal ion concentration vs. time

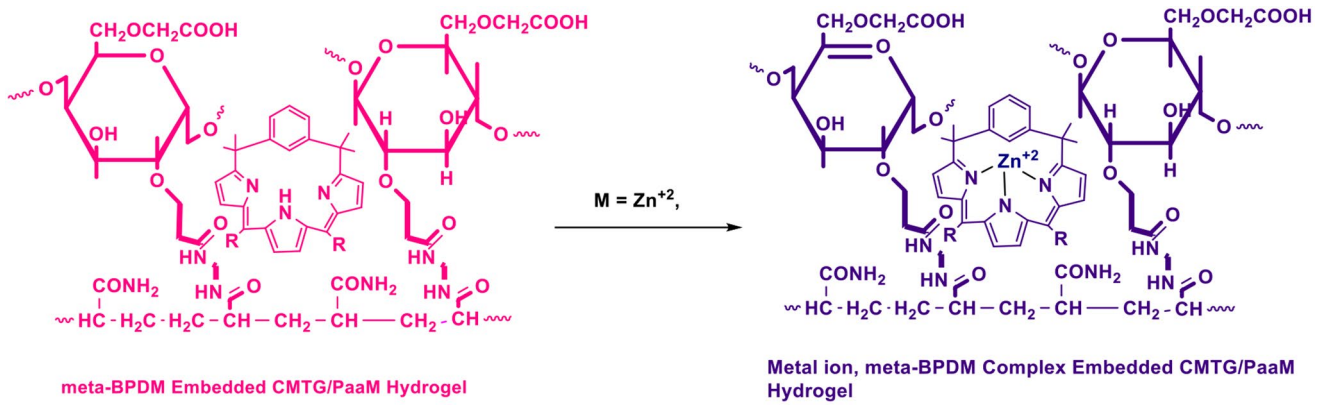


Fig. 10 Mechanism of sensing

Table 3 Different sensor probes for sensing Zn^{+2}, Cd^{+2}, and Hg^{+2}

Materials	Metals	Detection Limit	Response Time	Ref.
TiO_2/poly(acrylamide-co methylenbisacrylamide) nanocomposite	Hg^{+2}	10 ppm	15 min	[60]
PVA based hydrogel	Co ions	0.001 ppm	3 h	[61]
AgNPs/PVA nanocomposite hydrogels	Hg^{+2}	25nM.	1 h	[62]
CMG/PaaM based composite	Zn^{+2},	0.5 ppm	1 h	[57]
	Cd^{+2},	1 ppm	2 h	
	Hg^{+2}	2 ppm	3 h	
Enzyme based hydrogel	Hg^{+2}	50 µM	2 h	[63]
poly(NVP-co-AMPSA)-based hydrogels	Hg^{+2}	0.55ppm	-	[64]
Present work	Zn^{+2}	0.5 ppm	1 h	
	Cd^{+2}	1 ppm	2.5 h	
	Hg^{+2}	2 ppm	18 min	

of xylose, galactose, and glucose units in a ratio of 2:1:3, having hydrophilic groups in its structure that are responsible for water retention, and PaaM provides strength to the matrix, while m-BPDM is an inorganic moiety that is used as a filler in the hydrogel matrix. m-BPDM is responsible for sensing of Zn^{+2}, Cd^{+2}, and Hg^{+2}. When m-BPDM embedded CMTG/PaaM was submerged in salt solution then, the N group present in the ring of m-BPDM binds with the metal present in solution ($M^{+2} = Zn^{+2}$, Cd^{+2}, and Hg^{+2}) and imparts color changes from pinkish red to blue as presented in Fig. 10.

3.9 Different Sensing Materials for Sensing of Zn^{+2}, Cd^{+2}, and Hg^{+2}

The study also compared the detection limit and response time of some of the newly reported sensor probes of similar material for the sensing of Zn^{+2}, Cd^{+2}, and Hg^{+2} presented in Table 3. The results showed that the sensor probe used

in the present study was highly effective for sensing Zn^{+2}, Cd^{+2}, and Hg^{+2} in short time frame compared with other sensors.

3.10 Impact on Industrial Wastewater

To check the utility of synthesized sensors in actual field applications. Usually, the real effluent has a pH ranging from 6 to 9 [65]. To perform this study, wastewater was collected from the water treatment plant of Delhi Technological University (DTU), India, and industrial effluent from Dharuhera, Rewari, India. The effluent samples were utilized as received without any prior treatment. A test was performed to assess the synthesized hydrogel's ability in which the sensor was added directly to the effluent sample. No color change was observed in a sample collected from DTU. It may be attributed to the fact that no metal ions were present in a sample collected from DTU. Hence, synthesized hydrogel can be utilized to check the presence or absence of metal ions in water bodies.

To study the potential of synthesized hydrogel in industrial effluent, the pH of the sample was checked, and it was found to be 6.5. Further, a hydrogel disc was dipped into collected industrial effluent, and the solution was kept in an incubator at 100 rpm for 1 h at ambient temperature. It shows that the color changes from pinkish red to dark blue. This is attributed to the presence of Zn^{+2}, Cd^{+2}, and Hg^{+2} ions in the collected sample of wastewater represented in Fig. 11. Further, the presence of $Zn^{+2,}$ Cd^{+2}, and Hg^{+2} ions in the pollutant was quantitatively determined by AAS. The results revealed the concentration (ppm) of Zn^{+2} (8), Cd^{+2} (0.035), and Hg^{+2} (1.4), respectively. Hence, the concentration of Zn^{2+} and Hg^{2+} is higher than the sensor's detection limit, attributed to the fact that both are responsible for color change.

3.11 Colorimetric Response of Zn^{+2} in *E. Coli.* Bacterial Cells

Cell lysis is a method that is used to break down cell membranes so that they can release protein, DNA, etc., from a cell, resulting in the formation of voids in the cell, which shows complete cell lysis. Surfactants are commonly used in cell lysis processes and may be ionic, non-ionic, or zwitterionic. Cell lysis can be attained by denaturing cell proteins with potent ionic detergents such as sodium dodecyl sulfate (SDS) in a few seconds [66]. The choice of surfactant must be appropriate because it affects the speed of cell lysis. In this study, we have observed the colorimetric response of m-BPDM-modified CMTG/Paam hydrogel as a sensor for the presence of Zn^{+2} ions in *E. coli* cells. Escherichia coli NCIMP-1 (*E. coli*) was used to perform this study. At first, the cells were nurtured in a medium containing 2% Luria broth at 37 °C and grow for 24 h. Further, the cells were

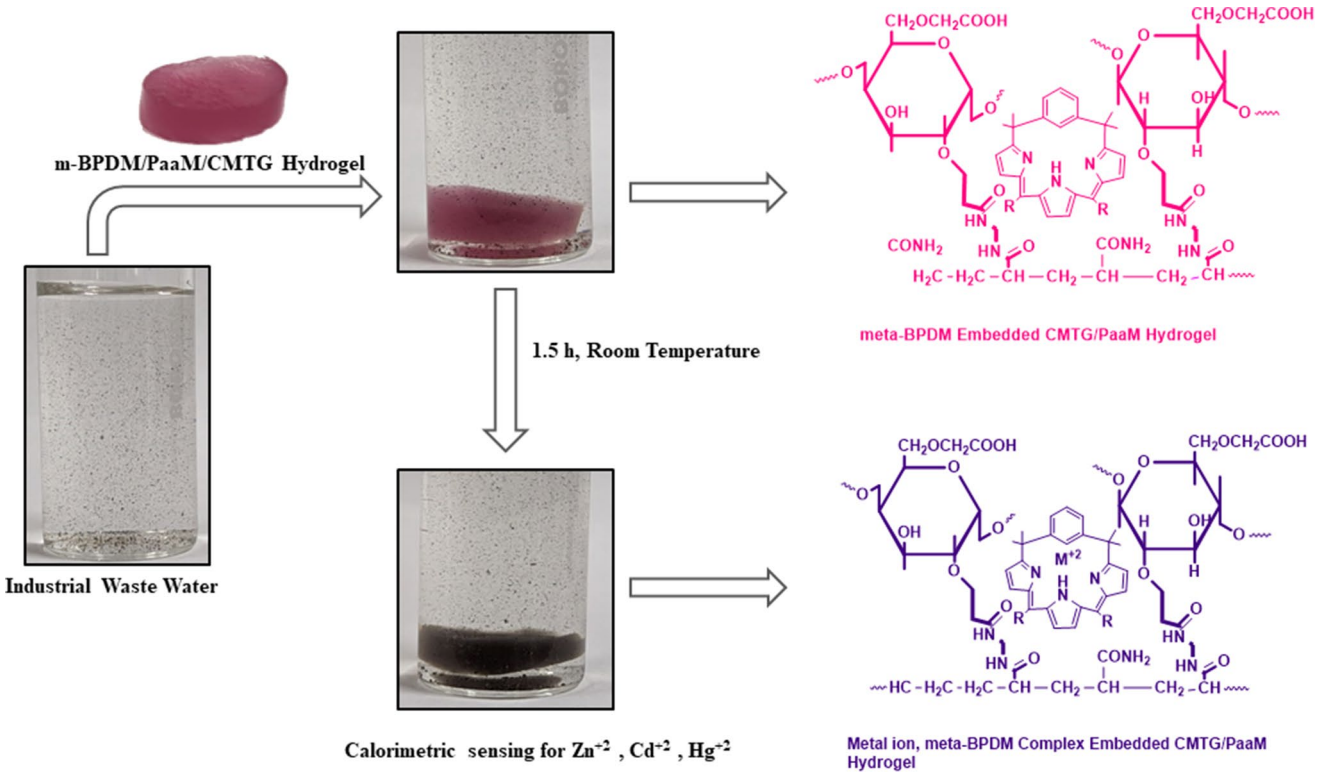


Fig. 11 Colorimetric response in Industrial wastewater; all the data are presented by the mean ± standard deviation, ($P < 0.05$)

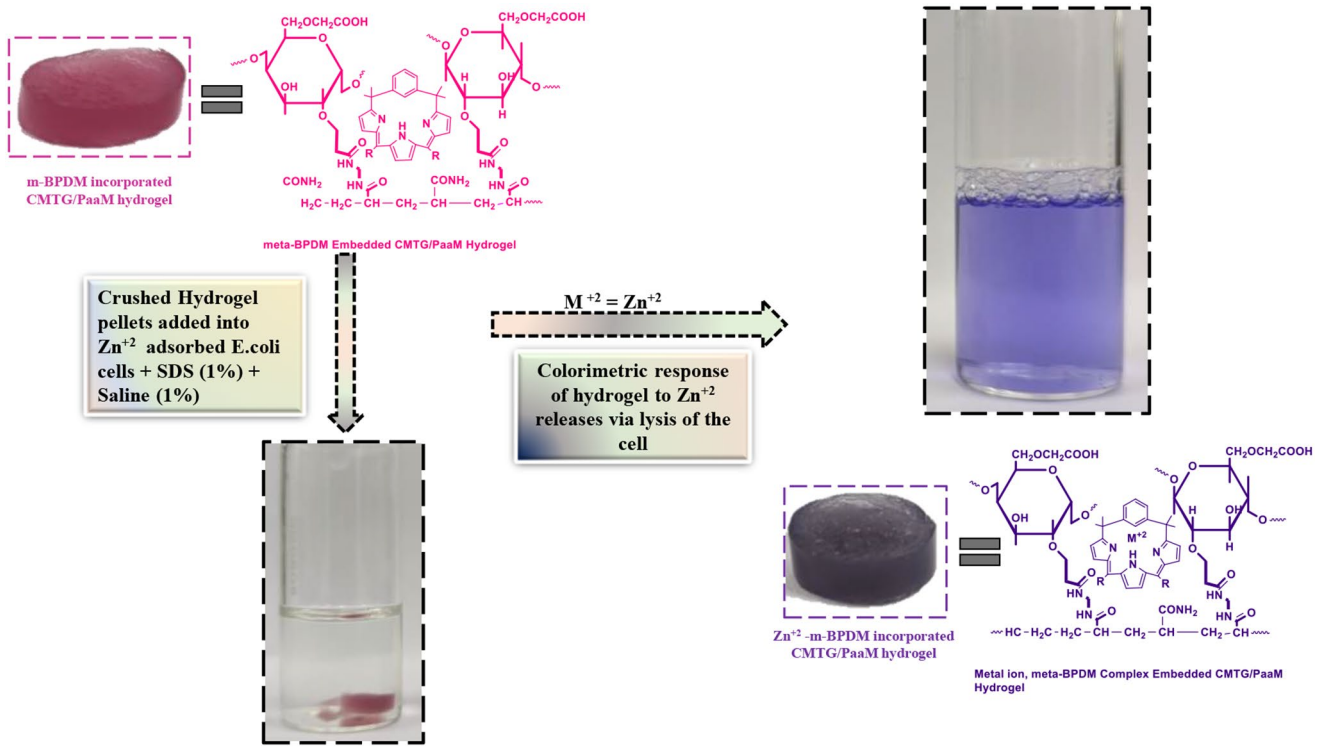


Fig. 12 Colorimetric response for Zn ions absorbed by E. Coli cells; all the data are presented by the mean ± standard deviation, ($P < 0.05$)

removed from the media in a culture tube and washed with Phosphate buffer solution (PBS) (3 × 1 mL). Then, to examine the effects of Zn^{+2}, the cells were incubated for 15 h in PBS solution (1%, 1 mL) with Zn^{+2} aqueous solution ($ZnCl_2$, 1 ppm, 5 mL). Then, the cells were incubated and washed with PBS (3 × 1 mL), and the leftover Zn^{+2} solution was removed from the cells. The synthesized probe (2 mg of dried broken hydrogel) was added and incubated with the pretreated cells with zinc ions (1 ppm) in saline solution (1 mL) in two different culture tubes. Then, 1 mL of surfactant (1% SDS) was added in one of the culture tubes, while no surfactant was added in the second test tube. The m-BPDM-modified hydrogel matrix absorbs Zn^{2+} from *E. Coli* cells and exhibits a colorimetric response from pinkish red to dark blue within three hours. This suggested that Zn^{+2} was present in the aqueous media, indicating that cell lysis may have triggered the release of Zn^{+2} from the cell shown in Fig. 12. The hydrogel's color did not change without the surfactant's solution. Hence, it was confirmed that no cell lysis was obtained in the controlled test [50]. The colorimetric response is unique to Zn^{+2} ions with a wide range of physiologically significant mono and divalent metal ions, including Na^{+}, K^{+}, Ca^{+2}, Mg^{+2}, etc. In spite of additional metal ions being present, m-BPDM specifically binds to Zn^{+2}, causing the sensor to change color. Therefore, Zn^{+2} in plant and animal cells can be sensed this way.

4 Conclusion

In the present study, a novel m-BPDM modified CMTG/PaaM hydrogel-based sensor was successfully developed by in situ insertion of m-BPDM. Since m-BPDM is a hydrophobic bioinorganic moiety, it hinders their utility in aqueous media. To overcome its hydrophobicity, it was dispersed in a hydrogel matrix using a 1:1 ratio of acetone and water, and further colorimetric response was observed for Zn, Cd, and Hg metal ions in water. It was observed that the synthesized modified hydrogel could be applied as a sensor probe for successful sensing of Zn, Cd, and Hg metal ions in water and shows a colorimetric response from pinkish red to dark blue. The sensor recognized that the detection limit for Hg ions is up to 0.5 mg L^{-1}, comparable to the above permissible limit, and the detection limit for Zn and Cd was 1 mg L^{-1} and 2 mg L^{-1}, respectively. Moreover, it has been concluded that as the temperature increases, sensing time decreases, and the concentration of metal ions is inversely proportional to the sensing time. In addition, it was implemented to sense the Zn^{+2}, Cd^{+2}, and Hg^{+2} from industrial effluent and shows color change upon the presence of Zn, Cd, and Hg ions. It was further confirmed by quantitatively using AAS. Further, the probe was also used to sense Zn^{+2} ion in *E. Coli* bacterial cells with high selectivity and showed color change upon metalation. Hence, it can be concluded that the probe is a cost-effective and

eco-friendly biopolymer-based matrix that senses the spectroscopically silent (Zn and Cd, and Hg metal ions) from an aqueous medium.

Author Contributions Priyanka Yadav: executed experiment, writing; Sudhir G. Warkar: Conceptualization and Editing; Anil Kumar: Editing and Conceptualization.

Funding The authors declare that no funds and grants were received during the preparation of this manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

1. F. Ahmed, I. Ali, S. Kousar, S. Ahmed, The environmental impact of industrialization and foreign direct investment: empirical evidence from Asia-Pacific region. Environ. Sci. Pollut Res. **29**, 29778–29792 (2022). https://doi.org/10.1007/s11356-021-17560-w
2. S. Pandey, E. Makhado, S. Kim, M. Kang, Recent developments of polysaccharide based superabsorbent nanocomposite for organic dye contamination removal from wastewater — a review. Environ. Res. **217**, 114909 (2023). https://doi.org/10.1016/j.envres.2022.114909
3. K.H. Hama Aziz, F.S. Mustafa, K.M. Omer, S. Hama, R.F. Hamarawf, K.O. Rahman, Heavy metal pollution in the aquatic environment: efficient and low-cost removal approaches to eliminate their toxicity: a review. RSC Adv. **13**, 17595–17610 (2023). https://doi.org/10.1039/d3ra00723e
4. N. Saruchi, V. Kumar, D. Bhatt, S. Pandey, A.A. Ghfar, Synthesis and characterization of silver nanoparticle embedded cellulose-gelatin based hybrid hydrogel and its utilization in dye degradation. RSC Adv. **13**, 8409–8419 (2023). https://doi.org/10.1039/d2ra03885d
5. S. Pandey, S. Kim, Y.S. Kim, D. Kumar, M. Kang, Fabrication of next-generation multifunctional LBG-s-AgNPs@ g-C3N4 NS hybrid nanostructures for environmental applications. Environ. Res. **240**(117540) (2024). https://doi.org/10.1016/j.envres.2023.117540
6. A.A.S. Alahl, H.A. Ezzeldin, A.A. Al-Kahtani, S. Pandey, Y.H. Kotp, Synthesis of a Novel Photocatalyst based on Silicotitanate nanoparticles for the removal of some Organic Matter from Polluted Water. Catalysts. **13** (2023). https://doi.org/10.3390/catal13060981
7. J. Briffa, E. Sinagra, R. Blundell, Heliyon Heavy metal pollution in the environment and their toxicological effects on humans, *Heliyon*, vol. 6, pp. e04691, 2020, https://doi.org/10.1016/j.heliyon.2020.e04691
8. R.A. Hauser-Davis, N. Wosnick, Climate Change Implications for Metal and Metalloid Dynamics in aquatic ecosystems and its context within the Decade of Ocean Sciences. Water (Switzerland), **14**, https://doi.org/10.3390/w14152415
9. M. Balali-Mood, K. Naseri, Z. Tahergorabi, M.R. Khazdair, M. Sadeghi, Toxic mechanisms of five heavy metals: Mercury, lead, Chromium, Cadmium, and Arsenic. Front. Pharmacol. **12**, 1–19 (2021). https://doi.org/10.3389/fphar.2021.643972
10. K. Jomova et al., Essential metals in health and disease. Chem. Biol. Interact. **367**, 110173 (2022). https://doi.org/10.1016/j.cbi.2022.110173
11. H. Ali, E. Khan, I. Ilahi, Environmental chemistry and ecotoxicology of hazardous heavy metals: environmental persistence, toxicity, and bioaccumulation. J. Chem. vol. (2019). https://doi.org/10.1155/2019/6730305
12. A. Anandababu, S. Anandan, A. Syed, N. Marraiki, M. Ashokkumar, Upper rim modified calix[4]arene towards selective turn-on fluorescence sensor for spectroscopically silent metal ions. Inorganica Chim. Acta. **516**(120133) (2021). https://doi.org/10.1016/j.ica.2020.120133
13. M. Zaynab et al., Health and environmental effects of heavy metals. J. King Saud Univ. - Sci. **34**(101653) (2022). https://doi.org/10.1016/j.jksus.2021.101653
14. M.S. Braga, R.F.V.V. Jaimes, W. Borysow, O.F. Gomes, W.J. Salcedo, Portable multispectral colorimeter for metallic ion detection and classification. Sensors, **17**, https://doi.org/10.3390/s17081730
15. S. Xia et al., Ratiometric fluorescent and colorimetric BODIPY-based sensor for zinc ions in solution and living cells. Sens. Actuators B Chem. **258**, 1279–1286 (2018). https://doi.org/10.1016/j.snb.2017.11.129
16. R.K. Sharma, L.K. Gajanan, M.S. Mehata, F. Hussain, A. Kumar, Synthesis, characterization and fluorescence turn-on behavior of new porphyrin analogue: meta-benziporphodimethenes. Spectrochim Acta - Part. Mol. Biomol. Spectrosc. **169**, 58–65 (2016). https://doi.org/10.1016/j.saa.2016.06.020
17. D. Ahluwalia, A. Kumar, S.G. Warkar, M.M. Deshmukh, Effect of substitutions on the geometry and intramolecular hydrogen bond strength in meta-benziporphodimethenes: a new porphyrin analogue. J. Mol. Struct. **1220**, 128773 (2020). https://doi.org/10.1016/j.molstruc.2020.128773
18. S. Mitra et al., Impact of heavy metals on the environment and human health: novel therapeutic insights to counter the toxicity. J. King Saud Univ. - Sci. **34**(101865) (2022). https://doi.org/10.1016/j.jksus.2022.101865
19. P.N. Obasi, B.B. Akudinobi, Potential health risk and levels of heavy metals in water resources of lead–zinc mining communities of Abakaliki, Southeast Nigeria. Appl. Water Sci. **10**, 1–23 (2020). https://doi.org/10.1007/s13201-020-01233-z
20. Y. Dai, K. Yao, J. Fu, K. Xue, L. Yang, K. Xu, A novel 2-(Hydroxymethyl)quinolin-8-ol-based selective and sensitive fluorescence probe for Cd2 + ion in water and living cells. Sens. Actuators B Chem. **251**, 877–884 (2017). https://doi.org/10.1016/j.snb.2017.05.103
21. V. Vinod Kumar, S.P. Anthony, Silver nanoparticles based selective colorimetric sensor for cd 2+, Hg2 + and Pb2 + ions: tuning sensitivity and selectivity using co-stabilizing agents. Sens. Actuators B Chem. **191**, 31–36 (2014). https://doi.org/10.1016/j.snb.2013.09.089
22. M. Hong, X. Lu, Y. Chen, D. Xu, A novel rhodamine-based colorimetric and fluorescent sensor for Hg2 + in water matrix and living cell. Sens. Actuators B Chem. **232**, 28–36 (2016). https://doi.org/10.1016/j.snb.2016.03.125
23. D. Prabhakaran, H. Nanjo, H. Matsunaga, Naked eye sensor on polyvinyl chloride platform of chromo-ionophore molecular assemblies: a smart way for the colorimetric sensing of toxic metal ions. Anal. Chim. Acta. **601**, 108–117 (2007). https://doi.org/10.1016/j.aca.2007.08.032
24. S. Pandey, N. Son, S. Kim, D. Balakrishnan, M. Kang, Locust Bean gum-based hydrogels embedded magnetic iron oxide nanoparticles nanocomposite: Advanced materials for environmental and

energy applications. Environ. Res. **214**(114000) (2022). https://doi.org/10.1016/j.envres.2022.114000
25. T. Hu, Q. Lai, W. Fan, Y. Zhang, Z. Liu, Advances in Portable Heavy Metal Ion Sensors. Sensors, **23**, https://doi.org/10.3390/s23084125
26. A. Kumar, C.H. Hung, S. Rana, M.M. Deshmukh, Study on the structure, stability and tautomerisms of meta-benziporphodimethene and N-Confused isomers containing γ–lactam ring. J. Mol. Struct. **1187**, 138–150 (2019). https://doi.org/10.1016/j.molstruc.2019.03.064
27. H. Liu et al., A selective colorimetric and efficient removal strategy for mercury (II) using mesoporous silver-melamine nanocomposites synthesized by controlled supramolecular self-assembly. J. Hazard. Mater. **388**(121798) (2020). https://doi.org/10.1016/j.jhazmat.2019.121798
28. N. Kwon, Y. Hu, J. Yoon, Fluorescent chemosensors for various analytes including reactive oxygen species, Biothiol, Metal ions, and toxic gases. ACS Omega. **3**, 13731–13751 (2018). https://doi.org/10.1021/acsomega.8b01717
29. M. Li et al., Current trends in the detection and removal of heavy metal ions using functional materials. Chem. Soc. Rev. **52**, 5827–5860 (2023). https://doi.org/10.1039/d2cs00683a
30. N. De Acha, C. Elosúa, J.M. Corres, F.J. Arregui, Fluorescent sensors for the detection of heavy metal ions in aqueous media. Sensors, **19**, https://doi.org/10.3390/s19030599
31. K. Radhakrishnan, S. Sivanesan, P. Panneerselvam, Turn-On fluorescence sensor based detection of heavy metal ion using carbon dots@graphitic-carbon nitride nanocomposite probe. J. Photochem. Photobiol Chem. **389**(112204) (2020). https://doi.org/10.1016/j.jphotochem.2019.112204
32. O. Yadav, A. Varshney, A. Kumar, R.K. Ratnesh, M.S. Mehata, A2B corroles: fluorescence signaling systems for sensing fluoride ions. Spectrochim Acta - Part. Mol. Biomol. Spectrosc. **202**, 207–213 (2018). https://doi.org/10.1016/j.saa.2018.05.051
33. N. Nataraj, P. Dash, R. Sakthivel, Y.C. Lin, H.W. Fang, R.J. Chung, Simultaneous electrochemical and colorimetric detection of tri-heavy metal ions in environmental water samples employing 3D-MOF/nickel selenide as a synergistic catalyst. Chem. Eng. J. **485**(1), 149965 (2024). https://doi.org/10.1016/j.cej.2024.149965
34. M. Shahbaz et al., Recent advances in the fluorimetric and colorimetric detection of cobalt ions. RSC Adv. **14**(14), 9819–9847 (2024). https://doi.org/10.1039/d4ra00445k
35. M. Zareh Jonaghani, H. Zali-Boeini, Highly selective fluorescent and colorimetric chemosensor for detection of Hg2 + ion in aqueous media. Spectrochim Acta - Part. Mol. Biomol. Spectrosc. **178**, 66–70 (2017). https://doi.org/10.1016/j.saa.2017.01.065
36. B. Molecules, Gold Nanoparticle-Based Colorimetric Sensors: Properties, 2023
37. X. Zeng, X. Li, W. Sun, Highly selective and sensitive colorimetric chemosensor based on tricarboyanine for detection of ag + in industrial wastewater. J. Leather Sci. Eng. **2**(1) (2020). https://doi.org/10.1186/s42825-020-00031-2
38. T.H. Nguyen, T. Sun, K.T.V. Grattan, A Turn-On fluorescence-based Fibre Optic Sensor for the detection of Mercury. Sens. (Basel). **19**(9) (2019). https://doi.org/10.3390/s19092142
39. D. Ahluwalia, A. Kumar, S.G. Warkar, Recent developments in meta -benziporphodimethene: a new porphyrin analogue. J. Mol. Struct. **1228**, 129672 (2021). https://doi.org/10.1016/j.molstruc.2020.129672
40. R.K. Sharma, A. Maurya, P. Rajamani, M.S. Mehata, A. Kumar, meta-benziporphodimethenes: New Cell-Imaging Porphyrin Analogue molecules. ChemistrySelect. **1**(13), 3502–3509 (2016). https://doi.org/10.1002/slct.201600812
41. D. Chauhan, A. Kumar, S.G. Warkar, Synthesis, characterization and metal ions sensing applications of meta-benziporphodimethene-embedded polyacrylamide/carboxymethyl guargum polymeric hydrogels in water. Environ. Technol. **43**, 991–1002 (2022). https://doi.org/10.1080/09593330.2020.1812730
42. Khushbu, S.G. Warkar, Potential applications and various aspects of polyfunctional macromolecule- carboxymethyl tamarind kernel gum. Eur. Polym. J. **140**(110042) (2020). https://doi.org/10.1016/j.eurpolymj.2020.110042
43. J. Ham, K. Lim, Y.J. Kim, J. Yeong Kim, J.W. Oh, W.G. Koh, Colorimetric detection of ammonia using an adhesive, stretchable hydrogel patch. Chem. Eng. J. **479**(147596) (2024). https://doi.org/10.1016/j.cej.2023.147596
44. J. Wang, D. Wang, Z. Su, Y. Song, J. Zhang, Y. Xiahou, Green synthesis of chitosan/glutamic acid/agarose/Ag nanocomposite hydrogel as a new platform for colorimetric detection of Cu ions and reduction of 4-nitrophenol. Int. J. Biol. Macromol. **259**(129394) (2024). https://doi.org/10.1016/j.ijbiomac.2024.129394
45. Z. Peng, H.R. Yu, J.Y. Wen, Y.L. Wang, T. Liang, C.J. Cheng, A novel ion-responsive photonic hydrogel sensor for portable visual detection and timely removal of lead ions in water. Mater. Adv. 5393–5405 (2022). https://doi.org/10.1039/d2ma00232a
46. Y. Hong, Z. Lin, Y. Yang, T. Jiang, J. Shang, Z. Luo, Biocompatible conductive hydrogels: applications in the field of Biomedicine. Int. J. Mol. Sci., **23**, https://doi.org/10.3390/ijms23094578
47. P. Dallas, V. Georgakilas, *NU*. Elsevier B.V., 2015. https://doi.org/10.1016/j.cis.2015.07.008
48. M. Munisamy et al., β-Cyclodextrin-encapsulated rhodamine derivatives core–Shell microspheres—based fluorescent sensor for Au3 + and Template for Generating microplates of Gold. Micromachines, **14**, https://doi.org/10.3390/mi14071443
49. X. Zheng, Y. Gao, X. Ren, G. Gao, Polysaccharide-tackified composite hydrogel for skin-attached sensors. J. Mater. Chem. C **9**, 3343–3351 (2021). https://doi.org/10.1039/d0tc05589a
50. D.C.A.K.S.G. Warkar, An efficient adsorbent for the removal of zn 2 + cd 2 + and hg 2 + from the real industrial effluents. Int. J. Environ. Sci. Technol. **19**, 1483–1494 (2022). https://doi.org/10.1007/s13762-021-03615-5
51. F.J. Arregui, Z. Ciaurriz, M. Oneca, I.R. Matias, An experimental study about hydrogels for the fabrication of optical fiber humidity sensors. Sens. Actuators B Chem. **96**, 165–172 (2003). https://doi.org/10.1016/S0925-4005(03)00520-3
52. U. Baruah, D. Chowdhury, Functionalized graphene oxide quantum dot-PVA hydrogel: a colorimetric sensor for Fe2+, Co2 + and Cu2 + ions. Nanotechnology, **27**, https://doi.org/10.1088/0957-4484/27/14/145501
53. S. Rana, R.K. Sharma, N. Fridman, A. Kumar, Structural characterization and bioimaging of Zn2 + using meta-benziporphodimethene analogue. Luminescence. **38**, 1268–1274 (2023). https://doi.org/10.1002/bio.4382
54. A. Zarbab, A. Sajjad, A. Rasul, F. Jabeen, M. Javaid Iqbal, Synthesis and characterization of Guar gum based biopolymeric hydrogels as carrier materials for controlled delivery of methotrexate to treat colon cancer. Saudi J. Biol. Sci. **30**, 103731 (2023). https://doi.org/10.1016/j.sjbs.2023.103731
55. R. Kajal, P. Kumar, Meena, S.G. Warkar, Development and characterization of pH-responsive CMTKG/PAM/PEG hydrogel for oral administration of etophylline. Colloid Polym. Sci. **301**, 1313–1323 (2023). https://doi.org/10.1007/s00396-023-05152-8
56. P. Meena, P. Singh, S.G. Warkar, Development and assessment of carboxymethyl tamarind kernel gum-based pH-responsive hydrogel for release of diclofenac sodium. Eur. Polym. J. **197**(112340) (2023). https://doi.org/10.1016/j.eurpolymj.2023.112340
57. D. Chauhan, A. Kumar, S.G. Warkar, Modified polymeric hydrogels for the detection of Zn in E. Coli bacterial cells and zn, cd and hg in industrial effluents. Environ. Technol. 1–8 (2021). https://doi.org/10.1080/09593330.2021.1928294

58. Y. Tushar, P. Saraswat, Meena, S.G. Warkar, Synthesis and characterization of novel xanthan gum-based pH-sensitive hydrogel for metformin hydrochloride release. Colloid Polym. Sci. **301**, 1147–1158 (2023). https://doi.org/10.1007/s00396-023-05135-9
59. R. Paolesse, S. Nardis, D. Monti, M. Stefanelli, C. Di Natale, Porphyrinoids for Chemical Sensor Applications. Chem. Rev. **117**, 2517–2583 (2017). https://doi.org/10.1021/acs.chemrev.6b00361
60. R. Sedghi, B. Heidari, M. Behbahani, Synthesis, characterization and application of poly(acrylamide-co-methylenbisacrylamide) nanocomposite as a colorimetric chemosensor for visual detection of trace levels of hg and pb ions. J. Hazard. Mater. **285**, 109–116 (2015). https://doi.org/10.1016/j.jhazmat.2014.11.049
61. D. Kim, J. Kim, S. Jo, T.S. Lee, Colorimetric detection of Co ions by a poly(vinyl alcohol)-based hydrogel using color coordinate. Dye Pigment. **197**(109894) (2022). https://doi.org/10.1016/j.dyepig.2021.109894
62. A. Jeevika, D.R. Shankaran, Functionalized silver nanoparticles probe for visual colorimetric sensing of mercury. Mater. Res. Bull. **83**, 48–55 (2016). https://doi.org/10.1016/j.materresbull.2016.05.029
63. Y. Zhang, T. Ren, H. Tian, B. Jin, J. He, Hydrogel-encapsulated enzyme facilitates colorimetric Acute Toxicity Assessment of Heavy Metal ions. ACS Appl. Mater. Interfaces. **10**, 26705–26712 (2018). https://doi.org/10.1021/acsami.8b08949
64. A. Diacon et al., Dual-responsive hydrogels for Mercury Ion Detection and removal from Wastewater. Gels. **10** (2024). https://doi.org/10.3390/gels10020113
65. Y. Zhang et al., PH Effect on Heavy Metal Release from a polluted sediment. J. Chem. (2018). https://doi.org/10.1155/2018/7597640
66. M.S. Islam, A. Aryasomayajula, P.R. Selvaganapathy, A review on macroscale and microscale cell lysis methods. Micromachines, **8**, https://doi.org/10.3390/mi8030083

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

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

Check for updates

Received: 16 May 2024 | Revised: 16 July 2024 | Accepted: 30 July 2024

DOI: 10.1002/cta.4234

WILEY

ORIGINAL ARTICLE

Mathematical modeling and stress-aware stability analysis of a nonideal multiport Single Inductor DC–DC converter for renewable energy

Ashutosh Gupta | **Dheeraj Joshi**

Delhi Technological University, Delhi, India

Correspondence
Ashutosh Gupta, Delhi Technological University, Delhi, India.
Email: ashutoshgupta.nitm@gmail.com

Abstract

This paper proposes a novel configuration for a multiport boost converter (MPBC) with a single inductor (SI), accounting for equivalent series resistances (ESRs) and minimizing input switching stress. The MPBC performance is evaluated and compared with other established topologies. The proposed MPBC interfaces two unidirectional input DC power ports and a rechargeable port for an energy storage element (ESE) with two output ports. The design integrates two renewable sources with the ESE as a third source. One output is for higher voltage, linked to a single-phase inverter for AC loads. The other output is for lower DC voltage, used for DC loads. The configuration can be adjusted based on requirements. This converter has numerous applications in renewable energy systems, electric vehicles, and agriculture. The steady-state and small signal modeling of MPBC has been done to derive the mathematical expressions for analyzing stability, stresses (both voltage and current), and performance considering ESRs. A 240 W, MPBC is fabricated along with improved switching strategies using DSP TMS320F28379D. Experimental and simulation results are compared to show the effectiveness of proposed scheme on stability, stresses, and efficient power transfer. Output power is regulated effectively by sharing the input power thereby reducing voltage stress on switches.

KEYWORDS

hybrid DC–DC power converter, isolated microgrid multiport converter, parasitics, small signal modeling, switch stress, topology Derivation

1 | INTRODUCTION

The importance of implementing clean and sustainable energy solutions is becoming increasingly evident as the world's population continues to grow. Fossil fuels, upon which the societies have relied for so long, are no longer a viable option due to escalating oil and natural gas prices, ecological damage, and the depletion of resources. Distributed renewable energy sources such as photovoltaics (PV) and fuel cells are rapidly gaining momentum as a solution to these issues. Wind energy is also making significant contributions to the energy transition. The installation of PV capacity has risen from kW to MW to GW in less than two decades, and it is expected to reach 1 TW of installed global capacity soon. A shift toward clean and green energy is essential to ensure a sustainable future.[1-5] Because these energy sources are uncertain and intermittent, they are unreliable for stand-alone use. To address this, renewable resources

 © 2024 John Wiley & Sons Ltd.

are often combined with each other or with fuel cells, supercapacitors, and energy storage systems (ESEs). Recently, multi-input converters (MICs) have been used to connect multiple renewable energy sources. These converters offer simple circuit designs, low manufacturing costs, centralized control, and compact size.[2-7] In literature,[8] a novel isolated multi-input DC–DC converter based on the concept of flux additivity uses phase-shifted pulse width modulation (PWM) control on two input current-fed full-bridge DC–DC converters. This approach offers several advantages, including the ability to deliver power separately and simultaneously from multiple DC sources, soft switching technology, and electric isolation via a transformer. In literature,[9] a modular non-isolated multi-input step-up high-gain converter topology with continuous current conduction provides high gain using the same number of devices and has the least number of device counts. The switches and diodes' peak inverse voltage (PIV) are controlled within acceptable limits to maintain high voltage gain.

In literature,[10] a new bidirectional DC–DC converter is presented for grid-connected renewable energy systems and hybrid electric vehicles (EVs). It eliminates the need for an additional transformer, combining two parallel buck converters with series switches for input sources to avoid short circuits, reducing the need for inductors and capacitors, enhancing cost efficiency11,12. However, it faces power flow control challenges. In many applications, integrating power electronics into multiple output converters (MOCs) is crucial for reducing size, weight, maintenance needs, and costs, as explained in references.13[14,15] A comprehensive classification of fuel cell-based hybrid electric vehicles (FCHEVs). A key component in these FCHEVs is the DC–DC power converter. A multi-input DC–DC power converter is employed to connect multiple energy sources, simplifying the system and enhancing its overall efficiency.[16]

Voltage regulation for each DC output in converters is crucial. Post-regulation is used to ensure robustness against line or load variations. Multiple loads require different low voltages, so efficiently supplying multiple-and-low output voltages is essential.[17] The literature[18] presents an accurately regulated multiple outputs (MO) ZVS DC–DC converter with operation and control analysis. Isolated multiport DC–DC converters are heavy and expensive and utilize transformers for electrical isolation. Non-isolated multi-input DC–DC converters provide a practical and cost-effective solution without the need for transformers or high-frequency components, resulting in improved efficiency.[19]

Multi-input multi-output (MIMO) converters provide a medium of interaction between different renewable resources, energy-storing elements, and advanced electric loads with a centralized control system. A generalized form of MIMO converter is shown in Figure 1. Integrating MIMO systems with sources and loads provides compact packaging, relatively easier control, simultaneous power management, and voltage regulation. The authors[20,21] presented a topology that can interface between diversified renewable energy sources and diode-clamped multilevel inverters for AC loads. As the sources are intermittent, ESE is a must, and the impossibility of energy transfer between sources is a drawback of this converter.

The literature[21–28] discuss a non-isolated MIMO converter with individual inductors for each source but lacks power transfer between inputs. The individual inductors make the converter larger, and a cascading approach increases device count and complexity.

A novel MIMO converter is introduced in literature,[29] capable of distributing power among sources and the load. Dynamic analysis is conducted on a 2-in-2-out system with two inputs (35 and 48 V), two outputs (40 and 80 V), and 236 W output power using ideal parameters. The paper explores n-input and m-output configurations, but increasing the number of sources leads to switching stress and duty cycle issues with the in-line switches. The paper proposes a lead–lag compensator to regulate the system by adjusting poles and zeros, but this approach may be unreliable when

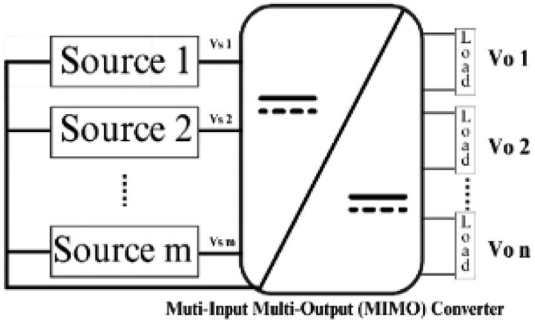


FIGURE 1 Generalized MIMO converter.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

faced with uncertainties in load power variations. In literature,[30] discusses the design, implementation, and control of a scalable DC microgrid that uses a hybrid of renewable energy sources. It employs a multi-input, multi-output dual active half-bridge (DAHB) converter.

This paper introduces a modified MPBC that forms the backbone in hybrid renewable energy systems. The novelty of this paper as follows:

- The proposed non-isolated converter utilizes a single inductor (SI) to reduce electromagnetic interference, size, and cost.
- ESRs are considered which are neglected in previous MIMO papers. Rigorous analysis has been implemented as it may affect the stability of the proposed converter.
- Improved switching strategies using TMS320F28379D DSP with faster response.
- It enables seamless addition or removal of input sources with least switching stress. The MPBC configuration includes two input sources, an ESE, and two DC outputs—one integrated with an inverter for AC loads and the other for auxiliary DC loads.

The paper follows a structured as follows: Section 2 provides a comparative evaluation of existing non-isolated multiport converter topologies with the proposed MPBC. Section 3 explains the structure and operation modes of the proposed MPBC along with comparative study between different multiport converters. Section 4 presents dynamic modeling, considering circuit parasitics and component analysis. Sections 5 analyzes the voltage and current stress on switches in the proposed converter and showcases simulation, hardware implementation, and experimental results. Finally, Section 6 concludes the paper, summarizing the work and analysis of the proposed MPBC. The paper provides insights into the design and performance of the modified converter for hybrid RES and culminates the paper with work and analysis of the MPBC.

2 | TOPOLOGY DERIVATION AND COMPARISON STUDY

The researchers[5,19,21,22,27–29] have proposed numerous non-isolated multiport converters, emphasizing the need for more efficient and cost-effective solutions for hybrid renewable energy systems as shown in Table 1. Specific parasitic effects can influence the converter's characteristics, including its components' equivalent series resistances (ESRs). The crucial question to address is whether and how these ESRs impact the converter's input resistance. Some of these converters have redundant switching components, and most of the converters are shown ideally without parasitic effects, so the topology should be optimized by analyzing the voltages across the switch components, including practical behavior with parasitics. Two topologies from the family of non-isolated multi-port converters in Section 2 have been selected to create the proposed converter having three inputs including ESEs, which can store energy in the event of low power demand at the load side, and two outputs with their parasitics. The charging and discharging switching states of converters are presented in Tables 2 and 3. Figures 2–4 show the three topologies selected from Table 1 of non-isolated DC–DC converters for three input and two output non-isolated DC–DC converters with their parasitic elements. The switching behavior of the three topologies showing their discharging and charging characteristics is tabulated in Tables 2 and 3, respectively. These three topologies use five MOSFETs—$S_1, S_2, S_3, S_4,$ and S_5—to perform the circuit operation and their parasitics are $r_{ds1}, r_{ds2}, r_{ds3}, r_{ds4,}$ and r_{ds5}, respectively. ESR of switches, inductors, and capacitors are included in all the topologies showing the practical behavior of the circuit. The averaged and dynamic mathematical derived symbols along with specifications are listed in Table 4.

To evaluate the advantages of the proposed converter, a comparison was conducted between the proposed converter and various MIMO converters from previous studies, as outlined in Table 1. The comparison primarily focused on factors such as the number of switches, capacitors, inductors, voltage stress, control complexity, and expandability.

The converters found in references[5,19,21,27,28] were deemed unsuitable for connecting renewable energy sources to the power grid due to their low voltage outputs except.[29] On the other hand, the proposed converter offers several advantages. Its control system is straightforward, and its design is cost-effective. Additionally, it allows for seamless addition or removal of input sources, and its output terminals function independently. Notably, the converter operates without the need to worry about switch transition times, which enhances its expandability and reduces design complexity. A comparative analysis between the proposed converter and existing SI MIMO converters as shown in Table 1 and in Figures 2–4 clearly highlights the benefits of the proposed design.

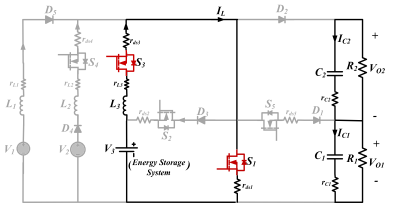
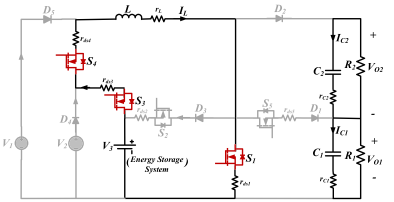
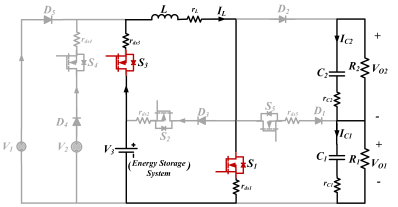
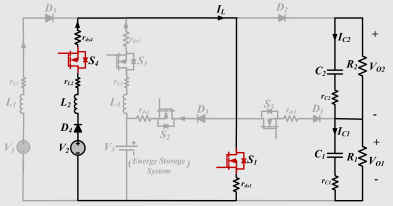
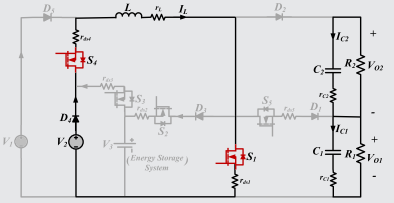
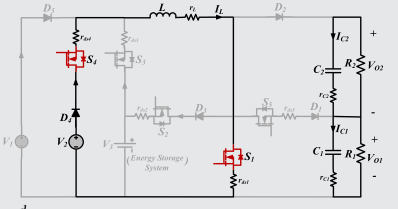
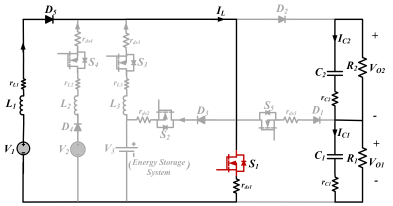
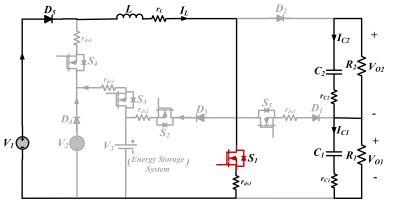
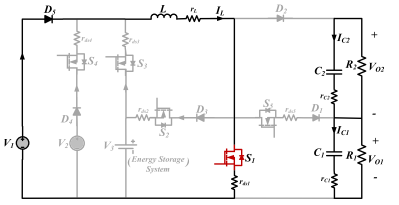
1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Comparison analysis of different converter topologies.

Converter		[5]	[19]	[21]	[22]	[27]	[28]	[29]	Proposed converter
Number of	Inputs	M	M	M	M	M	M	M	M
	Output	N	N	N	N	N	N	N	N
	Switch	M + N	M + N	M + N	M	M + N	M + N	M + N	M + N
	Diodes	M + N	M + N	M + N	N	M + N + 1	M + N + 1	M + N	M + N
	Inductor	1	1	1	M	1	1	1	1
	Capacitor	1	1	1	N	1	1	N	N
Input current		Discontinuous	Discontinuous	Discontinuous	Continuous	Discontinuous	Discontinuous	Discontinuous	Discontinuous
Frequency		40 kHz	100 kHz	40 kHz	20 kHz	20–100 kHz	100 kHz	10 kHz	40 kHz
Input voltage		25 V, 20 V	20 V, 9–40 V, 20 V	30 V, 20 V	159.3 V, 150 V	24 V, 20 V	18 V, 22 V	35 V, 48 V	12 V, 24 V, 60 V
Output voltage		22 V, 11 V	5.25 V, 3–15 V, 32 V	21 V, 8 V	155 V	12 V, 8–10 V	12 V, 8 V	80 V, 40 V	60 V, 60 V
Output power		29.5 W	30.4 W	35.4 W	—	35 W	33 W	236 W	240 W
Output voltage level		Low	Low	Low	High	Low	Low	Medium	Medium
Max. voltage across switch		Medium	Medium	Medium	Low	Medium	Medium	High	Low
Efficiency		—	88%	90%	—	92.1%	—	80%	91.6%
Extension capability		Medium	High	Medium	Medium	Medium	Medium	Medium	High
Cost		Medium	Low	Medium	High	Medium	Medium	Medium	Low

10970007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

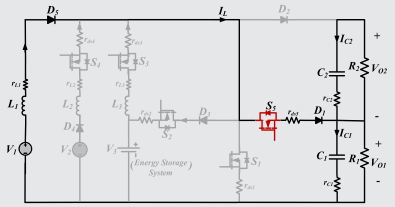
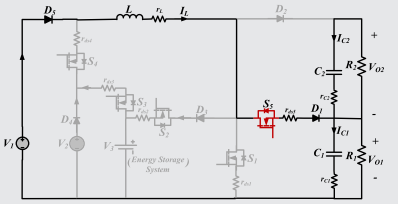
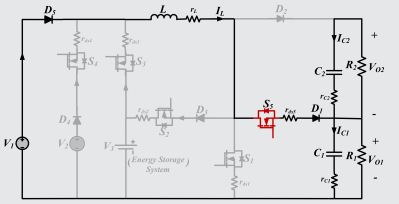
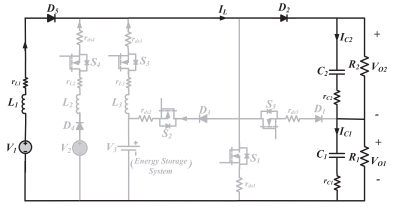
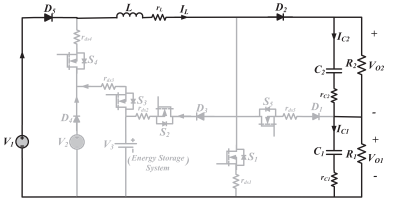
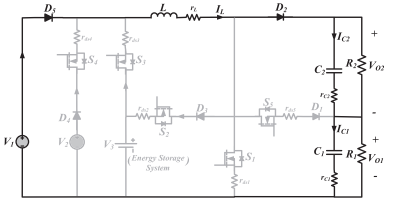
TABLE 2 Switching states of each converter topology showing discharging operation mode of ESES along with power transfer to loads.

Switching states (discharging)	Topology 1[22]	Topology 2[29]	Topology 3 (proposed converter)
Switching states 1 ($0 < t < D_3T$)			
	$L_3\frac{d_{i_{L3}}}{dt} = v_3 - i_{L3}(r_{ds3} + r_{ds1} + r_{L3})$	$L\frac{d_{i_L}}{dt} = v_3 - i_L(r_{ds3} + r_{ds4} + r_{ds1} + r_L)$	$L\frac{d_{i_L}}{dt} = v_3 - i_L(r_{ds3} + r_{ds1} + r_L)$
	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$
Switching states 2 ($D_3T < t < D_4T$)			
	$L_2\frac{d_{i_{L2}}}{dt} = v_2 - i_{L2}(r_{ds4} + r_{ds1} + r_{L2})$	$L\frac{d_{i_L}}{dt} = v_2 - i_L(r_{ds4} + r_{ds1} + r_L)$	$L\frac{d_{i_L}}{dt} = v_2 - i_L(r_{ds4} + r_{ds1} + r_L)$
	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$
Switching states 3 ($D_4T < t < D_1T$)			
	$L_1\frac{d_{i_{L1}}}{dt} = v_1 - i_{L1}(r_{ds1} + r_{L1})$	$L\frac{d_{i_L}}{dt} = v_1 - i_L(r_{ds1} + r_L)$	$L\frac{d_{i_L}}{dt} = v_1 - i_L(r_{ds1} + r_L)$
	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$

(Continues)

10970070x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

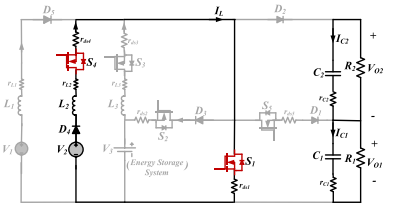
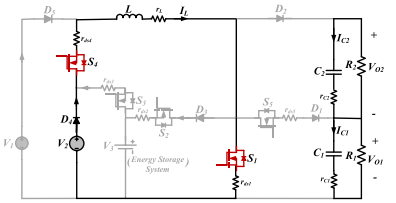
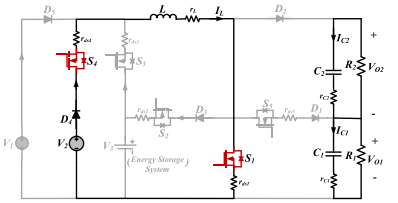
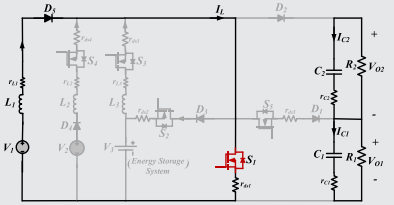
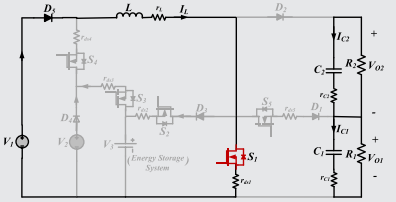
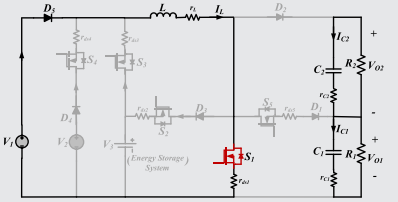
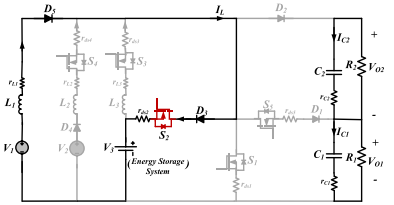
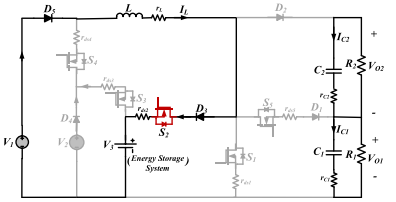
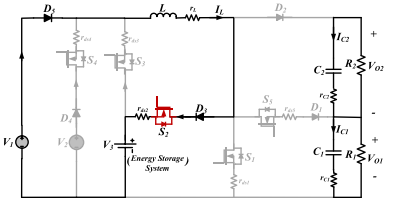
TABLE 2 (Continued)

Switching states (discharging)	Topology 1[22]	Topology 2[29]	Topology 3 (proposed converter)
Switching states 4 ($D_1T < t < D_5T$)			
	$L_1\frac{d_{iL1}}{dt} = v_1 - i_{L1}(r_{ds5} + r_{L1}) - v_{o1}$	$L\frac{d_{iL}}{dt} = v_1 - i_L(r_{ds5} + r_L) - v_{o1}$	$L\frac{d_{iL}}{dt} = v_1 - i_L(r_{ds5} + r_L) - v_{o1}$
	$C_1\frac{dv_{o1}}{dt} = i_{L1} - \frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = i_{L1} - \frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$
Switching states 5 ($D_5T < t < T$)			
	$L_1\frac{d_{iL1}}{dt} = v_1 - i_{L1}r_{L1} - (v_{o1} + v_{o2})$	$L\frac{d_{iL}}{dt} = v_1 - i_L r_L - (v_{o1} + v_{o2})$	$L\frac{d_{iL}}{dt} = v_1 - i_L r_L - (v_{o1} + v_{o2})$
	$C_1\frac{dv_{o1}}{dt} = i_{L1} - \frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = i_{L1} - \frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = i_L - \frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = i_L - \frac{v_{o2}}{(R_2 + r_{C2})}$

Note: D_1 = duty ratio of switch 1 (S_1), D_2 = duty ratio of switch 2 (S_2), D_3 = duty ratio of switch 3 (S_3), D_4 = duty ratio of switch 4 (S_4), and D_5 = duty ratio of switch5 (S_5), T = time period, and SS = switching state.

10970007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

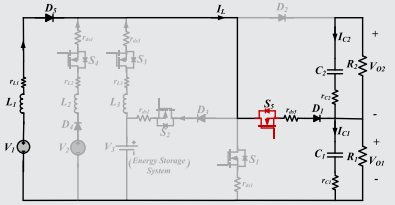
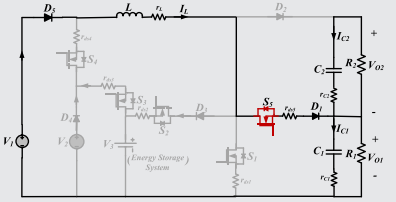
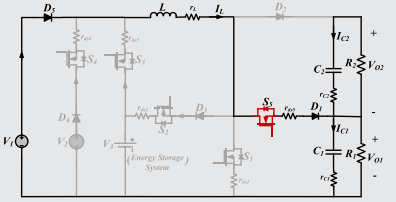
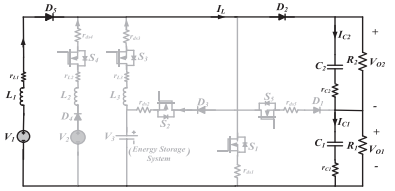
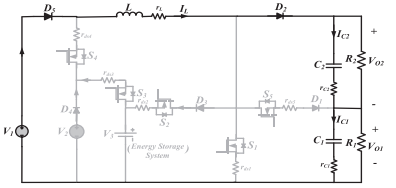
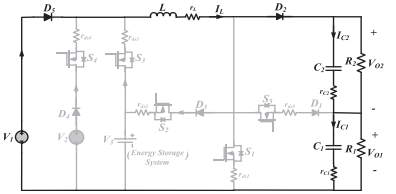
TABLE 3 Switching states of each converter topology showing charging operation mode of ESES along with power transfer to loads.

Switching states (charging)	Topology 1[22]	Topology 2[29]	Topology 3 (proposed converter)
Switching states 1 ($0 < t < D_4T$)			
	$L_2\frac{di_{L2}}{dt} = v_2 - i_{L2}(r_{ds4} + r_{ds1} + r_{L2})$	$L\frac{di_L}{dt} = v_2 - i_L(r_{ds4} + r_{ds1} + r_L)$	$L\frac{di_L}{dt} = v_2 - i_L(r_{ds4} + r_{ds1} + r_L)$
	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$
Switching states 2 ($D_4T < t < D_1T$)			
	$L_1\frac{di_{L1}}{dt} = v_1 - i_{L1}(r_{ds1} + r_{L1})$	$L\frac{di_L}{dt} = v_1 - i_L(r_{ds1} + r_L)$	$L\frac{di_L}{dt} = v_1 - i_L(r_{ds1} + r_L)$
	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$
Switching states 3 ($D_1T < t < D_2T$)			
	$L_1\frac{di_{L1}}{dt} = v_1 - v_3 - i_{L1}(r_{ds2} + r_{L1})$	$L\frac{di_L}{dt} = v_1 - v_3 - i_L(r_{ds2} + r_L)$	$L\frac{di_L}{dt} = v_1 - v_3 - i_L(r_{ds2} + r_L)$
	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$	$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})}$
	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$

(Continues)

10970070x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 (Continued)

Switching states (charging)	Topology 1[22]	Topology 2[29]	Topology 3 (proposed converter)
Switching states 4 ($D_2T < t < D_5T$)	 $L_1\frac{d_{iL1}}{dt} = v_1 - i_{L1}(r_{ds5} + r_{L1}) - v_{o1}$ $C_1\frac{dv_{o1}}{dt} = i_{L1} - \frac{v_{o1}}{(R_1 + r_{C1})}$ $C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	 $L\frac{d_{iL}}{dt} = v_1 - i_L(r_{ds5} + r_L) - v_{o1}$ $C_1\frac{dv_{o1}}{dt} = i_{L1} - \frac{v_{o1}}{(R_1 + r_{C1})}$ $C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$	 $L\frac{d_{iL}}{dt} = v_1 - i_L(r_{ds5} + r_L) - v_{o1}$ $C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})}$ $C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})}$
Switching states 5 ($D_5T < t < T$)	 $L_1\frac{d_{iL1}}{dt} = v_1 - i_{L1}r_{L1} - (v_{o1} + v_{o2})$ $C_1\frac{dv_{o1}}{dt} = i_{L1} - \frac{v_{o1}}{(R_1 + r_{C1})}$ $C_2\frac{dv_{o2}}{dt} = i_{L1} - \frac{v_{o2}}{(R_2 + r_{C2})}$	 $L\frac{d_{iL}}{dt} = v_1 - i_L r_L - (v_{o1} + v_{o2})$ $C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})}$ $C_2\frac{dv_{o2}}{dt} = i_L - \frac{v_{o2}}{(R_2 + r_{C2})}$	 $L\frac{d_{iL}}{dt} = v_1 - i_L r_L - (v_{o1} + v_{o2})$ $C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})}$ $C_2\frac{dv_{o2}}{dt} = i_L - \frac{v_{o2}}{(R_2 + r_{C2})}$

10970007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

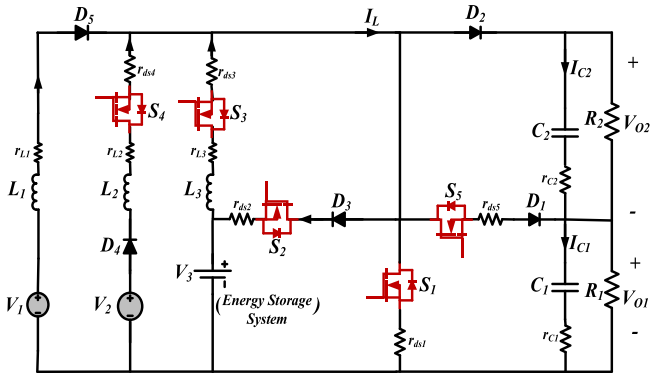


FIGURE 2 Topology 1: non-isolated DC–DC converter with multiple inductor on source with parasitics.[22]

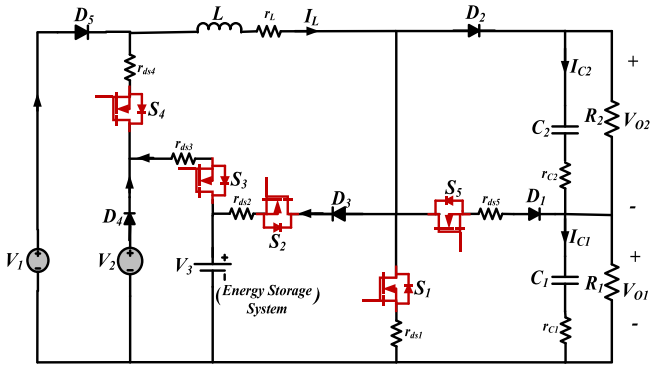


FIGURE 3 Topology 2: non-isolated single inductor DC–DC converter with parasitics.[29]

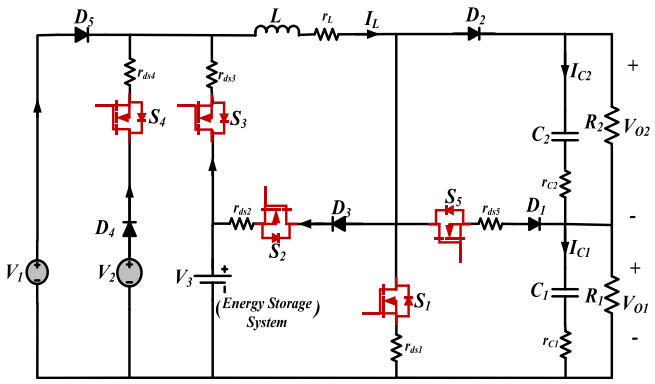


FIGURE 4 Topology 3: proposed three input two output converter with parasitics in circuit.

3 | CONVERTER STRUCTURE AND OPERATION MODES

This paper introduces a highly efficient non-isolated SI multiport DC–DC converter that combines state-of-the-art research.[7,19,29] Figure 5A illustrates the multiport extension of the proposed converter for interfacing multiple inputs and outputs. As shown in Figure 5B, the proposed MPBC structure facilitates the seamless integration of three input power sources V_1, V_2, and V_3, where $V_1 < V_2 < V_3$. The converter achieves superior performance with just one inductor, two output capacitors, and five switches. The load resistances R_1 and R_2 represent the power feeding to a DC load or

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Hardware component and simulation parameters of proposed system.

Simulation and prototype parameters	Symbols	Specification(s)
Output power	P_{o1}, P_{o2}	120 W, 120 W
Switching frequency	fs	40 kHz
Output voltage	V_T, V_{O1}, V_{O2}	120 V, 60 V, 60 V
DC source 1 and 2	V_1, V_2	12 V,24 V
DC source 3 (battery)	V_3	60 V of 12 V, 7 $Ah_{(5)}$
Inductor	L	2 mH
Inductor ESR	r_L	0.33 Ω
Capacitors	C_1	1000 μF (150 V)
	C_2	1000 μF (150 V)
Capacitor ESR	r_{C1}, r_{C2}	100 m Ω each
Load resistances	R_1	0–500 Ω
	R_2	0–500 Ω
MOSFET	SPW20N60C3	650 V, 20.7 A
On-state resistance	$r_{ds1}, r_{ds2}, r_{ds3}, r_{ds4}, r_{ds5}$	0.19 Ω
Diode	STTH3012	1200 V, 30 A

single-phase motor as AC loads. Proper switching of the MOSFETs enables precise control of power flow between input sources and voltage boosting. The converter can produce output voltages suitable for connection to a single-phase inverter, with each output having varying or equal voltage levels. The controllable switches $S_1, S_2, S_3, S_4,$ and S_5 are the primary elements that regulate the converter's power flow and output voltages. ESRs are added for MOSFETs, inductors, and capacitors to understand the circuit practically.

This enhancement significantly improves the system's practical applicability. Notably, the converter's unique design allows V_1 and V_2 to supply power to V_3, creating a perfect fit for EV applications. The configuration optimizes the layout, placing solar PV in the V_1 position, fuel cell (FC) at V_2, and energy storage elements (ESEs) like batteries or supercapacitors (SC), which can be recharged, in the V_3 slot.

In this paper, solar PV and FCs serve as robust power sources, while the battery operates as an efficient ESS. The proposed converter introduces two distinct power operation modes based on the battery's usage status, ensuring optimal system performance. During high load power demands, all input sources contribute to the load power, with S_2 remaining inactive while $S_1, S_3, S_4,$ and S_5 are engaged. In scenarios of low load power and a need to charge V_3, V_1 and V_2 provide power for both the loads and V_3, with $S_1, S_2, S_4,$ and S_5 active and S_3 in a standby state.

To maintain system stability, it is essential to minimize the input current fluctuations in both PV and FCs. The output voltage's dependency on input current necessitates precise power balance between input sources and the load. This paper thoroughly investigates the converter's steady state and dynamic behavior in Continuous Conduction Mode (CCM). It is important to note that each input source can function independently, allowing the converter to operate as a single-input dc-dc converter, including a standalone mode.

Tables 2 and 3 provide a comprehensive overview of the switching states for each converter during ESE charging and discharging, including their inherent parasitic elements, which are neglected.[22,29] It will give better insight for the performance comparison. It is evident that topology 1, with inductors in each source for energy accumulation, results in a bulky and lossy converter. This setup introduces significant electromagnetic interference in the system and elevates the number of components, ultimately reducing system reliability and complicates modeling, analysis, and control.

On the other hand, topology 2 proposes a unified structured single inductor multiport converter[29] with three inputs and two outputs to hybridize power sources. During the discharging mode, switching state 1 of this converter reveals that three switches are ON at a time, which can increase up to n switches with the addition of n sources. As a result, ON time of inner source will be very limited, and the duty cycle range for higher-order sources is limited. This will also create stresses on higher order source switches, which are also verified experimentally as shown in Section 5. Cascading of switches will increase losses due to the internal resistance of switches, as can be observed from the inductor differential equation in switching state 1 of topology 2 during the discharging mode.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

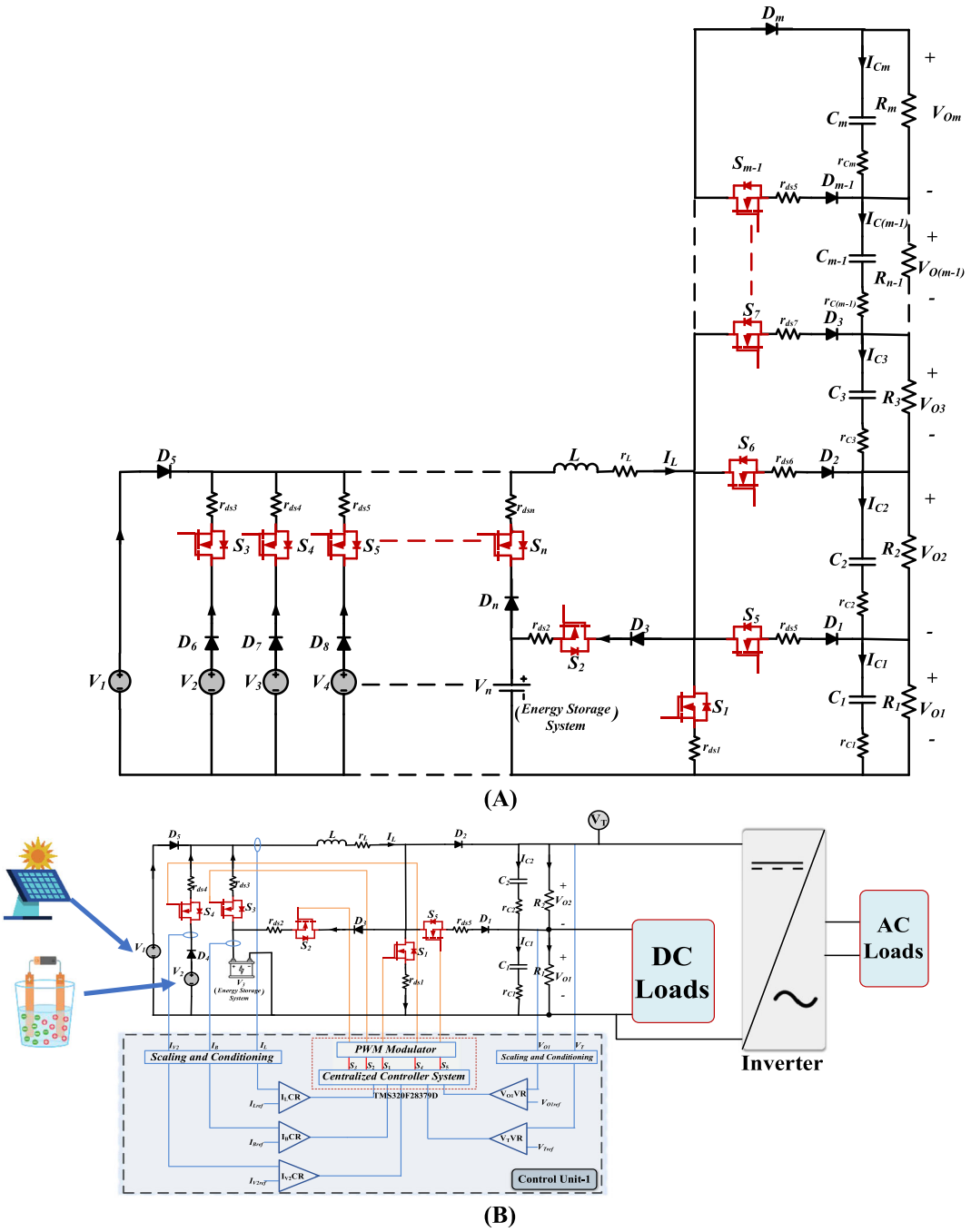


FIGURE 5 (A) Multiport extension for the proposed converter to interface multiple inputs and outputs. (B) Circuit diagram of proposed MPBC with DC load, integrated single phase inverter with AC load and control unit.

$$L\frac{d_{i_L}}{dt} = v_3 - i_L(r_{ds3} + r_{ds4} + r_{ds1} + r_L) \tag{1}$$

Controlling of these switches will affect the output response as the recovery rate of each switch in switching state 1 of discharging mode is dependent on each other. Cascading of switches will also delay the applied control technique. From Equation (1), it can be concluded that for n number of sources, differential equation will have $(n-1)$ ESR of switches.

To address the drawbacks of these converters, the proposed solution connects sources in parallel with their switches, eliminating the possibility of multiple switches being activated simultaneously. The proposed MPBC is shown in Figure 4, whose parameters are depicted in Table 4. Further, its switching states for discharging and charging are listed

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

under topology 3 of Tables 2 and 3, respectively. The proposed MPBC includes two renewable sources integrated with ESEs as a third source and two outputs, where a higher voltage output that is connected to a single-phase inverter for AC loads and a lower DC voltage output that can be used for DC loads or vice versa depending upon the practical requirements as shown in Figure 5. The switch for source V_3, which corresponds to the ESEs, is isolated and connected in parallel with sources V_1 and V_2 in the circuit.

Instead of considering ideal switches and elements, switching states are analyzed considering the effect of the parasitics. This design not only offers improved performance and reliability but also enhances the overall safety of the system.

The proposed converter is expected to significantly impact the development of future renewable energy systems by addressing some of the significant challenges associated with existing converter designs. Converter structure, averaged and small signal modeling, voltage and current stress analysis, stability, and control structure are explained in detail in the subsequent section. Table 5 shows the three operating modes: stand-alone, discharging, and charging modes, respectively. Here, the converter has been studied in two primary operational modes: battery discharging and charging mode.

3.1 | First operation mode (discharging mode)

In this configuration, source V_3, the battery, along with two input sources, V_1 and V_2, is solely responsible for delivering power to the loads. The switches S_2 is completely OFF, while S_1, S_3, S_4, and S_5 are actively engaged. Switches S_3 and S_4 are responsible for actively regulating the source currents to a desired value. Ensuring the total output voltage V_T, the sum of V_{O1} and V_{O2}, attains the selected value is controlled by switch S_1. Switch S_5 regulates the output voltage V_{O1}. By effectively regulating V_T and V_{O1}, the output voltage V_{O2} is also controlled to the desired value.

During state 1 ($0 < t < D_3T$), switches S_1 and S_3 are turned ON. Because S_1 is ON, diodes D_1 and D_2 are reversely biased, so switch S_5 is turned OFF. Furthermore, since $V_1 < V_2 < V_3$ and S_3 is ON and S_4 is OFF, diode D_5 becomes reversely biased. The equivalent circuit of the proposed converter in this state is shown in Figure 6A where source V_3 charges inductor L, increasing inductor current I_L. Additionally, capacitors C_1 and C_2 discharge their stored energy to the load resistances R_1 and R_2, respectively. In this mode, the inductor current and capacitor voltages considering ESRs are expressed as follows:

$$L\frac{d_{i_L}}{dt} = v_3 - i_L(r_{ds3} + r_{ds1} + r_L) \tag{2}$$

$$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})} \tag{3}$$

$$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})} \tag{4}$$

Steady-state waveforms are shown in Figure 8.

During state 2 ($D_3T < t < D_4T$), switch S_4 is turned ON, S_1 remains ON, and S_3 is turned OFF. Since S_1 is ON, diodes D_1 and D_2 are reverse-biased, so switch S_4 remains OFF. The equivalent circuit of the proposed converter during this state is illustrated in Figure 6B. During this state, source V_2 charges inductor L, leading to an increase in inductor current. Additionally, capacitors C_1 and C_2 are discharged, releasing their stored energy to load resistances R_1 and R_2, respectively. The equations governing the inductor's and capacitors' behavior considering ESRs during this state are provided below.

$$L\frac{d_{i_L}}{dt} = v_2 - i_L(r_{ds4} + r_{ds1} + r_L) \tag{5}$$

$$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})} \tag{6}$$

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Operating modes with switching patterns of proposed multiport converter.

Operation modes	V_1	V_2	V_3 (battery)	V_{O1}	V_{O2}	S_1	S_2	S_3	S_4	S_5	SS
Stand-alone mode (mode 1)		Active		Active	Active	$(1 - D_1)T$	X	X	$(1 - D_4)T$	X	2
	Active			Active	Active	$(1 - D_1)T$	X	X	$(1 - D_4)T$	X	3
	Active			Active	Active	X	X	X	X	$(1- - D_5)T$	4
	Active			Active	Active	X	X	X	X	X	5
Battery discharging mode (mode 2)			Active	Active	Active	$(1 - D_1)T$	X	$(1 - D_3)T$	X	X	1
		Active		Active	Active	$(1 - D_1)T$	X	X	$(1 - D_4)T$	X	2
	Active			Active	Active	$(1 - D_1)T$	X	X	$(1 - D_4)T$	X	3
	Active			Active	Active	X	X	X	X	$(1\text{-}D_5)T$	4
	Active			Active	Active	X	X	X	X	X	5
Battery charging mode (mode 3)		Active		Active	Active	$(1 - D_1)T$	X	X	$(1 - D_4)T$	X	1
	Active			Active	Active	$(1 - D_1)T$	X	X	X	X	2
			Active	Active	Active	X	$(1 - D_2)T$	X	X	X	3
	Active			Active	Active	X	X	X	X	$(1 - D_5)T$	4
	Active			Active	Active	X	X	X	X	X	5

Note: X = inactive, T = time period, and SS = switching state.

10970007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

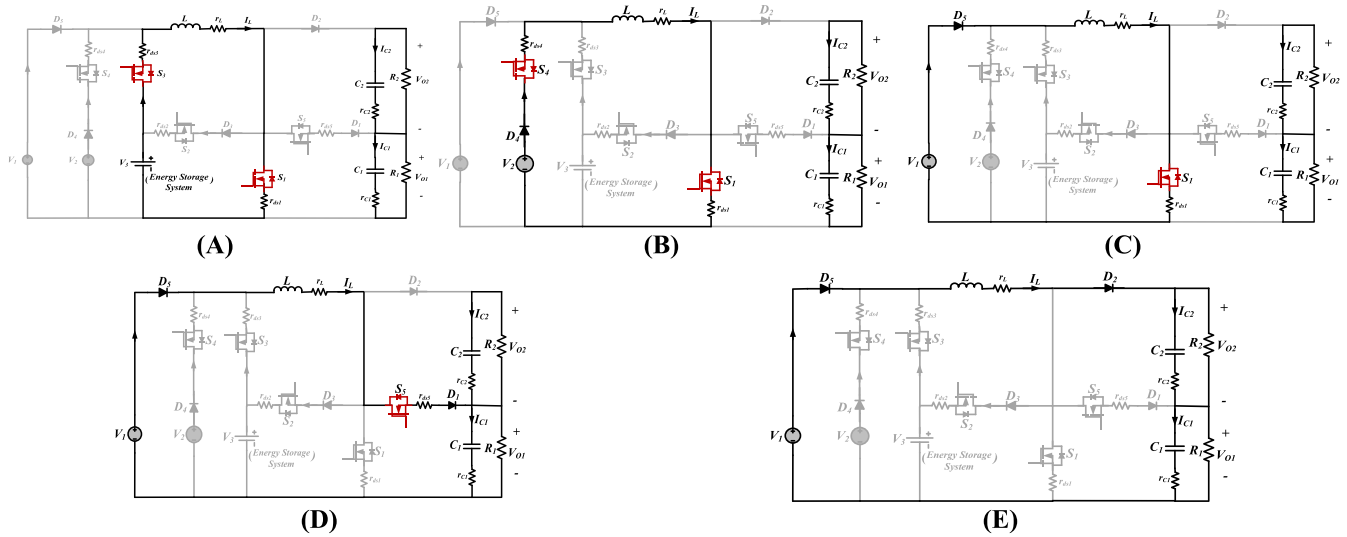


FIGURE 6 Equivalent circuit of battery discharging mode considering ESRs: (A) switching state 1, (B) switching state 2, (C) switching state 3, (D) switching state 4, and (E) switching state 5 of proposed MPBC.

$$C_2\frac{dv_{o2}}{dt}=-\frac{v_{o2}}{(R_2+r_{C2})} \tag{7}$$

During state 3 ($D_4T < t < D_1T$), switch S_1 is still ON, and S_4 and S_3 are turned OFF. Because S_1 is ON, diodes D_1 and D_2 are reversely biased, so switch S_5 is still OFF. The equivalent circuit of the proposed converter in this state is shown in Figure 6C. V_1 charges inductor L, so the inductor current increases. In addition, capacitors C_1 and C_2 are discharged and deliver their stored energy to load resistances R_1 and R_2, respectively. The inductor current and capacitor voltages considering ESRs in this mode are as follows:

$$L\frac{d_{i_L}}{dt}=v_1-i_L(r_{ds1}+r_L) \tag{8}$$

$$C_1\frac{dv_{o1}}{dt}=-\frac{v_{o1}}{(R_1+r_{C1})} \tag{9}$$

$$C_2\frac{dv_{o2}}{dt}=-\frac{v_{o2}}{(R_2+r_{C2})} \tag{10}$$

During state 4 ($D_1T < t < D_5T$), the proposed converter operates as follows: Switch S_1 is turned OFF, while switches S_3 and S_4 remain OFF. Additionally, switch S_5 is turned ON, causing diode D_2 to be reversely biased. The equivalent circuit of the converter during this state is depicted in Figure 6D. In this mode, inductor L discharges and transfers its stored energy to C_1 and R_1, decreasing the inductor current. Meanwhile, capacitor C_1 is charged, and capacitor C_2 is discharged, delivering its stored energy to the load resistance R_2. The expression governing the energy storage elements L, C_1, and C_2 considering ESRs in this mode are as follows:

$$L\frac{d_{i_L}}{dt}=v_1-i_L(r_{ds5}+r_L)-v_{o1} \tag{11}$$

$$C_1\frac{dv_{o1}}{dt}=i_L-\frac{v_{o1}}{(R_1+r_{C1})} \tag{12}$$

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$C_2\frac{dv_{o2}}{dt}=-\frac{v_{o2}}{(R_2+r_{C2})} \tag{13}$$

None of the switches is active during state 5 ($D_5T < t < T$), causing diode D_2 to be forward-biased. During this phase, inductor L is discharged, transferring its stored energy to capacitors C_1, C_2, and load resistances R_1 and R_2. Additionally, capacitors C_1 and C_2 are charged. Figure 6E shows the equivalent circuit of the proposed converter in this state. Expressions governing the inductor current and capacitor voltages considering parasitics during this mode are as follows:

$$L\frac{d_{i_L}}{dt}=v_1-i_Lr_L-(v_{o1}+v_{o2}) \tag{14}$$

$$C_1\frac{dv_{o1}}{dt}=i_L-\frac{v_{o1}}{(R_1+r_{C1})} \tag{15}$$

$$C_2\frac{dv_{o2}}{dt}=i_L-\frac{v_{o2}}{(R_2+r_{C2})} \tag{16}$$

Figure 7 shows the steady-state voltage and current waveforms of the inductor as well as different output voltages obtained during battery discharging modes. It also shows the switching strategies of the proposed MPBC.

These five states will combine to form the average model and will be further analyzed for stability and stress determination of proposed MPBC. Slope of the inductor current in Equations (2), (5), (8), (11), and (14) shows that parasitics plays major role causing transfer of energy in the circuit components.

3.2 | Second operation mode (battery charging)

In this operational mode, sources V_1 and V_2 supply loads and provide power to V_3 (battery). This occurs when the load power is low, and the battery requires immediate charging. To accomplish this, switches $S_1, S_2, S_4,$ and S_5 are activated, while switch S_3 is deactivated. Like the previous operational mode of the converter, each switch also has a specific duty to perform in this mode. Switch S_1 regulates the overall output voltage $V_T = V_{O1} + V_{O2}$ to the desired level, while switch S_2 regulates the charging current I_b.

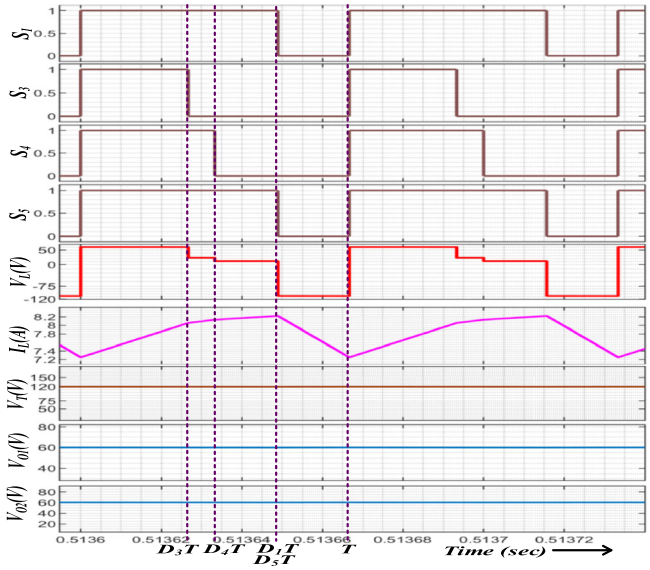


FIGURE 7 Battery discharging mode steady-state waveform of proposed converter.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Additionally, switch S_5 controls the output voltage V_{O1}. It is evident that the output voltage V_{O2} is also controlled while regulating V_T and V_{O1}. Depending on the various switch states, five different operational modes are discussed throughout one switching period considering the parasitics.

During state 1 ($0 < t < D_4T$), switch S_1 is turned ON, while S_2 and S_5 are reverse-biased and cannot be activated. Diode D_2 is also reversely biased and does not conduct. The equivalent circuit of the proposed converter in this state is shown in Figure 8A. In this mode, V_2 charges inductor L, causing an increase in inductor current. Capacitors C_1 and C_2 are discharged, delivering their stored energy to load resistances R_1 and R_2, respectively. Inductor current and capacitor voltages during this mode are as follows:

$$L\frac{d_{i_L}}{dt} = v_2 - i_L(r_{ds4} + r_{ds1} + r_L) \tag{17}$$

$$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})} \tag{18}$$

$$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})} \tag{19}$$

During state 2 ($D_4T < t < D_1T$), again switch S_1 is turned ON, while S_2 and S_5 are reverse-biased and cannot be turned ON. Diode D_2 is also reversely biased and does not conduct. The equivalent circuit of the proposed converter in this state is shown in Figure 8B. In this mode, V_1 charges inductor L, increasing inductor current. Capacitors C_1 and C_2 are discharged, delivering their stored energy to load resistances R_1 and R_2, respectively. Expressions governing the inductor current and capacitor voltages during this mode are as follows:

$$L\frac{d_{i_L}}{dt} = v_1 - i_L(r_{ds1} + r_L) \tag{20}$$

$$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})} \tag{21}$$

$$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})} \tag{22}$$

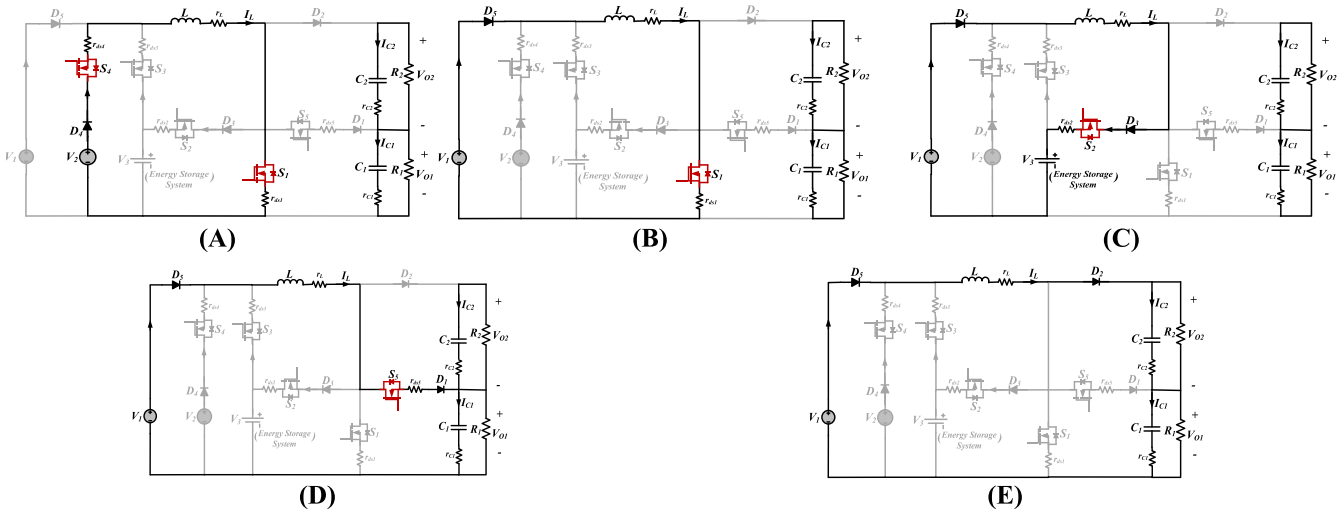


FIGURE 8 Equivalent circuit of battery charging mode considering ESRs: (A) switching state 1, (B) switching state 2, (C) switching state 3, (D) switching state 4, and (E) switching state 5 of proposed MPBC.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

During state 3 ($D_1T < t < D_2T$), switch S_1 is turned OFF, while switch S_2 is turned ON. Diodes D_1 and D_2 are reversely biased, and as a result, S_5 remains OFF. The equivalent circuit of the proposed converter in this state is shown in Figure 8C. Since $V_1 < V_2 < V_3$, the inductor current decreases during this period, and the inductor delivers its stored energy to the battery V_3. Capacitors C_1 and C_2 are discharged, delivering their stored energy to load resistances R_1 and R_2, respectively. Expressions governing the inductor current and capacitor voltages during this mode are as follows:

$$L\frac{d_{i_L}}{dt} = v_1 - v_3 - i_L(r_{ds2} + r_L) \tag{23}$$

$$C_1\frac{dv_{o1}}{dt} = -\frac{v_{o1}}{(R_1 + r_{C1})} \tag{24}$$

$$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})} \tag{25}$$

During state 4 ($D_2T < t < D_5T$), switch S_1 is still OFF, while switch S_2 is turned OFF and switch S_5 is turned ON. Diode D_2 is reversely biased. The equivalent circuit of the proposed converter in this state is shown in Figure 8D. In this mode, inductor L is discharged, and its stored energy is delivered to C_1 and R_1, resulting in a decrease in inductor current.

Capacitor C_1 is charged, while capacitor C_2 is discharged, delivering stored energy to load resistance R_2. Expressions governing the energy storage elements (L, C_1, and C_2) during this mode are shown as follows:

$$L\frac{d_{i_L}}{dt} = v_1 - i_L(r_{ds5} + r_L) - v_{o1} \tag{26}$$

$$C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})} \tag{27}$$

$$C_2\frac{dv_{o2}}{dt} = -\frac{v_{o2}}{(R_2 + r_{C2})} \tag{28}$$

During state 5 ($D_5T < t < T$), all switches are OFF, causing diode D_2 to be forward biased. The equivalent circuit of the proposed converter in this state is shown in Figure 8E. In this mode, inductor L is discharged, and its stored energy is delivered to capacitors C_1, C_2, and load resistances R_1 and R_2. Capacitors C_1 and C_2 are charged during this mode. Inductor current and capacitor voltages during this mode are as follows:

$$L\frac{d_{i_L}}{dt} = v_1 - i_L r_L - (v_{o1} + v_{o2}) \tag{29}$$

$$C_1\frac{dv_{o1}}{dt} = i_L - \frac{v_{o1}}{(R_1 + r_{C1})} \tag{30}$$

$$C_2\frac{dv_{o2}}{dt} = i_L - \frac{v_{o2}}{(R_2 + r_{C2})} \tag{31}$$

4 | AVERAGED AND DYNAMIC ANALYSIS OF PROPOSED MPBC ALONG WITH CONTROLLER DESIGN

Switches $S_1, S_2, S_3, S_4,$ and S_5 are used to operate the proposed MPBC, as mentioned in the previous section. Each switch performs a specific function. By carefully controlling the switches' duty cycle, the output voltages' charging and

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

discharging current of batteries can be adjusted. Dynamic model needs to be obtained to construct the converter's closed-loop controller, while steady-state performance may be analyzed for stress and performance parameter computation.

Present paper focuses on discharging mode of operation, and thus, small signal model is developed. Averaging of the state matrices of proposed MPBC having five switching configuration is shown in Figure 6 and can be written as follows:

$$\begin{aligned} L\frac{di_L}{dt} &= (v_3 - i_L(r_{ds3}+r_{ds1}+r_L))d_3 + (v_2 - i_L(r_{ds4}+r_{ds1}+r_L))(d_4-d_3) + (v_1 - i_L(r_{ds1}+r_L))(d_1-d_4) \\ &\quad + (v_1 - i_L(r_{ds5}+r_L) - v_{o1})(d_5-d_1) + v_{o2}d_5 + (v_1 - i_L r_L - (v_{o1}+v_{o2}))(1-d_5) \end{aligned} \tag{32}$$

$$\begin{aligned} C_1\frac{dv_{o1}}{dt} &= -\frac{v_{o1}}{(R_1+r_{C1})}.d_3 - \frac{v_{o1}}{(R_1+r_{C1})}.(d_4-d_3) - \frac{v_{o1}}{(R_1+r_{C1})}.(d_1-d_4) \\ &\quad + \left(i_L - \frac{v_{o1}}{(R_1+r_{C1})}\right)(d_5-d_1) + \left(i_L - \frac{v_{o1}}{(R_1+r_{C1})}\right)(1-d_5) \end{aligned} \tag{33}$$

$$\begin{aligned} C_2\frac{dv_{o2}}{dt} &= -\frac{v_{o2}}{(R_2+r_{C2})}.d_3 - \frac{v_{o2}}{(R_2+r_{C2})}.(d_4-d_3) - \frac{v_{o2}}{(R_2+r_{C2})}.(d_1-d_4) \\ &\quad - \frac{v_{o2}}{(R_2+r_{C2})}(d_5-d_1) + \left(i_L - \frac{v_{o2}}{(R_2+r_{C2})}\right)(1-d_5) \end{aligned} \tag{34}$$

Small perturbation around the equilibrium point is considered and represented as follows:

$$\begin{aligned} i_L &= I_L + \widehat{i}_L, v_1 = V_1 + \widehat{v}_1, v_2 = V_2 + \widehat{v}_2 \\ v_3 &= V_3 + \widehat{v}_3, v_{o1} = V_{o1} + \widehat{v}_{o1}, v_{o2} = V_{o2} + \widehat{v}_{o2} \\ d_1 &= D_1 + \widehat{d}_1, d_2 = D_2 + \widehat{d}_2, d_3 = D_3 + \widehat{d}_3 \\ d_4 &= D_4 + \widehat{d}_4, d_5 = D_5 + \widehat{d}_5, d_5 = D_5 + \widehat{d}_5 \end{aligned} \tag{35}$$

where $I_L, V_1, V_2, V_3, V_{o1}, V_{o2}, D_1, D_3, D_4,$ and D_5 are the steady-state variables and $\widehat{i}_L, \widehat{v}_1, \widehat{v}_2, \widehat{v}_3, \widehat{v}_{o1}, \widehat{v}_{o2}, \widehat{d}_1, \widehat{d}_3, \widehat{d}_4,$ and $\widehat{d}_5$ are the perturbation in respective steady-state variables.

After substituting Equation (35) into Equations (32)–(34) and neglecting second-order terms, the small signal model is written as follows:

$$\begin{aligned} L\frac{d\widehat{i}_L}{dt} &= (V_{o1} - I_L(r_{ds1}+r_L) + I_L(r_{ds5}+r_L))\widehat{d}_1 + (V_3 - V_2 - I_L(r_{ds3}+r_{ds1}+r_L) \\ &\quad + I_L(r_{ds4}+r_{ds1}+r_L))\widehat{d}_3 + (V_2 - V_1 - I_L(r_{ds4}+r_{ds1}+r_L) + I_L(r_{ds1}+r_L))\widehat{d}_4 \\ &\quad + (V_{o2} - I_L(r_{ds5}+r_L) + I_L r_L)\widehat{d}_5 + D_3\widehat{v}_3 + (D_4 - D_3)\widehat{v}_2 + (1-D_4)\widehat{v}_1 + (D_1-1)\widehat{v}_{O1} \\ &\quad + (D_5-1)\widehat{v}_{O2} + (D_1(r_{ds5}-r_{ds1}) + D_3(r_{ds4}-r_{ds3}) - D_4 r_{ds4} - D_5 r_{ds5} - r_L)\widehat{i}_L \end{aligned} \tag{36}$$

$$C_1\frac{d\widehat{v}_{o1}}{dt} = (1-D_1)\widehat{i}_L - I_L\widehat{d}_1 - \frac{\widehat{v}_{o1}}{(R_1+r_{C1})} \tag{37}$$

$$C_2\frac{d\widehat{v}_{o2}}{dt} = (1-D_5)\widehat{i}_L - I_L\widehat{d}_5 - \frac{\widehat{v}_{o2}}{(R_2+r_{C2})} \tag{38}$$

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

State space model is represented as

$$\begin{aligned}\dot{X} &= AX + BU \\ Y &= CX + DU\end{aligned} \tag{39}$$

where matrix X contains state variables, U contains control variables, and Y contains system outputs.

$$X = \begin{bmatrix} \widehat{i}_L \\ \widehat{v}_{o1} \\ \widehat{v}_{o2} \end{bmatrix} Y = \begin{bmatrix} \widehat{v}_{o1} \\ \widehat{v}_T \\ \widehat{i}_b \\ \widehat{i}_{v2} \end{bmatrix} U = \begin{bmatrix} \widehat{d}_1 \\ \widehat{d}_3 \\ \widehat{d}_4 \\ \widehat{d}_5 \\ \widehat{v}_1 \\ \widehat{v}_2 \\ \widehat{v}_3 \end{bmatrix} \tag{40}$$

$$A = \begin{bmatrix} \frac{D_1(r_{ds5}-r_{ds1})+D_3(r_{ds4}-r_{ds3})-D_4r_{ds4}-D_5r_{ds5}-r_L}{L} & \frac{(D_1-1)}{L} & \frac{(D_5-1)}{L} \\ \frac{(1-D_1)}{C_1} & \frac{-1}{(R_1+r_{C1})C_1} & 0 \\ \frac{(1-D_5)}{C_2} & 0 & \frac{-1}{(R_2+r_{C2})C_2} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{V_{o1}+I_L(r_{ds5}-r_{ds1})}{L} & \frac{V_3-V_2+I_L(r_{ds4}-r_{ds3})}{L} & \frac{V_2-V_1-I_Lr_{ds4}}{L} & \frac{V_{o2}-I_Lr_{ds5}}{L} & \frac{(1-D_4)}{L} & \frac{(D_4-D_3)}{L} & \frac{(D_3)}{L} \\ \frac{-I_L}{C_1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-I_L}{C_2} & 0 & 0 & 0 \end{bmatrix} \tag{41}$$

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ D_3 & 0 & 0 \\ D_4-D_3 & 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_L & 0 & 0 & 0 & 0 & 0 \\ 0 & -I_L & I_L & 0 & 0 & 0 & 0 \end{bmatrix}$$

The A, B, C, and D matrices using Equations (36)–(38) and (40) are represented in Equation (41), where $V_1, V_2,$ and V_3 are input voltages and V_{o1} and V_{o2} are output voltages. All the parameter except duty cycle of switches D_1, D_3, D_4, D_5 and steady-state value of inductor current I_L are known and can be calculated from steady-state values.

The transfer function matrix Q of the converter is obtained from small signal modeling as follows:

$$\begin{aligned} y &= Qu \\ \text{where } \quad Q &= C(sI-A)^{-1}B + D \end{aligned} \tag{42}$$

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Q is a 4×4 matrix:

$$\begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \\ q_{41} & q_{42} & q_{43} \end{bmatrix} \tag{43}$$

where y and u are system outputs and inputs vectors:

$$\begin{aligned} \frac{\widehat{v}_{o1}(s)}{\widehat{d}_5(s)} &= q_{11} \\ \frac{\widehat{v}_T(s)}{\widehat{d}_1(s)} &= q_{22} \\ \frac{\widehat{i}_b(s)}{\widehat{d}_3(s)} &= q_{33} \\ \frac{\widehat{i}_{v2}(s)}{\widehat{d}_4(s)} &= q_{44} \end{aligned} \tag{44}$$

where the corresponding open-loop transfer function's pole–zero values are descripted in Table 6. The four open-loop transfer functions for Equation (44) correspond to an output variable concerning their duty cycle. To assess their stability and formulate a controller design, performing frequency domain root locus and Bode plot analysis using MATLAB is essential. The root locus plot of the proposed converter for q_{11}, q_{22}, q_{33}, and q_{44} with and without controller is shown in Figure 9A-D for $R_1 = R_2 = 30\Omega$. The denominator of the transfer functions q_{22} shows that the imaginary poles of forward path transfer function of MPBC lie in left half of s-plane and it is shifted away from the imaginary axis with application of controller; thus, the stability of the system improves, whereas as seen from numerator of the transfer function q_{11}, an extra zero appears in the left half of s-plane because of controller; hence, the stability of the system improves with respect to open-loop transfer function as shown in Figure 9A. The objective is to achieve a controller that minimizes steady-state error, ensures an adequate phase margin, promotes high stability, and facilitates a broad bandwidth. Employing MATLAB, we have successfully obtained open-loop Bode diagrams for q_{11}, q_{22}, q_{33}, and q_{44}, considering the simulation parameters outlined in Table 4, and the resulting Bode diagrams are presented in Figure 9E–H.

Upon examining the Bode plots, it is apparent that both phase margin in q_{11} and both gain and phase margin in q_{22} fall short, which can be seen from Table 7, indicating system instability. To rectify this, an improved PID controller needs to be devised to enhance system stability based on Equation (45).

TABLE 6 Pole–zero value of transfer function.

	Open loop		Close loop	
T/F	**Zero**	**Pole**	**Zero**	**Pole**
$q_{11} = \frac{\widehat{v}_{o1}(s)}{\widehat{d}_5(s)}$	-66.67	-33.33 $-16.665 \pm 156.020i$	-94.29 -66.67	$-48.9 \pm 21.1374i$ $-236.7 \pm 222.200i$
$q_{22} = \frac{\widehat{v}_T(s)}{\widehat{d}_1(s)}$	738.7 -33.33	-33.33 $-16.665 \pm 156.020i$	-748.7	-11.56 $-10 \pm 154.536i$
$q_{33} = \frac{\widehat{i}_b(s)}{\widehat{d}_3(s)}$	-33.33 $-121.95 \pm 129.492i$	-33.33 $-16.665 \pm 156.020i$	-294.7 $-121.95 \pm 129.49i$	-36.66 $-30.56 \pm 177.52i$
$q_{44} = \frac{\widehat{i}_{v2}(s)}{\widehat{d}_4(s)}$	-33.33 $-35.135 \pm 156.8933i$	-33.33 $-16.665 \pm 156.020i$	$-35.135 \pm 156.8933i$	-208 $-28.98 \pm 167.54i$

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

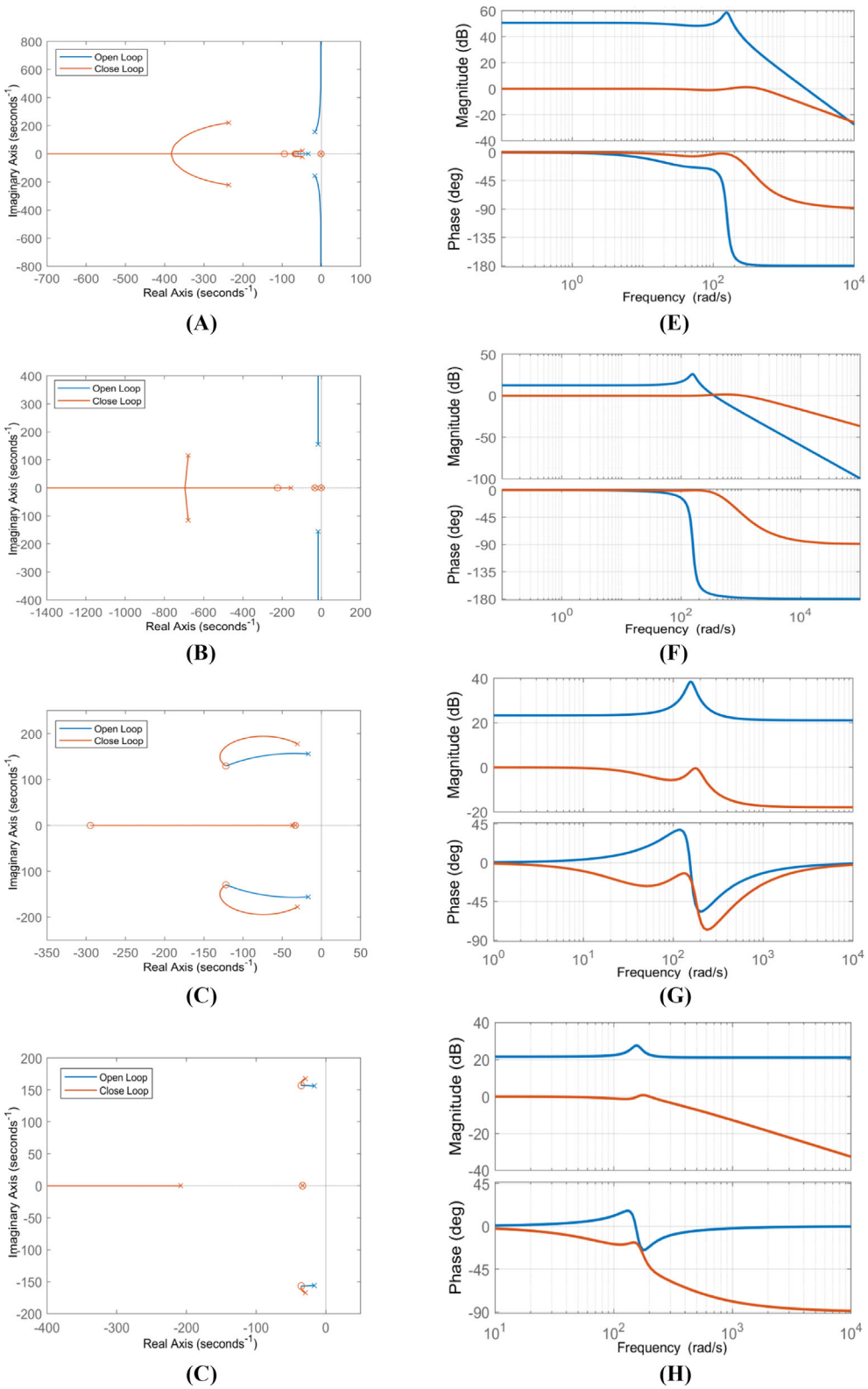


FIGURE 9 Root locus for (A) q_{11}, (B) q_{22}, (C) q_{33}, and (D) q_{44} for open- and close-loop transfer function of proposed MPBC and Bode plot for (E) q_{11}, (F) q_{22}, (G) q_{33}, and (H) q_{44} with and without applying controller of proposed MPBC at $R_1 = R_2 = 30\Omega$.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 Stability analysis of the proposed converter.

	Open loop		Close loop	
T/F	**Gain margin (dB)**	**Phase margin (degree)**	**Gain margin (dB)**	**Phase margin (degree)**
$q_{11}=\frac{\widehat{v_{o1}}(s)}{\widehat{d_5}(s)}$	Inf	0.00591	Inf	135
$q_{22}=\frac{\widehat{v_T}(s)}{\widehat{d_1}(s)}$	−50.7	−86.1	5.25	136
$q_{33}=\frac{\widehat{i_b}(s)}{\widehat{d_3}(s)}$	Inf	Inf	Inf	Inf
$q_{44}=\frac{\widehat{i_{v2}}(s)}{\widehat{d_4}(s)}$	Inf	Inf	Inf	137

$$K(s)=K_p+K_i\left(\frac{1}{s}\right)+K_d s \tag{45}$$

Controlled closed-loop transfer function for Equation (44) using Equation (45) is represented in Table 6. Considering the closed-loop response, Bode diagrams are shown in Figure 9. Upon analyzing the controller stability on the proposed converter system, it is found that for q_{22} closed-loop gains, margin is 5.25d B and the phase margin is 136°, which is under the stable region for $K_P=0.0128, K_I=3.78,$ and $K_D=0$, which are determined using modified genetic algorithm (MGA).[31–33]

5 | RESULTS AND DISCUSSIONS

The proposed MPBC's reliability is analyzed by implementing simulation in MATLAB/Simulink and then verified using experimental findings. Table 4 explains all the parameters associated with the proposed MPBC. As previously discussed, switches $S_1, S_3, S_4,$ and S_5 actively participate in the discharging operation of the converter. The converter is integrated with the inverter for powering single-phase AC loads, which is an addition achievement, and it also has charging capability when load demand is less.

This paper focuses on the stress analysis and performance of the MPBC considering the parasitics whose simulation results are explained in the subsequent subsections and verified with the experimental work. Following the stability analysis conducted in Section 4, the simulation and experimental results are assessed, considering a 30% ripple current and a 2%–3% ripple voltage.

5.1 | Profile analysis of prototype converter

The analysis of a prototype converter encompasses assessing and scrutinizing its characteristics, performance, and behavior, including input voltage and current from sources, inductor behavior, and the profile of output voltage. Figure 10 shows the proposed converter's inductor current, voltage, and output voltages. The step behavior of inductor voltage on the positive side shows the switching of different sources. As the input sources change, the inductor current rate also changes as mention in Equations (2), (5), (8), (11), and (14), respectively, while output voltages V_{T} and V_{O1} are stiff to 120 and 60 V, respectively. Switching operation results in the discontinuity of source currents, as depicted in Figure 11. The switches are controlled based on the power demands of the loads. Specifically, switches S_3 and S_4 facilitate power flow among the three sources.

The converter is tested to check its stability and voltage regulation under various load conditions. These tests aim to see how well the converter handles changing AC and DC loads, which is important for its reliable performance in real-world applications where loads can vary unpredictably.

To simulate real scenarios, both AC and DC loads are applied in different ways. At 0.3 s (point A), a DC load is added, testing how the converter deals with sudden load changes. The system quickly stabilizes within 0.003 s with a

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

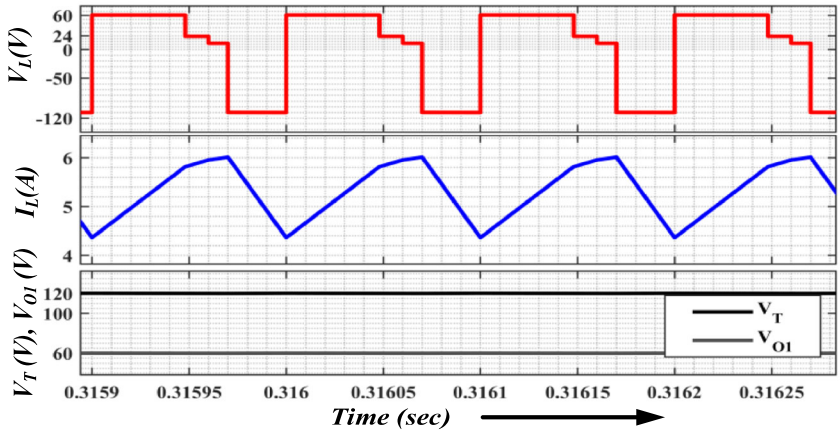


FIGURE 10 Inductor profile and output voltage waveform of MPBC at $R_1 = R_2 = 30\Omega$.

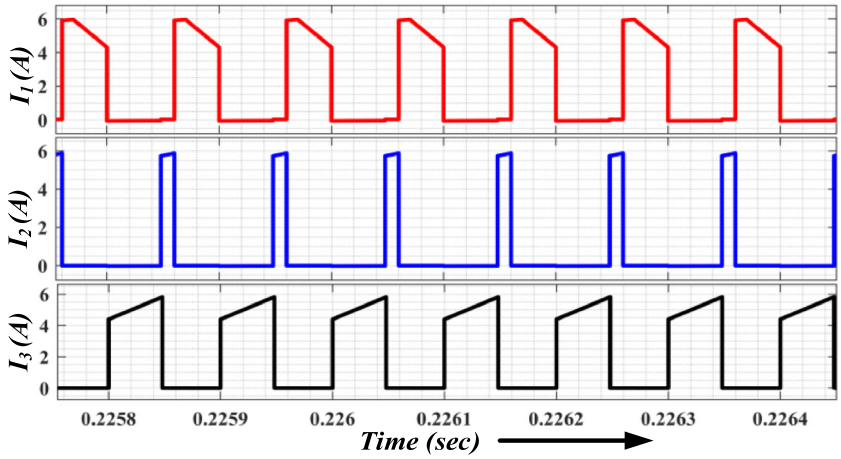


FIGURE 11 Input current waveform of MPBC at $R_1 = R_2 = 30\Omega$.

0.8 V steady-state error, showing that the converter's controls effectively manage the sudden load without major fluctuations. This fast response is crucial for applications needing a constant power supply to avoid performance issues.

At 0.7 s (point B), a single-phase AC motor is added. This tests the converter's ability to handle AC loads, which can be more challenging due to varying current and potential harmonics. The system stabilizes in 0.002 s with a 1.2 V steady-state error. This small error indicates that while there's a slight deviation from the ideal output, the converter adjusts well to the AC load, maintaining performance effectively.

In a more complex test at 0.5 s (Figure 12B), both AC and DC loads switch simultaneously. This tests the converter's ability to handle multiple load changes at once. The system stabilizes within 0.006 s and keeps a 1.3 V steady-state error. This fast response and consistent error rate show the converter's robustness and flexibility, indicating it can handle environments with unpredictable load changes.

Overall, the converter's ability to quickly stabilize with minimal error after both individual and simultaneous load changes highlights its efficiency and reliability for delivering steady power in varying load conditions (Table 8).

5.2 | Voltage and current stress on switching devices.

Switching devices play a crucial role in controlling the flow of electrical energy. However, their operation involves significant changes in voltage and current characteristics, leading to stresses that can affect device performance and

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

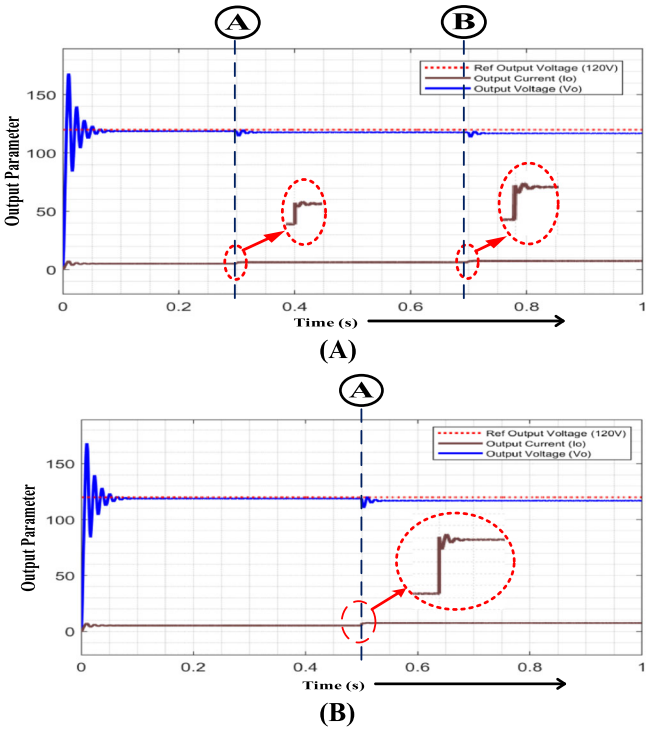


FIGURE 12 Load variation (A) when DC load is connected at A and AC load at B and (B) when both DC and AC load connected at A.

TABLE 8 Performance indices on load change for Figure 12.

		Rise time	Peak overshoot	Settling time	Steady-state error
Figure 12A	Point A	2 ms	0.7%	3 ms	0.8 V wrt 120 V (ref)
	Point B	1.5 ms	0.9%	2 ms	1.2 V wrt 120 V (ref)
Figure 12B	Point A	4.3 ms	1.2%	6 ms	1.3 V wrt 120 V (ref)

TABLE 9 The voltage stress on diodes and power switches.

States	SS-1	SS-2	SS-3	SS-4	SS-5
V_{D1}	$-V_{O1}$	$-V_{O1}$	$-V_{O1}$	$-V_{O1}+V_d$	$-V_{O1}$
V_{D2}	$-V_T$	$-V_T$	$-V_T$	$-V_T$	$-V_T$
V_{D3}	$-V_3$	$-V_3$	$-V_3$	$-V_3$	$-V_3$
V_{D4}	$-V_3+V_2$	V_d	$-V_2$	$-V_2$	$-V_2$
V_{D5}	$-V_3+V_1$	$-V_2+V_1$	V_d	V_d	V_d
V_{S1}	0	0	0	V_T	V_T
V_{S2}	0	0	0	V_T	V_T
V_{S3}	0	V_3-V_2	V_3-V_1	V_3-V_1	V_3-V_1
V_{S4}	0	0	V_3-V_1	V_T-V_3	V_T-V_3
V_{S5}	0	0	0	V_T	V_T

Abbreviation: SS, switching states.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 10 The current stress on diodes and power switches.

States	SS-1	SS-2	SS-3	SS-4	SS-5
I_{D1}	0	0	0	0	0
I_{D2}	0	0	0	0	$I_L-(I_{C2}+I_O)$
I_{D3}	0	0	0	0	0
I_{D4}	0	I_L	0	0	0
I_{D5}	0	0	I_L	I_L	I_L
I_{S1}	I_L	I_L	I_L	0	0
I_{S2}	0	0	0	0	0
I_{S3}	I_L	0	0	0	0
I_{S4}	0	I_L	0	0	0
I_{S5}	0	0	0	0	0

Abbreviation: SS, switching states.

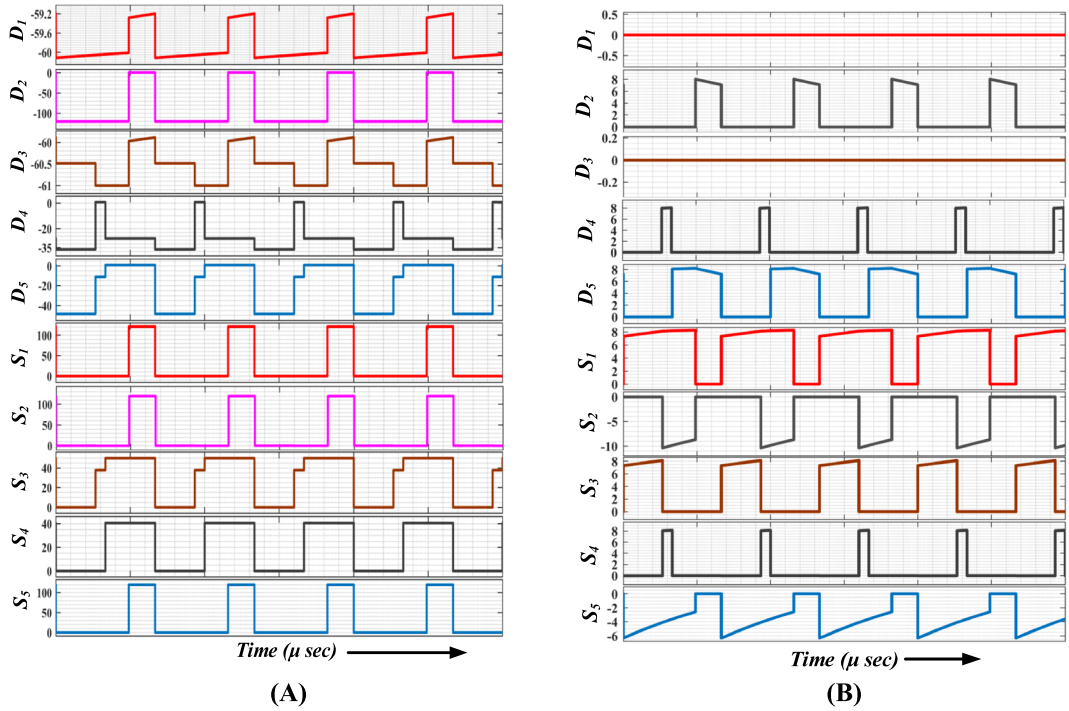


FIGURE 13 Stresses on semiconductor switches of MPBC at $R_1=R_2=30\Omega$. (A) Voltage stress on each diode and switches. (b) Current stress on each diode and switches.

reliability. To find the proper rating of semiconductor's safe operation and improve the system's reliability determination of voltage and current stress is very helpful.

Voltage across the switches in each switching state in discharging mode of operation is shown in Table 9 and average current through them is shown in Table 10.

Voltage stress on all the diodes $D_1, D_2, D_3, D_4,$ and D_5 and switches $S_1, S_2, S_3, S_4,$ and S_5 are shown in Figure 13A. It has lower stress considering the topology 2[29] as sources are connected directly to inductor via switches. So, the proposed converter is better as compared to topology 2 in terms of switch stresses. Current stresses on each individual switches are shown in Figure 13B. The impact on S_3 is notably diminished in the proposed MPBC, a validation confirmed in the following section through experiments. Reduced voltage stress on switches leads to minimized switching losses.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5.3 | Experimental results

A 240 W converter is designed in the laboratory to analyze the prototype's effectiveness. Different profiles of the converter are experimented for developed prototype and shown in Figure 14. The improved switching strategies are implemented using the TMS320F28379D DSP. Table 4 presents the hardware specification of the developed prototype. In the experimental setup, the DC power supply is configured to maintain constant voltage levels of 12 and 24 V, representing fuel cell (FC) and photovoltaic sources, respectively.

Experimental setup is evaluated under two distinct operation modes of the converter. Detailed descriptions of the experimental results for battery discharging are provided in this section. The prototype system integrates several input sources: a dual-channel DC power supply delivering 12 and 24 V, and a 60-V battery composed of five series-connected 12-V, 7 Ah lead-acid batteries. The single inductor L is connected such that it stores energies of all the sources within their time interval in the proposed converter. During the experiments, specific values are defined for the battery current and output voltages. Battery current is denoted as $I_b = 1.9\,\text{A}$, while the output voltages are specified as $V_{O1} = 60\text{V}$ and $V_{O2} = 60\text{V}$, respectively. The resistances R_1 and R_2 are approximately equal to $30\,\Omega$. Stress comparison for topology 2 and proposed converter, which is discussed in Section 2, is verified from experimental

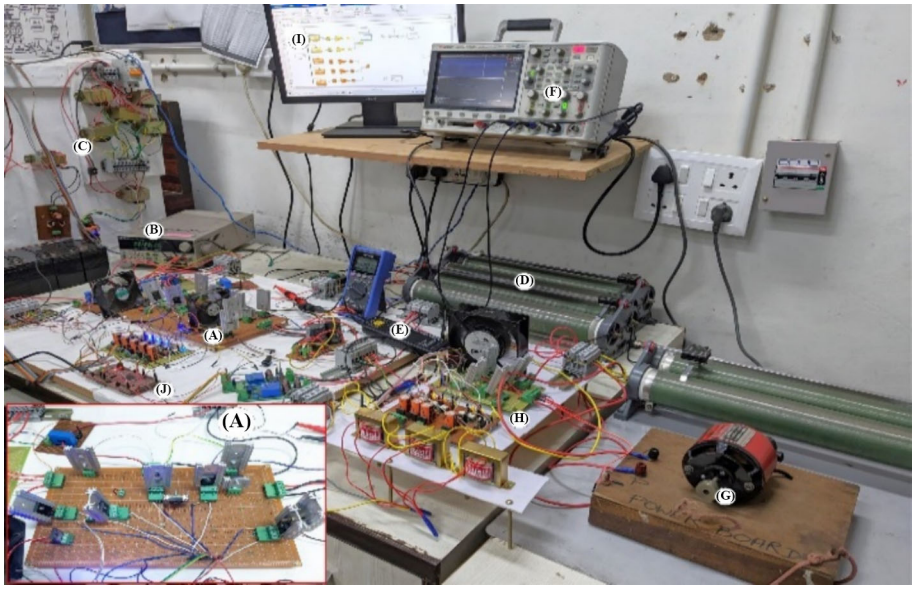


FIGURE 14 Laboratory setup for converter's prototype. (A) Proposed MIMO converter. (B) Input power supply for sources and ESE. (C) Controller and sensor power circuit. (D) DC resistive load. (E) Probes for measurements. (F) Digital storage oscilloscope. (G) Motor setup. (H) Inverter configuration. (I) Computer setup. (J) TMS320F28379D DSP.

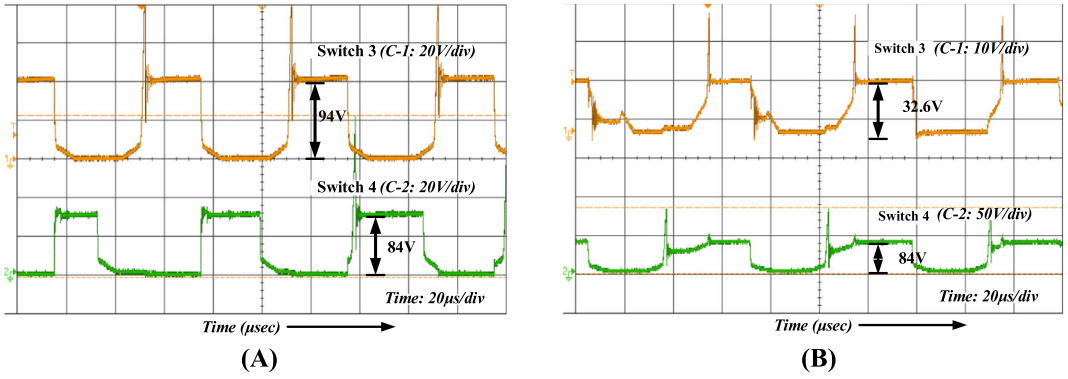


FIGURE 15 (A) Topology 2 switching stress on S3 and S4. (B) Proposed converter switching stress on S3 and S4 at $R_1 = R_2 = 30\Omega$.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 11 Experimental setup specifications.

Components	Part number	Specification(s)
MOSFET	SPW20N60C3	650 V, 20.7 A
Diode	STTH3012	1200 V, 30 A
Voltage sensor	$LV-25P$	500 V
Current sensor	$LA-55P$	50 A
DSP	TMS320F28379D	200 MHz, Dual Core, 32 bit
Optocoupler	TLP350	For isolation
Voltage regulator	IC7815 and IC7819	15 V and −15 V
Rheostat	R_1 R_2	0–500 Ω 0–500 Ω
Inductor	EE65/32/27 Ferrite Core	2 mH
Capacitors	C_1 C_2	1000 μF (150 V) 1000 μF (150 V)

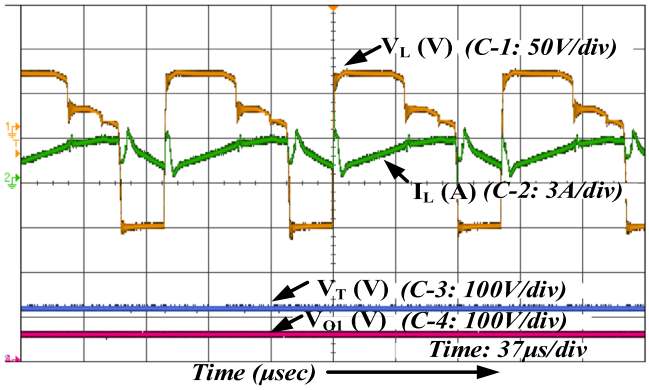


FIGURE 16 Inductor and output voltage waveforms of MPBC at $R_1 = R_2 = 30\Omega$.

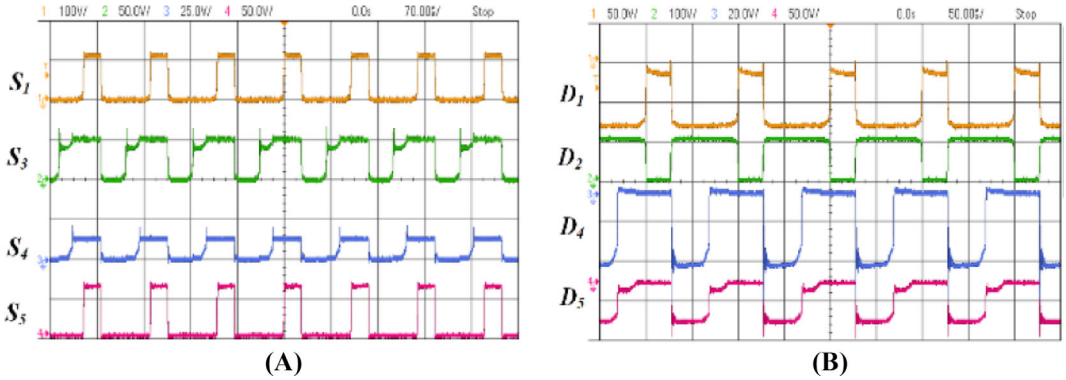


FIGURE 17 Stresses on semiconductor devices of MPBC at $R_1 = R_2 = 30\Omega$. (A) Voltage stress on switches. (B) The voltage stress on diodes.

waveform in Figure 15 where voltage stress on switch S_3 is higher in topology 2 (94 V) due to cascading of switches. The peak-to-peak stress on switch S_3 of proposed converter is 32.6 V, which is 65.31% less than earlier topology (Table 11).

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Experimental results of discharging mode are shown in Figures 16–20. The $V_T = 120V$ is achieved with the duty ratio of switch 1 (S_1), which is approximately the same as that of the results obtained in the simulation. In battery discharging mode, $P_{in1} \cong 22W, P_{in2} \cong 60W, P_b \cong 180W, P_{O1} \cong 130W, and\, P_{O2} \cong 110W$ with an efficiency of $\eta = 91.6\%$.

Figure 16 depicts the converter's output waveforms during steady-state conditions, showcasing the inductor voltage, inductor current, the combined output voltage $V_T = V_{O1} + V_{O2}$, and output voltage V_{O1}. The stepped behavior of

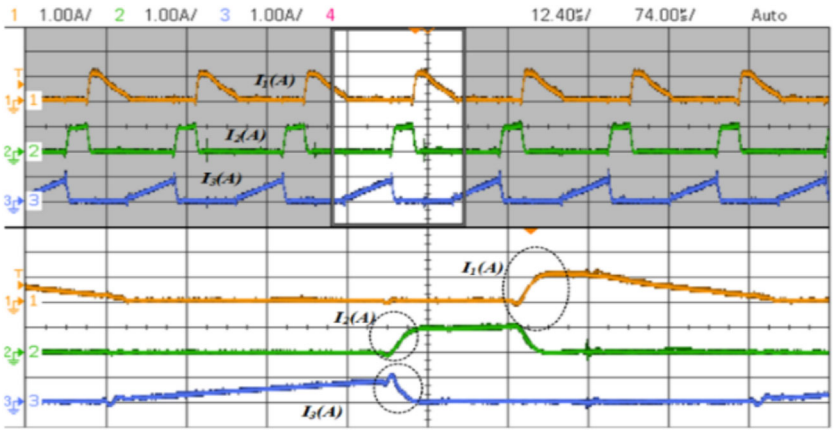


FIGURE 18 Input currents conduction of all sources ($V_1, V_2,$ and V_3) during discharging modes.

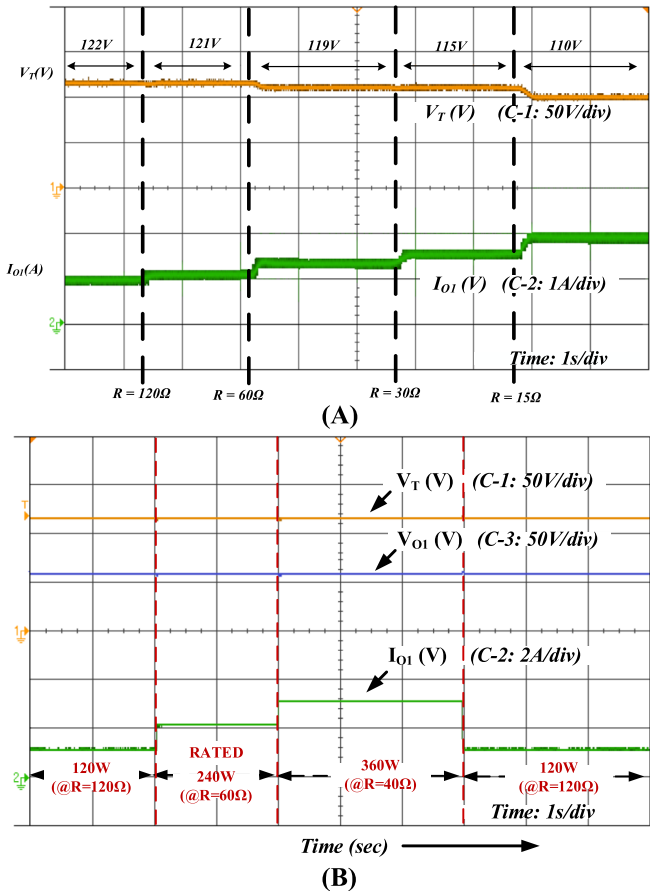


FIGURE 19 Effect of DC load variation on terminal voltage during discharging mode (A) under open loop and (B) under closed loop operation.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

inductor voltage on the positive side is due to the interaction of different sources and ESE to the inductor while switching. Inductor voltage peaks at 62 V due to discharging behavior of ESE. In Figure 17, different stresses on semiconductors are illustrated to assess the switch's reliability, which is within a significant range. S_1 and D_2 are subjected to a maximum voltage stress of -120 V each. Figure 18 displays current sharing among different sources, including ESE, revealing that the load power is distributed based on their duty ratio. Additionally, the waveforms allow observation of the impact of parasitics. Figure 18 also shows the drawn current from input sources V_1, V_2, and ESE, V_3.

Unidirectional port and ESE source currents achieve maximum values of 1.5, 1.3, and 1.9 A, respectively, with an average output current of approximately 1.8 A. Regulation of battery current allows proper distribution of loads power between input sources. The three sources combine to charge the inductor for energy storage. The behavior of terminal voltage V_T is observed on varying load to twice the rated condition using a 500 Ω rheostat, which is illustrated in Figure 19A. A slight dip is observed at twice the rated condition, but the converter is still stable under open-loop condition. After applying the controller to the converter, the output terminal voltage remains stable and fixed at 120 V on load variation, which can be seen from Figure 19B with a steady-state error of less than 5%.

The experimental setup for the proposed MPBC is linked to both a DC load and an AC load. Figure 20A shows the section containing the inverter connected to the MPBC, which, in turn, is linked to a single-phase motor acting as a load. PWM switching is executed to ensure a correct AC output for the load. In Figure 20B, a peak-to-peak voltage of 220 V is observed along with the gating pulse to the inverter. The PWM is created through the TMS320F28379D DSP controller, operating at a carrier frequency of 5 kHz and a fundamental frequency of 50 Hz. Experimental data are obtained for best combination of duty cycle to obtain maximum efficiency, which is plotted in Figure 21A. Duty cycle with $D_1 = 60, D_3 = 40, D_4 = 50, D_5 = 60$ shows maximum efficiency of 90.1% at rated condition of 240 W. Figure 21B shows efficiency plot for the closed loop-controlled output of the proposed MPBC showing various loading conditions, peaking at 91.6% efficiency at 240 W.

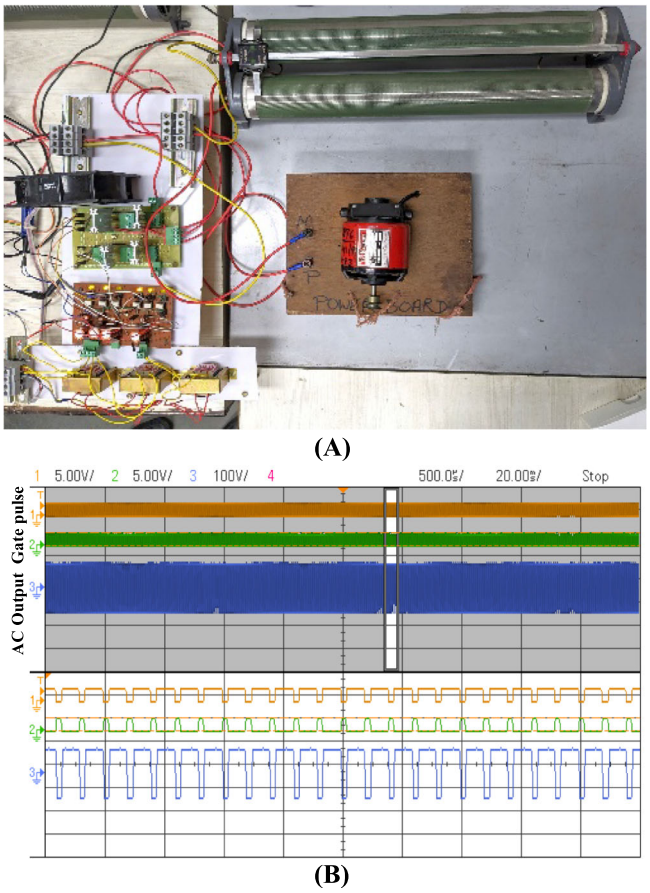


FIGURE 20 (A) Inverter section to the proposed converter loaded with single phase motor. (B) PWM inverter switching and output AC voltage waveform.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

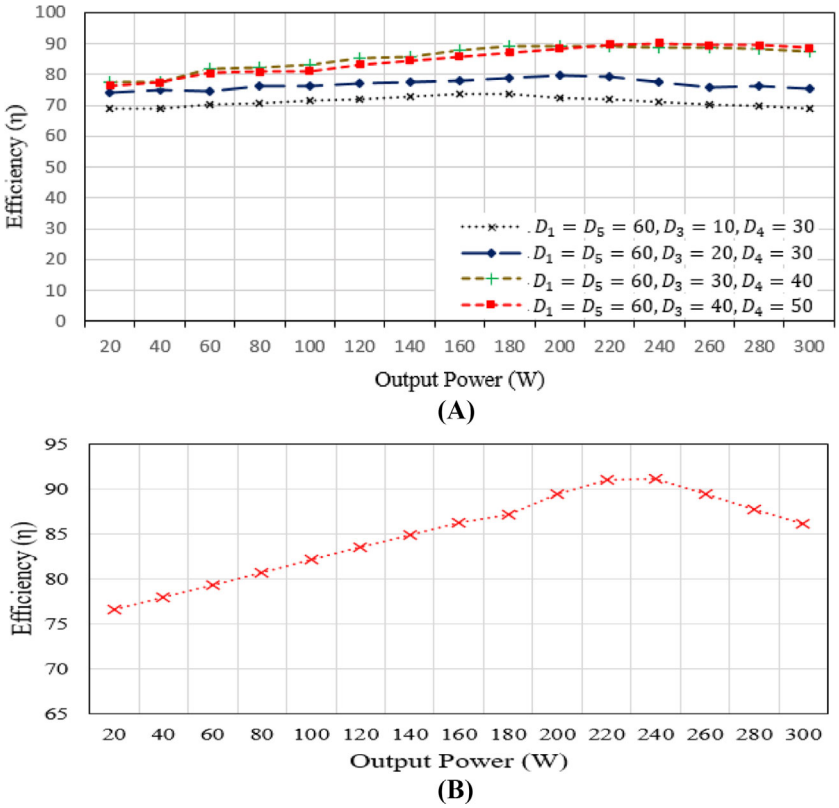


FIGURE 21 (A) Efficiency plot for different duty cycle combination of proposed MPBC at different loading conditions and (B) efficiency plot for the controlled output of the proposed MPBC showing various loading conditions.

6 | CONCLUSION

This paper presents a novel MIMO DC–DC boost converter (MPBC) with three inputs and two outputs. Evaluation and comparison against established topologies show the MPBC's efficiency, having multiple input ports, a rechargeable port for energy storage, and balanced shared output ports. The converter features a unified structure with an SI that integrates multiple power sources in renewable energy systems. The proposed converter is a versatile solution for hybridizing power sources, where FC, PV, and battery are considered power sources and ESE, respectively, and enables efficient energy transfer between them. It operates in two main modes: battery discharging mode, where both input sources deliver power to the output, and battery charging mode, where the input sources not only supply loads but also deliver power to the battery. The converter output offers varied DC voltage levels and executes them for connection to multilevel inverters. These different voltage levels can be employed to power AC motors through integrated inverters and provide auxiliary DC power as needed. To validate the operation of the proposed converter, experimental verification has been conducted using a low-power range prototype. Significant contributions of this paper include thorough steady-state and small signal modeling of the MPBC, considering ESRs for stability, stress analysis (voltage and current), and performance evaluation. A fabricated 240 W MPBC, coupled with improved switching strategies utilizing DSP TMS320F28379D, highlights the efficiency through comparative experimental and simulation results. The MPBC's design effectively regulates output power by efficiently distributing input power from ESE and RES, thereby reducing voltage stress on switches. The converter's capabilities were experimentally verified, and its potential applications include hybrid power systems, grid-connected PV systems, and reducing torque ripple in induction motors through multilevel inverters.

CONFLICT OF INTEREST STATEMENT

On behalf of all authors, the corresponding author states that there is no conflict of interest.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Ashutosh Gupta https://orcid.org/0000-0002-1379-8707

REFERENCES

1. Haegel NM, Kurtz SR. Global progress toward renewable electricity: tracking the role of solar. *IEEE Journal of Photovoltaics*. 2021;11(6):1335-1342. doi:10.1109/JPHOTOV.2021.3104149
2. Palanisamy R, Seenithangam J, Palanisamy R. A hybrid output multiport converter for standalone loads and photovoltaic array integration. *Int Trans Electr Energ Syst*. 2020;30(7):e12410. doi:10.1002/2050-7038.12410
3. Tarassodi P, Adabi J, Rezanejad M. A power management strategy for a grid-connected multi-energy storage resources with a multiport converter. *Int J Circ Theor Appl*. 2023;51(6):2744-2759. doi:10.1002/cta.3540
4. Malik MZ, Zhang S, Ali A, Farooq A. Design and implementation of a multi-port bidirectional converter for electric vehicle applications. *Int J Circ Theor Appl*. 2023;51(10):4728-4753. doi:10.1002/cta.3649
5. Behjati H, Davoudi A. Power budgeting between diversified energy sources and loads using a multiple-input multiple-output DC-DC converter. 49th IEEE/IAS Industrial & Commercial Power Systems Technical Conference, Stone Mountain, GA, USA; 2013, pp. 1–12. 10.1109/ICPS.2013.6547333.
6. Kwasinski A. Identification of feasible topologies for multiple-input DC–DC converters. *IEEE Trans Power Electron*. 2009;24(3):856-861. doi:10.1109/TPEL.2008.2009538
7. Matsuo H, Shigemizu T, Kurokawa F, Watanabe N. Characteristics of the multiple-input DC-DC converter, Proceedings of IEEE Power Electronics Specialist Conference – PESC '93, Seattle, WA, USA; 1993, pp. 115–120. 10.1109/PESC.1993.472079.
8. Bizhani H, Techato K, Noroozian R, Muyeen SM, Blaabjerg F. A novel multiport converter switching scheme to integrate multiple renewable energy sources. *Int J Circ Theor Appl*. 2022;50(5):1750-1769. doi:10.1002/cta.3227
9. Jayaram N, Halder S, Panda KP, Pulavarthi SV. A novel design with condensed component of multi-input high gain non-isolated DC–DC converter for performance enhancement in carbon neutral energy application. *IEEE J Emerg Sel Top Ind Electron*. 2023;4(1):37-49. doi:10.1109/JESTIE.2022.3211779
10. Khaligh A, Cao J, Lee Y-J. A multiple-input DC–DC converter topology. *IEEE Trans Power Electron*. 2009;24(3):862-868. doi:10.1109/TPEL.2008.2009308
11. Prabhala VAK, Somayajula D, Ferdowsi M. Power sharing in a double-input buck converter using dead-time control. 2009 IEEE Energy Conversion Congress and Exposition, San Jose, CA, USA, IEEE; 2009, pp. 2621–2626. 10.1109/ECCE.2009.5316355.
12. Li Z, Onar O, Khaligh A, Schaltz E. Design and control of a multiple input DC/DC converter for battery/ultra-capacitor based electric vehicle power system. 2009 Twenty-Fourth Annual IEEE Applied Power Electronics Conference and Exposition, Washington, DC, USA, IEEE; 2009, pp. 591–596. 10.1109/APEC.2009.4802718.
13. Luo P, Liang T-J, Chen K-H, Chen S-M, Chen J-F. Syntheses of three-port DC–DC converters. *IEEE Trans Power Electron*. 2023;38(12):16196-16207. doi:10.1109/TPEL.2023.3309892
14. Kamaraj V, Chellammal N, Chokkalingam B, Munda JL. Minimization of cross-regulation in PV and battery connected multi-input multi-output DC to DC converter. *Energies*. 2020;13(24):6534. doi:10.3390/en13246534
15. Zheng Y, Guo J, Leung KN. A single-inductor multiple-output Buck/boost DC–DC converter with duty-cycle and control-current predictor. *IEEE Trans Power Electron*. 2020;35(11):12022-12039. doi:10.1109/TPEL.2020.2988940
16. Pathak PK, Yadav AK, Padmanaban S, Alvi PA, Kamwa I. Fuel cell-based topologies and multi-input DC-DC power converters for hybrid electric vehicles: a comprehensive review. *IET Gener Transm Distrib*. 2022;16(11):2111-2139. doi:10.1049/gtd2.12439
17. Wu H, Wan C, Sun K, Xing Y. A high step-down multiple output converter with wide input voltage range based on quasi two-stage architecture and dual-output LLC resonant converter. *IEEE Trans Power Electron*. 2015;30(4):1793-1796. doi:10.1109/TPEL.2014.2349917
18. Zhang Y, Xu D. Design and implementation of an accurately regulated multiple output ZVS DC–DC converter. *IEEE Trans Power Electron*. 2007;22(5):1731-1742. doi:10.1109/TPEL.2007.904206
19. Li XL, Tse CK, Lu DDC. Single-inductor multi-input multi-output DC-DC converter with high flexibility and simple control. 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain, IEEE; 2020, pp. 1–5. 10.1109/ISCAS45731.2020.9180788.
20. Behjati H, Davoudi A. A MIMO topology with series outputs: an interface between diversified energy sources and diode-clamped multilevel inverter. 2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC), Orlando, FL, USA, IEEE; 2012, pp. 1–6. 10.1109/APEC.2012.6165790.
21. Behjati H, Davoudi A. A multiple-input multiple-output DC–DC converter. *IEEE Trans Ind Appl*. 2013;49(3):1464-1479. doi:10.1109/TIA.2013.2253440
22. Salimi M, Radmand F, Firouz MH. Dynamic modeling and closed-loop control of hybrid grid-connected renewable energy system with multi-input multi-output controller. *J Mod Power Syst Clean Energy*. 2021;9(1):94-103. doi:10.35833/MPCE.2018.000353
23. Saadatizadeh Z, Heris PC, Mantooth HA. Modular expandable multiinput multioutput (MIMO) high step-up transformerless DC–DC converter. *IEEE Access*. 2022;10:53124-53142. doi:10.1109/ACCESS.2022.3175876

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

24. Veerachary M, Trivedi A. Linear matrix inequality-based multivariable controller design for boost cascaded charge-pump-based double-Input DC–DC converter. *IEEE Trans Ind Appl*. 2022;58(6):7515-7528. doi:10.1109/TIA.2022.3201173
25. Rostami S, Abbasi V, Parastesh M. Design and implementation of a multiport converter using Z-source converter. *IEEE Trans Ind Electron*. 2021;68(10):9731-9741. doi:10.1109/TIE.2020.3022538
26. Filsoof K, Lehn PW. A bidirectional multiple-input multiple-output modular multilevel DC–DC converter and its control design. *IEEE Trans Power Electron*. 2016;31(4):2767-2779. doi:10.1109/TPEL.2015.2448112
27. Sadeghpour D, Bauman J. Integrated three-port converter for solar-charged electric vehicle applications. *IEEE Trans Ind Electron*. 2024;71(7):6907-6917. doi:10.1109/TIE.2023.3301520
28. Wang B, Zhang X, Ye J, Gooi HB. Deadbeat control for a single-inductor multiple-input multiple-output DC–DC converter. *IEEE Trans Power Electron*. 2019;34(2):1914-1924. doi:10.1109/TPEL.2018.2832243
29. Nahavandi A, Hagh MT, Sharifian MBB, Danyali S. A non-isolated multiinput multioutput DC–DC boost converter for electric vehicle applications. *IEEE Trans Power Electron*. 2015;30(4):1818-1835. doi:10.1109/TPEL.2014.2325830
30. Anees M, Nasir M, Ahmad Zaffar N, Abbas Khan H, Vasquez JC, Guerrero JM. Scalable architecture of DC microgrid implemented with multi-input multi-output converter. *IET Power Electron*. 2020;13(19):4480-4489. doi:10.1049/iet-pel.2020.0375
31. Sundareswaran K, Sreedevi VT. Boost converter controller design using queen-bee-assisted GA. *IEEE Trans Ind Electron*. 2009;56(3):778-783. doi:10.1109/TIE.2008.2006026
32. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput*. 2002;6(2):182-197. doi:10.1109/4235.996017
33. Kumar V, Behera RK, Joshi D, Bansal R. *Power Electronics, Drives, and Advanced Applications*. CRC Press; 2020. doi:10.1201/9781315161662

How to cite this article: Gupta A, Joshi D. Mathematical modeling and stress-aware stability analysis of a nonideal multiport Single Inductor DC–DC converter for renewable energy. *Int J Circ Theor Appl*. 2024;1-32. doi:10.1002/cta.4234

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4234 by Delhi Technological University, Wiley Online Library on [26/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

Classification of Hateful Memes by Multimodal Analysis using CLIP

Samarth Munzni
Department of Computer Engineering
Delhi Technological University
New Delhi, India
smunzni@gmail.com

Shikhar Dixit
Department of Computer Engineering
Delhi Technological University
New Delhi, India
shikhardixit11@gmail.com

Dr. Aruna Bhat
Department of Computer Engineering
Delhi Technological University
New Delhi, India
aruna.bhat@dtu.ac.in

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10626270

***Abstract*—Due to the widespread usage of social media platforms, memes and other hateful contents have become a major concern as they propagate harmful stereotypes, and often inciting violence and social tensions in societies. To counter it, this research paper proposes a novel approach for the detection of such hateful memes using multimodal techniques by using the CLIP (Contrastive Language-Image Pre-training) model. As the image and textual parts of the memes are seldom semantically aligned, thus we need to integrate both these information to capture the correct context of the memes to detect the hateful memes. Through extensive testing on a large dataset of memes provided by Meta we demonstrate the accuracy of this approach. Through this paper we contribute to the development and refinement of more robust and ethical solutions to tackle this growing problem of hate speech to promote a healthy and safe discourse on social media.**

Keywords—Hateful memes, Multimodal analysis, CLIP model, Online platforms, Hate speech

I. INTRODUCTION

The widespread usage of social media has led to the rapid spread of memes, which are a type of online content combining images and some caption. But, apart from humour, this medium of content has become a serious problem as it has led to the spread of hateful content. These memes which target minority groups, organizations, race or religion present a serious danger to the discourse as they disseminate stereotypes and even real world violence and polarisation of society.

Normal methods of moderation are not enough to counter the nuanced nature of this hateful content. These approaches are inadequate given the sheer volume of this content but also the complexity because of which manual reviews and keyword filtering of the content is almost impossible. There is a need to automate this process by making systems which are able to understand the nuances of this multimodal information.

In this paper we propose a novel approach for the detection of hateful memes. Using both text and image cues we aim to capture the subtle context and its nuances which allows us to do more accurate detection of harmful content. Using empirical evaluation and analysis we show the accuracy of our approach which can contribute to development of more robust systems allowing us to tackle the hate speech epidemic and promoting healthy, inclusive and friendly online discourse.

II. RELATED WORK

A large amount of research has been done on the detection of hateful memes using both unimodal and multimodal techniques. We will discuss some of these especially the multimodal approaches by different researchers.

[1] Sethi et al. (2021) recognize the difficulty of detection of hateful content in memes as it involves both textual and visual cues. They used multiple pre-trained models like VGG19, Xception, Support Vector Machines and Naïve Bayes for the classification of these hateful memes. The integrated stacked models, early fusion techniques and GloVe Embeddings were also notably studied by the researchers in the multimodal technique.

[2] Bhat et al. (2023) for the detection and classification of hateful memes also focused on the multimodal approaches. An innovative structure was made that uses the object tags and image details in combination with the textual analysis of the captions. For the detection of hateful memes, they are subsequently merged in order to train the classification model. This method was able to properly evaluate both the visual as well as textual contexts that comprise the memes for effective detection.

[3] Zhou, Chen, and Yang (2021) focus on an interesting new feature for identification of the hateful memes. They recognize the difficulty of incorrectly understood or irrelevant textual or image features of the memes. As it is assumed that the texts and the images are semantically aligned, this proves to be a fatal flaw because, in this type of content, it might not be necessarily true resulting in an incorrect understanding of the context. So, the model should have a deep contextual understanding as well as justification across different modalities. Thus, they use the image captioning process in the classification process. Using this, the classification model understands the content at a deeper level beyond the basics, thereby creating an efficient and robust approach to detect hateful content much more accurately. They are able to show the robustness of this approach by conducting experiments on many large diverse datasets of multimodal memes.

[4] Fang et al. (2022) deals with the issue that our present algorithms have a lesser understanding of the image modality semantic, where we are unable to understand the image modality as much as the textual modality, due to the great complexity of an image. Image caption supervision (ICS) auxiliary approach is used, which uses the text modality to better understand the image modality. ICS auxiliary method does so by using the image caption which is then used to supervise the feature learnings for the image. The ICS approach shows much better results compared to both unimodal and multimodal models tested on large and diverse Facebook Hateful meme challenge dataset. The above approach shows that its very effective to use the textual

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:31 UTC from IEEE Xplore. Restrictions apply.

modalities to better understand the images of the memes and thus it is a very robust and precise model for the classification of offensive content.

The above-mentioned researches are evident of the increased interest and importance of creating effective and robust classification models for the detection of hateful content especially in the form of memes in the online discourse. Deep learning, natural language processing and computer vision are being used by researchers all over the world in order to provide a solution for this. This could help people create better content moderation strategies in order to provide everyone a safe, healthy and inclusive discourse on the internet.

III. PROPOSED WORK

A. Architecture

CLIP is a very innovative and interesting model designed for multimodal analysis, that takes both images and texts to process through its dual encoders. The first encoder handles and process all the image data, extracting any semantic feature present using convolutional neural networks that are pretrained on some datasets like ImageNet. The other encoder processes text and generates embeddings via transformer architectures such as BERT and GPT. Using multimodal contrastive learning, it learns to draw correlations within similar concepts across different modalities while it separates different ones. While training, it is trained on pairs of images and its text description, which together represents semantically related content and removes unrelated ones.

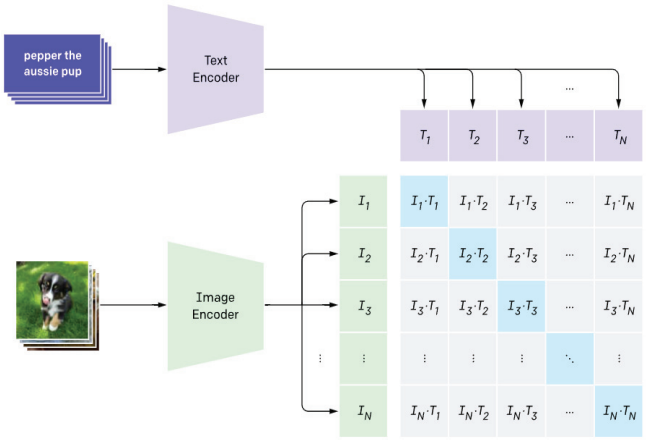


Fig. 1. Contrastive Pre-Training [5].

For CLIP, contrastive pre-training works by grouping together topics with a same semantic meaning in both pictures and text and separating those with a different meaning. The complex CLIP training program aims at teaching the system an appropriate alignment of the textual and visual representations inside one shared embedding space. When training we display photo paired together with captions in an attempt to ensure similarity of relevant pair embeddings and dissimilarity of unrelated pair embeddings. An architecture of a neural network with two encoders, one for pictures and one for text, is used by CLIP during contrastive pre-training. In order to efficiently capture the semantic similarity between multimodal inputs, these encoders are trained together to map text and pictures into a single embedding space. Tasks like image-text retrieval, zero-shot classification, and multimodal comprehension are made easier by CLIP's ability to comprehend and link visual and textual concepts through this common embedding space.

The Dataset Classifier is the main part of CLIP that provides zero-shot prediction by building connections between textual descriptions and particular image classes. Through this training process, the Dataset Classifier learns to relate textual inputs with the correct image classes by bridging the semantic gap that existed between words and images.

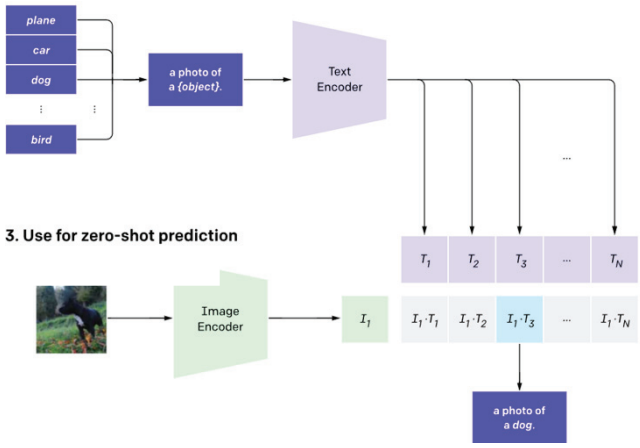


Fig. 2. Dataset Classifier from labelled text and using it for zero shot prediction [5]

The Labelled Text Dataset is produced to train the Dataset Classifier. In this dataset, every text sample is connected to one or more class labels that signify the picture class or classes which it outlines. For example, Text samples describing cats can be tagged with labels such as "cat," "animal," and "pet." After that, Dataset Classifier is instructed to use this annotated text as training data, which is usually a supervised learning technique, in which the model is taught to identify the class labels from the annotated descriptions. The Dataset Classifier returns the most probable class labels, which are also linked with text, when given fresh textual descriptions during prediction time. In a zero-shot prediction, CLIP uses the output of the class labels as an input and then creates embeddings of the textual description and the predicted class labels. Then it determines how similar or distant they are in order to show the picture classes that may be attributed to them.

B. Data Preprocessing

We analyze text descriptions connected to memes and identify the most often stated objects in each category using Natural Language Processing (NLP) techniques. This preprocessing stage not only facilitates feature engineering and model training by illuminating significant themes and subjects in the data, but it also facilitates comprehension of the memes' content.

Starting with an NLP toolkit, we tokenize and parse the textual meme descriptions. This is the first step in the preparation phase. We can identify entities like individuals, groups, locations etc, using a feature of pre-trained NLP models called Named Entity Recognition (NER). We are able to get a list of objects (individuals, minority groups) mentioned in both the hateful and non-hateful meme groups using the NER feature.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:31 UTC from IEEE Xplore. Restrictions apply.

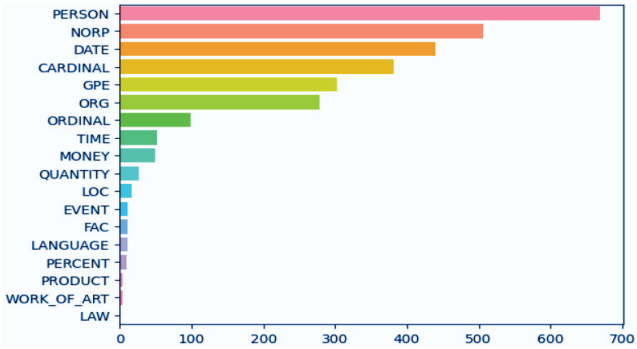


Fig. 3. Most mentioned entity types of non-hateful memes

Using this we check which words appear where and how often by doing a frequency analysis. Counting the occurrences of each item and identifying the ones that appear most often is what this involves. Comparing the occurrence of these objects in both these groups, we can develop a rough idea of what the prevalent themes of Hateful and Non-Hateful memes are, in general. This study sheds information on the entities that matter most for classifying hate speech as opposed to non-hate speech, which in turn helps with feature selection and model training, two further steps in the preprocessing pipeline.

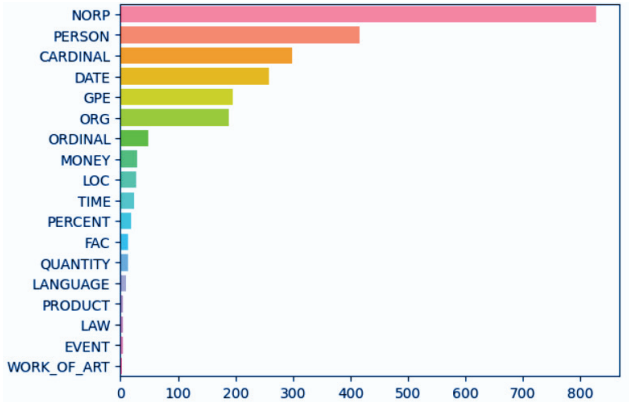


Fig. 4. Most mentioned entity types of hateful memes

To properly understand the thematic between hateful and non-hateful memes, we additionally look at the frequency of certain entity types including races, political affiliations, and religious groups. From this analysis we can understand the dataset in much more detail especially by focusing on such entities whose presence in the two groups is very different.

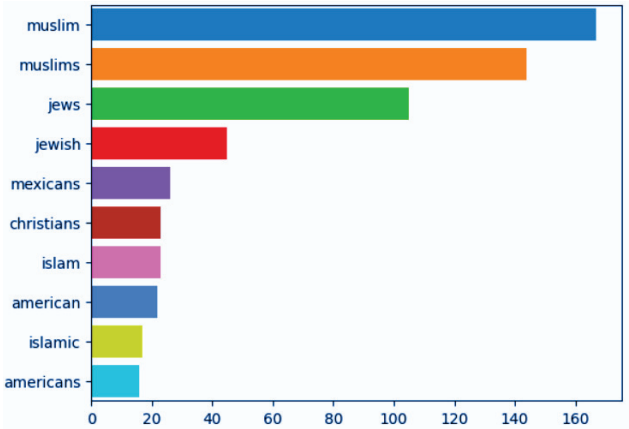


Fig. 5. Most mentioned NORP entities of hateful memes

After using natural language processing techniques like NER for entity recognition and frequency analysis, we are able to understand and grasp all the themes and nuances which are present in both the hateful and non-hateful meme datasets. This understanding helps us in feature selection and model training allowing us to create good models for the detection of these hate memes. These insights also inform further phases of analysis.

C. *Detecting hateful content using labelled data*

We used the pre-trained model along with labeled data for fine-tuning the model to detect hateful memes using both text and images. The varied dataset had wide range of diverse subjects like race, religion and other minute details that provided a good variance of context and similar fine distinctions. Due to this robust preprocessing, we can detect hateful content very accurately and effectively during training. Due to its recursive nature, CLIP sees multiple iterations of updates to its params, which eventually helps to correctly judge hateful and non-hateful memes with a great accuracy. To evaluate the performance of the model on new and unseen data, we used validation.json, which is used to finetune and refinement of model. We also used multiple metrics like accuracy, precision, recall, and F1 score to measure the effectiveness and soundness of the fine-tuned model. This model can be used for a positive experience by fostering safe positive and inclusive internet for all internet users.

IV. Results

While the main aim of this paper was the CLIP model, in our study we used many other state-of-the-art models like Text CNN, TF-IDF with SVM/NB, BERT and VisBERT. All these models showed varying levels of performance in detecting hateful content which gave us very insightful findings.

Text-CNN (Convolutional neural network) specializes in textual data. This achieved an accuracy of 63.1% in classification of hateful content. Since it is unable to combine the image features of the memes, it fails to completely understand all the context and the nuances of the meme.

TF-IDF with SVM and NB, is a classic approach using term frequency-inverse document frequency representations. We combine this with support vector machines and naïve bayes, where it produced accuracies of 61.8% and 60% respectively. This gave a decent performance especially in those memes where the semantic understanding was less critical.

BERT which was a transformer-based model gave an accuracy of 58.3% in detection of hateful memes.

VisBERT, which is able to take in visual inputs also by incorporating visual embeddings along with textual ones performs much better with an accuracy of 67.3%. VisBERT is properly able to use the multimodal information which allows it to enhance its detection abilities.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:31 UTC from IEEE Xplore. Restrictions apply.

TABLE I. RESULTS

Model	Accuracy(%)	Recall(%)	Precision(%)	F1
TF-IDF+SVM	61.8	18.2	63.3	0.28
TF-IDF+NB	60	13.2	57.1	0.21
BERT	58.3	35.5	63.6	0.46
Text-CNN	63.1	45.7	50.4	0.48
Visual BERT+Tags	67.3	58.8	66	0.62

Moving on to the main subject of this paper, CLIP which is a multimodal model trained on a vast number of image and texts surpasses expectations. In our experiments it achieves a very high level of accuracy of 81.98% in the detection of hateful memes. It is able to integrate the visual and textual information smoothly which allows it to understand complex semantic associations as well as contextual nuances making it highly effective.

The reason, behind CLIP's high performance is in its approach to learning, which allows it to understand and combine visual and textual information seamlessly. By training across a variety of datasets CLIP develops representations that grasp the core meanings in both images and text improving its ability to properly understand the contextual nuances of the meme. Moreover, refining its skills through targeted training on the labeled data has increased CLIP's accuracy in identifying hateful content. With fine-tuning and thorough evaluation processes CLIP has proven its applicability for real world scenarios.

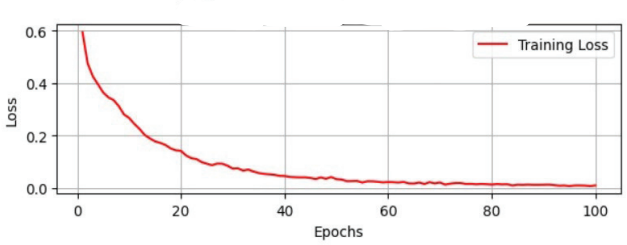


Fig. 6. Training loss vs epochs

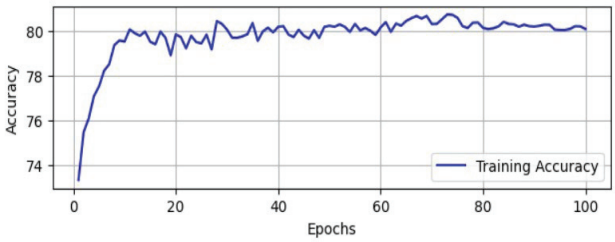


Fig. 7. Training accuracy vs epochs

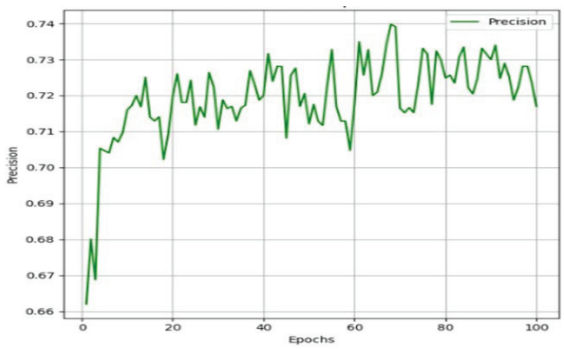


Fig. 8. Precision for CLIP

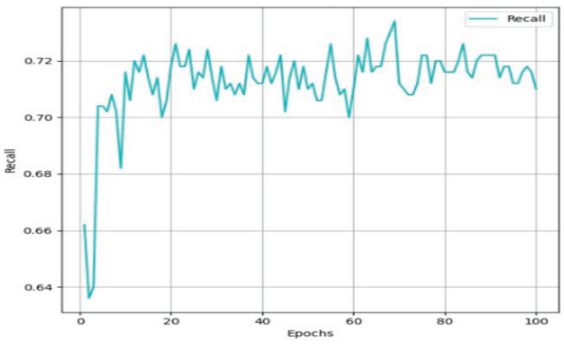


Fig. 9. Recall for CLIP

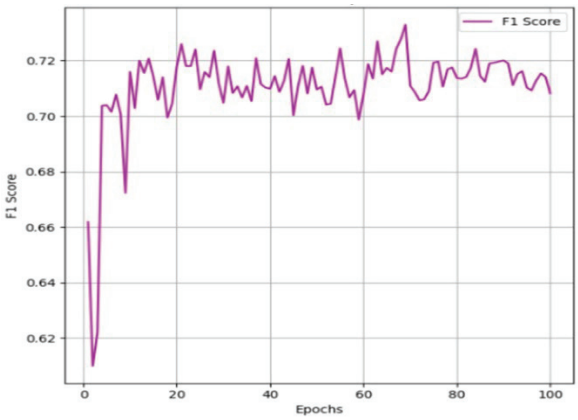


Fig. 10. F1 score for CLIP

In all the 3 graphs, we can observe the metrics (precision, recall, and F1 score) for the CLIP model throughout 100 epochs of training. For each of the metrics, we observe an abrupt rise initially, after which it tends to stabilize from around 20 epochs. The first graph shows the precision, which tends to stabilize between 0.70 and 0.74, meaning that the model was able to accurately predict the positives throughout the epochs. The second graph shows the recall, which also stabilizes between 0.70 and 0.72, indicating that the model was able to consistently detect relevant instances of hateful memes. The third graph shows the F1 score, which follows the patterns of both precision and recall, as it balances the two and tends to stabilize between 0.70 and 0.72. From these graphs, it can be understood that the CLIP model reached reliable performance after initial periods of training, in both epochs and time, for the meme dataset.

V. FUTURE WORK AND CONTRIBUTION

Enhanced Multimodal Techniques: Utilization of the recent advanced methodologies for multimodal fusion with attention mechanisms and graph neural network can also prove to be very important for increasing the accuracy and the robustness of our model. Moreover, to improve the classification outcomes, it will be feasible to study other contexts of information, including user tags/attributes and spread patterns of memes.

Real-Time Detection Systems: With regard to hatred, there is a need to design reaction routines to recognize shameful memes the moment they are posted on the internet. Part of this is designing huge architectures using distributed computing and stream processing philosophies, whereby detecting procedures will not be slow, and such systems can be integrated seamlessly with social networking platforms for efficient content moderation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:31 UTC from IEEE Xplore. Restrictions apply.

Ethical and Fairness Considerations: Concerning proper usage of this technique, it is absolutely necessary to ensure the fair and safe usage of these memes without the curtailment of the right to speech. There is a serious threat that these tools could potentially be used to restrict the right to speech of people. In particular, we will set up and define the measures of bias, distinguishing the features of bias audits, clearing up the means of methodologies with regard to bias, and outlining reasonable ethical regulations on bias in consideration of fairness, transparency, and accountability. It involves reporting to the stakeholders with the focus of addressing the mentioned biases to meet legal and ethical standards. Steps should be taken to maintain data security to protect the user as well as their private data and meeting the requirements of the data privacy acts.

VI. Conclusion

This research is dedicated to displaying a revolutionary way of tackling the escalation of hateful memes on social media platforms. Memes have turned into the most popular way for the spreading misconceptions and even calling for violence among people. Our research suggests an approach that utilizes multimodal analysis with CLIP to be a very effective method that detects hateful content.

The traditional extent of moderation fails to fully address the contextualism of hateful memes mostly due to their staggering number and fast-changing attributes. Such systems, which are making it possible to process the multiplexity of multimodal information including the understanding of context are the building blocks of this project. CLIP with its architecture based on how the model looks and multimedia contrastive learning, is a new, creative solution.

The study is carried out by using of large dataset provided by Meta against the proposed method have its effectiveness proven. This mechanism that combines both the use of textual and visual cues caused CLIP to achieved a great accuracy of 81.98% in identifying hateful memes. The manner in which it effortlessly unifies data from pictures and written words enables it to be remarkably sensitive to the different nuances and semantic meanings which in turn makes it most suited to identifying harmful content.

Comparison with popular top-of-the-line models explicitly emphasizes the benefits of CLIP for generating hashtags in this task. While the normal Text boasted some decent accuracy, CLIP still outperforms these by a significant margin non-multimodal approaches such as BERT are also shown to be decent models but it is CLIP that shows how multimodal analysis can deliver more soulful results.

The strength of CLIP in relation with generalization is attributed to the diverse datasets which the model has been trained on aforehand. This helps envisage the fundamental meanings, both visually in pictures and textually in descriptions. As its purely deep learning-based model, it can intelligently differentiate between offensive and light-hearted content.

References

[1] Sethi, Anushka, Utkarsh Kuchhal, Anjum, and Rahul Katarya. 2021. "Study of Various Techniques for the Classification of Hateful Memes," August. https://doi.org/10.1109/rteict52294.2021.9573926.

[2] Bhat, Aruna, Vaibhav Vashisht, Vaibhav Raj Sahni, and Sumit Meena. 2023. "Hate Speech Detection Using Multimodal Meme Analysis," May. https://doi.org/10.1109/icaaic56838.2023.10140393.

[3] Zhou, Yi, Zhenhao Chen, and Huamin Yang. 2021. "Multimodal Learning for Hateful Memes Detection." ArXiv (Cornell University), July. https://doi.org/10.1109/icmew53276.2021.9455994

[4] Fang, Huaicheng, Fuqing Zhu, Jizhong Han, and Songlin Hu. 2022. "Multimodal Hateful Memes Detection via Image Caption Supervision," December. https://doi.org/10.1109/smartworld-uic-atc-scalcom-digitaltwin-pricomp-metaverse56740.2022.00221.

[5] "CLIP: Connecting Text and Images." n.d. Openai.com. https://openai.com/research/clip.

[6] Bhat, Aruna, Vaibhav Varshney, Varun Bajlotra, and Vishesh Gupta. 2022. "Detection of Hatefulness in Memes Using Unimodal and Multimodal Techniques." IEEE Xplore. May 1, 2022. https://doi.org/10.1109/ICICCS53718.2022.9788376.

[7] Macrayo, Genevive, Wilfredo Casino, Jerecho Dalangin, Jervin Gabriel Gahoy, Aaron Christian Reyes, Christian Vitto, Mideth Abisado, Shekinah Lor Huyo-a, and Gabriel Avelino Sampedro. 2023. "Please Be Nice: A Deep Learning Based Approach to Content Moderation of Internet Memes," February. https://doi.org/10.1109/iceic57457.2023.10049865.

[8] "Hateful Memes." n.d. Ai.meta.com. Accessed April 29, 2024. https://ai.meta.com/tools/hatefulmemes/.

[9] Jain, Raghav, Krishanu Maity, Prince Jha, and Sriparna Saha. 2023. "Generative Models vs Discriminative Models: Which Performs Better in Detecting Cyberbullying in Memes?," June. https://doi.org/10.1109/ijcnn54540.2023.10191363.

[10] Constantin, Mihai Gabriel, Dan-Stefan Pârvu, Cristian Stanciu, Denisa Ionascu, and Bogdan Ionescu. 2021. "Hateful Meme Detection with Multimodal Deep Neural Networks." IEEE Xplore. July 1, 2021. https://doi.org/10.1109/ISSCS52333.2021.9497374.

[11] Paul, Chayan. 2023. "Hate Speech in Social Networks and Detection Using Machine Learning Based Approaches," February. https://doi.org/10.1109/isacc56298.2023.10084222.

[12] Burbi, Giovanni, Alberto Baldrati, Lorenzo Agnolucci, Marco Bertini, and Del Bimbo. 2023. "Mapping Memes to Words for Multimodal Hateful Meme Classification." ArXiv (Cornell University), October. https://doi.org/10.48550/arxiv.2310.08368.

[13] Kumar, Gokul Karthik, and Karthik Nandakumar. 2022. "Hate-CLIPper: Multimodal Hateful Meme Classification Based on Cross-Modal Interaction of CLIP Features." ArXiv (Cornell University), October. https://doi.org/10.48550/arxiv.2210.05916.

[14] OpenAI. 2022. "CLIP." GitHub. December 10, 2022. https://github.com/openai/CLIP.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:31 UTC from IEEE Xplore. Restrictions apply.

Comparative Evaluation of Satellite- Based Merra-2 against Ground-based Data for $PM_{2.5}$ and TC Concentrations in Ghaziabad, India

Rishika Sinha, *Preha* Madan, *Rahul* Singh, and *Lovleen* Gupta*

Department of Environmental Engineering, Delhi Technological University, Delhi, India

* *Corresponding author: lgupta@dce.ac.in*

Abstract. This study compares satellite-based observations of $PM_{2.5}$ and Total Carbon (TC) concentrations derived from NASA's MERRA-2 reanalysis dataset for the Ghaziabad region with ground-based observations at two sites (Site A and Site B) from June 2018 to May 2019. Results reveal consistent underestimation by MERRA-2, with daily mean biases of -85.34 (Site A) and -111.31 (Site B) for $PM_{2.5}$, and -54.77 (Site A) to -59.08 (Site B) for TC, alongside monthly mean biases of -81.30 to -103.74 for $PM_{2.5}$ and -54.77 to -59.08 for TC. The absolute error indicates a 49.01% to 53.85% underestimation for both $PM_{2.5}$ and TC. Daily FAC2 values show around 39% agreement for PM2.5, reducing to 0.27 (Site A) and 0.3 (Site B) monthly, indicating reduced agreement over time. For TC, daily FAC2 is notably low at 0.04 (Site A) and 0.06 (Site B), with no monthly agreement within a factor of 2 of ground-level data. Strong correlations (R^2=0.68-0.84 for $PM_{2.5}$; R^2=0.94-0.96 for TC) between biases and ground-level data are observed, indicating proportional relationships. However, discrepancies increase with higher $PM_{2.5}$ mass concentration, highlighting MERRA-2's limitations during elevated pollution periods. Significant monthly variations are observed in GLC-$PM_{2.5}$ concentrations at Site A ($F= 61.42$, $p= 1.68\times10^{-11}$) and Site B ($F=25.15$, $p=2.96\times10^{-06}$), and For GLC-TC concentrations at Site A ($F=103.85$, $p=4.10\times10^{-16}$) and Site B ($F=55.70$, $p=7.54\times10^{-11}$). Both GLC and MERRA-2 $PM_{2.5}$ and TC concentrations follow a consistent monthly pattern, with higher levels during post-monsoon and winter seasons and lower levels during monsoon and pre-monsoon periods. This study underscores MERRA-2's limitations in estimating $PM_{2.5}$ and TC concentrations compared to ground-based observations and emphasises the necessity for further refinement and validation of the MERRA-2 model to enhance accuracy across different spatial and temporal scales.

1 Introduction

The Indo-Gangetic Plain (IGP) which is situated in Southeast Asia is identified as a prominent global aerosol hotspot. The IGP area is characterised by a remarkable diversity in the composition of aerosol and is marked by significant spatial, temporal, and vertical

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

fluctuations [1,2]. These changes are primarily linked to be due to several anthropogenic activities such as waste and biomass combustion, emissions from power plants that are coal-fired, and from vehicular exhaust. In addition to local emissions, the pollutants' transportation over long distances has an important role in determining the variability of aerosol composition and also in impacting rainfall patterns across India [3].

Ghaziabad, situated in North India is located within the Indo-Gangetic Plain and borders the western side with New Delhi. The city is a highly industrialised urban centre [4] and encompasses approximately 12 industrial zones, as reported by the Ministry of Micro Small and Medium Enterprises in 2012, indicating a notable concentration of air pollution sources. The concentration of industrial activity signifies a notable presence of air pollution sources. The World Air Quality Report 2019 composed by IQAir states Ghaziabad peaks the 30 most contaminated cities throughout the world. The city's proximity to Delhi, combined with prevailing wind patterns, positions Ghaziabad as a key area of study. Considering its downwind location, Ghaziabad is anticipated to experience elevated PM concentrations [5]. Despite Ghaziabad facing distinct air quality challenges within the broader NCR region, it has received comparatively less attention and resources than Delhi. Furthermore, the availability of ground-level and satellite data for the Ghaziabad region offers a valuable opportunity to enhance our understanding of local dynamics contributing to air quality degradation.

The rising impact of carbonaceous aerosols on the climate of a region, visibility, air quality, water resources and even public health has spurred widespread scientific interest [6–9]. These tiny airborne particles, composed primarily of organic and elemental carbon (OC and EC), play a complex role in Earth's radiative balance. Traditionally, OC was believed to cool the planet by reflecting sunlight back into space, while EC was seen as a warming agent due to its strong absorption of solar radiation [10]. However, the true optical impact of EC on global climate change remains a subject of active debate [11], a likely consequence of the diverse character of carbonaceous components.

Aerosols are formed either directly as primary gases emitted from the combustion process or indirectly through the chemical oxidation of gaseous components, converting them into particulate matter (secondary aerosols). Elemental carbon (EC), a product of incomplete combustion, emerges from diverse sources primarily fueled by carbon. These include residential wood burning, transportation-related fossil fuel emissions, and coal-fired power plant discharges. Organic carbon (OC) can enter the atmosphere directly as particulate matter and transform from volatile organic compounds (VOCs) present in the air through secondary processes. The ambient concentrations of OC and EC, serve as essential parameters for assessment of the direct impacts as well as the indirect impacts that aerosols have on the regional radiative forcing [12]. Elemental carbon encompasses all thermally transformed carbon materials arising from fossil fuel combustion and burning of biomass [13].

The exploration of carbonaceous aerosols and $PM_{2.5}$ is crucial due to its diverse implications. Firstly, these particles pose a significant threat to the health of the public, along with adverse effects on respiratory and cardiovascular systems upon inhalation [5]. Monitoring and understanding their characteristics are vital for assessing and mitigating these health risks. Furthermore, a significant fraction of fine particle matter is influenced by carbonaceous aerosols which comprise both organic and elemental carbon. This affects the overall air quality. Ground-based observations offer a way to monitor $PM_{2.5}$ and related carbon fractions, which makes it possible to analyse changes, trends, and effects of these pollutants in any given area.

MERRA-2 i.e. The Modern-Era Retrospective Analysis for Research and Applications Version 2, developed by the Global Modeling Assimilation Office (GMAO) [14], is a detailed global dataset that accurately records surface $PM_{2.5}$ mass concentration. This dataset includes key components of $PM_{2.5}$. In the context of our research study, we utilised the MERRA-2 dataset to extract data pertaining to specific variables, namely BCSMASS (Black Carbon Surface Mass Concentration), DUSMASS25 (Dust Mass Surface Mass Concentration - PM2.5), OCSMASS (Organic Carbon Surface Mass Concentration), SO4 MASS (SO4 Surface Mass Concentration), and SSSMASS25 (Sea Salt Surface Mass Concentration - PM2.5). These variables are essential for calculating $PM_{2.5}$ concentrations, forming a crucial basis for our study.

In our investigation into the air quality of Ghaziabad, our research is centred on conducting a comprehensive comparative analysis between ground-level observations and MERRA-2 reanalysis data. This includes a detailed examination of ground-level concentrations of TC and $PM_{2.5}$ at two sites in Ghaziabad–Site A and B, and their comparison with the satellite-observed TC and $PM_{2.5}$ concentrations which have been derived from MERRA-2. We are evaluating biases and performance statistics associated with MERRA-2 TC and $PM_{2.5}$ concentrations. This objective holds significant importance in assessing the accuracy of satellite data in capturing the dynamics of carbonaceous aerosols (CA) and providing a detailed insight into the distribution of aerosols and their behaviour within the area under study.

2 Methodology

2.1 Description of the Study Site

Keeping in mind the financial constraints and the logistical obstacles regarding the simultaneous speciation of Particulate Matter across numerous locations within a city such as Ghaziabad, this research paper concentrates on collecting samples at two specific sites (Site A and Site B) within the city. The selection of these two sites (Figure 1) is based on the prevailing wind direction as observed all along the year, as outlined in preceding studies[5,15,16].

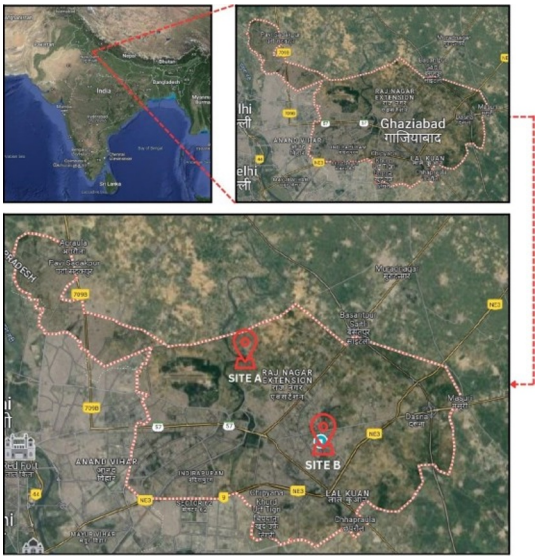


Fig. 1. Map of the Study Site with Locations of Site A and Site B in Ghaziabad

2.2 Sampling Period

The sampling of aerosol was done from June 2018 to May 2019. The $PM_{2.5}$, the organic carbon fractions—OC1, OC2, OC3, and OC4, the elemental carbon fractions—EC1, EC2, and EC3, and pyrolysed carbon—OP were collected at two sites—Site A and Site B; boht located in Ghaziabad. Twenty-four-hour integrated (9:00 AM to 9:00 AM) samples were collected on every third day [15].

A comprehensive understanding of the aerosol composition at both Site A and Site B is done by this sampling strategy which incorporates organic, elemental carbon along with pyrolyzed carbon. This method considers both temporal and spatial approaches, ensuring the accuracy of data collected. It allows for a detailed study of how aerosol particles change over the specified time frame and geographical locations. The observations of the OC as well as EC fractions, were utilised to further study the concentrations of Total Carbon.

2.3 Estimation of TC

Through the sampling, we obtained the concentrations of the organic carbon components(OC1, OC2, OC3, OC4), elemental carbon fractions (EC1, EC2, EC3), and pyrolysed carbon (OP). The Organic Carbon (OC) fraction is defined as the summation of individual components, i.e. the sum of the organic carbon fractions, and OP as shown in Equation 1. It's significant to highlight that the OC fraction is corrected by removing the overlapping contribution of OP from the Elemental Carbon (EC) fraction. Thus to calculate OC, the following equation has been used:

$$OC = OC\,1 + OC\,2 + OC\,3 + OC\,4 + OP \qquad (1)$$

The Elemental Carbon (EC) fraction is precisely characterised by the combination of its individual components, identified as EC1, EC2, and EC3. This process includes subtractingOP to avoid any duplication in accounting for carbonaceous components.

For the calculation of EC, equation thus used is:

$$EC = EC1 + EC2 + EC3 - OP \quad (2)$$

The calculation of Total Carbon (TC) is undertaken by aggregating the individual components of both the OC fractions and EC fractions. The determination of TC is done as the cumulative sum of OC1, OC2, OC3, OC4, EC1, EC2, and EC3 i.e. as the sum of OC and EC (Equations 3 and 4). This approach to estimating TC ensures a comprehensive representation of the TC contents, encompassing both organic and elemental carbonaceous elements. The thorough breakdown of OC and EC fractions, followed by their combination to derive TC, is fundamental for a nuanced understanding of the overall carbonaceous composition. These refined calculations contribute to the reliability and accuracy of our analytical framework, enabling a comprehensive assessment of carbonaceous aerosol fractions in our study area.

$$TC = OC1 + OC2 + OC3 + OC4 + EC1 + EC2 + EC3 \quad (3)$$

$$TC = OC + EC \quad (4)$$

2.4 MERRA-2 Reanalysis Data

The Modern-Era Retrospective Analysis for Research and Applications Version 2 or MERRA-2 which was introduced in 1980 serves as an upgrade to its predecessor, MERRA. It has been developed by the GMAO i.e. Global Modelling and Assimilation Office of NASA. MERRA-2 incorporates sophisticated assimilation capabilities for contemporary observational datasets. The reanalysis also integrates the ozone profile observations of NASA since late 2004. Both MERRA and MERRA-2 have a similar spatial resolution in the latitudinal direction of about 50 km. Also, MERRA-2 relies on aerosol optical depth (AOD) data at 550 nm obtained from NASA's Aqua and Terra satellites' Moderate Resolution Imaging Spectroradiometer (MODIS) as its primary input [14,17].

The various datasets in MERRA-2 can be accessed through MDISC which is managed by GES DISC (Goddard Earth Sciences Data and Information Services Center). For this study, the MERRA-2 hourly aerosol products (M2T1NXAER or tavg1_2d_aer_Nx) were obtained from NASA's GSFC portal at https://disc.gsfc.nasa.gov/. This dataset in MERRA-2 provides hourly averaged 2-dimensional information. It includes assimilated aerosol data, such as the column mass density and surface concentrations of various components of aerosol.

In this study, we extracted the variables namely DUSMASS25 (Dust Mass Surface Mass Concentration - PM2.5), OCSMASS (Organic Carbon Surface Mass Concentration), BCSMASS (Black Carbon Surface Mass Concentration), SSSMASS25 (Sea Salt Surface Mass Concentration – PM2.5) and SO4SMASS (SO4 Surface Mass Concentration) from the M2T1NXAER dataset for the location Ghaziabad (77.238˚W, 28.528˚S, 77.612˚E, 28.757˚N). In our study, we used these variables to calculate the MERRA-2 TC and PM2.5 concentrations. In order to calculate the Total carbon (TC) concentrations as observed by MERRA-2, the OC and BC surface mass concentrations obtained from MERRA-2 were used:

$$TC_{MERRA-2} = OC + BC \quad (5)$$

The following equation was used to find out the MERRA-2 $PM_{2.5}$ values [18]:

$$PM2.5_{MERRA-2} = 1.375 \times [SO4] + 1.8 \times [OC] + [Dust_{2.5}] + [SS_{2.5}] + [BC] \quad (6)$$

Here, SO4, BC, OC, Dust2.5 and SS2.5 stand for sulphate, organic carbon, black carbon, dust, and sea-salt particulate matter with a diameter of less than 2.5μm from the GOCART aerosol module. Equation (6) has been used for calculating $PM_{2.5}$ concentrations and it involves the addition of the contributions from dust and sea-salt which are in size bins smaller than 2.5 μm, as well as hydrophilic and hydrophobic OC and BC, and sulphate. The sulphate concentration in MERRA-2 datasets is in the form of neutralised ammonium sulphate [$(NH_4)_2SO_4$]. Thus, to get the 'true' sulphate concentration, we use a factor of 1.375 as suggested by previous studies [17,19]. We can convert organic carbon (OC) to particulate organic matter (POM) using a factor ranging from 1.2-2.6 [17,20,21]. This considers the influence of other elements associated with organic matter. For our study, we use a factor of 1.8 and multiply OC by this conversion factor. This conversion factor is used to account for the elements other than carbon present in aerosols and the specific value depends on the sampling location which could be an urban area, a rural area, or a high-altitude region [3].

3 Results and Discussion

3.1 Comparison of Ground-Level Concentrations with Satellite-ObservedConcentrations

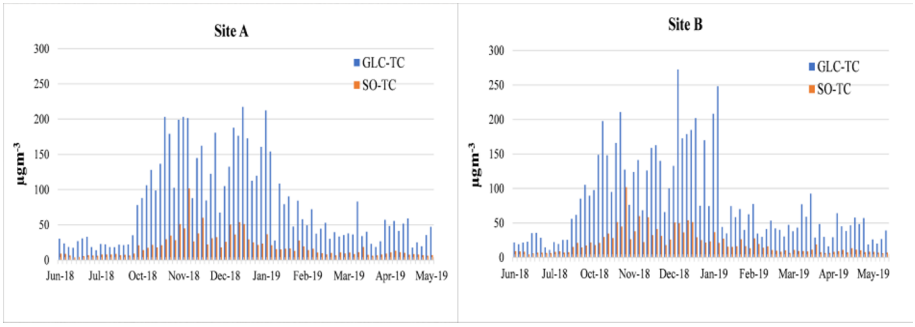
The ground-level concentrations i.e. GLC of Total Carbon and $PM_{2.5}$ denoted as GLC-TC, GLC-$PM_{2.5}$, obtained from the two sampling sites were compiled and compared with the concentrations of TC and $PM_{2.5}$ derived from MERRA-2, denoted as SO-TC and SO-$PM_{2.5}$ (where SO represents Satellite Observed). This comparative analysis was conducted for the period June 2018 to May 2019. Fig 2. presents these comparisons graphically.

A significant difference between the ground-level concentrations of both sampling sites and MERRA-2 reanalysis data is observed (i.e. between GLC-TC and SO-TC and between GLC and SO $PM_{2.5}$). Fig. 2 (i) shows that the SO concentrations of TC at both Site A and Site B significantly and also constantly underestimate the concentrations of GLC-TC and GLC-TC. We observe a similar pattern of underestimation by SO-$PM_{2.5}$ concentrations as well (Fig. 2 (ii)). Similar underestimations have been reported by previous studies as well. In a study carried out over the Yangtze River Basin, SO-$PM_{2.5}$ was underestimated compared to ground-level-based measurements, which was partly attributed to biases in MERRA-2 Aerosol Optical Depth (AOD) assimilation [22]. Several studies have reported similar results for the components of TC i.e. Black Carbon and Organic Carbon. A study across four sites in the Indo-Gangetic Plain highlights discrepancies in MERRA-2 reanalysis data. BC mass concentrations are underestimated in the North-Western region and overestimated in the central area. Additionally, MERRA-2 consistently underestimates OC across all these sites, attributed to emission-related ambiguities in the North-West and the ageing process of carbonaceous aerosols in the central area. The study also suggests that the underestimation of organics may be linked to secondary organic aerosol formation from regionally transported gas-phase precursors [23].

In our study, while the graphs illustrating the comparisons between GLC-TC and SO-TC, as well as GLC-$PM_{2.5}$ and SO-$PM_{2.5}$, indicate a general tendency of the MERRA-2 model to underestimate these concentrations, there are slight exceptions. Notably, in June

2018 (Site A and Site B) and July 2018 (Site A) and November 2018 (Site A and Site B), there are instances of slight overestimation of $PM_{2.5}$ concentrations by satellite observed data as seen in Fig. 2 (ii). The overestimation in these months could be due to a number of factors. In a study by [24], BC concentrations were found to be overestimated by MERRA-2, particularly in central and eastern China, attributing it to uncertainties in emission types and strengths, while also noting significant underestimation of organic carbon (OC) during winter months.

(i)



(ii)

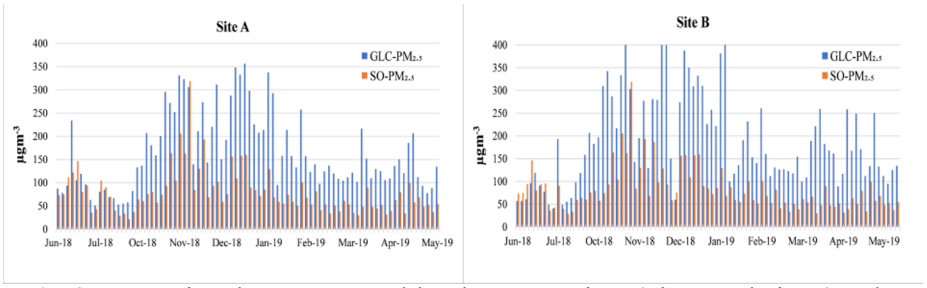


Fig. 2. Comparison between ground-level concentrations (Site A and Site B) and MERRA-2reanalysis concentration (i) TC (ii) $PM_{2.5}$

3.2 The Correlation between GLC and Bias

Table 1. Mean bias as observed for different size ranges of $PM_{2.5}$ and TC at Site A and Site B.

PM 2.5			TC		
	Mean Bias			Mean Bias	
Range	Site A	Site B	Range	Site A	Site B
<100	19.72	2.62	<20	11.06	8.95
100-150	65.47	63.58	20-35	17.99	17.3
>150	133.01	169.35	>35	79.27	74.76

In this study, differences in mean bias across various concentration ranges of ground-level concentrations (GLC) of $PM_{2.5}$ and total carbon were revealed through the analysis of MERRA-2 data at both Site A and Site B. At lower $PM_{2.5}$ concentrations, a negative bias was observed, indicating that the satellite data overestimated GLC compared to ground observations. This trend reversed at higher $PM_{2.5}$ concentrations. However, TC exhibited positive bias across all concentration ranges indicating a higher GLC value.

Further investigation identified a proportional relationship between the bias and GLC-

PM2.5 for both Site A ($R^2 = 0.68$) and Site B ($R^2 = 0.84$). This relationship was also evident for TC at both sites, with even stronger correlations (R^2 SiteA= 0.96, R^2 SiteB = 0.94).

Table 1 details the mean bias for different size ranges of $PM_{2.5}$, and TC at each site. The trend of escalating disparity with rising $PM_{2.5}$ mass concentration was in alignment with findings from prior research [25], thereby emphasising the shortcomings of MERRA-2 in simulating elevated $PM_{2.5}$ levels during periods of heightened pollution. Comparable patterns were noted for TC.

In Figure 3, the relationship between GLC and bias with MERRA-2 simulated concentrations for $PM_{2.5}$ and TC is depicted for both sites. The strong coefficient of determination (R2) between bias and ground-level concentration (GLC) of $PM_{2.5}$ in the plots (Fig. 3) is indicative of MERRA-2's notable tendency to consistently underestimate observed values. This trend extends to Total Carbon (TC), indicating MERRA-2's limitation in accurately simulating concentrations compared to ground-based observations.

These findings underscore the necessity for caution when relying on MERRA-2 for accurate estimations of $PM_{2.5}$, and TC concentrations, particularly in light of the consistent underestimation trend identified.

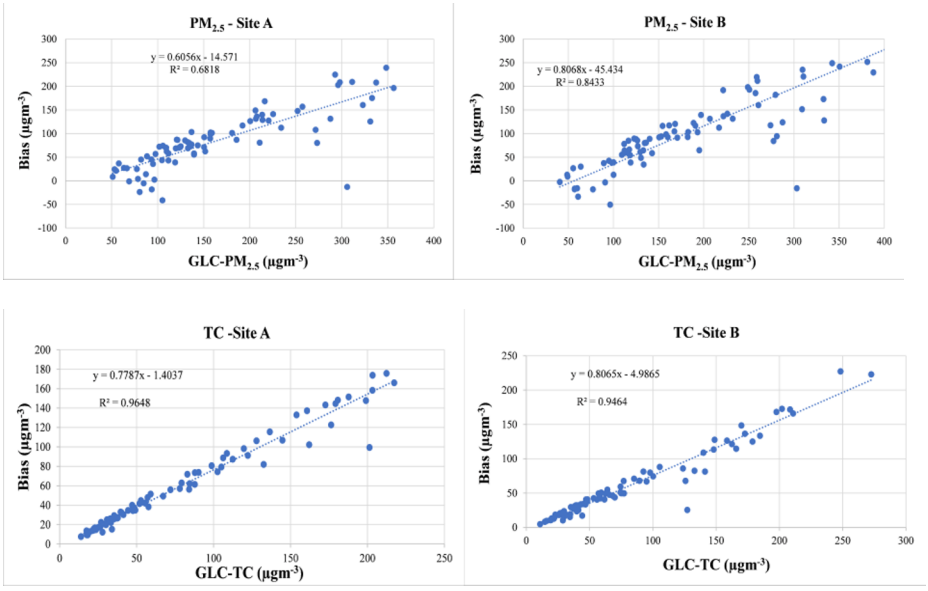


Fig. 3. The correlation between the Ground Level Concentration (GLC) and the bias, assessed for $PM_{2.5}$, and Total Carbon (TC) for Site A and Site B where, Bias = GLC – MERRA-2 Simulated Concentration

3.3 Study of Performance statistics of MERRA-2 TC and $PM_{2.5}$ concentrationcompared to Ground-level TC and $PM_{2.5}$ concentrations at Site A and Site B

Table 2. Performance Metrics of MERRA-2 TC and PM $_{2.5}$ Concentration compared to GLC-TC and $PM_{2.5}$ at Site A and Site B

Site	Statistical Metrics	Daily TC	Monthly TC	Daily $PM_{2.5}$	Monthly $PM_{2.5}$
Site A	Mean Bias	-59.08	-55.37	-85.34	-81.30
	Mean Fraction	0.26	0.25	0.52	0.51
	Correlation coefficient	0.83	0.92	0.69	0.77
	FAC2	0.04	0	0.39	0.27
	RMSE	75.69	67.35	104.63	92.95
	MAE	59.08	55.37	87.86	81.30
	Absolute Error (%)	73.47	74.40	50.69	49.01
Site B	Mean Bias	-58.62	-54.77	-111.31	-103.74
	Mean Fraction	0.28	0.26	0.52	0.49
	Correlation coefficient	0.71	0.89	0.48	0.65
	FAC2	0.06	0	0.38	0.3
	RMSE	77.26	65.07	153.17	121.23
	MAE	58.62	54.77	115.38	105.94
	Absolute Error (%)	71.13	73.47	53.85	53.84

Various statistical metrics were computed for both daily and monthly average TC and $PM_{2.5}$ concentrations to assess the performance of MERRA-2 TC and $PM_{2.5}$ concentrations by comparing it with the Ground-level TC and $PM_{2.5}$ concentrations observed at Site A and Site B (Table 2). Optimal performance in assessing TC and $PM_{2.5}$ concentrations is characterised by lower values in statistical metrics such as mean absolute error percentage,
MAE, RMSE, and mean bias (B), while a mean fraction (F) closer to one indicates accuracy[25]. The following Statistical Metrics (Equations. 7-11; where SO is for the Satellite Observed i.e. MERRA-2 concentrations, and GLC for Ground-level concentrations) were evaluated to understand the accuracy of SO i.e. MERRA-2 TC and $PM_{2.5}$ concentrations compared to ground-level concentrations:

$$\overline{B} = \overline{SO - GLC} \tag{7}$$

$$\overline{F} = \left(\frac{\overline{SO}}{GLC}\right) \tag{8}$$

$$RMSE = \sqrt{\sum_{i=0}^{n}\left(\frac{GLC_i - SO_1}{n}\right)^2} \tag{9}$$

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|GLC_i - SO_i\right| \tag{10}$$

$$AbsolutePercentageError = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(\left|\frac{GLC_i - SO_i}{GLC_i}\right|\right) \times 100 \tag{11}$$

The performance statistics reveal notable differences in the accuracy of MERRA-2 TC and $PM_{2.5}$ concentrations at Site A and Site B. Both sites have a considerable mean bias indicating a systematic underestimation of TC and $PM_{2.5}$ levels by MERRA-2. For Site A, the monthly mean bias of -81.30 for $PM_{2.5}$ and -55.37 for TC indicates underestimation, with monthly mean fractions of 0.51 ($PM_{2.5}$) and 0.25 (TC) and monthly correlation coefficients of 0.77 ($PM_{2.5}$) and 0.92 (TC) indicating improved performance on longer time scales. Site B shows a larger daily mean bias for both $PM_{2.5}$ (-111.31) and TC (-58.62), lower daily mean fractions (PM2.5: 0.49, TC: 0.26), and moderate daily correlation coefficients ($PM_{2.5}$: 0.48, TC: 0.71). Monthly metrics exhibit a similar trend with lower mean fractions and higher correlation coefficients.

Site A shows RMSE values of 104.63 ($PM_{2.5}$) and 75.69 (TC) daily, and 92.95 ($PM_{2.5}$) and 67.35 (TC) monthly, indicating moderate discrepancies. Site B exhibits higher RMSE values of 153.17 ($PM_{2.5}$) and 77.26 (TC) daily, and 121.23 ($PM_{2.5}$) and 65.07 (TC) monthly, suggesting larger deviations between model predictions and actual data. MAE values show similar trends with Site A having lower values ($PM_{2.5}$: 87.86, TC daily: 59.08; $PM_{2.5}$,: 81.30, TC monthly: 55.37) compared to Site B ($PM_{2.5}$: 115.38, TC daily: 58.62; $PM_{2.5}$: 105.94, TC monthly: 54.77), indicating slightly better model accuracy at Site A. The Absolute Error percentage further highlights these differences, with Site A showing lower error percentages (50.69% $PM_{2.5}$, 73.47% TC daily; 49.01% $PM_{2.5}$, 74.40% TC monthly) than Site B (53.85% $PM_{2.5}$, 71.13% TC daily; 53.84% $PM_{2.5}$, 73.47% TC monthly),
suggesting a higher level of accuracy in predicting PM2.5 concentrations at Site A. These metrics collectively demonstrate the varying levels of accuracy and deviations between MERRA-2 model predictions and ground-level observations, emphasising the need for further refinement and validation, particularly at Site B, to enhance model performance and reliability.

For evaluating the accuracy of concentrations observed by MERRA-2, the FAC2 metric was also used in this study. The Factor of Two (FAC2) metric is utilised to evaluate the performance of air quality models and is carried out by estimating the proportion of simulated data which falls within a factor of 2 of the observed data. FAC2 examines the ratio of concentrations that are simulated (Cs i.e. SO-$PM_{2.5}$) to observed concentrations (Co i.e. GLC-$PM_{2.5}$), ensuring that the value lies within the range of 0.5 to 2.0. This approach helps to mitigate the impact of extreme values and the errors present in the data [26]. FAC2 values closer to 0.50 or higher reflect better model performance [25]. As observed, for Site A, the monthly FAC2 for Site A reduces to 0.27 ($PM_{2.5}$), which suggests a lower proportion of agreement over longer time scales for $PM_{2.5}$. For TC at this site, the monthly FAC2 value of 0 suggests no agreement within a factor of 2 of GLC at Site A. However, the daily FAC2 values at this site are 0.39 for $PM_{2.5}$ and 0.04 for TC, which indicates that approximately 39% of the simulated data of $PM_{2.5}$ and only 4% of the TC data fall within a factor of 2 of the observed data on a daily basis. Similarly, for Site B, the daily FAC2 values are 0.38

($PM_{2.5}$) and 0.06 (TC), with the monthly FAC2 decreasing to 0.3 ($PM_{2.5}$) and 0 (TC). These values denote a comparable trend of higher agreement on a daily basis as compared to a monthly basis for both $PM_{2.5}$ and TC concentrations, highlighting areas where the model may require improvement for better replicating observed concentrations, particularly over longer time scales.

The evaluation of MERRA-2 $PM_{2.5}$ concentrations at Site A and Site B underscores the importance of considering multiple statistical metrics to comprehensively assess model performance, especially concerning real-world ground-level $PM_{2.5}$ concentrations. Further refinement and validation of the model may be necessary to enhance its accuracy across different spatial and temporal scales.

3.4 Analysis of Monthly Variation for TC and PM2.5 of Ground-level Concentration and MERRA-2 for SITE A and SITE B.

Ground-level concentrations of TC and $PM_{2.5}$ provide insights into air quality trends in Ghaziabad, an industrial area. Based on the GLC- $PM_{2.5}$ concentrations, throughout the study period, more than 90% of days at both Site A (95.12%) and Site B (91.76%) surpassed the standard prescribed by National Ambient Air Quality (NAAQ) of 60 μg/m³ for 24-hour $PM_{2.5}$ concentration, as designated for Industrial, Residential, Rural, and Other Areas. Additionally, 98.83% of days exceeded the NAAQ standard for 24-hour $PM_{2.5}$ based on the $PM_{2.5}$ concentrations derived from MERRA-2 data. This consistent trend signifies a prevalent issue of poor air quality [27].

Figure 4 highlights the consistent underestimation of pollutant concentrations in satellite-observed (SO) data, emphasizing the critical importance of ground-level monitoring for accurately assessing air quality. Additionally, it illustrates the seasonal variations in TC and $PM_{2.5}$ concentrations for both sites, comparing ground-level and satellite-level data across the four seasons.

Table 3 and Table 4 present the average seasonal concentrations for both ground-level (GLC) and satellite-observed (SO) concentrations, showcasing notable variations across different seasons: monsoon, post-monsoon, winter, and pre-monsoon. Notably, both sites exhibited higher $PM_{2.5}$ concentrations throughout post-monsoon and winter months (averaging 235.08μg/m³ and 199.28μg/m³ for Site A and averaging 267.59μg/m³ and 217.33μg/m³ for Site B), with lower levels observed during monsoon and pre-monsoon periods (averaging 86.90μg/m³ and 127.37μg/m³ for Site A and averaging 77.57μg/m³ and 157.32μg/m³ for Site B) for ground-level concentration (GLC), shown in Table 3. A similar trend is shown by satellite observed (SO) concentration with higher concentrations for post-monsoon and winter months (averaging 116.90μg/m³ and 76.93μg/m³ for Site A and averaging 119.16μg/m³ and 78.64μg/m³ for Site B), and lower levels observed during monsoon and pre-monsoon periods (averaging 71μg/m³ and 52.04μg/m³ for Site A and averaging 71.55μg/m³ and 53.25μg/m³ for Site B), shown in Table 4.

A parallel trend was observed for total carbon (TC) concentrations during post-monsoon and winter months (averaging 138.08μg/m³ and 102.67μg/m³ for Site A and averaging 129.92μg/m³ and 91.25μg/m³ for Site B), with lower levels observed during monsoon and pre-monsoon periods (averaging 23.29μg/m³ and 39.57μg/m³ for Site A and averaging 23.32μg/m³ and 42.71μg/m³ for Site B) for ground-level concentration (GLC), shown in Table 3. A similar trend is shown by satellite observed (SO) with higher concentrations for post-monsoon and winter months (averaging 34.70μg/m³ and 21.29μg/m³ for Site A and averaging 35.64μg/m³ and 21.84μg/m³ for Site B), and lower levels observed during monsoon and pre-monsoon periods (averaging 7.12μg/m³ and 9.26μg/m³ for Site A and averaging 7.23μg/m³ and 9.13μg/m³ for Site B), shown in Table 4.

To assess the statistical significance of these trends, we conducted ANOVA tests, summarized in Table 5 and Table 6, providing values for Site A ($F= 61.42$, $p= 1.68\times10^{-11}$ for GLC and $F=16.95$, $p=9.26\times10^{-05}$ for SO) and Site B ($F=25.15$, $p=2.96\times10^{-06}$ for GLC

and F=20.20, p=2.24×10^{-05} for SO) for $PM_{2.5}$ concentrations and for Site A (F=103.85, p=4.10×10^{-16} for GLC and F=56.20, p=7.77×10^{-11} for SO) and Site B (F=55.70, p=7.54×10^{-11} for GLC and F=57.44, p=4.45×10^{-11} for SO) for TC. These statistical analyses confirm the observed monthly variations, highlighting their relevance in air quality management strategies.

Table 3. Average Ground level concentration of TC and PM 2.5

AVERAGE GROUND LEVEL CONCENTRATION				
SEASON	**GLC-TC (µg/mt)**		**GLC-PM 2.5 (µg/mt)**	
	SITE A	**SITE B**	**SITE A**	**SITE B**
MONSOON	23.29	23.32	86.90	77.57
POST-MONSOON	138.08	129.92	235.08	267.59
WINTER	102.67	91.25	199.28	217.33
PRE-MONSOON	39.57	42.71	127.37	157.32

Table 4. Average Satellite observed concentration of TC and PM 2.5

AVERAGE SATELLITE OBSERVED CONCENTRATION				
SEASON	**SO-TC (µg/mt)**		**SO-PM 2.5 (µg/mt)**	
	SITE A	**SITE B**	**SITE A**	**SITE B**
MONSOON	7.12	7.23	71	71.55
POST-MONSOON	34.70	35.64	116.90	119.16
WINTER	21.29	21.84	76.93	78.64
PRE-MONSOON	9.26	9.13	52.04	53.25

Table 5. Analysis of variance (ANOVA) Test of Site A for GLC and SO

VALUES	**SITE A**			
	GLC		**SO**	
	$PM_{2.5}$	**TC**	**$PM_{2.5}$**	**TC**
F-value	61.42	103.85	16.95	56.20
p-value	1.68×10^{-11}	4.10×10^{-16}	9.26×10^{-05}	7.77×10^{-11}

Table 6. Analysis of variance (ANOVA) Test of Site B for GLC and SO

VALUES	**SITE B**			
	GLC		**SO**	
	$PM_{2.5}$	**TC**	**$PM_{2.5}$**	**TC**
F-value	25.15	55.70	20.20	57.44
p-value	2.96×10^{-06}	7.54×10^{-11}	2.24×10^{-05}	4.45×10^{-11}

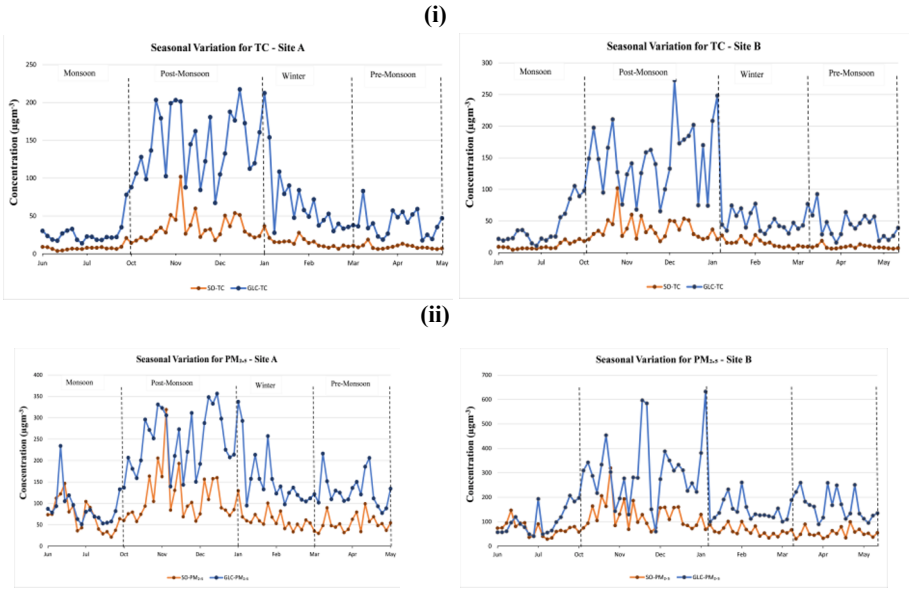


Fig 4. Seasonal Variation of (i) TC and (ii) $PM_{2.5}$ Concentrations

4 Conclusion

Conducted in the Ghaziabad region of India, this study, aimed to evaluate the accuracy of MERRA-2, a reanalysis data product from NASA/GMAO, in measuring concentrations of Total Carbon (TC) and $PM_{2.5}$ compared to ground-based observations of TC and $PM_{2.5}$. The comparative analysis demonstrated consistent trends of underestimation for both TC and $PM_{2.5}$ concentrations by MERRA-2 reanalysis data, with occasional overestimations of $PM_{2.5}$ noted in specific months. Also, a proportional relationship between bias and GLC indicated systematic underestimation by MERRA-2 across all components. On analysis limitations in accurately simulating elevated $PM_{2.5}$ levels during periods of heightened pollution were observed. Performance metrics like RMSE, MAE, and absolute error percentages showcased the tendency of MERRA-2 to underestimate the TC and the $PM_{2.5}$ concentrations compared to GLC. MERRA-2 underestimated $PM_{2.5}$ concentrations by approximately 50-60% at both Site A and Site B. The Factor of Two (FAC2) values revealed that approximately 39% of the simulated $PM_{2.5}$ data fell within a factor of 2 of the observed data on a daily basis at Site A, and 38% at Site B, which showed moderate agreement between the SO and GLC $PM_{2.5}$ data. At Site A, the monthly mean bias for $PM_{2.5}$ was -81.30, with a correlation coefficient of 0.77, which validated the systematic underestimation by MERRA-2. Similarly, at Site B, the monthly mean bias was -103.74, with a correlation coefficient of 0.65, further showing the tendency of MERRA-2 to underestimate $PM_{2.5}$ concentrations. Likewise, for the concentrations of TC, MERRA-2 underestimated the GLC-TC concentrations at both sites by approximately 70-75% with significantly low FAC2 values for daily TC concentrations (Site A: 0.04; Site B:0.06) and no monthly agreement within a factor of 2 of ground-level data at both the sites.

The assessment of monthly variations in ground-level concentrations underscores a discernible seasonal trend, with elevated pollution levels observed during the post-monsoonand winter months compared to the relatively cleaner air observed during the seasons of monsoon and pre-monsoons. ANOVA results confirm statistically significant monthly variations in TC and $PM_{2.5}$ concentrations, which show the importance of considering these

dynamics in air quality management. Similar seasonal trend was observed in the TC and$PM_{2.5}$ as observed by MERRA-2.

Analysis of the GLC-$PM_{2.5}$ revealed that for 24-hour $PM_{2.5}$ concentration, over 90% of days within the study period surpassed the National Ambient Air Quality (NAAQ) prescribed 60 $\mu g/m^3$ standard at both the sites. Likewise, the $PM_{2.5}$ concentrations derived from MERRA-2 showed that 98.83% of days exceed this standard. This indication of poor quality of air stresses the need for concerted efforts towards improving the air quality and toalso safeguard the public-health in the area.

In conclusion, although NASA's GMAO MERRA-2 satellite-based model provides valuable insights into the atmospheric composition and dynamics of aerosols, caution must be exercised while utilising it for accurate estimations of TC and $PM_{2.5}$ concentrations, especially in regions like Ghaziabad which are characterised by complex atmospheric processes and diverse emission sources. Continued validation and refinement of the MERRA-2 model, along with efforts at comprehensive monitoring of air quality, are crucial for effective air quality management and the protection of public-health in regions experiencing significant air pollution like Ghaziabad. Our findings highlight the urgent call for comprehensive air quality management strategies for the mitigation of the adverse health effects of air pollution in Ghaziabad.

References

1. S. Dey and L. Girolamo, Geophysical Research Letters - GEOPHYS RES LETT **38**, (2011)
2. M. Kumar, K. S. Parmar, D. B. Kumar, A. Mhawish, D. M. Broday, R. K. Mall, and T. Banerjee, Atmospheric Environment **180**, 37 (2018)
3. K. Ram and M. M. Sarin, Journal of Aerosol Science **41**, 88 (2010)
4. S. Kumar and S. K. Dwivedi, Environmental Research **200**, 111754 (2021)
5. L. Gupta, R. Dev, K. Zaidi, R. Sunder Raman, G. Habib, and B. Ghosh, Environ Monit Assess **193**, 735 (2021)
6. J. Hansen, T. Bond, B. Cairns, H. Gaeggler, B. Liepert, T. Novakov, and B. Schichtel, Eos, Transactions American Geophysical Union **85**, 241 (2004)
7. M. Jacobson, Letters to Nature 695 (2001)
8. J. L. Mauderly and J. C. Chow, Inhal Toxicol **20**, 257 (2008)
9. S. Menon, J. Hansen, L. Nazarenko, and Y. Luo, Science **297**, 2250 (2002)
10. A. Ackerman, O. Toon, D. Stevens, A. Heymsfield, V. Ramanathan, and E. Welton, Science (New York, N.Y.) **288**, 1042 (2000)
11. Intergouvernemental panel on climate change, editor , *Climate Change 2007: The Physical Science Basis* (Cambridge university press, Cambridge, 2007)
12. T. Novakov, S. Menon, T. W. Kirchstetter, D. Koch, and J. E. Hansen, Journal of Geophysical Research: Atmospheres **110**, (2005)
13. C. A. Masiello, Marine Chemistry **92**, 201 (2004)
14. C. A. Randles, A. M. da Silva, V. Buchard, P. R. Colarco, A. Darmenov, R. Govindaraju, A. Smirnov, B. Holben, R. Ferrare, J. Hair, Y. Shinozuka, and C. J. Flynn, Journal of Climate **30**, 6823 (2017)
15. L. Gupta, M. Bansal, P. Nandi, G. Habib, and R. Sunder Raman, Atmospheric Environment **298**, 119614 (2023)
16. L. Gupta, S. Joshi, G. Habib, and R. Sunder Raman, Atmospheric Environment **312**, 120020 (2023)
17. V. Buchard, A. M. da Silva, C. A. Randles, P. Colarco, R. Ferrare, J. Hair, C. Hostetler, J. Tackett, and D. Winker, Atmospheric Environment **125**, 100 (2016).

18. J. Hand, S. A. Copeland, C. Mcdade, D. Day, J. Moore, A. Dillner, M. Pitchford, H. Indresand, B. Schichtel, W. Malm, and J. Watson, *Spatial and Seasonal Patterns and Temporal Variability of Haze and Its Constituents in the United States, IMPROVE Report V* (2011)
19. Z. Song, D. Fu, X. Zhang, Y. Wu, X. Xia, J. He, X. Han, R. Zhang, and H. Che, Atmospheric Environment **191**, 70 (2018)
20. W. C. Malm, B. A. Schichtel, and M. L. Pitchford, Journal of the Air & Waste Management Association **61**, 1131 (2011)
21. B. J. Turpin and H.-J. Lim, Aerosol Science and Technology **35**, 602 (2001)
22. L. He, A. Lin, X. Chen, H. Zhou, Z. Zhou, and P. He, Remote Sensing **11**, 460 (2019)
23. A. Soni, A. Mandariya, P. Rajeev, S. Izhar, G. Singh, V. Choudhary, A. Qadri, A. Gupta, A. Singh, and T. Gupta, Environmental Science: Atmospheres (2021)
24. X. Ma, P. Yan, T. Zhao, X. Jia, J. Jiao, Q. Ma, D. Wu, Z. Shu, X. Sun, and B. Habtemicheal, Remote Sensing **13**, (2021)
25. C. D. Navinya, V. Vinoj, and S. K. Pandey, Aerosol Air Qual. Res. **20**, 1329 (2020)
26. J. Chang and S. Hanna, Meteorology and Atmospheric Physics **87**, 167 (2004)
27. Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, (n.d.)
28. K. Ali, A. Panicker, G. Beig, S. Reka, and P. Acharja, Journal of Atmospheric Chemistry 73, (2015)
29. Atmospheric Environment 37, 1451 (2003)
30. J. J. Cao, F. Wu, J. C. Chow, S. C. Lee, Y. Li, S. W. Chen, Z. S. An, K. K. Fung, J. G. Watson, C. S. Zhu, and S. X. Liu, Atmospheric Chemistry and Physics 5, 3127 (2005)
31. J. J. Cao, S. C. Lee, K. F. Ho, K. Fung, J. C. Chow, and J. G. Watson, Aerosol Air Qual. Res. 6, 106 (2006)
32. J. C. Chow, J. G. Watson, Z. Lu, D. H. Lowenthal, C. A. Frazier, P. A. Solomon, R. H. Thuillier, and K. Magliano, Atmospheric Environment 30, 2079 (1996)
33. Y. M. Han, Z. W. Han, J. J. Cao, J. C. Chow, J. G. Watson, Z. S. An, S. X. Liu, and R. J. Zhang, Atmospheric Environment 42, 2405 (2008)
34. Y. Chen, G. Sheng, X. Bi, Y. Feng, B. Mai, and J. Fu, Environmental Science & Technology 39, 1861 (2005)
35. L.-Y. He, M. Hu, X.-F. Huang, B.-D. Yu, Y.-H. Zhang, and D.-Q. Liu, Atmospheric Environment 38, 6557 (2004)
36. J. C. Chow, J. G. Watson, H. Kuhns, V. Etyemezian, D. H. Lowenthal, D. Crow, S. D. Kohl, J. P. Engelbrecht, and M. C. Green, Chemosphere 54, 185 (2004)
37. J. Cao, S. Lee, J. Chow, J. Watson, K. F. Ho, R. Zhang, Z. D. Jin, Z. Shen, G. Chen, Y. Kang, S. Zou, L. Zhang, S. Qi, M. Dai, Y. Cheng, and K. Hu, Journal of Geophysical Research D112, 22 (2007)
38. Global Modeling and Assimilation Office (GMAO) (2015), MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly,Time-averaged,Single-Level,Assimilation,Aerosol Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: Sept 4, 2023, 10.5067/KLICLTZ8EM9D.

Comprehensive assessment of air quality dynamics in Delhi: a tri-phased analysis spanning pre, during, and post-lockdown periods

Satvik Popli[1], *Esha* Yadav[1], *Aastha* Singh[1], and *Geeta* Singh[1]

[1]Environmental Engineering Department, Delhi Technological University, India

Abstract: To mitigate the spread of the global COVID-19 pandemic, various regions, including India, implemented stringent measures such as nationwide lockdowns to curb the spread of the virus. The air quality dynamics during the COVID-19 lockdown period revealed diverse trends across various pollutants all over the world. This study explores the impact of these lockdowns on air quality in Delhi, one of India's megacities, across distinct phases: pre-lockdown, during lockdown, and post-lockdown periods. PM 10 was observed to show a substantial decline in concentration, particularly in April 2020, serving as a testament to the immediate effectiveness of implementing lockdown measures to control particulate matter concentrations. However, these levels returned to their previous peak in subsequent years post lockdown. PM 2.5 followed a similar pattern, experiencing an initial decline during the lockdown, but as industries and factories resumed their activities post June 2020, the levels shot up back to their original baseline, highlighting the transient nature of low PM concentrations. Other pollutants like Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), and Carbon Monoxide (CO) levels showed varying degrees of mild to steep decline during the lockdown, but they too returned to their original concentrations as few months after the lockdown was lifted. There was, contrary to expectations, an increase in ozone (O_3) concentrations after the implementation of lockdown, challenging our assumptions. Spatial mapping was done to observe the variation of concentration in different tehsils of Delhi.

1 Introduction

The extensive and rapid spread of the SARS-CoV-2 virus and the thousands of deaths that resulted from it led the World Health Organisation to declare it a pandemic on 12 March 2020[1]. Over 210 nations worldwide were affected, including India which saw upwards of 45 million positive cases and a death toll over 500,000. Due to COVID-19's impact on the majority of the global populace, there was a public health emergency of global concern, raising the alarm for medical and political organizations all over the world. COVID 19 is the first coronavirus following the Spanish Flu (1918–1919), to have had an impact this sizable on India's economy, psychology, and healthcare system[2, 3].

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

The initial cases of the COVID-19 viral disease were reported in the city of Wuhan, China, towards the end of 2019. As people traveled from China to other parts of the world, the virus rapidly spread across the globe. To protect citizens and limit the transmission of the disease through contact with infected individuals, governments worldwide implemented nationwide lockdowns and social distancing measures in their respective countries. India reported its first case of COVID-19 in Kerala, on January 30, 2020. Recognizing the potential for community spread, the Indian government imposed a nationwide lockdown on March 25, 2020, which was scheduled until April 14, 2020. This initial 21-day lockdown was followed by subsequent extensions, with Lockdown 2.0 spanning from April 15 to May 3, and Lockdown 3.0 from May 1 to May 17. After a 75-day period of complete lockdown, a phased reopening began on June 8, 2020[4]. The guidelines issued by the government during the nationwide lockdown in India mandated the suspension of all domestic and international air travel, train services, public transportation, individual vehicle movement, educational institutions, markets, construction activities, industrial and commercial operations, and other activities that required social interaction[5].

The nationwide social lockdown and the abrupt suspension of economic activities drastically altered the way citizens went about their daily lives. Although this proved to be a significant deterrent in the nation's economic growth[6], the opposite was true for the environment. There were significant improvements in the ecosystems of urban megacities, which had previously been the epicenters of noise, water and air pollution. India in particular, experienced an unprecedented improvement in air quality, showing substantial reductions in primary pollutants like fine particulate matter (42–60%), coarse particulate matter (24–62%), nitrogen dioxide (46–61%), ozone (22–56%) and Carbon Monoxide (16–46%)[7]. This sudden positive change was largely due to the cessation of operations of most large-scale factories as well as the lack of vehicular emissions contributing to the poor air quality.

Delhi, officially recognized as the National Capital Territory (NCT) of Delhi, constitutes both a city and a union territory within India, housing New Delhi, the nation's capital. Positioned at approximately 28.7°N and 77.1°E, the NCT spans an expanse of 1,484 square kilometers (573 sq mi). The 2011 census reported a population exceeding 11 million within the city limits and around 16.8 million in the NCT. Delhi experiences a dry winter humid subtropical climate intermingled with a hot semi-arid climate. To systematically monitor air quality, 40 Continuous Ambient Air Quality Monitoring Stations (CAAQMS) have been strategically established across Delhi.

This research wishes to answer the following questions:

1. How did the concentration levels of PM 2.5, PM 10, SO_2, NO_2, CO, and O_3 change during the pre-lockdown, lockdown, and post-lockdown periods in Delhi?
2. Which pollutants exhibited the most pronounced changes in concentration levels during the lockdown period?
3. What were the potential factors contributing to the observed changes in air pollution levels during the lockdown?
4. How did the spatial distribution of pollutant concentrations vary across different tehsils (administrative subdivisions) of Delhi?

By analyzing the variation in levels of air pollution during the COVID-19 lockdown, this study aims to provide valuable insights into the effectiveness of lockdown measures in reducing air pollution and inform policy decisions to mitigate air pollution in Delhi and other similarly affected cities. The findings of this study will also contribute to a better understanding of the direct and indirect correlation between human activities, air pollution,

and public health, ultimately informing strategies to improve air quality and protect human health in the face of future pandemics or environmental crises.

2 Analysis

The ten reviewed papers collectively investigate the impact of COVID-19 lockdowns on air quality across various regions, focusing on key pollutants such as PM 2.5, PM 10, NO_2, SO_2, CO, and O_3. The findings consistently highlight significant reductions in particulate matter and nitrogen dioxide during lockdown periods, indicating the positive influence of restricted anthropogenic activities on air quality. Noteworthy improvements are observed in major polluted cities, emphasizing the potential of strategic lockdowns at pollution hotspots. The global scope of some studies enhances the understanding of the pandemic's impact on air quality worldwide.

However, a common limitation is the temporary nature of improvements, with pollutant levels rebounding post-lockdown. Additionally, there's a tendency to overlook the role of meteorological conditions and the complex interplay of atmospheric chemistry in pollutant variations. Future studies should delve deeper into the long-term effects of lockdowns, considering secondary pollutants, regional disparities, and the persistence of cleaner air beyond immediate lockdown periods. Comprehensive research incorporating meteorological factors, secondary pollutants, and sustained policy measures is crucial for effective long-term air quality management.

3 Data sources and methodology

3.1 Study area, period of study and data sources

This research is aimed at undertaking a detailed assessment on the evolution of air quality in Delhi over five years, from 1st January 2018 to 31st December 2023. The study mainly involves an analysis of daily mean concentration data for six primary criteria pollutants viz. PM 10, PM 2.5, NO_2, O_3, SO_2 and CO within NCT of Delhi.

The study obtained daily average concentrations for these parameters from Continuous Ambient Air Quality Monitoring Stations (CAAQMS) in Delhi. This data was accessed through the Central Control Room portal CCR on the CAAQMS section platform on this website. The CCR portal managed by Central Pollution Control Board (CPCB) acts as a one-stop-shop for ambient air quality data from CAAQMS installations across the country.

Delhi has a network of forty CAAQMS that are purposefully installed in different locations under the supervision of various bodies including CPCB, DPCC, IMD and IITM.

Unlike previous studies that occasionally excluded stations due to data irregularities, our approach is characterized by an inclusive strategy, maximizing data points wherever feasible to enhance precision and reliability. The longer time horizon facilitates a more nuanced and detailed exploration of the long-term fluctuations in air quality trends, including baseline conditions, the immediate repercussions of the lockdown implementation, and the resulting trajectory over the post-lockdown periods.

3.2 Methodology

This research conducts an in-depth study of the ambient air quality levels of the NCT of Delhi, over a five year period, from 2018 to 2023, emphasizing on six key criteria of air pollutants, namely: PM 10, PM 2.5, Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Carbon

Monoxide (CO) and Ozone. The data set used for this analysis consists of the daily average concentration records taken from the official website of the Central Pollution Control Board (CPCB). This data is recorded by the extensive network of 40 monitoring stations dispersed strategically across the city to ensure comprehensiveness and representativeness in the collected data.

The process of data collection and processing involved a thorough five-year data extraction process conducted in a methodical way, leading to a detailed and extensive account of the atmospheric conditions of Delhi during the period of study.

The daily average concentrations were then systematically compiled into monthly averages through comprehensive measures involving extensive data processing. This methodological refinement was performed with the intent to reduce inherent local perturbations in day-by-day observations and well as to increase analytical relevance by utilizing long-term trends.

4 Data analysis, comparison and results

4.1 Graphical observations

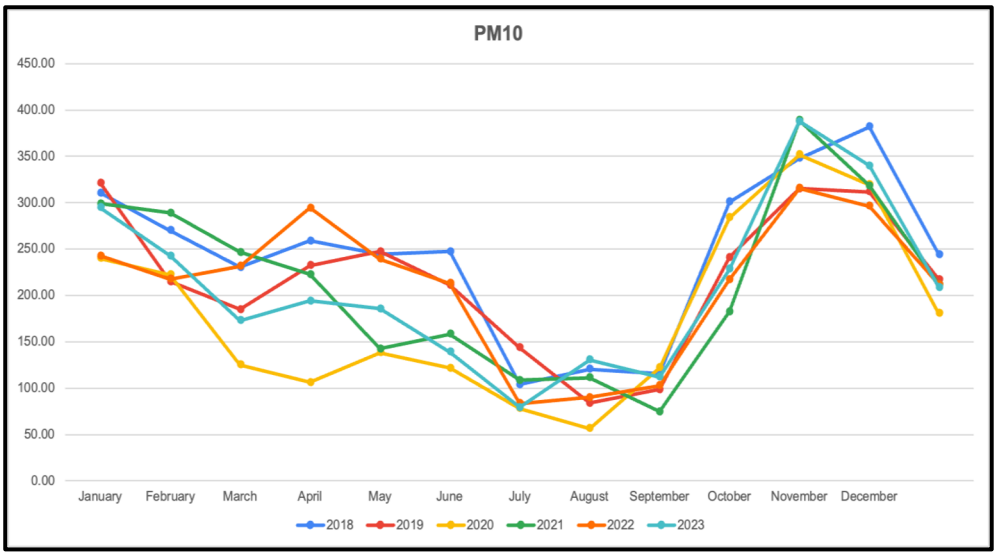


Fig 1. Graph Representing variation of PM 10 concentration with time

The PM 10 graph (Fig 4.1) displayed a significant drop during the lockdown in 2020, indicating an improvement in air quality compared to previous years. PM 10 levels dropped dramatically from 232 to 106 µg/m³ in April 2020, indicating a significant drop during the lockdown. This significant drop highlights how well the policies put in place to reduce particulate matter levels in the atmosphere are working.

Even following the lifting of the lockdown, a sustained relief in PM 10 concentrations persisted for an additional three months. This protracted improvement suggests a lasting impact on air quality beyond the specific lockdown period, emphasizing the enduring benefits of reduced anthropogenic activities on particulate matter levels. Within the same year (2020), PM 10 levels rebounded to typical winter norms despite the earlier reduction during the

lockdown period. This seasonal fluctuation underscores the complex interplay of meteorological conditions and human activities on particulate matter concentrations.

However, in 2021 and 2022, PM 10 levels reverted to pre-lockdown norms, indicating a lack of sustained pollution reduction compared to the distinctive improvements observed during the lockdown period. This suggests that the temporary relief witnessed in the immediate aftermath of the lockdown was not a lasting trend, emphasizing the need for continued efforts in pollution control to maintain air quality improvements. 2023, however, saw a decline in PM 10 levels.

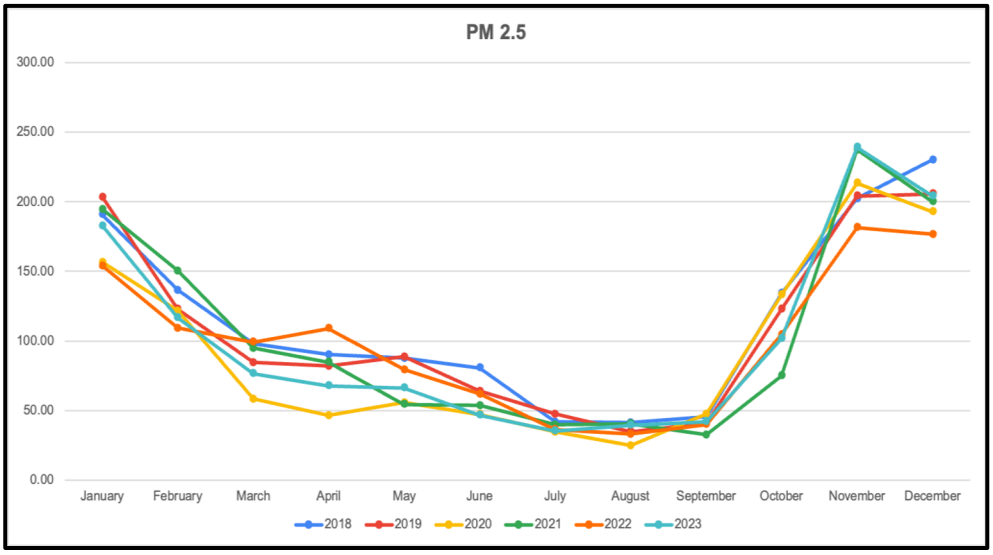


Fig .2 Graph Representing variation of PM 2.5 concentration with time

The PM 2.5 graph (Fig 4.2) exhibited a notable 43.44% decrease immediately upon the imposition of the lockdown (from 81.9 µg/m³ in April 2019 to 46.32 µg/m³ in April 2020), indicating a prompt response to reduced anthropogenic activities. However, unlike PM 10, no sustained relief in PM 2.5 concentrations was observed beyond the lockdown period. Even after the lifting of restrictions, PM 2.5 levels returned to pre-lockdown levels, underscoring the transient nature of the air quality improvements seen during the lockdown for this fine particulate matter.

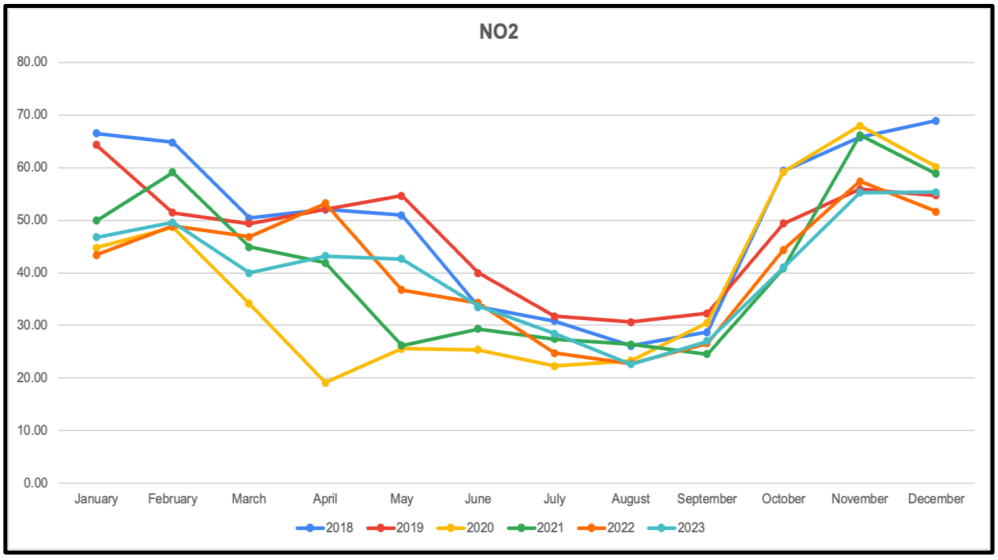


Fig 3. Graph Representing variation of NO_2 concentration with time

The NO_2 (Nitrogen Dioxide) graph (Fig 4.3) exhibited a substantial reduction following the initiation of the lockdown, indicative of the immediate impact of decreased anthropogenic activities.

However, post-lockdown, NO_2 levels gradually returned to their original concentrations, illustrating a resilience to sustained improvements after the easing of restrictions. This temporal pattern underscores the dynamic relationship between human activities and NO_2 emissions, emphasizing the transient nature of air quality enhancements during the lockdown period.

Subsequent to the lockdown period, the NO_2 trend persisted unaltered in the subsequent years, maintaining typical concentration levels. This steadfastness in the atmospheric levels of NO_2 underscores the resilience of pollutant concentrations to sustained changes in human activities, highlighting the need for continuous efforts in mitigating emissions to effect lasting improvements in air quality.

There was a minor decrease observed in NO_2 levels for 2021, 2022, and 2023, especially in May. Barring this slight decline, the general trend remained consistent with previous patterns, highlighting the consistency of nitrogen dioxide concentrations throughout this time period. This subtle dip implies that the overall trajectory of NO_2 concentrations seems to be consistent throughout time, despite possible small oscillations.

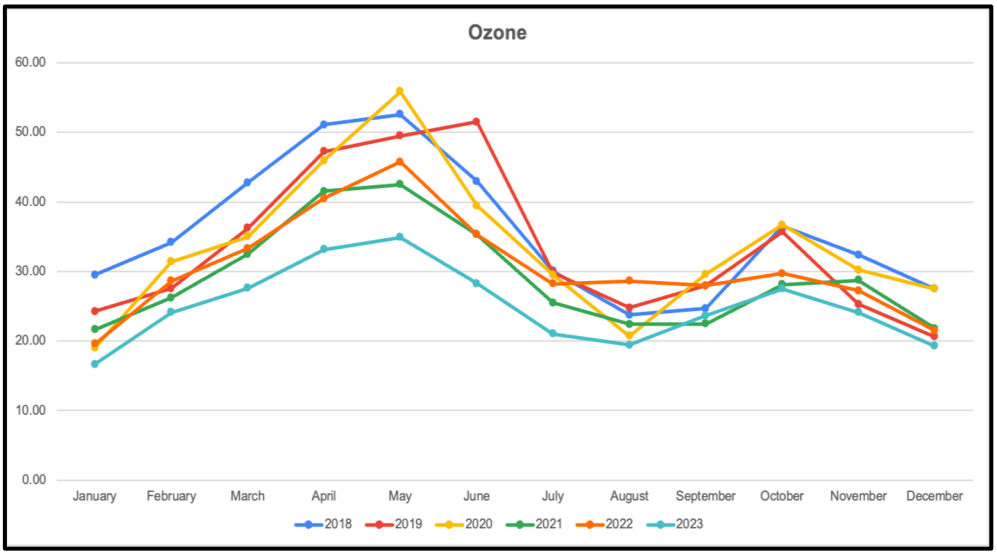


Fig 4. Graph Representing variation of ozone concentration with time

Contrary to expectations, the ozone graph (Fig 4.4) did not exhibit a decrease during the lockdown period; instead, it showed an unexpected increase in levels. This counterintuitive finding challenges our conventional assumptions about reduced human activities leading to lower ozone levels in the atmosphere. The observed rise in ozone levels throughout the lockout period highlights the necessity for a comprehensive research in interpreting the dynamics of air quality and the intricate interactions between precursor emissions and atmospheric chemistry.

The increase in ozone (O_3) during the COVID-19 lockdown, despite reduced pollutant emissions, can be attributed to a complex interplay of factors. The study conducted in Changzhou, China[8], found that the significant decrease in human activities during the lockdown led to lesser emissions of nitrogen oxides (NOx) and volatile organic compounds (VOCs), key ozone precursors. Surprisingly, this reduction led to a drop in VOC concentrations, especially aromatics, while NOx reduction was inappropriate. The imbalance in NOx and VOC reductions shifted O_3 formation from a VOC-sensitive regime to a VOC- and NOx-limited regime, resulting in an increase in O_3 during the lockdown. This counterintuitive finding underscores the need to carefully manage the reduction ratio of anthropogenic VOCs and NOx to mitigate ozone pollution effectively.

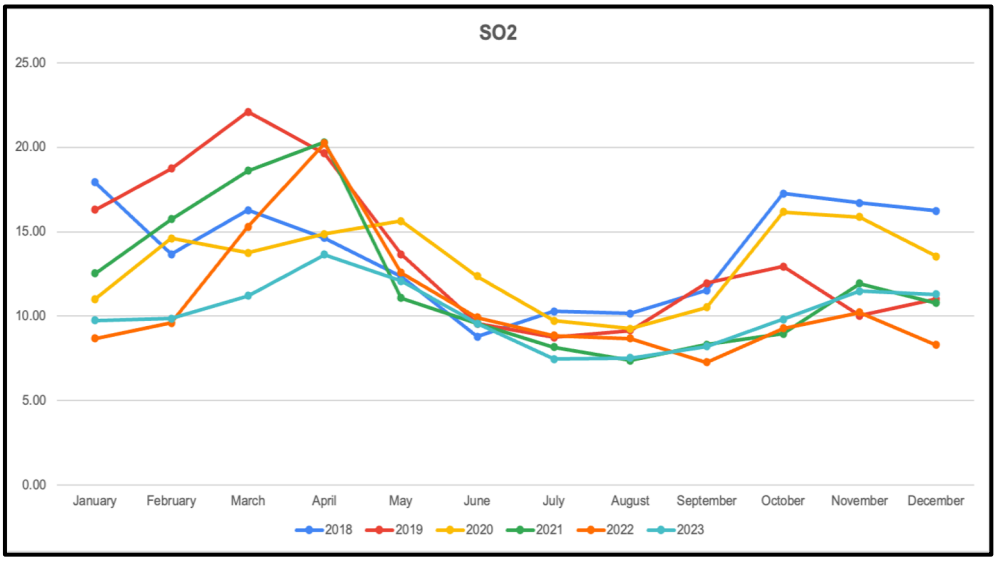


Fig 5. Graph Representing variation of SO_2 concentration with time

The SO_2 (Sulphur Dioxide) graph (Fig 4.5) revealed lower values during March and April, coinciding with the initial period of the lockdown. However, in the subsequent months, no prominent changes were observed, indicating a relative stability in SO_2 concentrations. This suggests that while short-term reductions occurred during the lockdown, the levels rebounded to baseline, highlighting the need for ongoing strategies to address sulfur dioxide emissions and maintain sustained air quality improvements.

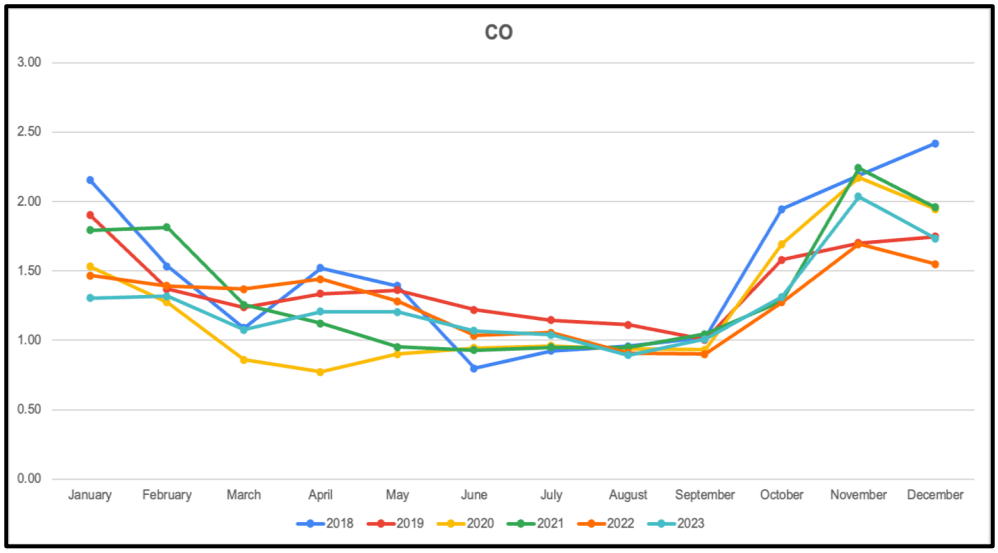


Fig 6. Graph Representing variation of CO concentration with time

Consistently, in both 2021 and 2022, SO_2 levels exhibited a similar pattern, remaining relatively stable and comparable to the baseline levels observed in the pre-lockdown period.

This sustained consistency underscores the persistent nature of sulfur dioxide concentrations, emphasizing the ongoing challenge of effectively managing and mitigating SO_2 emissions for lasting air quality improvements.

The CO graph (Fig 4.6) exhibited a substantial decline initially during the lockdown period, reflecting the immediate impact of reduced human activities. While a residual effect persisted post-lockdown, particularly noticeable, CO concentrations reverted to baseline levels by the winter months. This temporal pattern underscores the transient nature of the reductions, emphasizing the dynamic relationship between human activities and carbon monoxide emissions, with the return to typical levels indicating a resumption of regular sources of CO in the atmosphere.

In the subsequent years, 2021 and 2022, the CO graph showed no deviation from the established pattern, maintaining a consistent trend with no notable fluctuations. This stability in carbon monoxide concentrations suggests a resilience to sustained changes in human activities during these years.

4.2 Spatial mapping

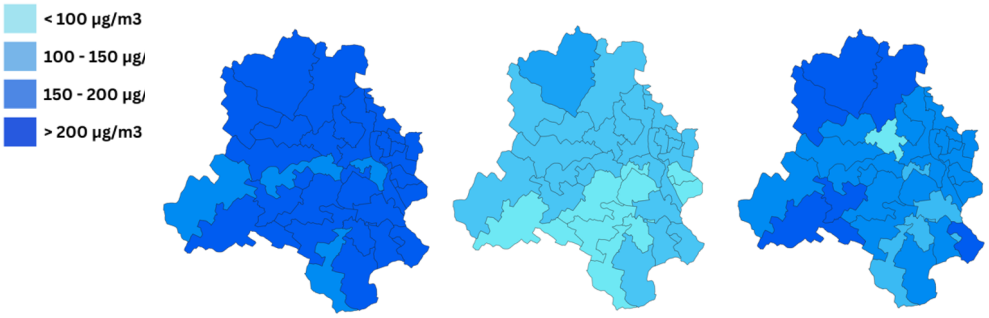


Fig 7. Maps representing spatial variation of PM 10 concentration across Delhi during pre (left), during (middle) and post (right) lockdown periods

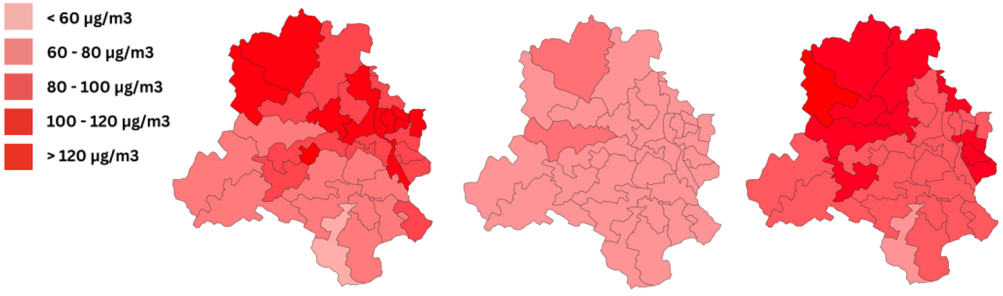


Fig 8. Maps representing spatial variation of PM 2.5 concentration across Delhi during pre (left), during (middle) and post (right) lockdown periods

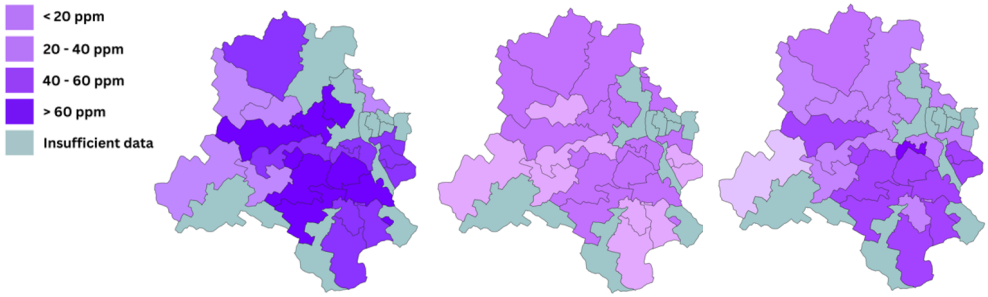


Fig 9 Maps representing spatial variation of NO_2 concentration across Delhi during pre (left), during (middle) and post (right) lockdown periods

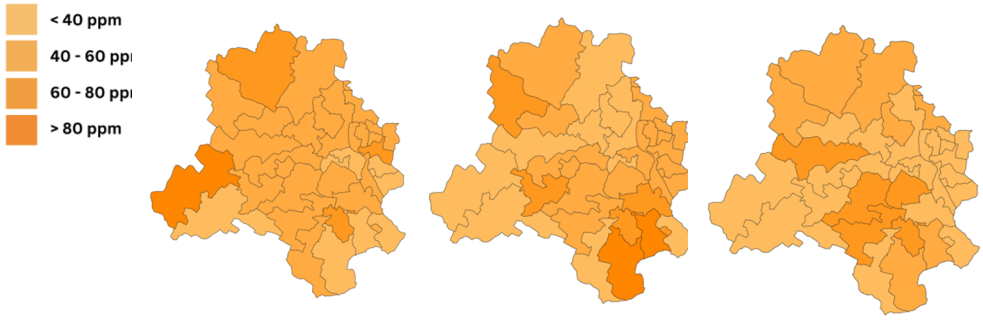


Fig 10. Maps representing spatial variation of O_3 concentration across Delhi during pre (left), during (middle) and post (right) lockdown periods

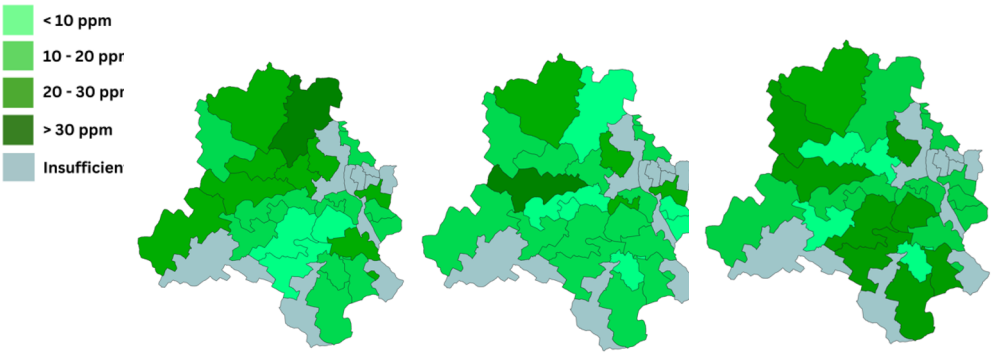


Fig 11. Maps representing spatial variation of SO_2 concentration across Delhi during pre (left), during (middle) and post (right) lockdown periods

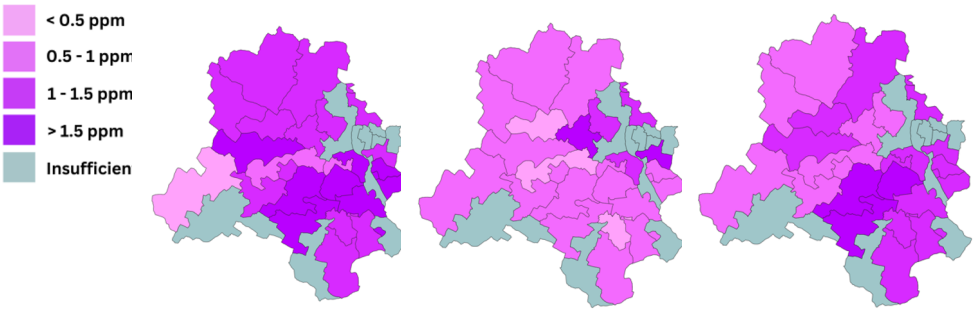


Fig 12. Maps representing spatial variation of CO concentration across Delhi during pre (left), during (middle) and post (right) lockdown periods

The spatial mapping shows how pollution levels changed in Delhi's various tehsils over the course of three separate time periods: before the lockdown, during the lockdown, and after the lockdown. A discernible pattern is revealed by contrasting the concentration levels shown on the maps.

In the pre-Lockdown period, the maps reveal comparatively high levels of pollution in the majority of Delhi's tehsils prior to the implementation of lockdown restrictions. This is to be expected, given regular levels of human activity—such as driving, manufacturing, and construction—contribute significantly to air pollution. In the lockdown period, the maps show a discernible decrease in pollution concentrations in all tehsils during the lockdown periods, when severe limitations on travel and business operations were in effect. This decline is explained by the significant drop in sources of emissions during the lockdowns, including factories, cars, and construction activity. Post the lockdown period, pollutant concentrations on the maps show a minor increase in comparison to the lockdown period, however overall levels remained lower than pre-lockdown conditions as lockdown restrictions were progressively lifted and economic activity restarted.

According to this rebound effect, although the lockdowns gave a little reprieve, once individuals went back to their regular schedules and businesses resumed operations, the fundamental causes of pollution soon reappeared.

5 Conclusion

Different pollutants exhibited different trends in air quality dynamics during COVID-19 lockdown period. PM 10 significantly reduced, particularly in April 2020, indicating the immediate effectiveness of locking down to mitigate levels of particulate matter. Nevertheless, this improvement was short lived and the levels reverted back to the baseline in subsequent years. Similarly, PM 2.5 also declined at the beginning of lockdown; however, it was evident that there were no lasting reductions past this time period for fine particulate matter.

Nitrogen Dioxide (NO_2) levels decreased significantly throughout the lockdown but eventually made a gradual recovery towards original concentrations, showing us its durability against long-term improvements. This pattern persisted in subsequent years thereby underlining that measures aimed at abating emissions must be continuous if lasting improvement on air quality is to be achieved Ozone levels actually increased during the lockdown, contrary to conventional expectations. Such surges were associated with disproportionately large cutbacks in nitrogen oxides (NOx) and volatile organic compounds

(VOCs), thereby transforming ozone formation paradigms, as per the study conducted in Changzhou, China[8].

6 Future Scope

The inferences extracted from this research lead to promising areas for further research on the air quality dynamics in urban settlements and the correlation of fluctuations in pollutant levels with anthropogenic activities. First and foremost, it is possible to carry out comprehensive apportionment studies to find out the sources of pollutants, thereby identifying whether regional or local factors are responsible for worsening air quality situations. The findings from these works will be useful in designing more targeted and efficient emission mitigation measures. On top of that, a deeper analysis of public health effects caused by air pollution is essential where correlation between pollutant levels and health outcomes is researched in Delhi. By doing so, comprehensive health impact assessment would be an effective tool for guiding policy makers to put measures which are not only meant to improve air quality but also safeguard public health. All together these future research activities could lead to development of policies based on evidence which will promote sustainable lifestyle and reduce air pollution related problems.

References

1. M. Ciotti, M. Ciccozzi, A. Terrinoni, W. C. Jiang, C. B. Wang, S. Bernardini, The COVID-19 pandemic. Critical reviews in clinical laboratory sciences, 57(6), 365–388 (2020). https://doi.org/10.1080/10408363.2020.1783198
2. S. Kaushik, S. Kaushik, Y. Sharma, R. Kumar, J. P. Yadav, The Indian perspective of COVID-19 outbreak. Virusdisease, 31(2), 146–153. (2020) https://doi.org/10.1007/s13337-020-00587-x
3. S. Ambika, U. Basappa, A. Singh, V. Gonugade, R. Tholiya, Impact of social lockdown due to COVID-19 on environmental and health risk indices in India. Environmental research, 196, 110932. (2021) https://doi.org/10.1016/j.envres.2021.110932
4. S. Mahato, S. Pal, K. G. Ghosh, Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. The Science of the total environment, 730, 139086. (2020) https://doi.org/10.1016/j.scitotenv.2020.139086
5. J. Mandal, S. Samanta, A. Chanda, S. Halder, Effects of COVID-19 pandemic on the air quality of three megacities in India. Atmospheric research, 259, 105659. (2021) https://doi.org/10.1016/j.atmosres.2021.105659
6. S. C. Pal, I. Chowdhuri, A. Saha, M. Ghosh, P. Roy, B. Das, R. Chakrabortty, M. Shit, COVID-19 strict lockdown impact on urban air quality and atmospheric temperature in four megacities of India. Geoscience frontiers, 13(6), 101368. (2022) https://doi.org/10.1016/j.gsf.2022.101368
7. S. Mahato, S. Pal, Revisiting air quality during lockdown persuaded by second surge of COVID-19 of megacity Delhi, India. Urban climate, 41, 101082. (2022) https://doi.org/10.1016/j.uclim.2021.101082
8. K. Zhang, Z. Liu, X. Zhang, Q. Li, A. Jensen, W. Tan, L. Huang, Y. Wang, J de Gouw, L Li, Insights into the significant increase in ozone during COVID-19 in a typical urban city of China, Atmos. Chem. Phys., 22, 4853–4866, (2022) https://doi.org/10.5194/acp-22-4853-2022.

Colloid and Polymer Science
https://doi.org/10.1007/s00396-024-05313-3

RESEARCH

Crosslinking of polyvinyl alcohol with di, tri, and tetracarboxylic acids: an experimental investigation

Leela Gautam[1,2] · Manish Jain[1] · Sudhir G. Warkar[1]

Received: 24 April 2024 / Revised: 14 August 2024 / Accepted: 20 August 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

In this study, three polycarboxylic acids with varying numbers of carboxyl groups were employed to crosslink polyvinyl alcohol (PVA): malonic acid as a diacid (MA), citric acid as a triacid (CA), and 1,2,3,4-butane tetracarboxylic acid as a tetraacid (BTCA). The crosslinking abilities of these acids were compared using physical, chemical, mechanical, morphological, thermal, and swelling measurements to assess their impact on the physicochemical properties of the resulting films. Based on the degree of crosslinking, mechanical strength, and thermal stability, tetra acid demonstrated superior crosslinking performance compared to di and tri acids. The highest strength was observed in the tetra acid crosslinked film, which exhibited a 127% increase over neat PVA. Regarding thermal stability, the decomposition temperature followed the order of tetraacid (450 °C) > triacid (378 °C) > diacid (350 °C). However, in terms of swelling behavior, triacid-crosslinked film, i.e., P-CA, exhibited the highest swelling. Further, the tetra acid-crosslinked film exhibited the lowest crystallinity and a higher contact angle (104.9°) than diacid and triacid crosslinked films. Additionally, the crosslinked films displayed enhanced elasticity compared to pure PVA, with the elasticity order being diacid > triacid > tetraacid, possibly due to differences in available crosslinking sites among the crosslinkers.

Keywords PVA · Crosslinking · Physicochemical properties · Polycarboxylic acids · Esterification

Introduction

Chemical cross-linking is commonly used to modify the properties of polyvinyl alcohol (PVA), aiming to improve its water sensitivity and moisture barrier [1]. Among various cross-linkers like aldehydes [2], diisocyanate [3], boric acid [4], and poly(carboxylic) acids [5], the poly(carboxylic acids) are particularly promising due to their eco-friendliness, cost-effectiveness, non-toxicity, biodegradability, and lack of accumulation [6, 7]. Furthermore, the crosslinking of PVA with carboxylic acids involves an esterification process, with water being the primary by-product. This aspect contributes to the overall environmental friendliness of both the crosslinkers and the process itself, as the reactions generate no hazardous or toxic by-products harmful to the environment [8].

Numerous studies over the years have highlighted the advantages of using polycarboxylic acids for crosslinking PVA in various applications [9–11]. A recent review comprehensively summarizes the applications and properties associated with polycarboxylic acid-crosslinked PVA [12].

Additionally, in existing literature, certain comparative studies have investigated how the diversity of carboxylic acids used as crosslinkers can result in differences in molecular interactions, thus directly affecting the physicochemical properties of PVA [13–15]. For instance, a comparative investigation involving suberic acid as an aliphatic molecule and terephthalic acid as an aromatic molecule indicated the superior compatibility of aliphatic molecules for crosslinking compared to aromatic ones [16]. In another study, the positioning of carboxylic groups in the aliphatic chain was examined, revealing that the trans position of the carboxyl group in fumaric acid facilitated enhanced crosslinking compared to maleic acid [17]. Similarly, an analysis of phthalic acid, isophthalic acid, and terephthalic acid demonstrated that the para position in terephthalic acid was most effective

✉ Sudhir G. Warkar
sudhirwarkar@gmail.com

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

[2] Department of Chemistry, Zakir Husain Delhi College (University of Delhi), New Delhi, India

Published online: 30 August 2024

for crosslinking. In contrast, the ortho position was less favorable [18]. Furthermore, the research highlighted the impact of carbon chain length in crosslinking, with succinic acid and adipic acid in the dicarboxylic series showing that longer carbon chains create more space within the PVA matrix, leading to increased water permeability [19]. Overall, these studies emphasize the significance of considering factors such as molecular structure, carboxylic group positioning, and carbon chain length when selecting crosslinkers for specific applications to achieve desired properties in PVA films.

While independent reports exist on the properties of PVA films crosslinked with di, tri, and tetra acid crosslinkers [20–22], there is a lack of comparative studies conducted under identical experimental conditions to precisely determine the effect of the number of carboxyl groups of polycarboxylic acids on PVA film properties. This experimental work is therefore intended to fill this gap and is specifically aimed at investigating the impact of varying numbers of carboxylic acid groups on the physicochemical properties of PVA.

To investigate this, three different polycarboxylic acids, i.e., malonic acid (MA) representing dicarboxylic acids, citric acid (CA) for tricarboxylic acids, and 1,2,3,4-butanetetracarboxylic acid (BTCA) for tetracarboxylic acid were utilized for PVA crosslinking. The study involves a comprehensive comparative analysis of the chemical structure, hydrophilicity, swelling behavior, thermal stability, tensile strength, and morphology of the crosslinked PVA films. Various techniques such as ATR-FTIR, NMR, XRD, AFM, TGA, DSC, tensile testing, swelling measurements, and sessile drop contact angle analysis are employed for this assessment. The detailed findings from these investigations are presented and correlated with the structural characteristics of the di, tri, and tetracarboxylic acids.

Materials and method

Materials

The polyvinyl alcohol (PVA) with a molecular weight of 145,000, fully hydrolyzed, was sourced from Sigma-Aldrich, USA. Carboxylic acids, namely, malonic acid (MA), citric acid (CA), and 1,2,3,4-butanetetracarboxylic acid (BTCA), served as diacid, triacid, and tetraacid respectively, were procured from Loba Chemie and utilized without any purification. Deionized water was the solvent used for all experiments. A summary of the molecular structure and properties of the carboxylic acids is provided in Table 1 and Fig. 1.

Methodology

The crosslinked PVA films were prepared using the solution casting method as depicted in Fig. 2. Initially, an 8% w/v PVA solution was prepared by dissolving PVA in deionized water under stirring at 80 °C for approximately 1 h, ensuring the formation of a homogeneous and transparent solution. Subsequently, 20 mL of the PVA solution was poured into a round bottom flask equipped with a condenser, and 30% by weight of carboxylic acid was added. The resulting solution was stirred in a water bath for 1 h at 80 °C to achieve

Table 1 Description of molecular structure and properties of the carboxylic acids

Name of the acid	Nature of the acid	Carbon chain length	No. of carboxyl group	No. of hydroxyl group	Molecular weight (g/mol)	pKa values			
						pKa_1	pKa_2	pKa_3	pKa_4
Malonic acid	Dicarboxylic acid	3	2	0	104.062	2.83	5.69		
Citric acid	Tricarboxylic acid	6	3	1	192.125	3.13	4.76	6.40	
1,2,3,4-Butanetetracarboxylic acid	Tetracarboxylic acid	8	4	0	234.16	3.3	4.58	5.85	7.16

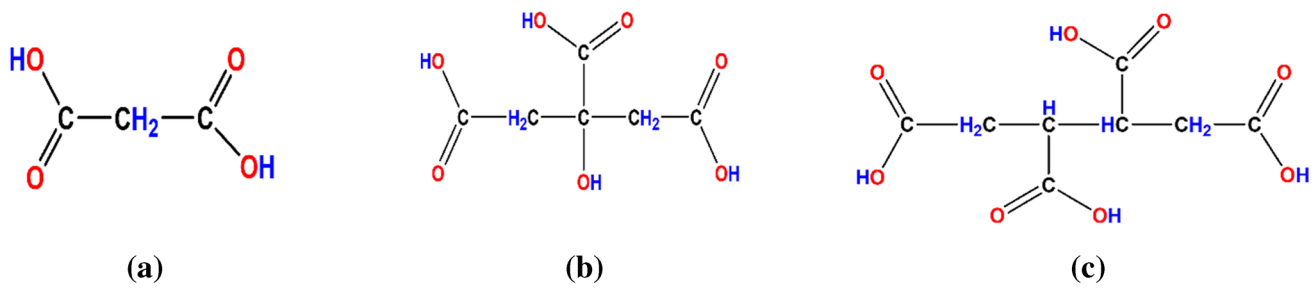


Fig. 1 Structure of polycarboxylic acids: **a** malonic acid, **b** citric acid, and **c** 1,2,3,4-butanetetracarboxylic acid

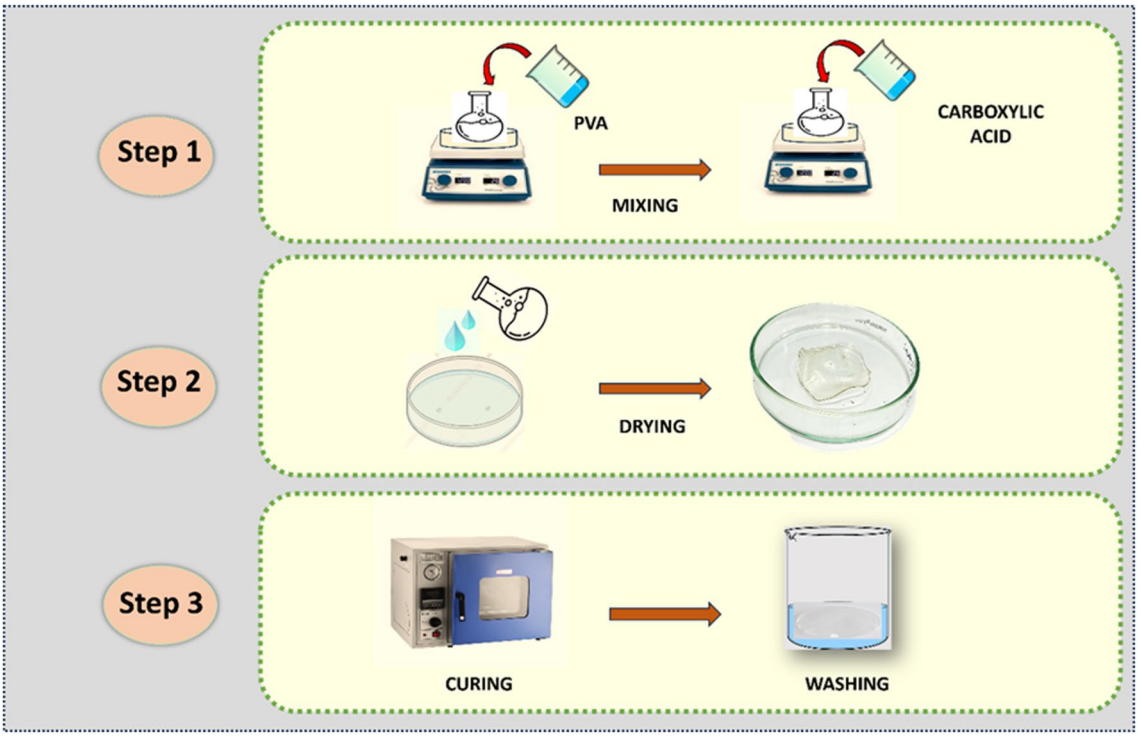


Fig. 2 Schematic diagram of the preparation of crosslinked PVA films

complete homogenization. The clear solution obtained was then poured into casting molds for the complete evaporation of water under ambient conditions, forming the film. This process was repeated for each acid. The dried films were carefully removed from the molds and subjected to curing to promote the crosslinking reaction. To remove any unreacted crosslinker, the cured films were washed multiple times by immersing them in distilled water to eliminate extracted acids. The washed crosslinked films were air-dried at room temperature and then stored in sealed plastic bags for further analysis. The films were named P-MA, P-CA, and P-BTCA, with their composition and formulation details provided in Table 2.

Characterization of the crosslinked films

Attenuated total reflectance (ATR)-FTIR spectroscopy

Confirmation of the presence of characteristic functional groups was verified by the Fourier-transform infrared (FTIR) technique. FTIR spectra of the samples were recorded using the Brukeroptik Spectrum instrument. The experiments were carried out with the attenuated total reflectance accessory (ATR) range of 4000–500 cm^{-1}. Origin lab software was used to plot the spectra from the obtained values.

Table 2 Compositions and formulation of crosslinked PVA films

S. No	Designation	PVA (mL)	Malonic acid (g)	Citric acid (g)	1,2,3,4-Butanetetracarboxylic acid (g)	Curing temperature (°C)	Curing time (Min.)
1	PVA-MA	20	0.48			140	10
2	PVA-CA	20		0.48		140	10
3	PVA-BTCA	20			0.48	140	10

Solid state ^{13}C nuclear magnetic resonance spectroscopy (^{13}C-NMR)

The ^{13}C-NMR spectrum was obtained using a Brüker AVANCE 400 MHz spectrometer equipped with a 4-mm probe, operating at a spinning frequency of 10 kHz. Chemical shifts were referenced against tetramethylsilane (TMS) as the standard. The spectrum was recorded with spinning sideband suppression to enhance clarity and accuracy.

Atomic force microscopy (AFM)

AFM images were recorded using Asylum research MFP3D-BIO atomic force microscope to observe the 3-D image and estimate the films' roughness.

X-ray diffraction (XRD) analysis

To study the crystalline structure of the PVA crosslinked films, XRD patterns of all crosslinked films were recorded using the X'Pert PRO model, diffractometer (PANalytical, Netherlands) [λ(Cu-Kα) = 1.54 A]. Scattered radiations were recorded in the range of 2θ ranging from 5 to 60°.

Thermal gravimetric analysis (TGA)

To observe the thermal stability and decomposition temperature of neat PVA and crosslinked PVA samples, TGA and DTGA were recorded using PerkinElmer TGA 4000 with an N_2 atmosphere in the range of 30–700 °C at the rate of 10 °C/ min.

Differential scanning calorimetry (DSC)

DSC analysis was performed to examine the change in glass transition temperature (T_g) and melting point in the PVA crosslinked films. The heat flow measurement was recorded using the PerkinElmer model DSC8000 analyzer. The samples of nearly 10 mg mass of pure PVA and crosslinked PVA films were heated from 30 to 250 °C at a rate of 10 °C/min.

Water contact angle measurement (WCA)

The hydrophilicity of the PVA crosslinked films was investigated by a Kruss advance 1.6.2.0 sessile drop contact angle Goniometer at ambient temperature. A deionized water droplet of 2 µL was dropped by a micro syringe onto the top surface of the films. For each film, a minimum of 5 points were measured to get the average WCA.

Swelling percentage and crosslinking density

Small sections of crosslinked PVA films, with different dicarboxylic acids, were cut into 1 cm × 1 cm pieces. They were weighed and placed in deionized water at room temperature. The swollen films were taken out after regular intervals, excess water was wiped off with filter paper, weighed, and then, they were put back into the water. This process continued until the films reached equilibrium swelling. The degree of swelling was calculated using Eq. (1).

$$S\% = \frac{W_S - W_D}{W_D} \times 100 \tag{1}$$

where S, W_S, and W_D are the degree of swelling in percent, the weight of the films in the swollen state, and the weight of the film in dry states, respectively.

To better understand the swelling properties of the crosslinked films, parameters like the number average molecular mass between crosslinks (Mc), cross-link density (q), and mesh size (ξ) were calculated. The Flory-Rehner equation, provided below, was used to calculate Mc [8, 23, 24].

$$\frac{1}{Mc} = -\frac{\frac{v}{V}([\ln 1 - \varphi] + \varphi + \chi\varphi^2)}{\varphi^{\frac{1}{3}} - \frac{1}{2}\varphi} \tag{2}$$

Here, V represents the molar volume of water ≃molar mass of water (18.1 g/mol). The Flory Huggins polymer–water interaction parameter, denoted as χ, ≃0.498. The specific volume of the polymer (v) is 0.788 cm^3/g. The polymer volume fraction in the swollen state, denoted as φ, was estimated from equilibrium swelling results using the provided relation given in Eq. (3) [8].

$$\varphi = \left[1 + \frac{d_p}{d_s}\left(\frac{m_a}{m_b}\right) - \frac{d_p}{d_s}\right]^{-1} \tag{3}$$

To determine φ using Eq. (3), the calculation involved the following parameters: the mass of the film before swelling (m_b), the mass of the film after swelling (m_a), the density of the film (d_p), and the density of the solvent (d_s). The density of the films was calculated using Eq. (4).

$$d_p = \frac{W}{S \times X} \tag{4}$$

In the given context, X denotes the average thickness of the film, S represents the cross-sectional area, and W corresponds to the weight of the film. The density (ρ) of the films was determined using the following equation 5.

$$\rho = \frac{1}{v.Mc} \tag{5}$$

The mesh size (ξ) was calculated using the Flory characteristic ratio of the polymer C by the following equation 6 [23]:

$$\xi = \varphi^{\frac{-1}{3}} \left[c_n \left(\frac{M_c}{M_r} \right) \right]^{\frac{-1}{2}} .l \tag{6}$$

Here, Cn is the Flory ratio (with a value of 8.3), l represents the bond angle between the C − C bond (measured at 1.54 Å), M_r is the molecular weight of repeating units of PVA, and M_C represents the number average molecular mass between crosslinks.

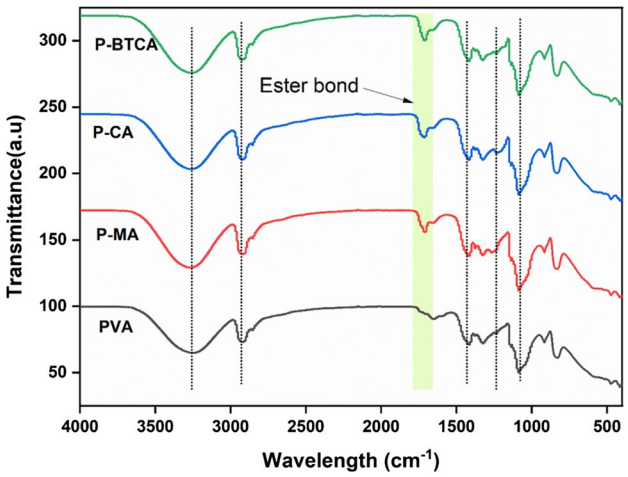


Fig. 3 FTIR spectra of the PVA and PVA crosslinked films with di, tri, and tetra acids

Results and discussions

Film formation and crosslinking mechanism

Cross-linked PVA films were prepared with by crosslinking 30% w/w concentration of MA, CA, and BTCA, at a curing temperature of 140 °C and a curing time of 10 min. The curing time and temperature were determined based on preliminary studies. PVA crosslinked films were obtained through an esterification crosslinking mechanism. Here, homogeneous solutions of PVA and acids were dried at room temperature. During drying, the reactive functional groups of PVA and acid molecules were aligned and potentially formed H-bonds and some of the ester bonds. A similar mechanism was also proposed in previous studies [8, 25]. To enhance the crosslinking reaction, the films were further cured by heating them in an oven. The supplied thermal energy during crosslinking facilitated the formation of more ester bonds between the previously aligned COOH and OH groups. Additionally, water evaporation occurred, promoting the esterification reaction in the forward direction and lowering the pH of the prepared films, thereby catalyzing esterification [25]. Washing the cured films in water helped leach out most of the unreacted acid, leaving behind ester-bonded and strongly H-bonded acids. The prepared films were further characterized by ATR-FTIR to prove the formation of ester bonds, discussed later.

Attenuated total reflectance (ATR)-FTIR spectroscopy

Figure 3 displays the ATR-FTIR spectra for both neat PVA and crosslinked PVA. In the PVA spectrum, the broadband at 3257 cm^{-1} corresponds to the hydroxyl group's stretching, and the peak at 2929 cm^{-1} is a result of symmetric and asymmetric vibrations of the C-H bond. The 1419 cm^{-1} peak represents the C-H bending of the methylene group, and the 1664–1668 cm^{-1} peaks are linked to the C-O bending associated with the hydroxyl group in PVA. In the crosslinked films, a distinct and sharp peak emerges in the 1712–1720 cm^{-1} range, indicating ester carbonyl stretching. Additionally, peaks in the 1230–1260 cm^{-1} region confirm asymmetric stretching of C–O–C in ester bonds, verifying an esterification reaction between PVA and acids. This reaction results in chemical crosslinking via ester linkage and induces C = O stretching. During the crosslinking process, the existing methylene groups (1419 cm^{-1}) in the films remain unchanged, maintaining the intensity of the methylene group peak. The quantification of crosslinking can be achieved using the band index method, calculating the area ratio of the C = O (ester bond) peak to the CH_2 (methylene group) peak [26]. Results indicated a slightly higher ratio for the P-BTCA film (0.57) compared to P-CA (0.54) and P-MA (0.52) among the three films, suggesting a greater contribution of tetracarboxylic acids to in-situ crosslinking in PVA. The enhanced cross-linking ability of BTCA over CA and MA can be linked to its extra carboxylic groups, which can cross-link with multiple OH groups of PVA. Therefore, a direct association was identified between the extent of crosslinking and the number of carboxylic groups.

Solid state ^{13}C-NMR analysis

Figure 4 illustrates the solid-state ^{13}C-NMR spectrum of PVA and its crosslinked films with polycarboxylic acids such as P-MA, P-CA, and P-BTCA. In the spectrum of pure PVA, distinct resonance peaks appear at 45.90 ppm (b), corresponding to the methine carbon and within the 65–78 ppm range (a), representing carbons associated with hydroxyl groups. These peaks are characteristic of the

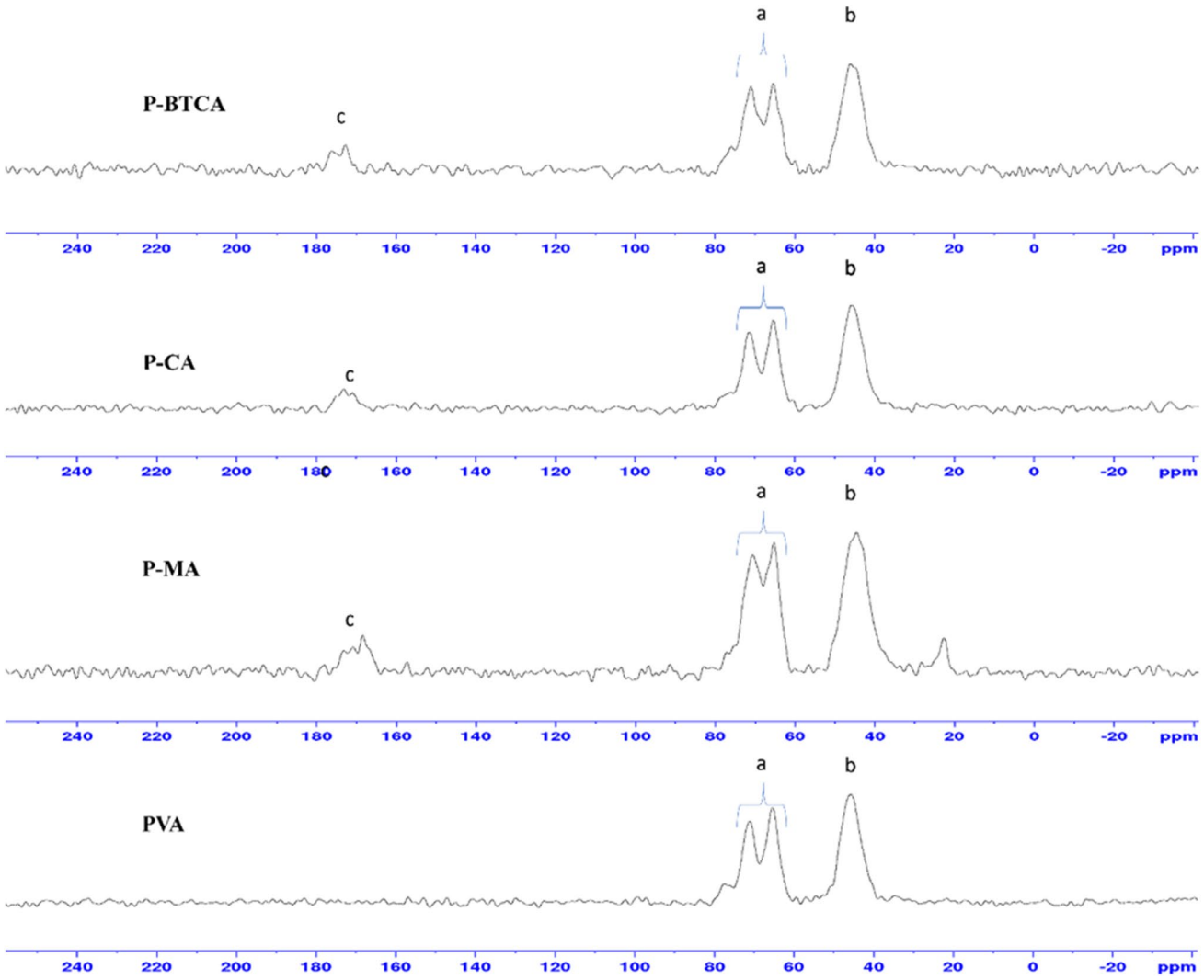


Fig. 4 Solid state ^{13}C-NMR analysis of the PVA and PVA crosslinked films with di, tri, and tetra acids

PVA structure. However, after crosslinking with polycarboxylic acids, a new resonance peak emerges between 167 and 173 ppm (c), attributed to the carbonyl carbon in ester groups. The absence of this peak in the pure PVA spectrum and its appearance in the crosslinked films strongly indicate the successful formation of ester crosslinks, confirming that the polycarboxylic acids have chemically bonded with the PVA.

Swelling analysis

The swelling behavior or hydrophilicity of PVA is primarily influenced by the abundance of free OH groups which attract water molecules and the ease with which water molecules can diffuse between the chains. The crosslinking of PVA with carboxylic acids is widely recognized for enhancing water resistance. As OH groups are consumed during crosslinking, PVA chains become covalently linked, reducing the available space for water molecules to penetrate. However, the reduction in the degree of swelling varies with the type of crosslinker used. The impact of di, tri, and tetracarboxylic acids as crosslinkers on the swelling behavior of PVA is illustrated in Fig. 5. All three crosslinked films maintained their dimensional stability and did not disintegrate in water throughout the study, confirming the effectiveness of crosslinking and the formation of numerous ester groups. The swelling behavior of the crosslinked films followed the order of P-CA > P-MA > P-BTCA with swelling percentages of 192%, 132%, and 106%, respectively.

The variation in swelling can be explained based on the structural aspects of MA, CA, and BTCA, as detailed in Table 1. PVA-CA exhibited the maximum swelling, despite having more crosslinking than P-MA, possibly due to the presence of one OH group in its structure, facilitating greater

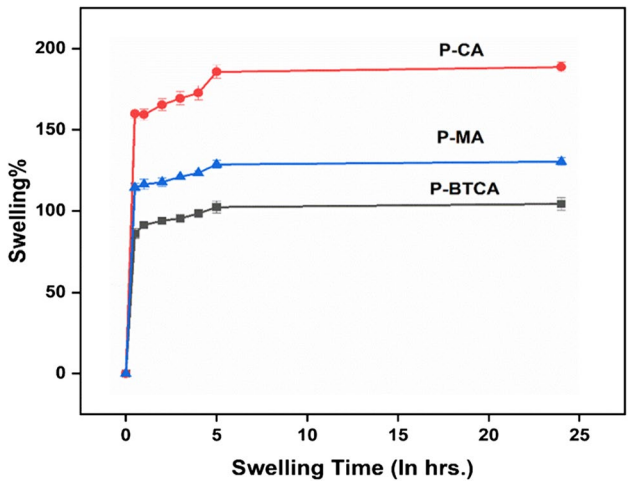


Fig. 5 Swelling analysis of the crosslinked PVA films with di, tri, and tetraacids

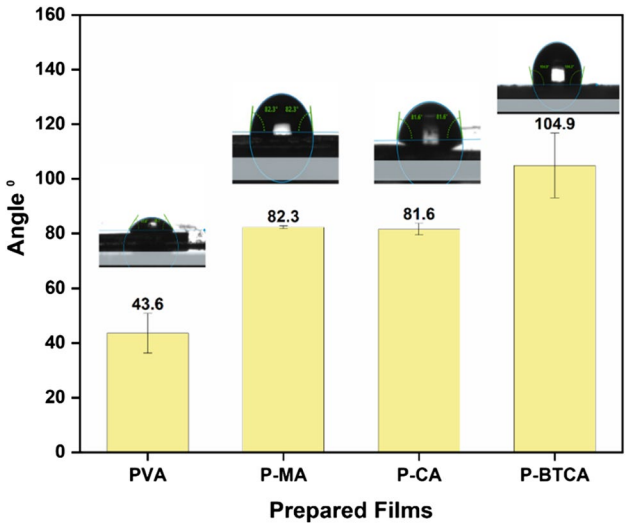


Fig. 6 Variation of contact angle of PVA and PVA crosslinked films with di, tri, and tetraacid

water attraction. Similar results were reported by Ahmet Cay et al. [27].

Furthermore, various parameters like MC, ξ, and ρ are used to further explain the swelling behavior of the films. These parameters can easily be determined from the Flory–Rehner swelling equilibrium theory [28]. On the application of this theory, it has been observed that the M_C and ξ values are found to be higher in cross-linked PVA films having low ρ as shown in Table 4. The crosslinking density is particularly crucial in understanding the swelling behavior, as a lower crosslinking density indicates greater distances between crosslinking points [28].

Additionally, the mesh size provides insights into the available space for water molecules to permeate. As can be seen in Table 3, the average molecular weight between crosslinks and mesh size values was found proportional to the swelling capacity of the film. The Mc, ρ, and ξ values obtained for crosslinked films were found aligned with swelling results.

Water contact angle measurement (WCA)

Water contact angle (WCA) serves as a crucial means to estimate the hydrophilicity or hydrophobicity of a surface. As depicted in Fig. 6, there is a significant reduction in the WCA of crosslinked films when compared to pure PVA film. This reduction is an indication of alterations in surface properties which is further elucidated by the corresponding images showing distinct behaviors of water droplets on the film surfaces. The presence of polar hydroxyl (OH) groups in PVA enhances its affinity for water, thereby leading to a lower contact angle. However, through the process of crosslinking with carboxylic acids, many of these polar OH groups are removed from the PVA chain, resulting in an increase in contact angle.

As shown in Fig. 6, the contact angles of PVA, P-MA, P-CA, and P-BTCA films were 43.6, 82.3, 81.6, and 104.9, respectively, which emphasize a significant transition in surface properties upon crosslinking. This transition signifies a shift from a highly hydrophilic state to a moderately hydrophobic state [22]. Notably, P-BTCA exhibits the highest contact angle among the films, indicating more effective crosslinking compared to the others. These findings are consistent with ATR-FTIR spectroscopy and swelling studies.

Table 3 Network parameters of crosslinked PVA films with di, tri and tetra acids

Crosslinked film	% swelling	φ	Mc (g/mL)	ρ (mol/cm^3)	ξ (Å)
P-MA	132	0.37	509.42	2.49×10^{-3}	21.02
P-CA	192	0.34	683.06	1.85×10^{-3}	25.04
P-BTCA	106	0.42	339.44	3.47×10^{-3}	16.45

X-ray diffraction (XRD) analysis

The XRD graphs of crosslinked films, namely, P-MA, P-CA, and P-BTCA, were analyzed and compared with the XRD graph of pure PVA film shown in Fig. 7. PVA is acknowledged as a semicrystalline polymer, characterized by both a crystalline phase and an amorphous matrix. The primary peak observed at around 2θ of 19.17° in the XRD graph of PVA is attributed to its orderly arrangement and crystalline phase, resulting from hydrogen bonding among the OH groups [29]. Post crosslinking, the intensity of PVA's characteristic peak at 19.17° was noted to diminish and broaden in all crosslinked films. This reduction in peak intensity

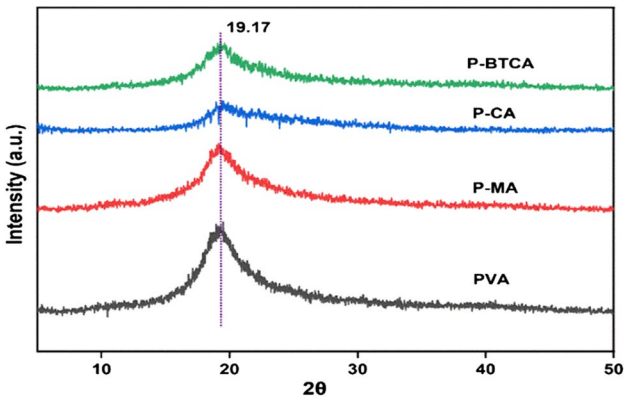


Fig. 7 XRD diffractograms of PVA and PVA crosslinked films with di, tri, and tetra acid

indicates a decrease in PVA's crystallinity after crosslinking due to the consumption of OH groups in PVA [18]. The decline in crystallinity is associated with the integration of large carboxylic acid moieties into the structure through interactions with free hydroxyl groups in PVA, leading to the formation of ester bonds. These alterations not only suggest changes in PVA's crystalline phase but also imply increasing molecular disorganization within the PVA matrix. The relative percent crystallinity of the crosslinked films was determined relative to the 19.17° peak of PVA using the following equation [19].

$$x\% = \frac{Ac}{Ac + A_A} \times 100$$

where A_C is the crystalline area and A_A is the amorphous area. The calculated relative percent crystallinity exhibited a trend of increasing crystallinity in the order of PVA (57.04%) > P-MA (53.60%) > P-CA (51.20%) > P-BTCA (50.03%), likely influenced by variations in the number of carboxylic acids. Importantly, these results on crystallinity are consistent with those obtained from ATR-FTIR analysis, reinforcing the conclusions of the study.

Surface analysis using AFM

Figure 8 illustrates AFM topography images, revealing distinct surface morphologies for each film. The pure PVA film exhibited a mean roughness of 9.30 nm and root mean

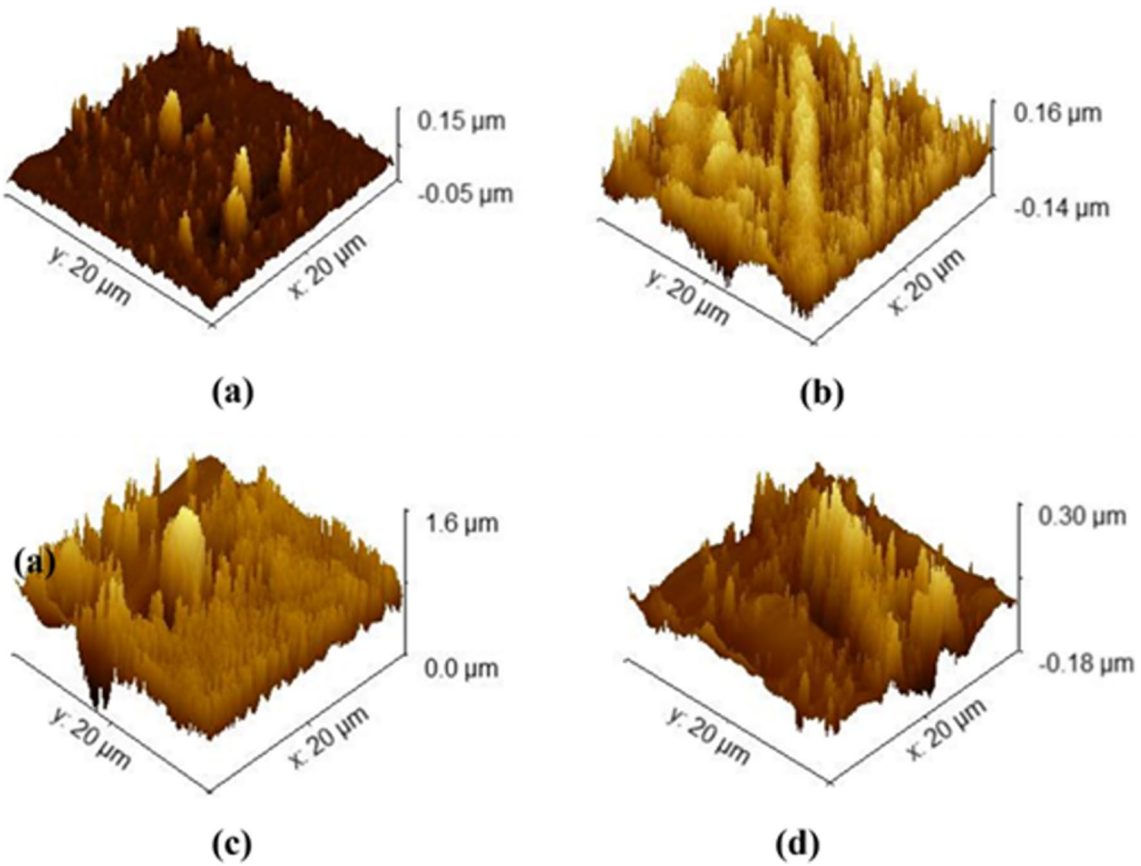


Fig. 8 AFM images of **a** PVA, **b** diacid crosslinked film, P-MA, **c** triacid crosslinked film, P-CA, and **d** tetraacid crosslinked film, P-BTCA

square (RMS) of 16.31 nm. In contrast, films crosslinked with MA, CA, and BTCA displayed mean roughness values of 49.14 nm, 126.30 nm, and 30.53 nm, with RMS values of 63.0 nm, 170.6 nm, and 45.88 nm, respectively. Additionally, these roughness values provide valuable insights into swelling studies. A consistent trend was observed, where a reduction in roughness followed the order of P-BTCA < P-MA < P-CA, correlating with a similar trend observed in the decrease in swelling. This relationship suggests that smoother surfaces exhibit more controlled swelling behavior due to reduced surface irregularities, which limits water absorption [30].

Thermal gravimetric analysis (TGA)

The TGA and DTG thermogram of the pure PVA and PVA-crosslinked films are shown in Figs. 9c and a. All TGA graphs show three-stage degradation. For PVA, initial mass loss is observed due to the water evaporation (80–220 °C, 8% loss), both bound water and adsorbed on the surface of the film. The second stage (265–498 °C, 85.5% loss) shows the main degradation and the largest weight loss due to the removal of hydroxyl groups and the formation of polyene macromolecules [31]. The third stage (500–690 °C, 5%) is because of the overall decomposition of the backbone of PVA [32].

For PVA crosslinked films, the first stage (85–210 °C) that corresponds to a minimal mass loss is due to the removal of moisture and the second stage at elevated temperatures (260–475 °C) which may be due to the breakage of ester linkage or decarboxylation of free carboxylic acids present in the film. The third stage should result from the degradation of the backbone of crosslinked films [32]. Additionally, there is a difference in decomposition rates of crosslinked films than PVA films which can be attributed to the different crosslinking degrees. At 350 °C, PVA film retains only

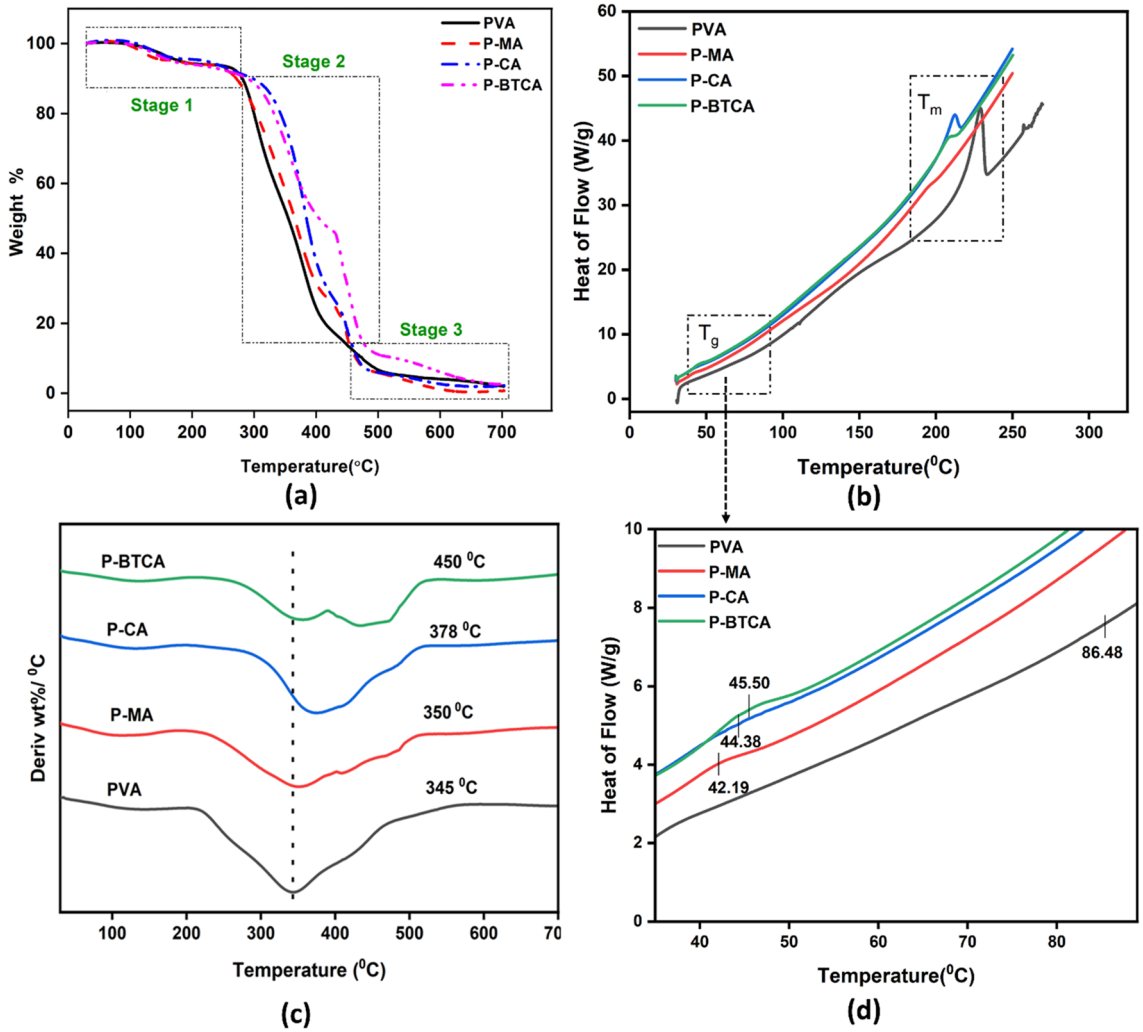


Fig. 9 **a** TGA, **c** DTG, and **b**, **d** DSC graphs of PVA and PVA crosslinked films with di, tri, and tetra acids

52.05% of its original weight, whereas P-MA, P-CA, and P-BTCA retain 57.65%, 71.48%, and 76.03%, respectively. Among all three crosslinked films, the P-BTCA film showed a slightly higher residue left at 700 °C than the other two crosslinked films.

In the DTG analysis shown in Fig. 9c, a noticeable shift in decomposition temperature was observed in all crosslinked films, indicating the formation of an ester network between the PVA chains, thereby imparting thermal stability to the crosslinked networks. Particularly, the P-BTCA films exhibited the highest decomposition temperature, with two decomposition peaks observed. The initial peak, above 350 °C, indicates the shift in decomposition temperature of PVA accompanied by its cyclization, while the subsequent peak at 450 °C can be attributed to the complete degradation of cyclized PVA [19]. The order results indicate that the crosslinked films display improved thermal stability compared to pure PVA.

Differential scanning calorimetry (DSC)

The DSC curves of both pure PVA and PVA films crosslinked with MA, CA, and BTCA are depicted in Fig. 9b, d. In the DSC curve of pure PVA, the glass transition temperature (T_g) was observed at 86.48 °C. The T_g values of all crosslinked PVA films were found to be lower than that of pure PVA film, following the order: pure PVA (86.48 °C) > P-BTCA (45.50 °C) > P-CA (44.38 °C) > PVA-MA (42.19 °C). To elucidate the aforementioned findings, several factors must be taken into account that can influence the T_g values of crosslinked polymers, such as the inherent rigidity of existing polymer chains, the density of crosslinking and the chemical structure of the crosslinking agent [3]. The presence of hydroxyl groups in PVA contributes to the polymer's stiffness through hydrogen bonding. Introducing a crosslinking agent reduces the number of hydroxyl groups and weakens hydrogen bonding interactions, resulting in decreased stiffness. Conversely, crosslinking enhances rigidity, and the characteristics of the newly introduced moieties can impact T_g differently, depending on their structure. In the current study, three distinct crosslinking agents were employed, each with varying numbers of crosslinking sites. Resulting, the bonds formed between PVA and BTCA may be more rigid due to the higher number of carboxyl groups and maximum crosslinking density. Conversely, MA and CA offer fewer carboxyl groups than BTCA, potentially resulting in a comparatively less rigid structure. In our case, the observed trend of T_g decreasing despite higher crosslinking density may be attributed to the compensatory effect of the weakening of intermolecular hydrogen bonding chemical network formation hence lower rigidity. Furthermore, the inclusion of acids may also contribute to some plasticizing effects [21]. Similar results were reported by Sheng Xu et al. [13].

Pure PVA is known for its semi-crystalline nature and demonstrates a higher melting temperature, with a sharp endothermic peak around 228 °C corresponding to crystalline melting (T_m). The peak area (ΔH_m) in the melting region serves as an indicator of the film's crystallinity. The introduction of crosslinkers results in lower melting temperatures and broader peaks, indicating effective crosslinking. This crosslinking disrupts the crystalline region, increasing the amorphous region and subsequently reducing T_m. The melting temperature (T_m) and ΔH_m values of the PVA and crosslinked PVA films follow the order: PVA (228.59 °C, 139.72 J/g) > P-CA (212 °C, 29.81 J/g) > P-BTCA (208.73 °C, 13.78 J/g) > P-MA (195.84 °C, 5.57 J/g). This suggests that the addition of a crosslinker has reduced the overall crystallinity of the films, consistent with the XRD results discussed earlier. However, P-MA exhibits the lowest T_m and ΔH_m values among the films, which could be attributed to structural changes in some parts of its polymer structure during crosslinking.

Tensile characteristics

The typical tensile stress (σ)–strain (ε) curves of PVA, P-MA, P-CA, and P-BTCA films presented in Fig. 10 highlight the significant changes in tensile properties resulting from the crosslinking process using di-, tri-, and tetra-carboxylic acids. For neat PVA, the recorded tensile stress (σ) and strain (ε) values are 39.62 ± 3.14 MPa and 148.06 ± 2.79 MPa, respectively. The summarized results in Table 4 reveal that all crosslinked films exhibited slightly higher tensile strength than pure PVA and displayed greater elongation.

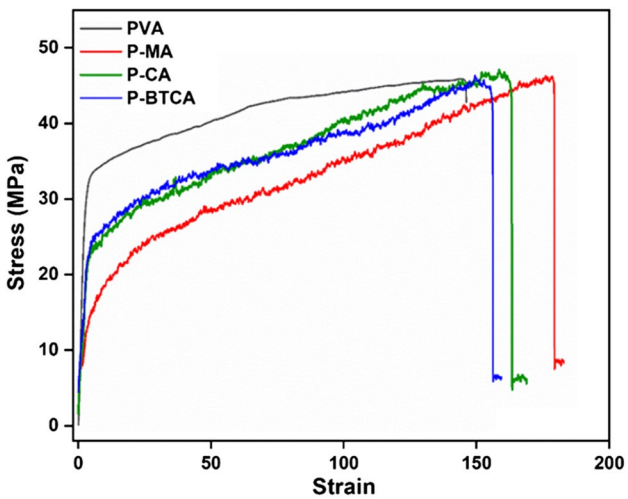


Fig. 10 Typical stress–strain curve of PVA and PVA crosslinked films with di, tri, and tetraacids

Table 4 Mechanical properties of PVA and PVA-crosslinked films with di, tri and tetra acids

S. No	Designation	Tensile strength (MPa)	Elongation (%)
1	PVA	39.62 ± 3.14	148.06 ± 2.79
2	P-MA	44.75 ± 4.74	183.28 ± 7.26
3	P-CA	43.62 ± 4.09	154.24 ± 5.93
4	P-BTCA	50.41 ± 1.45	150.72 ± 1.44

The above results can be better understood by applying the same arguments as discussed in thermal analysis, focusing on three competitive factors during the crosslinking process: variations in crosslinking density, alterations in intermolecular hydrogen bonding, and the structural characteristics of the crosslinkers. The higher tensile strength observed in crosslinked films can be attributed to the dense network structure formed by chemical ester bonds, which strengthen intermolecular forces within the film [23, 33]. Analysis of FTIR results and the Mc value reveals that BTCA creates more crosslinks compared to MA and CA, thereby contributing to the maximum tensile strength of P-BTCA films (127% more than neat PVA).

However, this increase in crosslinking comes at the cost of substituting hydrogen bonding, which decreases rigidity which in turn can impart more elongation in the films. Additionally, an inverse relationship was found between elongation and the number of carboxylic groups present in crosslinkers. Films containing crosslinkers with fewer carboxyl groups, such as MA and CA, afford greater flexibility to the overall structure of PVA, consequently exhibiting 122% and 102% more elongation compared to P-BTCA. These findings are supported by DSC and TGA analysis, highlighting the significant impact of both the crosslinking process and the structure of the crosslinker on the mechanical properties of the films.

Conclusion

The study aimed to compare the effect of variation of carboxyl groups of polycarboxylic acids, i.e., MA (diacid), CA (triacid), and BTCA (tetra acid) as crosslinkers on the properties of PVA films. The calculated crosslinking density of the tetraacid crosslinked film was noticed to be the highest among all crosslinked films and consequently exhibited the least swelling. In thermal analysis, the tetra acid crosslinked film demonstrated superior properties compared to the di and triacid crosslinked film. From DTG graphs, the order of decomposition temperature for crosslinked was found in order of P-BTCA (450 °C) > P-CA (378 °C) > P-MA (350 °C). XRD results indicated that crosslinking led to a decrease in the overall crystallinity of PVA films, resulting in lower percentages of crystallinity for all crosslinked films. Additionally, a more significant reduction in hydrophilicity was observed in the tetraacid crosslinked film compared to both di and triacid crosslinked PVA films. Also, all crosslinked films displayed enhanced elasticity and tensile in comparison to neat PVA, with the order of elasticity being P-MA > P-CA > P-BTCA probably because of the difference in crosslinking sites available in crosslinkers.

In summary, the collective findings emphasize the critical importance of considering the molecular structure of carboxylic acids during the crosslinker selection process. These considerations play a crucial role in tailoring PVA films to achieve specific properties for diverse applications. This study not only provides valuable insights into the influence of different crosslinkers on the physicochemical properties of PVA films but also contributes significantly to the comprehension of green crosslinking processes.

Acknowledgements The authors are grateful to the Delhi Technological University, New Delhi, India, and Zakir Husain Delhi College, University of Delhi, New Delhi, India.

Author contribution Conceptualization: [Manish Jain], [Sudhir G. Warkar], [Leela Gautam]; Methodology: [Leela Gautam] Formal analysis and investigation: [Leela Gautam]; Writing-original draft preparation: [Leela Gautam]; Writing-review and editing: [Manish Jain], [Sudhir G. Warkar]; Funding acquisition: NA; Resources: [Leela Gautam]; Supervision: [Sudhir G. Warkar], [Manish Jain].

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethical approval Not applicable.

Competing interests The authors declare no competing interests.

References

1. Kamoun EA, Chen X, Eldin MSM, Kenawy ERS (2015) Crosslinked poly (vinyl alcohol) hydrogels for wound dressing applications: A review of remarkably blended polymers. Arab J Chem 8(1):1–14. https://doi.org/10.1016/j.arabjc.2014.07.005
2. Nambiar AP, Pillai R, Vadikkeettil Y, Sanyal M, Shrivastav PS (2022) Glutaraldehyde-crosslinked poly (vinyl alcohol)/halloysite composite films as adsorbent for methylene blue in water. Materials Chem Phys 291:126752. https://doi.org/10.1016/j.matchemphys.2022.126752
3. Krumova M, Lopez D, Benavente R, Mijangos C, Perena JM (2000) Effect of crosslinking on the mechanical and thermal properties of poly (vinyl alcohol). Polymer 41(26):9265–9272
4. Itou T, Kitai H, Shimazu A, Miyazaki T, Tashiro K (2014) Clarification of cross-linkage structure in boric acid doped poly(vinyl alcohol) and its model compound as studied by an organized combination of x-ray single-crystal structure analysis, raman spectroscopy, and density functional theoretical calculation. J Phys Chem B 118(22):6032–6037. https://doi.org/10.1021/jp5026569

5. Gebresas GA, Szabó T, Marossy K (2023) A comparative study of carboxylic acids on the cross-linking potential of corn starch films. J Mol Struct 1277:134886. https://doi.org/10.1016/j.molstruc.2022.134886
6. Gebresas GA, Szabó T, Marossy K (2023) A comparative study of carboxylic acids on the cross-linking potential of corn starch films. J Mol Struct 5(1277):134886. https://doi.org/10.1021/acssuschemeng.1c06101
7. Pangon A, Saesoo S, Saengkrit N, Ruktanonchai U (2016) Multicarboxylic acids as environment-friendly solvents and in situ crosslinkers for chitosan/PVA nanofibers with tunable physicochemical properties and biocompatibility. Carbohydr Polym 138:156–165. https://doi.org/10.1016/j.carbpol.2015.11.039
8. Larrañeta E et al (2018) Synthesis and characterization of lignin hydrogels for potential applications as drug eluting antimicrobial coatings for medical materials. ACS Sustain Chem Eng 6(7):9037–9046. https://doi.org/10.1021/acssuschemeng.8b01371
9. Kudoh Y, Kojima T, Abe M, Oota M, Yamamoto T (2013) Proton conducting membranes consisting of poly(vinyl alcohol) and poly(styrene sulfonic acid): crosslinking of poly(vinyl alcohol) with and without succinic acid. Solid State Ionics 253:189–194. https://doi.org/10.1016/j.ssi.2013.09.047
10. Dattola E et al (2019) Development of 3D PVA scaffolds for cardiac tissue engineering and cell screening applications. RSC Adv 9(8):4246–4257. https://doi.org/10.1039/C8RA08187E
11. Xue YL, Huang J, Lau CH, Cao B, Li P (2020) Tailoring the molecular structure of crosslinked polymers for pervaporation desalination. Nature Commun 11(1):1461. https://doi.org/10.1038/s41467-020-15038-w
12. Gautam L, Warkar SG, Ahmad SI, Kant R, Jain M (2022) A review on carboxylic acid cross-linked polyvinyl alcohol: properties and applications. Polym Eng Sci 62(2):225–246. https://doi.org/10.1002/pen.25849
13. Xu S, Shen L, Li C, Wang Y (2018) Properties and pervaporation performance of poly(vinyl alcohol) membranes crosslinked with various dianhydrides. J Appl Polym Sci 135(17):15–19. https://doi.org/10.1002/app.46159
14 Suganthi S, Mohanapriya S, Raj V, Kanaga S, Dhandapani R, Vignesh S, KalyanaSundar J (2018) Tunable physicochemical and bactericidal activity of multicarboxylic-acids-crosslinked polyvinyl alcohol membrane for food packaging applications. ChemistrySelect 3(40):11167–11176. https://doi.org/10.1002/slct.201801851
15 Cadinelli BLS et al (2018) Synthesis and characterization by ellipsometry of cationic membranes for fuel cells. Mater Sci Forum 930:625–630. https://doi.org/10.4028/www.scientific.net/MSF.930.625
16. Sonker AK, Rathore K, Nagarale RK, Verma V (2018) Crosslinking of polyvinyl alcohol (PVA) and effect of crosslinker shape (aliphatic and aromatic) thereof. J Polym Environ 26(5):1782–1794. https://doi.org/10.1007/s10924-017-1077-3
17. Le Tran H, Vuu MND, Hoang QM, Nguyen NT, Mai PT (2016) Pervaporation dehydration of ethanol-water mixture using crosslinked poly(vinyl alcohol) membranes. Sci Technol Dev J 19(4):97–106. https://doi.org/10.32508/stdj.v19i4.787
18. Dlamini DS, Wang J, Mishra AK, Mamba BB, Hoek EMV (2014) Effect of cross-linking agent chemistry and coating conditions on physical, chemical, and separation properties of PVA-Psf composite membranes. Sep Sci Technol 49(1):22–29. https://doi.org/10.1080/01496395.2013.813040
19. Sonker AK, Tiwari N, Nagarale RK, Verma V (2016) Synergistic effect of cellulose nanowhiskers reinforcement and dicarboxylic acids crosslinking towards polyvinyl alcohol properties. J Polym Sci Part A Polym Chem 54(16):2515–2525. https://doi.org/10.1002/pola.28129
20. Gao CD, Ren JL, Wang SY, Sun RC, Zhao LH (2014) Preparation of Polyvinyl Alcohol/Xylan Blending Films with 1, 2, 3, 4-Butane Tetracarboxylic Acid as a New Plasticizer. J Nanomater 2014(1):764031. https://doi.org/10.1155/2014/764031
21. Jose J, Al-Harthi MA (2017) Citric acid crosslinking of poly (vinyl alcohol)/starch/graphene nanocomposites for superior properties. Iranian Polymer J 26:579–587. https://doi.org/10.1007/s13726-017-0542-0
22. Hrabalikova M, Holcapkova P, Suly P, Sedlarik V (2016) Immobilization of bacteriocin nisin into a poly(vinyl alcohol) polymer matrix crosslinked with nontoxic dicarboxylic acid. J Appl Polym Sci 133(28):1–10. https://doi.org/10.1002/app.43674
23. Pandit AH, Mazumdar N, Imtiyaz K, Rizvi MMA, Ahmad S (2019) Periodate-modified gum arabic cross-linked PVA hydrogels: a promising approach toward photoprotection and sustained delivery of folic acid. ACS Omega 4(14):16026–16036. https://doi.org/10.1021/acsomega.9b02137
24 Gupta P, Purwar R (2021) Influence of cross-linkers on the properties of cotton grafted poly (acrylamide-co-acrylic acid) hydrogel composite: swelling and drug release kinetics. Iranian Polymer J 30(4):381–391. https://doi.org/10.1007/s13726-020-00897-3
25. Liu T, Peng X, Chen Y, Zhang J, Jiao C, Wang H (2020) Solid-phase esterification between poly(vinyl alcohol) and malonic acid and its function in toughening hydrogels. Polym Chem 11(29):4787–4797. https://doi.org/10.1039/d0py00023j
26. Truong YB et al (2017) Functional cross-linked electrospun polyvinyl alcohol membranes and their potential applications. Macromol Mater Eng 302(8):1–9. https://doi.org/10.1002/mame.201700024
27. Çay A, AkçakocaKumbasar EP, Keskin Z, Akduman Ç, ŞendemirÜrkmez A (2017) crosslinking of poly (vinyl alcohol) nanofibres with polycarboxylic acids: biocompatibility with human skin keratinocyte cells. J Mater Sci 52(20):12098–12108. https://doi.org/10.1007/s10853-017-1370-5
28. Purkayastha S, Prakash M, Ghosh AK, Saha S (2022) Preparation and properties of crosslinked poly(vinyl alcohol)/nanofibrillated cellulose based biocomposite by subcritical water/CO2 process. Polym Compos 43(10):6887–6896. https://doi.org/10.1002/pc.26734
29. Liu G, Shi K, Sun H, Yang B, Weng Y (2023) Enhancing hydrophobicity and oxygen barrier of xylan/PVOH composite film by 1, 2, 3, 4-butane tetracarboxylic acid crosslinking. Polymers 15(13):2811
30 do Nascimento FC, de Aguiar LCV, Costa LAT, Fernandes MT, Marassi RJ, Gomes ADS, de Castro JA (2021) Formulation and characterization of crosslinked polyvinyl alcohol (PVA) membranes: effects of the crosslinking agents. Polym Bull 78(2):917–929. https://doi.org/10.1007/s00289-020-03142-2
31. Ghorpade VS, Dias RJ, Mali KK, Mulla SI (2019) Citric acid crosslinked carboxymethylcellulose-polyvinyl alcohol hydrogel films for extended release of water soluble basic drugs. J Drug Deliv Sci Technol 52(May):421–430. https://doi.org/10.1016/j.jddst.2019.05.013
32. Suganthi S, Vignesh S, Kalyana J, Vairamuthu S (2020) Fabrication of PVA polymer films with improved antibacterial activity by fine - tuning via organic acids for food packaging applications. Appl Water Sci 10(4):1–11. https://doi.org/10.1007/s13201-020-1162-y
33. Sonker AK, Teotia AK, Kumar A, Nagarale RK, Verma V (2017) Development of polyvinyl alcohol based high strength biocompatible composite films. Macromol Chem Phys 218(15):1–13. https://doi.org/10.1002/macp.201700130

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.

Electrical Engineering
https://doi.org/10.1007/s00202-024-02684-2

ORIGINAL PAPER

Decision-making in multi-objective DG planning for distribution system via Shannon's entropy

Shubham Gupta[1] · Vinod Kumar Yadav[1] · Madhusudan Singh[1] · Ashutosh K. Giri[2]

Received: 29 November 2023 / Accepted: 13 August 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

In the current paradigm, integration of distributed generation (DG) has become essential for ensuring the quality, reliability, and security of distribution network operations. Existing literature typically formulates multi-objective problems to quantify the techno-economic assessment of DG placement in a distribution system. Nevertheless, there is a notable lack of a suitable method to assign weight impartially to each objective for optimal decision-making. This paper introduces a novel technique for strategically placing DG units in the distribution network, employing weights calculated based on the importance level of various techno-economic objectives using Shannon's entropy. The proposed approach has been applied to a 38-node test system to illustrate its efficacy. The numerical findings from four distinct case studies reveal that changes in the physical attributes of the system correspondingly influence the significance of objectives in determining the optimal placement and size of DG. The results show significant reductions in active and reactive power losses and total annualized operational costs, with maximum reductions of 48.17%, 33.30%, and 42.96%, respectively. The minimum voltage magnitude improves from 0.9252 pu in the base case to 0.9384, 0.9695, 0.9369, and 0.9348 for Cases 1, 2, 3, and 4, respectively. Moreover, a comparative statistical analysis underscores the superiority of the proposed method over prevailing weight allocation strategies by achieving a 3.59% reduction in annual expenditure, while maintaining competitive network performance metrics in addressing the multi-objective DG placement problem.

Keywords Distributed generation · Distribution network planning · Load modeling · Multi-criteria decision-making · Shannon's entropy

List of symbols

U_ρ, U_σ, U_τ	Set of industrial, residential, and commercial type of load buses, respectively
U_t	Set of time intervals in 24 h
U_d	Set of number of days intervals in a year
U_dg	Set of DGs to be placed in distribution system
$U_\mathrm{b}, U_\mathrm{l}$	Set of buses (except for substation bus) and lines in the network, respectively

Variables

$P_\mathrm{Loss}(Q_\mathrm{Loss})$	Real (reactive) power loss
V_b^t	Bus voltage magnitude during tth time
$\overline{S}_l$	MVA flow in the lth line
P^dg	Real power generated by DG
O_{kl}	Index score of kth entity for lth criteria
λ_l	Weight value for lth criteria (objective)
η_k	Comprehensive score of kth entity (i.e., bus)

Parameters

$P_\mathrm{L}^\mathrm{cp}\left(Q_\mathrm{L}^\mathrm{cp}\right)$	Constant power real (reactive) load demand
$P_{\mathrm{L},i}\left(Q_{\mathrm{L},i}\right)$	Real (reactive) load demand at ith bus
$V_\mathrm{spec.}^t$	Voltage at root node at tth time interval
T	Total number of time intervals
S_l^max	MVA capacity of lth line
UIC^dg	Unit installation cost of DG ($/MW)

✉ Shubham Gupta
shubham.gupta6691@gmail.com

Vinod Kumar Yadav
vinodkumar@dtu.ac.in

Madhusudan Singh
madhusudan@dce.ac.in

Ashutosh K. Giri
ashu_ee2002@rediffmail.com

[1] Department of Electrical Engineering, Delhi Technological University, Bawana Road, New Delhi, Delhi 110042, India

[2] Department of Electrical Engineering, Government Engineering College, Bharuch, Gujarat 392002, India

Published online: 29 August 2024

UOC^{dg}	Unit O&M cost of DG ($/MW)
UC_{sub}	Unit cost of electricity purchased from grid ($/MW)

1 Introduction

The world is facing environmental issues such as the depletion of the ozone layer, ever-increasing pollution, the quality degradation of underground water, and GHG emissions, which are causing irreversible climate change. According to the International Energy Agency (IEA) report [1], three-quarters of the total GHGs are generated by the energy sector, which can only be tamed by using clean and efficient distributed energy technologies. The primary cause of the unprecedented growth in the integration of distributed generators (DGs) into the distribution networks (DNs) in recent years is to rapidly achieve the goal of Net Zero by 2050 for nations around the globe. Additionally, the idea of the self-sustainable micro-grid is evolving in the modern power system; hence, the effective use of DGs in the power distribution system has become crucial [2]. DGs integrated at the end-user side meet the fast-increasing electricity demand without installing new power lines and enhance system performance, namely by reducing power loss, upgrading system voltage, and enhancing loading capability [3, 4].

The transitioning of the power system from a vertically integrated structure to an unbundled one has yielded advantages, yet it has concurrently introduced intricacies in the planning, maintenance, and operational activities. Now, consumers have assumed the role of prosumers, as traditional distribution networks evolve into active distribution networks [5]. The advancement and cost-effectiveness of photovoltaic, wind, hydro, and other renewable energy technologies (RETs) are ongoing, leading to a substantial surge in their integration into power distribution networks. However, the escalating penetration of DGs has augmented the liabilities of Distribution Network Planners (DNPs). Consequently, a meticulous impact analysis of DGs on the distribution system has become imperative to prevent the deterioration of reliability and power availability [6]. Unplanned and arbitrary placement of DGs may result in adverse effects on DN quantities, including an upswing in power losses, system voltage drops, and diminished loading capability [7].

Numerous methods have been presented in the literature to deal with the optimal DG placement problem, which broadly focuses on node location, size, and type of DG. The methods to identify optimal node locations mainly were based on node sensitivity to different operational factors [8], for example, voltage, active or reactive power loss, and system load capacity, or reliability factors [9], for example, average failure rate, average annual outage time, and average outage time. The term DG placement index (DGPI) was coined [10], where the optimal location for DG placement was identified through the weighted sum of different indexes. In [11], various DN reliability factors were combined to form the expected interruption index for DG siting at optimal node locations. Numerous techniques were suggested to assess the optimal size in DG placement problems, such as analytical methods [12], hybrid methods [13], and artificial methods [14], where the problem was formulated as a single or multi-objective function. A detailed review can be found in [15–17]. DGs were classified based on their power factor (pf), and the optimal DG placement problem was solved for different DG types in [18].

Conventionally, minimizing power loss and enhancing node voltage emerge as two crucial objectives in the DG placement problem. Recently, the widespread integration of power electronics devices in DN has given rise to challenges such as recurring voltage fluctuations and poor power quality. Consequently, the DG placement problem has evolved into a multi-objective (MO) problem, aiming to address various technical and economic perspectives concurrently. Historically, numerous meta-heuristic optimization techniques have been proposed for solving MO-DG placement problems [16]. A hybrid meta-heuristic technique that combines grey wolf and particle swarm optimization presents a potent solution for attaining optimal DG allocations [19]. The multi-objective whale optimization algorithm has been proposed to address both technical and economic aspects of DG unit placement [20]. An improved golden jackal optimization algorithm was introduced in [21], primarily aimed at addressing global optimization challenges, particularly in deploying capacitor banks and DGs. Addressing the complexities inherent in dispatchable and non-dispatchable DGs, [22] introduced a dynamic multi-objective bidirectional co-evolutionary algorithm. In a two-stage procedure, a chaotic bat algorithm dynamically optimized DG size and location [23].

Additionally, [24] delves into a physics-inspired methodology for optimizing the siting and sizing of renewable energy sources. A fuzzy-based campus placement optimization algorithm was proposed in [25], minimizing voltage deviations and real and reactive power losses. A fuzzified firefly optimization algorithm was utilized to solve a multi-objective function to optimize DG integration [26]. Proposing an analytical framework, [27] put forth a method to evaluate the optimal size and location of DGs while considering various load scenarios. In [28], a multi-objective optimization approach utilized the analytical hierarchical process to estimate the weight associated with each attribute. [29] proposed a holistic framework for planning renewable energy sources and electric vehicle charging stations, with a focus on voltage stability, energy losses, and consumer satisfaction. Lastly, in [30], an improved analytical approach was introduced for optimal DG placement, emphasizing reduced power losses and enhanced voltage profile and stability.

Hence, the existing literature has some notable research gaps - namely, there remains a need for novel analytical and heuristic optimization methodologies that can efficiently handle the complexities of large networks while also ensuring robust convergence and avoidance of local optima. While the existing literature acknowledges that multi-objective optimization can evaluate multiple criteria to provide a more adaptable decision-making framework, there remains a notable absence of impartial weight selection methods that systematically account for the implications of varying weighting schemes on solution robustness and optimality. Addressing these gaps is essential for enhancing the reliability and applicability of multi-objective optimization methods in practical applications.

Motivated by the research gap highlighted above, this article aims to present a novel strategy for evaluating the weights impartially in a multi-objective problem of optimal decision-making for DG placement in a radial distribution network (RDN). The proposed approach utilizes Shannon's entropy (SE) to assess the weights for the considered objectives (i.e., impact indexes) in the problem. The SE formula quantifies these weights based on the degree of importance of an index in the placement of DG. The solution methodology identifies the technical and economic indexes to form a multi-objective DG placement problem. Then, the weighting factors are evaluated using the proposed approach to determine the optimal DG location and size in the second step. The efficacy of this approach is assessed by testing it on a 38-node test system, and the analysis of the results shows promising outcomes. Specifically, the proposed work has the following contributions:

1. Voltage-dependent load modeling analyzes the impact of loads on DG placement problems. The characteristics of the DN change as the type of load (i.e., residential, commercial, and industrial) changes.
2. Different objectives are considered to address various technical and economic issues like real or reactive power loss, voltage deviation, line capacity limit, installation, and operational cost of DGs.
3. The proposed approach quantifies the weights impartially of different impact indexes (i.e., objectives) to find the optimal location and size of the DG. These weights are quantified based on the degree of importance of an index in the DG placement problem.
4. The proposed approach is being applied to eliminate subjectivity in decision-making for the optimal solution in a multi-objective optimization problem.

The rest of the article is structured as follows; Sect. 2 discusses how the influence of an index is changed with the alteration in the network's physical properties. Section 3 discusses the modeling of voltage-dependent loads and different DN impact indexes. Section 4 delineates the proposed approach for quantifying the weights associated with the objectives in a multi-objective DG placement problem and presents the solution approach to the optimal placement of DG. Simulated results and discussion are reported in Sect. 5, followed by the concluding remarks in Sect. 6.

2 Impact assessment of power loss index in DG placement problem

It is known that the integration of DG significantly influences both technical and economic aspects within a DN. This section demonstrates how variations in DN configurations lead to differing impacts on these aspects. Consequently, their significance in a multi-objective assessment must be adjusted based on their varying impacts in the context of the DG placement problem. To accomplish this, we first derive the mathematical formula for the power loss index in equations (1)–(5) and subsequently apply it to evaluate two distinct configurations of a 7-bus distribution system, as depicted in Fig. 1.

$$P_{\text{loss}} = \sum_{n\in U_l}\left[r_{n,n+1}.\left(I_{n,n+1}\right)^2\right] \tag{1}$$

$$P_{\text{loss}} = \sum_{n\in U_l}\left[r_{n,n+1}.\left(\left(P_{n,n+1}\right)^2+\left(Q_{n,n+1}\right)^2\right)\right] \tag{2}$$

$$\begin{aligned} P_{\text{loss}}^{\text{dg}} = &\sum_{m\in U_l}\left[r_{m,m+1}.\left(\left(P_{m,m+1}-P^{\text{dg}}\right)^2+\left(Q_{m,m+1}\right)^2\right)\right] \\ &+\sum_{i=(n-m)\in U_l}\left[r_{i,i+1}.\left(\left(P_{i,i+1}\right)^2+\left(Q_{i,i+1}\right)^2\right)\right] \end{aligned} \tag{3}$$

$$P_{\text{loss}}^{\text{dg}} = P_{\text{loss}} - \sum_{m\in U_l}\left[r_{m,m+1}.\left(2.P_{m,m+1}.P^{\text{dg}}-\left(P^{\text{dg}}\right)^2\right)\right] \tag{4}$$

$$\begin{aligned} \%P_{\text{loss}}^{\Delta} &= \left|\frac{P_{\text{loss}}^{\text{dg}}}{P_{\text{loss}}}-1\right| \\ &= \frac{\sum_{m\in U_l}\left[r_{m,m+1}.\left(2.P_{m,m+1}.P^{\text{dg}}-\left(P^{\text{dg}}\right)^2\right)\right]}{\sum_{n\in U_l}\left[r_{n,n+1}.\left(\left(P_{n,n+1}\right)^2+\left(Q_{n,n+1}\right)^2\right)\right]} \end{aligned} \tag{5}$$

Figure 2 illustrates the power loss index computed when 100 kW of DG is successively placed at each potential node in each configuration. Additionally, we calculate the standard deviation for each configuration, revealing that DG placement in configuration two exerts a more pronounced impact on power loss compared to configuration one. Consequently, greater emphasis should be placed on the power loss index in formulating a multi-objective problem for configuration two.

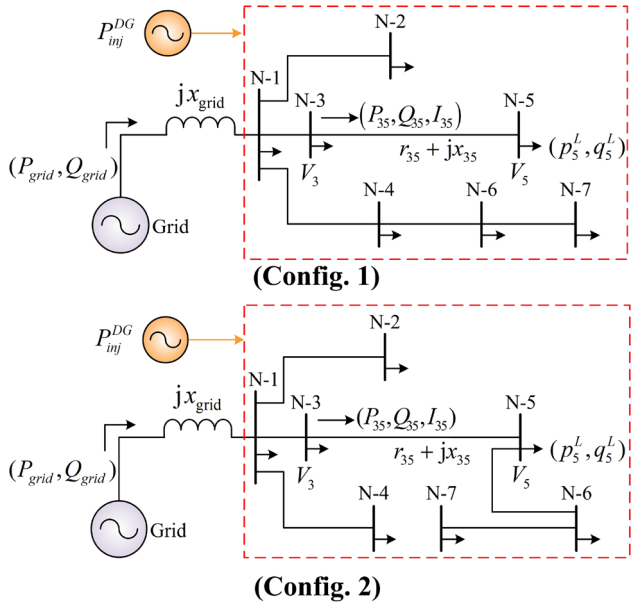


Fig. 1 7-bus distribution system depicting 0.10 MW of DG integration at each potential node in **a** config. 1, **b** config. 2

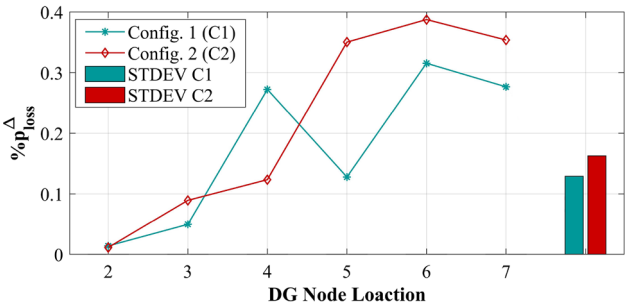


Fig. 2 Evaluated power loss index at each bus after placing 0.10 MW of DG in config. 1 and 2

3 Modeling of loads and MO-DG placement problem formulation

3.1 Mathematical modeling of load demand

The role of DNP is to mathematically analyze the techno-economic impact of DG on the DN before placement, taking into account the varying nature of loads. In the present analysis, voltage-dependent load modeling is considered, categorizing the loads into industrial, commercial, and domestic loads instead of constant power load demands (P_L^{cp} and Q_L^{cp}). Mathematically, the active ($P_{L,i}$) and reactive ($Q_{L,i}$) loads for the ith bus are presented as [31]:

$$P_{L,i} = P_{L,i}^{\text{cp}} . V^{\alpha} \tag{6}$$

$$Q_{L,i} = Q_{L,i}^{\text{cp}} . V^{\beta} \tag{7}$$

In (6) and (7), the α and β are the exponential components for the different types of loads. The values of α are 0.18, 0.92, and 1.51, and the values of β are 6.0, 4.04, and 3.40 for industrial, residential, and commercial loads, respectively. The total active and reactive power demands for a DN are expressed as (8) and (9), respectively.

$$P_{\text{TL}} = \sum_{b \in U_\rho} P_{L,b}^{\text{cp}} . V_b^{0.18} + \sum_{b \in U_\sigma} P_{L,b}^{\text{cp}} . V_b^{0.92} + \sum_{b \in U_\tau} P_{L,b}^{\text{cp}} . V_b^{1.51} \tag{8}$$

$$Q_{\text{TL}} = \sum_{b \in U_\rho} Q_{L,b}^{\text{cp}} . V_b^{6.0} + \sum_{b \in U_\sigma} Q_{L,b}^{\text{cp}} . V_b^{4.04} + \sum_{b \in U_\tau} Q_{L,b}^{\text{cp}} . V_b^{3.40} \tag{9}$$

It is typically expected that when a capacitor is placed in the network, the voltage profile improves, and demand from the substation decreases as losses reduce. However, experimental findings suggest that the demand from the substation increases as the voltage profile improves [32]. Therefore, in the present paper, voltage-dependent load modeling (8) and (9) investigates the effect of load models in DG placement planning.

3.2 Multi-objective DG placement problem

Mathematically, a multi-objective problem for DG placement in DN can be written as:

$$\text{minimize} \sum_{\phi=1}^{N} \xi_\phi . f_\phi (x, u) \tag{10}$$

$$s.t. \quad h(x, u) = 0 \tag{11}$$

$$g(x, u) \geq 0 \tag{12}$$

In (10), N represents the total objectives considered in the DG placement problem, while ξ_ϕ denotes the RI coefficient of the ϕth objectives. In (11)–(12), the equality and inequality constraints encompass power flow balance, node voltage limits, bounds on active and reactive power, as well as other operational and security limits. The proposed work considers both technical and economic objectives for techno-economic performance analysis, subject to constraints on line flow and variable bounds.

3.2.1 Technical objectives

It is a well-known fact that DG placement significantly influences various technical quantities within a DN, namely minimizes real and reactive power losses, improves the node voltage profile, and regulates line capacity. Consequently,

assessing the degree of change in these network parameters during the analysis of DG placement becomes imperative for practical application, considering numerous objectives. In this study, four distinct technical objectives are employed to examine the impact of DG integration within the network. These objectives are mathematically articulated as [33]:

$$\text{Active power loss:} \quad L_P = \sum_{t\in U_t}\sum_{n\in U_l}\left[r_{n,n+1}.\left(I^t_{n,n+1}\right)^2\right] \tag{13}$$

$$\text{Reactive power loss:} \quad L_Q = \sum_{t\in U_t}\sum_{n\in U_l}\left[x_{n,n+1}.\left(I^t_{n,n+1}\right)^2\right] \tag{14}$$

$$\text{Voltage deviation:} \quad V_D = \sum_{t\in U_t}\sum_{b\in U_b}\left(\frac{V^t_b - V^t_{spec.}}{V^t_{\max} - V^t_{\min}}\right)^2 \tag{15}$$

$$\text{Line capacity limit:} \quad I_C = \frac{1}{T}\sum_{t\in U_t}\max\left(\left|\overline{S}^{k,t}_l\right|\right)^{l\in U_l} \tag{16}$$

The central aim of these objectives is to identify potential nodes for DG placement. Consequently, the quantities within the DN can be optimized after DG placement, leading to increased reliability. In the presented work, the impact of DG is analyzed on one phase of a three-phase system, assuming that the DN system is balanced. The uncertainty in the time-varying nature of the load is not considered to reduce complexity without affecting the objective and contribution of the article.

3.2.2 Economic objective

In practice, the economic consideration in the DG placement problem is as important as the inclusion of various technical objectives. Therefore, an economical cost index (E_{Cost}) is devised and included in the proposed methodology, expressed as:

$$E_{cost} = C^{\text{dg}}_{\text{ins}} + \sum_{t\in U_t}\left[C^{\text{dg},t}_{o\&m} + C^t_{\text{sub}}\right] \tag{17}$$

$$\text{where} \quad C^{\text{dg}}_{\text{ins}} = \sum_{i\in U_{\text{dg}}}\left[\left(P^{\text{dg}}_i.\text{UIC}^{\text{dg}}_i.\text{CRF}\right)/365\right] \tag{18}$$

$$\text{and} \quad C^{\text{dg}}_{o\&m} = \sum_{i\in U_{\text{dg}}}\left(P^{\text{dg}}_i.UOC^{\text{dg}}_i\right) \tag{19}$$

$$\text{and} \quad C_{\text{sub}} = UC_{\text{sub}}.\left[eq.(4) + \left(P_{\text{TL}} - P^{\text{dg}}_T\right)\right] \tag{20}$$

$$\text{and} \quad \text{CRF} = \frac{\zeta(1+\zeta)^\gamma}{(1+\zeta)^\gamma - 1} \tag{21}$$

The capital recovery factor (CRF) in (21) represents the total installation cost paid in 'γ' equal installments at the interest rate of $\zeta\%$. However, many existing studies omit the economic aspect in the formulation of the DG placement problem. Yet, the financial operation of the system is the reality of today's competitive electricity market. Therefore, including an economic aspect in the DG placement problem is necessary alongside other technical objectives.

3.2.3 Constraints

Any optimization problem is subject to a set of constraints. Therefore, both equality and inequality constraints are used in this work to limit the variable and refine feasible solutions, as follows:

$$P^{\text{fr}}_{\text{br}} = P^{\text{to}}_{\text{br}} + \left(P^{\text{to}}_L - P^{\text{dg,to}}\right) + \left[r^{\text{fr}}\left(I^{\text{fr}}_{\text{br}}\right)^2\right] \tag{22}$$

$$Q^{\text{fr}}_{\text{br}} = Q^{\text{to}}_{\text{br}} + \left(Q^{\text{to}}_L - Q^{\text{dg,to}}\right) + \left[x^{\text{fr}}\left(I^{\text{fr}}_{\text{br}}\right)^2\right] \tag{23}$$

$$V^{\text{to}}_b = V^{\text{fr}}_b - \left[\left(r^{\text{fr}} + jx^{\text{fr}}\right).\left(I^{\text{fr}}_{\text{br}}\right)^*\right] \tag{24}$$

$$V^{i,\min}_b \le V^i_b \le V^{i,\max}_b \tag{25}$$

$$P^{\text{dg}}_{i,\min} \le P^{\text{dg}}_i \le P^{\text{dg}}_{i,\max} \tag{26}$$

$$\sum_{i\in U_{\text{dg}}} P^{\text{dg}}_i \le P^{\text{dg}}_{\max} \tag{27}$$

where (22)–(24) represent the power flow constraints on real and reactive power, (25) describes the voltage upper and lower limit, and (26)–(27) represent limits on individual and total install capacity of DG.

4 Solution methodology

Lately, numerous methods have been proposed for optimal DG placement focusing on multi-objective optimization. However, an appropriate strategy is not implemented to determine the weight factors associated with each objective for optimal decision-making. Consequently, getting a suboptimal solution to the DG placement problem is possible. The following subsection presents an approach that utilizes the application of Shannon's entropy to quantify the unbiased weight factors for the indexes based on their degree of influence in the multi-objective optimal DG placement problem.

4.1 Shannon's entropy for weight evaluation

Entropy is widely used in information theory to empirically measure the dispersion of a quantity in multi-criteria decision-making [34]. The underlying principle of entropy is that the higher the dispersion in the measured quantity, the greater the disparity of the criteria, resulting in more information being obtained. Moreover, greater weight (importance) should be given to that criterion, and vice versa [35]. Mathematically, the empirical weights are derived using

the principle of Shannon's entropy through the following steps.

1. Reconstruct techno-economic objectives to impact indexes for unitless measurement.

$$O_{kl} = 1 - \left(f_l^k / f_l^{\text{base}}\right) \tag{28}$$

2. Normalize the measured values (i.e., index)

$$\overline{O}_{kl} = \frac{O_{kl}}{\sum\limits_{k=1\to E} O_{kl}} \tag{29}$$

3. Evaluate Shannon's entropy value e_l for lth index, and is defined as

$$e_l = -e_0 \sum_{k=1}^{s} \overline{O}_{kl} \times \ln \overline{O}_{kl} \tag{30}$$

where e_0 is termed as entropy constant and is equal to $(\ln s)^{-1}$. Furthermore, the degree of diversification is obtained by $\delta_l = 1 - e_l$.

4. Compute the degree of importance (i.e., weight)

$$\lambda_l = {}^{\delta_l}/_{\sum_{l=1}^{r} \delta_l} \tag{31}$$

5. The comprehensive score for any kth entity can be obtained as

$$\eta_k = \sum_{k=1}^{s} \lambda_l \overline{O}_{kl} \tag{32}$$

In this paper, λ_l represents the weights in the multi-objective DG placement problem. The optimal DG location and(or) size can be obtained from the η_k; a higher score means higher priority.

4.2 Proposed approach for optimal DG planning

The weighted sum of different technical and economic objectives formulates the weighted multi-objective function for optimal DG placement in the distribution system. The weighted multi-objective index (IMW^k) is written as:

$$\text{IMW}^k = \left\{\lambda_1 \text{ILP}^k + \lambda_2 \text{ILQ}^k + \lambda_3 \text{IDV}^k + \lambda_4 \text{ICC}^k + \lambda_5 \text{ICE}^k\right\} \tag{33}$$

$$s.t. \quad (6) - (9) \quad \text{and} \quad (13) - (32) \tag{34}$$

$$\sum\nolimits_{i=1}^{m} \lambda_i = 1 \ \wedge \ \lambda_i \in [0, 1] \tag{35}$$

In (33), ILP^k represents the actual power loss index, ILQ^k the reactive power loss index, IDV^k the maximum voltage deviation index, ICC^k the line capacity limit index, and ICE^k the economic cost index for the kth distribution network configuration during tth time interval or peak hour of dth day. Generally, the weights define the degree of importance of technical and economic objectives in the DG placement problem. Therefore, this article utilizes the concept of Shannon's entropy to quantify the empirical weight values, as presented in the prior subsection.

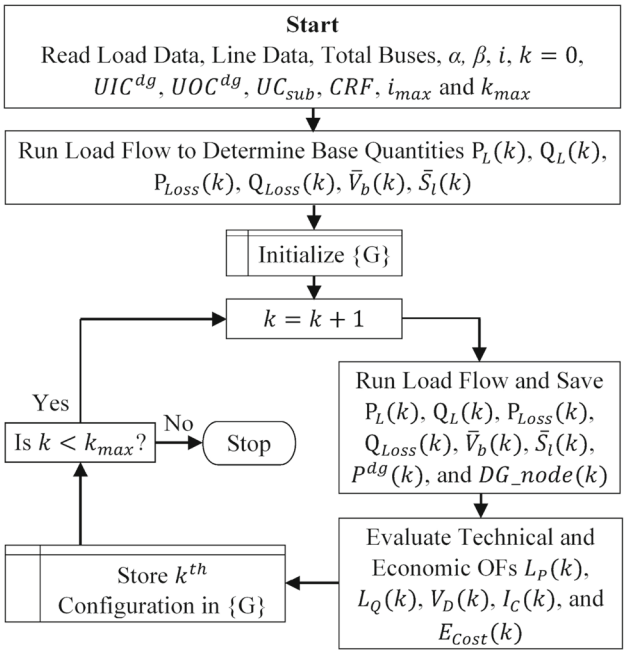


Fig. 3 Workflow to evaluate different techno-economic OFs

The workflow to evaluate different techno-economic objectives is illustrated in Fig. 3. The complete procedure for the proposed solution methodology is provided in the flowchart shown in Fig. 4. The efficacy of the proposed approach is examined on a 38-node test system in a MATLAB environment. The simulated results are presented and discussed in the ensuing section.

5 Results and discussion

The 38-node test system is comprised of the aggregate active and reactive power demand, amounting to 3501.6 kW and 1870.2 kVAr, respectively. Diverse load types, namely industrial, commercial, and residential, are incorporated into the system configuration (refer to Fig. 5). The simulation employs base quantities S_{base} and kV_{base} for analysis, set at 100 MVA and 12.66 kV, respectively. The numerical values for cost parameters used in computing the economic index (ICE) are detailed in Table 1. In the context of mixed loads, the base case real and reactive power losses are recorded as 152.8 kW and 101.5 kVAr, respectively.

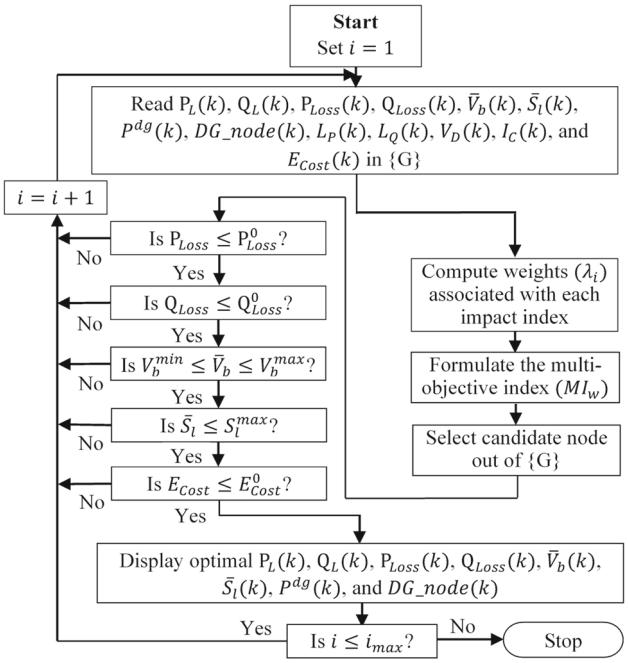


Fig. 4 Workflow for the proposed approach based on Shannon's entropy formula

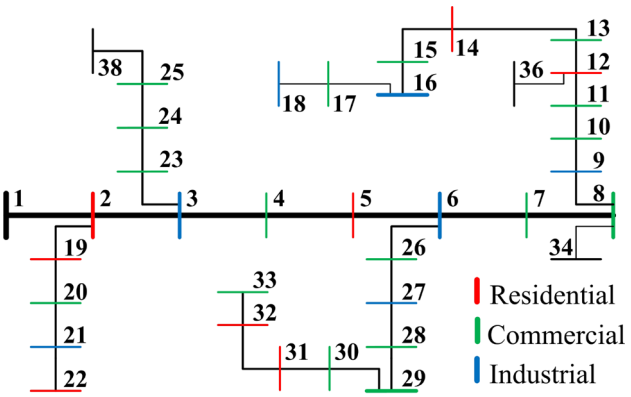


Fig. 5 Single-line diagram of 38-node test system depicting residential, commercial, and industrial loads

Table 1 Numerical values for cost parameters to evaluate ICE index

Parameters	Numerical values
C^{dg}_{ins}	750 $/kVA
$C^{dg}_{o\&m}$	0.04 $/kWh
C_{sub}	0.0656 $/kWh
ζ	10%
γ	18 years

In this article, the weights are evaluated impartially based on the degree of importance of an objective in the DG placement problem. Four different cases are delineated one by one in the following subsections to illustrate the efficacy of the proposed approach, based on the change in physical characteristics of the network. Finally, the proposed method is compared with the arbitrary weight allocation approach considered in past studies.

Table 2 Evaluated weights considering the change in DG location and size in 38-node test system

DG size (kW)	Weight factors				
	λ_1	λ_2	λ_3	λ_4	λ_5
150	0.25	0.27	0.38	0.09	0.01
300	0.22	0.24	0.33	0.20	0.01
600	0.17	0.19	0.24	0.40	0.00

5.1 Case 1: Effect on weights for change in DG location and size

In this case, every node is considered the potential location for DG placement, and three different DG sizes of 150, 300, and 600 kW are placed successively on each candidate node. The evaluated weights (λ_i) are reported in Table 2. The obtained weighted multi-objective index (IMW) score for each kth distribution network configuration is presented in Fig. 6. The following are the main findings of the simulation study.

1. *With the placement of 150 kW DG:* Bus 18 is identified as the optimal location for DG placement, as shown in Fig. 6. More dispersion can be observed for the voltage deviation index (IDV) than other indexes in Fig. 7. Thus, the impact will be high, and the quantified weight value for IDV, i.e., 0.38, is the highest among all indexes. The influence of the economic cost index (ICE) is negligible for this case, i.e., 0.01, as the dispersal is minimal shown in Fig. 7.
2. *With the placement of 300 kW DG:* Bus 17 is identified as the optimal location for DG placement, as shown in Fig. 6 in this case. The weight value is higher for the IDV than the other impact indexes, i.e., 0.33.
3. *With the placement of 600 kW DG:* The optimal node quantified to site the DG is bus 14 (refer to Fig. 6). For this DG size, the constraint of line capacity limit is violated in 12 configurations, as shown in Fig. 8. Therefore, it can be said that the influence of the line capacity limit index (ICC) is the most compared to other indexes. The computed weight value for ICC using SE, i.e., 0.40, is the highest among all indexes.

It is worth mentioning that the weight value associated with the ICC index is 0.09 when a DG of 150 kW is placed in the RDN. However, due to the more instances of line capacity limit violation as the DG size increases, the influence of the

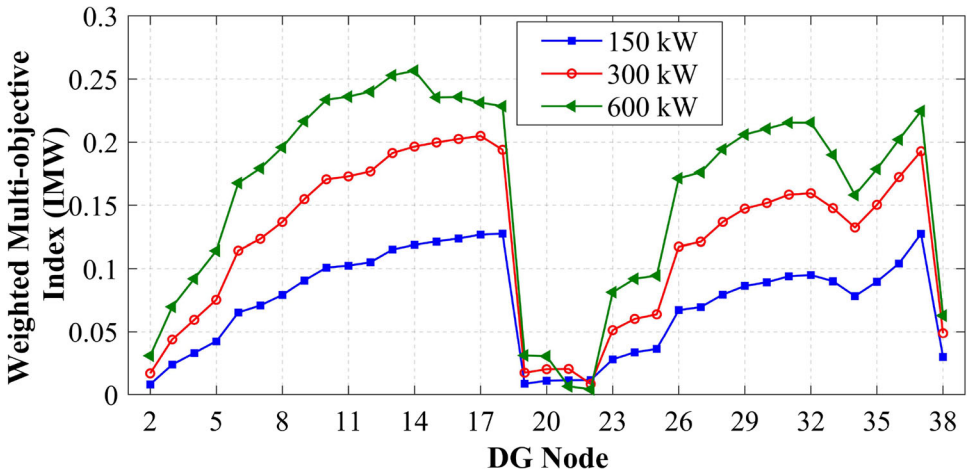


Fig. 6 Weighted multi-objective index for the 38-node test system with different DG sizes sited at each node of the circuit

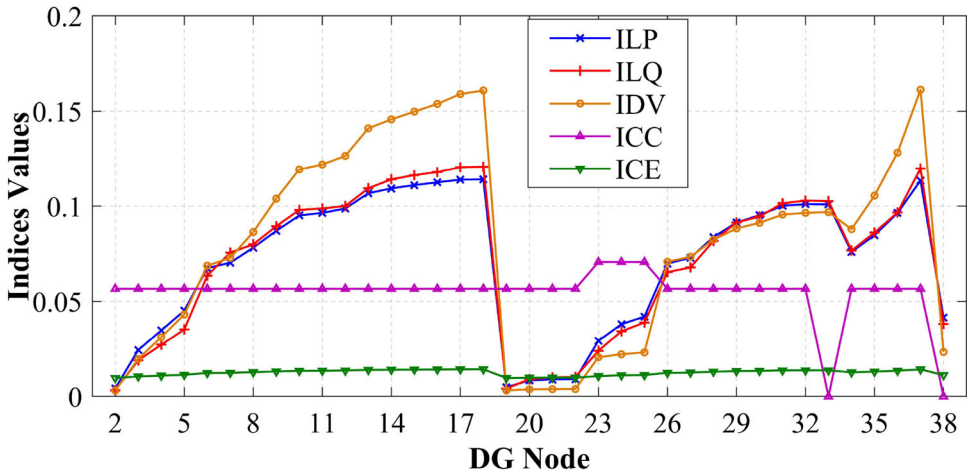


Fig. 7 Techno-economic indexes for 38-node test system with DG of 150 kW sited at each node of the circuit

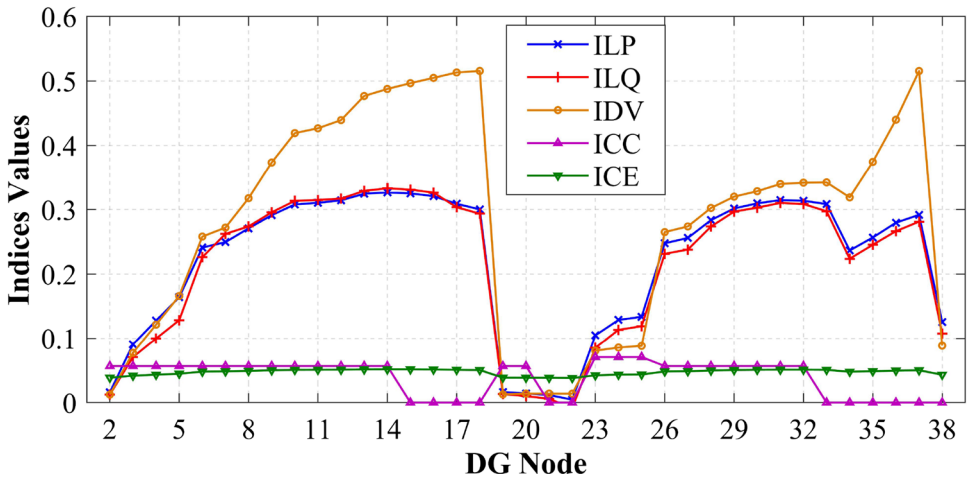


Fig. 8 Techno-economic indexes for 38-node test system with DG of 600 kW sited at each node of the circuit

ICC index increased *viz.*, 0.20 and 0.40, when DG of 300 and 600 kW is sited in the RDN, respectively. After the optimal placement of DG, the improvements in network quantities, e.g., power loss, voltage magnitude, and annual cost expenditure (DG installment cost, O&M cost, and power purchase from the grid), are reported in Appendix {see Table 8}.

5.2 Case 2: Effect on weights for varying the DG Size at 6th bus

The candidate bus location is kept fixed at 6th bus for DG siting, and the problem is formulated by varying the DG size from 0 to 4500 kW in an interval of 5 kW. The computed weight scores for different indexes are reported in Table 3.

Table 3 Evaluated weights when varying DG size at 6th bus in 38-node test system

Bus location	Weight factors				
	λ_1	λ_2	λ_3	λ_4	λ_5
6	0.19	0.20	0.25	0.00	0.36

Table 4 Evaluated weights considering change in DG location and power factor (pf) in 38-node test system

Power factor (pf)	Weight factors				
	λ_1	λ_2	λ_3	λ_4	λ_5
0.80	0.17	0.20	0.25	0.31	0.07
0.90	0.18	0.21	0.26	0.33	0.02
1.0	0.18	0.21	0.26	0.35	0.00

Figure 9 shows that dispersion measure is more for ICE (i.e., $\delta_{l=5} = 0.025$). As the DG size changes, consequently, the influence is higher. Therefore, the weight quantified for ICE, i.e., 0.36, is highest among all indexes. The influence of ICC is minimal as no deviation (i.e., $\delta_{l=4} \approx 0$) is traced during the increment in DG size.

Contrary to Case 1, where the weight associated with ICE possessed negligible influence in selecting optimal DG configuration, in this case, the ICE has a leading role in optimal DG selection. The quantified network quantities for before and after the DG placement are stated in Appendix {see Table 8}.

5.3 Case 3: Effect on weights with different power factor (pf)

In this case, the weight formulation for different impact indexes is analyzed with varying power factors. For this, each node in the RDN is considered a potential node for 500 kW of DG placement, and the simulation is conducted for three different pfs, i.e., 0.80, 0.90, and unity. The evaluated optimal weights (λ_i) are reported in Table 4. At all pf conditions, the impact of ICC is more significant than other indexes, and weight values associated with ICC are almost the same. The following are the main findings of the simulation study.

1. Bus 15 is identified as the potential node to place 500 kW of DG for all considered pf conditions. The quantified weight measures for each index in all pf conditions almost remain the same.
2. Besides ICC, the IDV influences the DG placement problem more than the ILP and ILQ. The ICE has the most negligible impact on the optimal placement of DG.
3. The computed network quantities after the optimal placement of DG are reported in Appendix {see Table 8}. It is concluded from the result that it is wise to operate the system on 0.9 pf condition for optimum performance.

Table 5 Evaluated weights considering the change in type of load buses in 38-node test system

Load type	Weight factors				
	λ_1	λ_2	λ_3	λ_4	λ_5
Constant power	0.08	0.09	0.13	0.70	0.00
Industrial load	0.21	0.24	0.32	0.22	0.01
Residential load	0.22	0.24	0.33	0.20	0.01
Commercial load	0.22	0.24	0.33	0.20	0.01
Mix-type load	0.22	0.24	0.33	0.20	0.01

Figure 10 presents the IMW scores when 500 kW of DG is sited at each potential node when different pf conditions are assumed.

5.4 Case 4: Effect on weights for change in bus load type

It is evident that the change in the type of load buses in a network impacts the DG placement problem. Therefore, the proposed approach is simulated, while the load bus type changes, as discussed briefly in Sect. 3. Each node in the RDN is a potential node for optimal placement of DG of 300 kW. The weight scores quantified under different types of loads are reported in Table 5. The following are the main findings of the simulation study.

1. ICC influences the DG placement problem the most in the case of constant power loads. However, the IDV has more impact than the other indexes in all voltage-dependent load models.
2. When a DG of 300kW is sited successively in the network to identify the optimal node, only 12 configurations are identified where line capacities are within the specified limit.
3. The impact of each index remains almost the same for all voltage-dependent load models, shown in Fig. 11. Therefore, the quantified weight measures for each index are also nearly the same.
4. For the same DG size, bus 18 is found the optimal node in the case of constant power load. On the contrary, bus 17 is the ideal location when voltage-dependent load buses are assumed.

Figure 11 represents the IMW scores when a DG of 300 kW is placed at each potential node considering the different load models. The calculated network quantities after the opti-

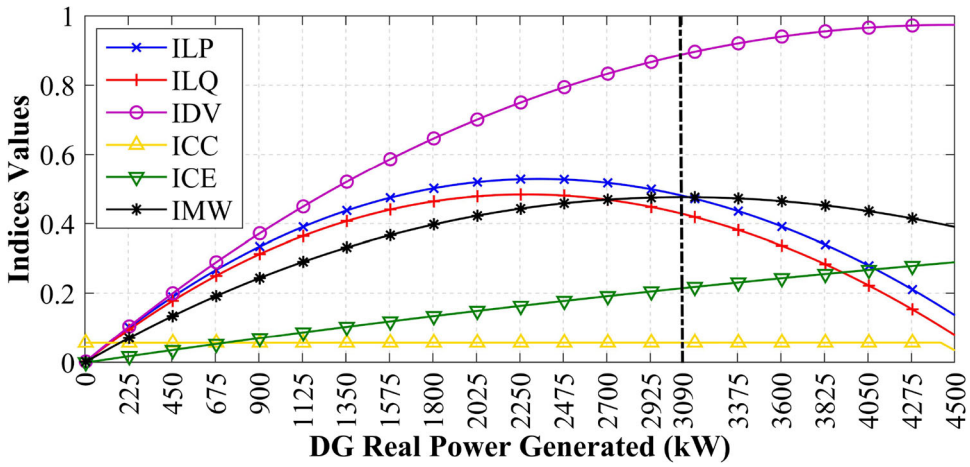


Fig. 9 Techno-economic indexes and multi-objective index for 38-node test system with varying DG placed at 6th bus

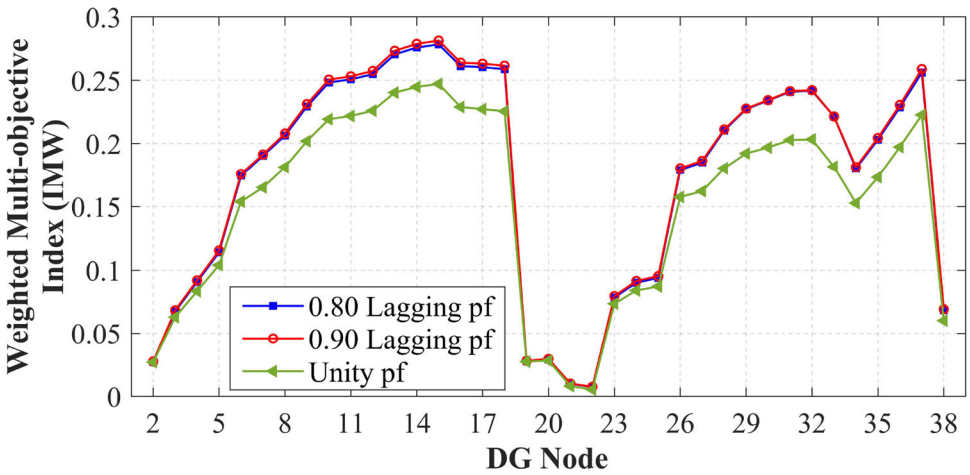


Fig. 10 Multi-objective index for each node considering different power factors in the 38-node test system (DG of 500 kW)

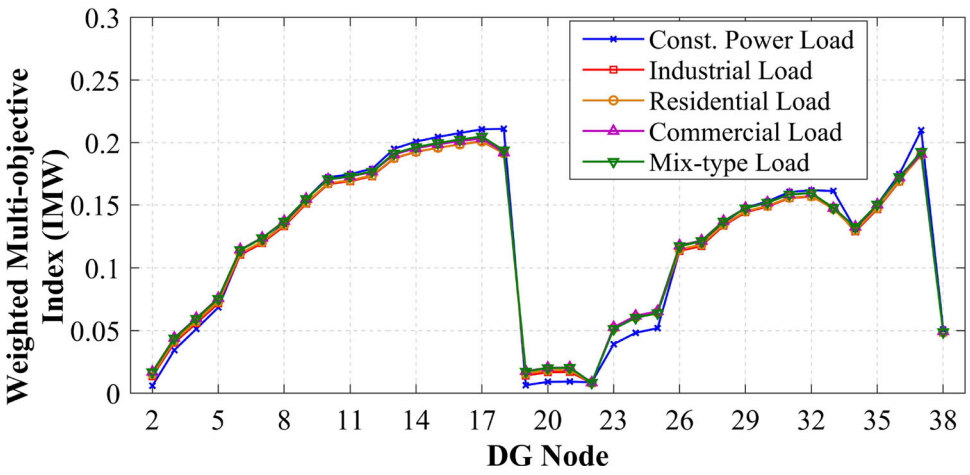


Fig. 11 The multi-objective index for each node considers different load buses in the 38-node test system (DG of 300 kW)

Table 6 Weight measures for the proposed approach and prevailing studies

Method	Weight factors				
	λ_1	λ_2	λ_3	λ_4	λ_5
WDNP [36]	0.35	0.20	0.10	0.20	0.15
AHP [37]	0.15	0.10	0.25	0.04	0.46
TOPSIS [38]	0.20	0.15	0.25	0.10	0.30
Proposed	0.19	0.20	0.25	0.00	0.36

mal placement of DG are reported in Appendix {see Table 9}.

5.5 Case 5: Comparison of the proposed approach to prevailing studies

A comparative study of the proposed method and existing approaches is conducted to highlight the importance of impartial weight evaluation in solving multi-objective DG placement problems. The results assessed in Case 2 are compared between the proposed method and previously published weight allocation methods for decision-making in multi-objective optimal DG planning, specifically: weights allocation based on the discretion of the distribution network planner (WDNP) [36], the analytical hierarchical process (AHP) [37], and the TOPSIS method [38]. Table 6 presents the quantified weight measures for this case, while the evaluated network quantity results are shown in Table 7. Some of the crucial findings from this analysis are discussed below.

1. The network quantities have significantly enhanced through all approaches compared to the base network configuration (i.e., without DG).
2. More improvements in technical quantities are observed using the WDNP technique; however, the proposed approach cut annual expenditure by a total amount of 61,310 USD while operating the system within the standard limits.
3. Decision-making using the AHP and TOPSIS methods reduces annual expenditure by 2.59% and 2.77%, respectively; compared to the proposed approach, it results in significantly higher active and reactive power losses, approximately 10.7%.
4. No dispersion is detected in the measured ICC index values; thus, the impact of the ICC is negligible in the DG placement problem. Therefore, the ICC index should assign a minimum (≈ 0) influence factor. Here, the past methods miss the fact above and allocate weights.

6 Conclusions

In conclusion, this article presents an approach that quantifies weights impartially in a weighted multi-objective problem for optimal DG placement within a distribution network. The efficacy of the proposed method is evaluated by implementing it on a 38-node test system while considering various technical and economic objectives and characterizing DN buses into voltage-dependent loads. Numerical results from four distinct case studies illustrate the adaptability of the proposed method to changes in system physical characteristics by eliminating subjectivity in weight selection. The simulation outcomes show significant reductions in active and reactive power losses, improvements in voltage stability, and notable decreases in total annualized investment costs. Specifically, after optimal DG placement, active power losses are reduced by up to 32.66%, 48.17%, 33.38%, and 22.06% for Cases 1, 2, 3, and 4, respectively. Reactive power losses are reduced by up to 33.30%, 21.58%, 34.19%, and 22.06% for Cases 1, 2, 3, and 4, respectively. Voltage magnitude is notably enhanced, with the minimum voltage magnitude increasing from 0.9252 pu in the base case to 0.9384, 0.9695, 0.9369, and 0.9348 for Cases 1, 2, 3, and 4, respectively. Total annualized investment costs are reduced by up to 5.15%, 42.96%, 4.38%, and 2.67% for Cases 1, 2, 3, and 4, respectively. Furthermore, a comparative study shows that the proposed method reduces annual expenditure by 3.59% compared to the WDNP method while maintaining competitive network performance metrics in power losses and voltage stability similar to those achieved by the AHP and TOPSIS methods. This underscores the effectiveness of our approach in optimizing cost efficiency without compromising critical technical performance metrics essential for distribution network planning.

Table 7 Comparison in network quantities results between the proposed and prevailing methods

Method	DG Size (kW)	P_{Loss} (kW)	Q_{Loss} (kVAr)	V_{min} (p.u.)	AEC ($)
WDNP [36]	2600	72.70	53.10	18/0.9628	1,708,179
AHP [37]	3450	87.9	64.0	18/0.9743	1,604,098
TOPSIS [38]	3475	88.7	64.5	18/0.9747	1,601,197
Proposed	3090	79.20	57.90	18/0.9695	1,646,869

Table 8 38-node test network quantities results obtained for Cases 1, 2, and 3 considered in Sect. 5

Case	Configuration	P_{Loss} (kW)	Q_{Loss} (kVAr)	V_{min} (p.u.)	AEC ($)
Base Configuration		152.80	101.50	18/0.9252	2,100,053
Case 1	150 kW	135.40	89.30	33/0.9316	2,070,118
	300 kW	121.90	80.20	33/0.9339	2,042,458
	600 kW	102.90	67.70	33/0.9384	1,991,708
Case 2	6th bus	79.20	57.90	18/0.9695	1,646,869
Case 3	0.80 pf	103.0	67.70	33/0.9383	2,062,490
	0.90 pf	101.80	66.80	33/0.9383	2,033,066
	1.0 pf	108.20	71.10	33/0.9369	2,008,025

Table 9 38-node test network quantities results obtained for Case 4 considered in Sect. 5

Load type	Configuration	P_{Loss} (kW)	Q_{Loss} (kVAr)	V_{min} (p.u.)	AEC ($)
Constant	Base	202.20	134.8	18/0.9133	2,251,048
Power	300 kW	167.20	110.60	33/0.9216	2,191,100
Industrial	Base	157.60	104.70	18/0.9240	2,205,888
Power	300 kW	124.20	81.60	33/0.9340	2,146,885
Residential	Base	154.20	102.40	18/0.9248	2,126,184
Power	300 kW	123.0	80.90	33/0.9337	2,068,420
Commercial	Base	148.60	98.60	18/0.9264	2,063,673
Power	300 kW	119.10	78.30	33/0.9348	2,006,900
Mix-type	Base	152.80	101.50	18/0.9252	2,100,053
Power	300 kW	121.90	80.20	33/0.9339	2,042,458

Appendix A

Tables 8 and 9 show the simulated network quantities after the optimal allocation of DG in the 38-node test system.

References

1. Cozzi L, Gul T (2021) Net zero by 2050 A roadmap for the global energy sector. International Energy Agency, USA. https://doi.org/10.1787/c8328405-en
2. Gupta Y, Doolla S, Chatterjee K, Pal BC (2021) Optimal DG allocation and Volt–Var dispatch for a droop-based microgrid. IEEE Trans Smart Grid 12(1):169–181
3. Singla A, Singh K, Yadav VK (2021) Optimization of distributed solar photovoltaic power generation in day-ahead electricity market incorporating irradiance uncertainty. J Mod Power Syst Clean Energy 9(3):545–560
4. Das S, Das D, Patra A (2019) Operation of distribution network with optimal placement and sizing of dispatchable DGs and shunt capacitors. Renew Sustain Energy Rev 113:109219
5. Li Z, Su S, Jin X, Chen H, Li Y, Zhang R (2021) A hierarchical scheduling method of active distribution network considering flexible loads in office buildings. Int J Electr Power Energy Syst 131:106768
6. Dranka GG, Ferreira P, Vaz AIF (2021) A review of co-optimization approaches for operational and planning problems in the energy sector. Appl Energy 304:117703
7. Manna D, Goswami SK (2020) Optimum placement of distributed generation considering economics as well as operational issues. Int Trans Electr Energy Syst 30(3):1–20
8. Aman MM, Jasmon GB, Bakar AHA, Mokhlis H (2014) A new approach for optimum simultaneous multi-DG distributed generation Units placement and sizing based on maximization of system loadability using HPSO (hybrid particle swarm optimization) algorithm. Energy 66:202–215
9. Arulraj R, Kumarappan N (2019) Optimal economic-driven planning of multiple DG and capacitor in distribution network considering different compensation coefficients in feeder's failure rate evaluation. Eng Sci Technol Int J 22(1):67–77
10. Memarzadeh G, Keynia F (2020) A new index-based method for optimal DG placement in distribution networks. Eng Rep 2(10):1–15
11. Heidari A et al (2017) Reliability optimization of automated distribution networks with probability customer interruption cost model in the presence of DG units. IEEE Trans Smart Grid 8(1):305–315
12. Ehsan A, Yang Q (2018) Optimal integration and planning of renewable distributed generation in the power distribution networks: a review of analytical techniques. Appl Energy 210:44–59
13. Lone RA, Javed Iqbal S, Anees AS (2023) Hybrid technique for optimal placement and sizing of distributed generation units considering real and reactive power injection. Int J Green Energy. https://doi.org/10.1080/15435075.2023.2288330
14. Almabsout EA, El-Sehiemy RA, An ONU, Bayat O (2020) A hybrid local search-genetic algorithm for simultaneous placement of DG units and shunt capacitors in radial distribution systems. IEEE Access 8:54465–54481
15. Bawazir RO, Cetin NS (2020) Comprehensive overview of optimizing PV-DG allocation in power system and solar energy resource potential assessments. Energy Rep 6:173–208
16. Pesaran HAM, Huy PD, Ramachandaramurthy VK (2017) A review of the optimal allocation of distributed generation: objectives, constraints, methods, and algorithms. Renew Sustain Energy Rev 75:293–312

17. Sultana U, Khairuddin AB, Aman MM, Mokhtar AS, Zareen N (2016) A review of optimum DG placement based on minimization of power losses and voltage stability enhancement of distribution system. Renew Sustain Energy Rev 63:363–378
18. Hung DQ, Mithulananthan N (2013) Multiple distributed generator placement in primary distribution networks for loss reduction. IEEE Trans Ind Electron 60(4):1700–1708
19. Akbar MI, Kazmi SAA, Alrumayh O, Khan ZA, Altamimi A, Malik MM (2022) A novel hybrid optimization-based algorithm for the single and multi-objective achievement with optimal DG allocations in distribution networks. IEEE Access 10:25669–25687
20. Hari Prasad C, Subbaramaiah K, Sujatha P (2023) Optimal DG unit placement in distribution networks by multi-objective whale optimization algorithm and its techno-economic analysis. Electric Power Syst Res. https://doi.org/10.1016/j.epsr.2022.108869
21. Elseify MA, Hashim FA, Hussien AG, Kamel S (2024) Single and multi-objectives based on an improved golden jackal optimization algorithm for simultaneous integration of multiple capacitors and multi-type DGs in distribution systems. Appl Energy. https://doi.org/10.1016/j.apenergy.2023.122054
22. Abbas G, Wu Z, Ali A (2024) Multi-objective multi-period optimal site and size of distributed generation along with network reconfiguration. IET Renew Power Gener. https://doi.org/10.1049/rpg2.12949
23. Mouwafi MT, El-Sehiemy RA, Abou El-Ela AA (2022) A two-stage method for optimal placement of distributed generation units and capacitors in distribution systems. Appl Energy. https://doi.org/10.1016/j.apenergy.2021.118188
24. Elseify MA, Kamel S, Nasrat L, Jurado F (2023) Multi-objective optimal allocation of multiple capacitors and distributed generators considering different load models using Lichtenberg and thermal exchange optimization techniques. Neural Comput Appl 35:11867–11899
25. Sunil A, Venkaiah C (2024) Multi-objective adaptive fuzzy campus placement based optimization algorithm for optimal integration of ders and DSTATCOMs. J Energy Storage. https://doi.org/10.1016/j.est.2023.109682
26. Anbuchandran S, Kalarathi M, Suresh S, Bhuvanesh A (2024) Multi-objective optimal distributed generators integration using firefly algorithm with fuzzy decision making. Electr Eng. https://doi.org/10.1007/s00202-024-02348-1
27. Kashyap M, Kansal S, Verma R (2022) Sizing and allocation of DGs in a passive distribution network under various loading scenarios. Electr Power Syst Res. https://doi.org/10.1016/j.epsr.2022.108046
28. Mewafy AA, Kaddah SS, Eladawy M (2023) Multi-objective optimal DG placement approach using analytical hierarchy process. Electr Power Comp Syst 51(15):1648-1663
29. Ali A, Shaaban MF, Awad ASA et al (2023) Multi-objective allocation of EV charging stations and RESs in distribution systems considering advanced control schemes. IEEE Trans Veh Technol 72(3):3146–3160
30. Lone RA, Javed Iqbal S, Anees AS (2023) Optimal location and sizing of distributed generation for distribution systems: an improved analytical technique. Int J Green Energy 21(3):682–700
31. Arif A, Wang Z, Wang J, Mather B, Bashualdo H, Zhao D (2018) Load modeling—a review. IEEE Trans Smart Grid 9(6):5986–5999
32. Rizy DT, Lawler JS, Patton JB, Nelson WR (1989) Measuring and analyzing the impact of voltage and capacitor control with high-speed data acquisition (distribution system). IEEE Trans Power Deliv 4(1):704–714
33. Ochoa LF, Padilha-Feltrin A, Harrison GP (2006) Evaluating distributed generation impacts with a multiobjective index. IEEE Trans Power Deliv 21(3):1452–1458
34. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(4):623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x
35. Gupta S, Yadav VK, Singh M (2021) Optimal allocation of capacitors in radial distribution networks using Shannons entropy. IEEE Trans Power Deliv. https://doi.org/10.1109/TPWRD.2021.3107857
36. Singh D, Singh D, Verma KS (2009) Multiobjective optimization for DG planning with load models. IEEE Trans Power Syst 24(1):427–436
37. Tanwar SS, Khatod DK (2017) Techno-economic and environmental approach for optimal placement and sizing of renewable DGs in distribution system. Energy 127:52–67
38. Galgali VS, Ramachandran M, Vaidya GA (2020) Multi-objective optimal placement and sizing of DGs by hybrid fuzzy TOPSIS and Taguchi desirability function analysis approach. Electr Power Comp Syst 48(19/20):2144–2155

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.

Demand Study and Impact of Electric Vehicle on grid system in Delhi

Chhavi Sehrawat, *Shubh* Mittal, *Krish* Shokeen, *Geeta* Singh

Department of Environmental Engineering, Delhi Technological University, Delhi, India

Abstract. Given the increasing influence of urban areas and worries about the environment, electric automobiles are being incorporated into the power grids of megacities. In this regard, this examine what happens to Delhi's grid when it adopts EVs which is one among many transportation-energy-challenged cities worldwide. The aim of this research is to show where and when people charge their electric vehicles (EVs) by means of studying overall demand patterns as well as impacts on electricity systems. It also looks at required infrastructure improvements including those related with peak loads during charging hours among others things such like stabilizing measures taken for ensuring such stability against any disturbances caused due increased utilization rates within these periods between different regions across Delhi city area served mainly by DUSIB substation(s). Trying our best during estimation process not only forecast activities but also analyze them separately considering each type according its purpose intended use or level expected intensity etc.; all these should be done while taking into account possible effects associated with renewable energy sources integration aiming towards sustainable development goals achievement in terms environmental friendliness.

Keywords—Electric Vehicles, Grid Integration, Charging Infrastructures, Environmental Impact, Megacities

1. INTRODUCTION

Globally, the market value of electric vehicles (EVs) is steadily rising, mostly in nations that have adopted tax ex- emptions, special parking zones, toll-free driving, etc. As a result, in the upcoming years, both the use of EVs and their impact on the grid will rise.

EVs are expanding quickly in the market due to environmental issues around internal combustion engine (ICE)- powered cars. EVs play a significant role in reducing air pollution and emission of greenhouse gases[1][2] . EVs are nevertheless becoming more and more popular despite their drawbacks, which include a short driving range, expensive batteries, a short battery life cycle, a lengthy charging period, and a lack of electricity to power them. Analysing the effects of charging station loads on the power grid network is therefore crucial. Random EV charging may have the following negative effects:

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

a decrease in voltage, an increase in unbalancing, an overloading of power system components such as transformers or cables, an increase in power losses, reduced reserve margin[3]. , an increase in harmonic distortion[4], and deterioration of voltage[5][6]and current transients under problematic circumstances, power quality issues[7]. However, putting the V2G idea into practice physically can improve the power grid's performance[8]. Electricity demand by EVs in 2030 is shown in "Fig. 1".

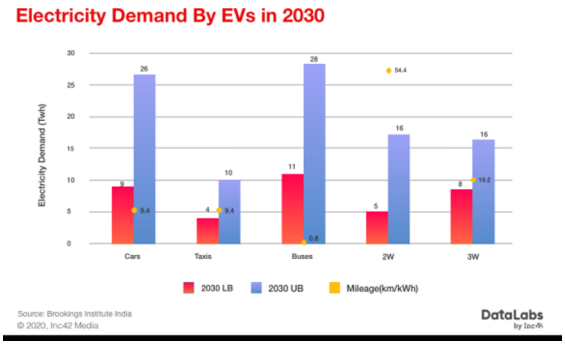


Fig. 1. Electricity demand by EVs in 2030.

This work's primary objective is to give a more detailed and comprehensive knowledge of the issue by reviewing current research on the effects of EV charging stations on the power grid network. Even though reports on related research highlight some key points, the paper's primary contributions might be focused as follows- An overview of the infrastructure for charging electric vehicles is provided by this paper.

The detailed impacts caused by electric vehicle charging stations on the grid, including voltage instability, power quality, and transformer loss of life.

Reviewing the several types of waste that are generated by EVs and identifying the most common hazardous waste kind that is detected. In addition, the study discusses potential future works that might be inferred in order to fully use the resources and offers some flexible methods to offset the challenges addressed.

Moreover, the research implies more work can be done in the future to make most of available resources and gives some adaptable ways of compensating for challenges. These may include among others; advanced charging technologies should be looked into, smart grids solutions implemented and renewable energy sources integrated with electric vehicle charging infrastructure developed while minimizing its effect on environment and grid.

Also, study about:

A. Explore Technological Solutions for Grid Integration:

Apart from identifying challenges, this study explores technological solutions that can address them. It does not only stop at designating but it also looks for methods of lowering negative impacts. This includes looking at new advancements in smart grid technology; demand response systems could be used too together with energy storage facilities so that power networks can withstand higher penetration levels brought about by more EVs.

B. Assess Economic Implications:

Besides technicalities involved here, economic aspects related to deployment and operation of electric vehicle charging stations are taken into account in this paper as well.

Different cost effective charging technologies will be analyzed along with their revenue generation potentials through service provision besides sustainable financial incentives which could promote wider adoption while integrating such onto grids.

C. Evaluate Regulatory and Policy Frameworks:

This research examines the regulatory and policy frame- works that control establishment and operation of EV charging systems. The study seeks to identify current rules on grid connectivity, positioning of charging stations, tariff regimes and rewards for adopting electric vehicles (EVs). Besides, it suggests modifications to such policies where there are gaps or inefficiencies with an aim of integrating EV into power grids seamlessly.

D. Focus On Environmental Sustainability:

Apart from cutting down on carbon emissions through adoption of electric vehicles, this article also looks at wider concerns about sustainability in relation to charging infrastructure for such cars. This includes looking into environmental effects caused by manufacturing batteries used in these cars, constructing charging stations as well as disposing them when they become obsolete; all while providing methods that can be used in reducing these effects like recycling old units back into new ones.

Overall, the document attempts to make a complete examination of the diverse problems and possibilities connected with EV charging station establishment and its assimilation into the electricity network. This investigation sets out to advise decisions making grounded on facts and foster strong transportation systems by tackling technical, economic, regulatory as well as ecological concerns.

2 Literature Review

The global move to electric vehicles (EVs) gives us a chance to reduce carbon emissions at large and small scales, as well as make progress in integrating power grids with the new technology. Some predictions indicate that EVs will represent about 25 percent of all cars on the road by 2040, thereby causing massive drops in city carbon outputs. Realizing what EVs can do, India has set its sights high under the National Electric Mobility Mission Plan by targeting an installation of six-to-seven million hybrid and battery electric vehicles by 2020.

This change in direction towards making more E.V.s common throughout India requires people to have a better understanding about how these practices work worldwide and their impact on electrical infrastructures at different levels – especially distribution systems which connect last mile consumers with power companies or other service providers. In this regard, it is necessary for us look at global trends regarding deployment of electric vehicle charging stations vis- a`-vis their effects on local electricity networks [3].

To generate effective strategies for the adoption of battery electric vehicles in India, the main areas to be considered are rewards, systems, charging stations and pilot schemes. This research seeks to provide recommendations that are relevant to the unique situations of the country's new EV market through a review of global best practices.

Though there is a wealth of knowledge on electric vehicle (EV) deployment and grid integration from various sources, few studies have been done on Delhi's distribution grids. The impact analysis of BEV integration into power supply conducted by this study was limited in scope and done as a preliminary case study with future investigations expected. Such further works should be aimed at accelerating adoption rates of electric vehicles among consumers while also fostering sustainable transport ecosystems through policy orientation shifts which advocate for integrated electric grid infrastructures in Indian

cities like Delhi where much attention is needed currently around these areas.

3 Methodology

A combination of primary and secondary data sources is used to conduct a comprehensive study on the "Impact of Electric Vehicles on Grid System in Delhi" and to examine the various factors that contribute to waste generation. On this note, secondary sources like published articles, essays, reports and personal experiences will be deployed to build on existing studies and shed light on different perspectives of EV-grid interactions and waste generation.

Moreover, Monte Carlo Simulation is conducted with secondary data analysis to predict future outcomes and deter- mine the strength of findings. This is an advanced statistical technique that makes it possible for us to model complex systems by generating random variables according to a given probability distribution function. Precisely when studying the effect of electric vehicles on the electrical power infrastructure in Delhi; Monte Carlo simulation helps us scrutinize multiple scenarios involving their associated probabilities.

Several guiding principles have been considered in modeling the simulation process so as to ensure its accuracy and relevance. To start off, it assumes that there are 1741 charging stations spread all over Delhi where electric vehicle (EV) owners could park their cars and start charging them immediately. This assumption is important because it helps answer the question of whether or not our grid can bear more EVs being plugged in without getting clogged up or causing any other problems with reliability.

Additionally, the model assumes that in the city itself, there is an adequate amount of generation capability within the existing electricity system to absorb all the additional loads that will result into number of electric cars rising. This consideration calls for factors such as MW (megawatts) as a measure of rating for current types of power plants, how many they are and spaces available at every location etc., also considering how much more power has to be generated so that more people can purchase EVs without compromising stability and security across entire Delhi networks.

Afterwards, when simulating charging power per hour, it approximates by rounding up to the next whole number. By using this approach, it is easier to model while still getting a fair estimate on how real charging should be done.

The research is expected to cover all aspects concerning electric cars' effects on Delhi's power grid using secondary data analysis coupled with Monte Carlo Simulation based on these stated assumptions; it will also consider waste production resulting from such impacts. Such findings would then serve as useful information for making policies as well planning infrastructures towards sustainable development that can foster successful incorporation of electric vehicles into transport systems within Delhi Metropolitan Region while promoting efficiency along with effectiveness in utilisation this technology.

4 Impact Of Electric Vehicles On Grid

The use of electric vehicles has been widely praised for reducing CO2 emissions and its benefits to the environment. However, it's important to note that electric vehicles impact our power grid by causing disruptions in voltage stability, power quality, peak load, and transformers' loss of life. The increasing number of electric vehicles that are now connected to the power grid has resulted in a number of new issues and complications, the most difficult of which are the security and economics of the present power system.

The first and most important of these is the issue of increased power loads caused by the electric vehicles' connection to the grid during peak hours, which will lead to additional difference between supply and demand and therefore cause more deadweight on the central grid. The charging schedule of the electric vehicles and the individual behaviour of their owners fluctuate greatly. This fluctuation creates an effect on the electrical grid's general functionality and makes it harder to maintain stability and control over it. Meanwhile, electricity quality is also impacted by the charging of electric cars. The "Fig. 2" below depicts the different factors that are affected by the EV charging.

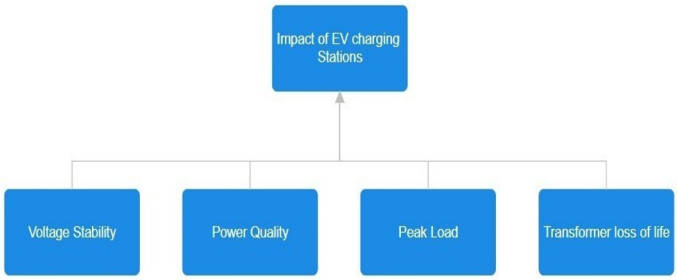


Fig. 2. Impact of EV on the Grid System

A. Voltage stability

Voltage stability refers to a power system's ability to sustain a fixed tolerable voltage at every single bus of the network under standard operating conditions, as well as after being subjected to a disruption.[9][10] Though their degree of penetration and matching connection places are yet unknown, it is anticipated that the extent of EV connectivity in residential regions of distribution networks will increase in the future. Due to single-phase loads in residential areas, EV chargers will be installed in accordance with that. This may result in a discrepancy in the three-phase distribution system's voltage. The reference to[11]came to the conclusion that using EV charging load degrades the stability of the distribution network. Regarding a view towards[12] , the author demonstrated how an EV charging load causes voltage fluctuations in the distribution system. Additionally, they displayed a competitive analysis of the effects of various pricing models (such as tariff-based and disorganised).

B. Power quality

Power quality is an important factor for power distribution networks, which rely on a stable and disturbance-free output within the voltage and frequency tolerance. Users of EVs that charge their vehicles erratically will negatively impact the system in a number of ways, including increased distortion, higher fault current, power quality degradation, line damage, downturns in distribution transformers, and overvoltage in the network of distribution transformers [13].

C. Loss of Transformers Life

Transformers play a significant role in power supply in the electric vehicle charging system. For the purpose of recharging the EVs, AC electricity needs to be transformed to DC power. High frequency converters and introduce harmonics into the grid, which lowers the grid's power quality. These harmonics can cause overloading issues in distribution transformers, which will shorten the transformer's lifespan. [14]

D. Peak load

Peak load is the highest amount of energy that is consumed or drawn from the grid in a set

period of time. The grid system is designed to sustain a certain load and peak load to meet standard operating conditions. McCarthy et al. investigated the impact of PHEV load on Australia's metropolitan distribution network in [15]. They came to the conclusion that 43% peak load shifting was necessary when there was uncoordinated charging and 100% PEV penetration. Putrus et al. examined the impact of unregulated electric vehicle charging on the daily load profile in [16]. They also demonstrated how adding synchronised charging improved the load profile. This can lead to an overload of the transformers, which affects the lifetime and efficiency of the power grid system. While the studies support that EV are promoted as a means of sustainable development, they also come with their own disadvantages. Impact on the power grid system by disrupting voltage and frequency, and reducing the lifetime of transformers can not be ignored.

This figure shows the percentage of studies that examine how EV charging stations affect peak load, power quality, voltage stability, and transformer performance: 34%, 34%, 15%, and 19%, respectively [14]. Even though there aren't many studies that examine how EV charging stations affect peak demand, it's been noted that more and more scholars have been drawn to this paradigm of study in recent years.

5 Factors Affecting Grid Demand

The integration of electric vehicles (EVs) in the grid brings in some factors that can have a significant effect to the power grid system. Understanding these parameters is crucial for efficient grid management as well as ensuring reliable and efficient operation of electrical infrastructure. These are the main factors shaping the power grid system regarding the widespread adoption of EVs:

Infrastructure and Charging Patterns–

- It is important to consider both distribution and accessibility of charging infrastructure. This may cause localised grid congestion when there are concentrated charges at specific places or times.
- Different patterns like peak loads during certain hours can overload the grid, hence, changes must be made to accommodate electricity usage differences.

Vehicle-to-Grid Technology (V2G)-

- V2G technology enables bi-directional electricity flow between EVs and the grid. It requires standardisation and regulatory support for large-scale adoption even if it provides avenues for grid support and energy storage.
- V2G capabilities can improve grid stability through such functions as load balancing, frequency regulation, and peak shaving.

Forecasting and Managing Loads-

- The unpredictability of EV charging practices makes accurate load forecasting more challenging. The grids operators need to update their forecast algorithms considering how unpredictable and dynamic Electric Vehicle charging is.
- By adopting demand response programs combined with smart grid technologies, grid load management will be made more efficient.

6 Waste Generation Due To Evs

Lithium-ion batteries (LIBs) have been regarded as a potential energy storage technology and widely used in EVs due to their portable high capacity for energy, and expanded cycle life (Lu et al. 2013). According to Williams and Lipman (2010), LIBs are generally deemed to have reached the end of their useful life in EV applications when they lose more than

20% of their initial capacity.
The average life of a LIB can range from five to ten years. (Dinger et al. 2010). As LIBs reach the end of their useful lives, a significant amount of trash EV LIBs are finding their way into the waste stream. Therefore LIBs are considered the most hazardous and major waste that is produced by the use of EVs. It is estimated that around 500,000 tons of LIBs waste was produced in 2020. This showcases an environmental concern that needs to be addressed.
Fundamental harm caused by LIB is to the environment and organisms. LIBs have the potential to overheat and catch fire when they are subjected to extremely high temperatures or other extreme conditions, which can be hazardous to human health and the sustainability of the environment.
More than a hundred chemicals are released during an electric vehicle fire, some of which emit highly toxic gases that are fatal to both people and the environment. These gases include hydrogen cyanide and carbon monoxide. Furthermore, the ecology will be impacted and worldwide air pollution will result from the transportation of electric vehicles. As an illustration, consider eutrophication and acidification, photo- chemical air pollution, or harmful consequences on human health including cytotoxicity, genetic material damage, oxidative stress, or allergens (PM10, SO2, and NOx). Waste batteries will also pollute water, impede the development and breeding of aquatic life, and present additional risks. For instance, Pimephales Promelas, Ceriodaphnia, Eliodaphia, and Elmia Clavaeformis are among the aquatic species whose development and development are significantly inhibited by modest concentrations of lithium [17]. Furthermore, plants will collect lithium in their water, which will harm their growth and development [18][19].

7 MANAGEMENT TECHNIQUES FOR WASTE GENERATED

Effective management techniques for the waste generated from electric vehicle (EV) adoption are crucial for mitigating environmental impacts and ensuring sustainable operation of the transportation ecosystem. Several strategies can be implemented to address this challenge, including:

- Recycling is highly underutilised. For the goal of ma- terial re-manufacturing, it employs chemical extraction techniques to remove expensive electrode materials like cobalt and lithium. In most cases, batteries get phased in when their performance falls to half or less of what it was initially. Street lighting, environmental protection, and energy storage systems can all benefit from the utilisation of these repurposed batteries. In order to maximise energy efficiency, achieve complete battery life utilisation, and limitlessly increase the secondary use potential of waste power batteries,
- Approximately 70 % of the vehicle's capacity remains functional after new energy vehicle power batteries are decommissioned, even though they no longer meet the performance standards of new energy sources. This is known as cascade utilisation and is applicable to applica-tions like energy storage, photovoltaic power generation, low- and high-speed electric vehicles, etc. Ladder utilisation is a kind of light scrapping that may be used to low-speed electric cars and energy storage devices, such the abandoned lithium iron phosphate battery in communication base station fields.
- Utilising ultrasonic recovery, active materials from composite electrodes may be quickly and easily removed using high-power ultrasound in a continuous flow process, allowing the electrode to break down in a matter of seconds. Ultrasonic treatment with low power (¡1000 W) has been applied for chemical process acceleration, cleaning, and mixing. A low-power ultrasonic vibration can be employed in lithium battery recycling to help with the electrode stacking process.

8 Waveform Distortion due to EV Charging

The harmonic emission of an EV charger is influenced by factors such as charging circuit design, power level, supply voltage distortion, and network impedance. Comparatively, newer chargers emit more harmonics than older models. The charging cycle comprises two stages: constant current (CC) and constant voltage (CV), with CC being dominant up to 80% state of charge (SOC) followed by CV until fully charged. Each stage exhibits distinct harmonic characteristics, with lower distortion during CC and increased distortion during CV. Variations in total harmonic distortion (THD) and individual harmonic currents among 18 different EVs have been observed, influenced by factors like trickle charger combination and battery SOC diversity. Fast chargers, though fixed on the grid, still face challenges in harmonizing due to EV and SOC diversity despite many stations using the same charger brand and model.

Deterministic and stochastic are the two main categories for conducting EV harmonic research. Charger modeling in time domain and frequency domain; termination effect, as well as background; study the effect of voltage on harmonic emission; Evaluating EV admittance in terms of harmonics is a popular area where we talk more about the deterministic approach. On the other hand, stochastic methods are used to estimate the diversity of electric vehicles or state costs (SOC), location in the grid, etc. introducing various uncertainties into the model, so that the results are closer to reality.

Alternatively, EV harmonic research can be divided into deterministic and stochastic research. Modeling the effect of cancellation in the charger in the time domain or frequency domain, the study of background voltage harmonic emissions, cancellation effects due to various combinations of HEE (harmonic emission estimation) - all these topics are widely discussed and have a decisive impact on this. issues. From another point of view, it offers more realistic results, because different types of electric vehicles (EVs) themselves, such as various, charge level conditions; The complexity of the SOC, and the different variations that exist at different points in the network, will add to the complexity needed to use stochastic methods.

Proper control of harmonic emissions from electric vehicle (EV) charging is a challenge due to charging circuit topology, power levels, background voltage distortion, and grid impedance. the new charger is less harmonious than the old device; however, harmonics are difficult to control when different EVs go through different charging stages. In addition, the concentration of slow chargers and the introduction of uncertainty by fast chargers further complicates the problem.

9 Comprehensive Strategies for EV Harmonics

To tackle harmonic emissions from EV charging, there must be a holistic plan that takes into account different aspects of harmonics' generation and transmission such as those related to charging stations (CSs). Below are several methods on how to reduce harmonics:

A. Charger Circuit Topology and Design

- Harmonic emissions can be greatly decreased by substituting older models of chargers with new ones that have better circuit topologies and power electronics. Filtering techniques and advanced control algorithms are usually used by modern chargers to limit harmonic distortion.
- One way to lower harmonics is by using multi-level converter topologies like cascaded H-bridge or modular multilevel converters which synthesize sinusoidal output voltages at low switching frequencies.
- Chargers should be designed with active power factor correction (PFC) circuits to ensure

good power quality through reducing reactive power consumption as well as harmonic distortion.

B. Smart Charging Technologies

- Algorithms and communication protocols are used by intelligent charging solutions to optimize charging sched- ules with respect to grid conditions, energy prices and user preferences.
- During peak periods, demand response strategies allow utilities to lower harmonic emissions that come with high load conditions by flexibly adjusting charging rates as well as schedules so as to relieve grid congestion.
- Integration of vehicle-to-grid (V2G) facilitates bidirectional power flow between EVs and the grid thereby enabling EVs serve as distributed energy resources while participating in grid frequency regulation which can help reduce harmonic distortions.

10 Results and Discussion

To evaluate the possibility of grid overload under diverse conditions, a Monte Carlo simulation was performed. The simulation considered electric vehicles which were about ten thousand in number that each charged at 0.2 kW over an eight- hour period and the total grid capacity was 100000 kWhs. It was carried out one thousand times. A chart (Fig. 3) is used to present the results whereby electricity demand per vehicle is plotted on x-axis while the y-axis shows total demand. This graphic representation helps in understanding how these variables interact within simulations for different scenarios, thereby giving us insights into possible grid overloads.

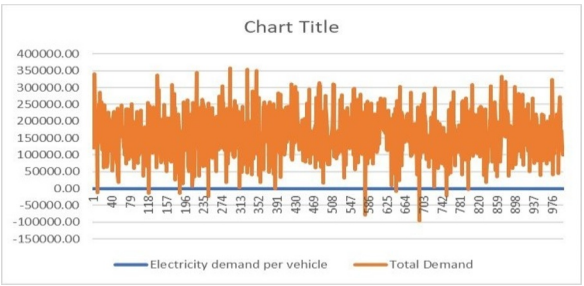


Fig. 3. Monte carlo simulation to represent analysis of grid capacity on different number of EVs

A few important discoveries were made while researching how electric vehicles (EVs) affect Delhi's grid system; this shed light on many factors that impact stability of the grid, power quality, and waste production.

Initially, the study determined voltage stability as a major problem when incorporating EVs into an electrical network. The research showed that voltage imbalances in three-phase distribution systems deteriorated the stability of the distribution networks due to an increase in residential-based EV penetrations. Furthermore, charging loads for electric cars were seen to cause variations in voltages which in turn affected power quality by increasing distortion levels leading to higher fault currents as well as overvoltage at distribution transformers.

According to Monte Carlo Analysis, on running 1000 simulations, this got a result of 838 times or 84 percent grid is overloaded and162 times or 16 percent Not Grid

Overloaded which is shown in Fig. 4, So that have to upgrade the grid system immediately to avoid failure in a grid system. Moreover, this research also discovered the destructive impacts of EV charging on transformers' life expectancy. The study showed that distribution transformers were overloaded because of harmonics brought about by electric vehicle chargers especially at the constant voltage (CV) phase of the charge cycle which led to their shorter lifespan.



Fig. 4. Monte carlo simulation analysis for grid capacity

The other thing which was addressed in this study is about peak load management in relation to EV charging. It was found out that if plug-in hybrid electric vehicles (PHEVs) and battery electric vehicles (BEVs) are charged without coordination during peak hours, then strategies for shifting peak loads should be employed so as not to overload grids. Among the identified methods were such things like synchronized charging which can help improve load profiles as well as reducing overloading of transformers.

Additionally, this research studied the garbage produced as a result of EV adoption with a particular emphasis on lithium- ion batteries (LIBs). The researchers discovered that when these batteries reach the end of their useful life in electric vehicles, they significantly contribute to waste generation which poses environmental and health risks due to their potential for overheating and emitting poisonous gases.

Consequently, the author suggested comprehensive methods for dealing with EV-related waste. These approaches involved recycling; cascade use of LIBs; and ultrasonic retrieval method which is aimed at maximizing secondary utilization potentials while minimizing environmental impacts caused by discarded cells.

Moreover, this study looked into waveform distortion from charging electric cars and identified charger circuit topology; power level; background voltage distortion; network impedance as major factors affecting harmonic emissions. To reduce harmonics it was recommended using modern charger designs with advanced filtering techniques coupled with active power factor correction also adoption smart charging technologies & vehicle-to-grid integration (V2G).

In general, the research results highlighted several different difficulties and prospects connected with merging electric cars into Delhi's grid system. These findings are useful for policy makers, power grid operators as well as stakeholders who can use them to come up with workable methods of ensuring sustainable waste management in an age where people move around using electricity and also maintaining constant power supply coupled with clean energy.

11 Conclusion

The aim of the study was to create an understanding about how electromobility may affect the power grid and identify areas where knowledge is lacking. Different aspects of service and power quality at the MV and LV networks were analysed from the TSO, DSO, utility and customer perspectives. These findings were then used as a basis for describing

their impli- cations and difficulties in various sections.

Waveform distortions including harmonics, interharmonics and supraharmonics have been observed during EV charging. This has been confirmed by measurements taken during this study. An investigation was carried out into how they aggregate and propagate in a low and medium voltage network with multiple EV integration. At first we looked at the PDFs of dominant lower order harmonics. It turned out that they are different for each charging cycle due to various uncertainties such as battery SOC, background distortion level, charging rate, power electronic technology used etc. Consequently there is need for new methods to deal with these issues.

The study on interharmonics aggregation has found out that the value of peak current in aggregation depends on the randomness of EVs charging. When a random number of Electric Vehicles is connected to power grid, the distribution takes an extreme value distribution which loses its skewness and becomes normal for constant number of EVs. Therefore, the standard summation rule fails in practice with random number of EVs. In terms of supraharmonics, unified measurements should be considered during collection along with immunity/standardization knowledge and interference information. Even though supra-harmonic currents have very low magnitudes; this does not prevent them from being added onto primary as well as secondary emissions as they pass through protective earth (PE), leading to unnecessary tripping off RCDs. It was thus also deduced that if capacitance is high while resistance remains low then those electric vehicles connected with TN-C-S earthing systems may encounter large leakages.

When it comes to power distribution, minor voltage changes cannot affect the charging of electric vehicles whose cables are laid underground in Sweden. Nonetheless, when fault-caused shallow dips occur, cable capacitance produces a significant initial transient. These dips on their way to the distribution grid may magnify, more so in high-impedance networks founded on resonant frequency. Even though these power swings only cause a phase jump of 20 degrees they can still considerably affect EV charging as well as battery life. Asymmetrical dips especially during discharge are riskier for EVs since they tend to result into high overvoltage and oscillations at the DC link of the charger, which are mainly observed during low EV charging activities.

To manage waste that arises from the adoption of electric vehicles, one must adopt a holistic approach that integrates battery recycling and innovative recovery methods for lithium- ion batteries. Techniques like ladder and cascade utilization can prolong the life span of a battery while ultrasonic recovery will improve recycling efficiency. These measures encourage resource conservation, mitigate environmental degradation and guarantee sustainable transportation systems.

Acknowledgment

The authors are thankful to Department of Environmental Engineering, Delhi Technological University for providing the necessary support.

References

1. Ma, Gang, et al. ”Study on the impact of electric vehicle charging load on nodal voltage deviation.” Archives of Electrical Engineering 66.3 (2017): 495-505.
2. Desai A, Patel CR. “Leveraging GIS to deploy demand-driven public charging infrastructure in an Indian Metropolitan city.” Spatial Information Research. 2023;31(4):467-474. doi:10.1007/s41324-023-00514-4.
3. Bilal M, Rizwan M. “Intelligent Algorithm based Efficient Planning of Electric Vehicle Charging Station: A Case Study of Metropolitan City of India.” Scientia

Iranica. 2021;0(0):0. doi:10.24200/sci.2021.57433.5238.
4. Chavhan S, Zeebaree SRM, Alkhayyat A, Kumar S. “Design of space efficient electric vehicle charging infrastructure integration impact on power grid network.” Mathematics. 2022;10(19):3450. doi:10.3390/math10193450.
5. Singh S, Chauhan P, Singh NJ. “Feasibility of Grid-connected Solar- wind Hybrid System with Electric Vehicle Charging Station.” Jour- nal of Modern Power Systems and Clean Energy. 2021;9(2):295-306. doi:10.35833/mpce.2019.000081.
6. Mateen S, Amir M, Haque A, Bakhsh FI. “Ultra-fast charging of electric vehicles: A review of power electronics converter, grid stability and opti- mal battery consideration in multi-energy systems.” Sustainable Energy, Grids and Networks. 2023;35:101112. doi:10.1016/j.segan.2023.101112.
7. Ahmad F, Shariff SM, Krishnamurthy M. “A Cost-Efficient Approach to EV Charging Station Integrated Community Microgrid: A case study of Indian Power market.” IEEE Transactions on Transportation Electrification. 2019;5(1):200-214. doi:10.1109/tte.2019.2893766.
8. Y. Wu and L. Zhang, ”Can the development of electric vehicles reduce the emission of air pollutants and greenhouse gases in developing coun- tries?”, Transportation Research Part D: Transport and Environment, vol. 51, pp. 129-145, 2017.
9. H. Yu and A. Stuart, ”Impacts of compact growth and electric vehicles on future air quality and urban exposures may be mixed”, Science of The Total Environment, vol. 576, pp. 148-158, 2017.
10. R. Diewvilai, R. Nidhiritdhikrai and B. Eua-arporn, ”Reserve margin evaluation for generation system using probabilistic based method,” The 8th Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology (ECTI) Association of Thailand - Confer- ence 2011, Khon Kaen, 2011, pp. 905-908.
11. P. C. Sekhar, R. A. Deshpande and V. Sankar, ”Evaluation and improve- ment of reliability indices of electrical power distribution system,” 2016 National Power Systems Conference (NPSC), Bhubaneswar, 2016, pp. 1-6.
12. T. Van Cutsem, ”Voltage instability: phenomena, countermeasures, and analysis methods”, Proceedings of the IEEE, vol. 88, no. 2, pp. 208- 227, 2000.
13. J. BIAN and P. RASTGOUFARD, ”Measures Of Voltage Instability And Voltage Collapse, PART I: Theoretical Development”, Electric Machines & Power Systems, vol. 23, no. 4, pp. 361-374, 1995.
14. T. J. Browne and G. T. Heydt, ”Power Quality as an Educational Opportunity,” in IEEE Transactions on Power Systems, vol. 23, no. 2, pp. 814-815, May 2008.
15. T. S. Ustun, C. R. Ozansoy and A. Zayegh, ”Implementing Vehicle- toGrid (V2G) Technology With IEC 61850-7-420,” in IEEE Transactions on Smart Grid, vol. 4, no. 2, pp. 1180-1187, June 2013.
16. P. Kundur, N. Balu and M. Lauby, Power system stability and control. New York : McGraw-Hill, 2009.
17. P. Kessel and H. Glavitsch, ”Estimating the Voltage Stability of a Power System,” in IEEE Power Engineering Review, vol. PER-6, no. 7, pp. 72-72, July 1986.
18. P. Kessel and H. Glavitsch, ”Estimating the Voltage Stability of a Power System,” in IEEE Power Engineering Review, vol. PER-6, no. 7, pp. 72- 72, July 1986.
19. H. Dharmakeerthi, N. Mithulananthan, and T. K. Saha, ”Impact of electric vehicle load on power system oscillatory stability,” Australasian Universities Power Engineering Conference (AUPEC), Hobart, TAS, Australia, pp. 1-6. IEEE, September 2013.
20. Ul-Haq, C. Cecati, K. Strunz, and E. Abbasi, ”Impact of EV Charging on Voltage Unbalance in an Urban Distribution Network,” Intelligent Industrial Systems, vol. 1,

pp. 51-60, June 2015.
21. P. Papadopoulos, L. M. Cipcigan and N. Jenkins, “Distribution Networks with Electric Vehicles”, Universities Power Engineering Conference. Glasgow: IEEE, 2009: 1-5.
22. Nikhil Garwa, Khaleequr Rehman Niazi, ”Impact of EV on Inte- gration with Grid System – A Review” published December 2019 in 2019 8th International Conference on Power Systems (ICPS), doi:10.1109/icps48983.2019.9067587.
23. McCarthy and P. Wolfs, ”The HV system impacts of large scale elec- tric vehicle deployments in a metropolitan area,” 2010 20th Australasian Universities Power Engineering Conference, Christchurch, 2010, pp. 1- 6.
24. G. A. Putrus, P. Suwanapingkarl, D. Johnston, E. C. Bentley and M. Narayana, ”Impact of electric vehicles on power distribution networks,” 2009 IEEE Vehicle Power and Propulsion Conference, Dearborn, MI, 2009, pp. 827-831.
25. L.A. Kszos, J.J. Beauchamp, A.J. Stewart, Toxicity of Lithium to Three Freshwater Organisms and the Antagonistic Effect of Sodium, Ecotoxicology 12(5) (2003) 427-437.
26. Shahzad, M. Tanveer, W. Hassan, A.N. Shah, S.A. Anjum, S.A. Cheema, I. Ali, Lithium toxicity in plants: Reasons, mechanisms and remediation possibilities – A review, Plant Physiol. Biochem. 107 (2016) 104-115.
27. M. Tanveer, L. Wang, Potential targets to reduce beryllium toxicity in plants: A review, Plant Physiol. Biochem. 139 (2019) 691-696.
28. Marra F, Yang G, Larsen E, Rasmussen C, You S, “Demand profile study of battery electric vehicle under different charging options”, IEEE. July 2012. doi:10.1109/pesgm.2012.6345063.
29. Sanchari Deb, Karuna Kalita, Pinakeshwar Mahanta, “Review of impact of electric vehicle charging station on the power grid”. IEEE Conference Publication — IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8397215.
30. Wang S, Zhang N, Li Z, Shahidehpour M. “Modeling and impact analysis of large scale V2G electric vehicles on the power grid”. IEEE. May 2012. doi:10.1109/isgt-asia.2012.6303372.
31. Carter R, Cruden A, Hall P. “Optimizing for Efficiency or Battery Life in a Battery/Supercapacitor Electric Vehicle”. IEEE Transactions on Vehic- ular Technology. 2012;61(4):1526-1533. doi:10.1109/tvt.2012.2188551.
32. Rizvi STH, Xin A, Masood A, Iqbal S, Jan MU, Rehman H. “Electric Vehicles and their Impacts on Integration into Power Grid: A Review.” IEEE. October 2018. doi:10.1109/ei2.2018.8582069.
33. Sayed MA, Atallah R, Assi C, Debbabi M. “Electric vehi- cle attack impact on power grid operation”. International Jour- nal of Electrical Power and Energy Systems. 2022;137:107784. doi:10.1016/j.ijepes.2021.107784.
34. Go´mez JC, Morcos MM. “Impact of EV battery chargers on the power quality of distribution systems.” IEEE Transactions on Power Delivery. 2003;18(3):975-981. doi:10.1109/tpwrd.2003.813873.
35. Mallig N, Heilig M, Weiß C, Chlond B, Vortisch P. “Modelling the weekly electricity demand caused by electric cars.” Future Generation Computer Systems. 2016;64:140-150. doi:10.1016/j.future.2016.01.014.
36. Bao K, Li S, Zheng H. “Battery charge and discharge control for energy management in EV and utility integration.” IEEE. July 2012. doi:10.1109/pesgm.2012.6344719.
37. Deilami S, Muyeen SM. “An Insight into Practical Solutions for Electric Vehicle Charging in Smart Grid.” Energies. 2020;13(7):1545. doi:10.3390/en13071545.
38. Tran QT, Nguyen-Van L. Integration of electric vehicles into an in- dustrial grid:

"Impact assessment and solutions." IEEE. July 2016. doi:10.1109/pesgm.2016.7741309.

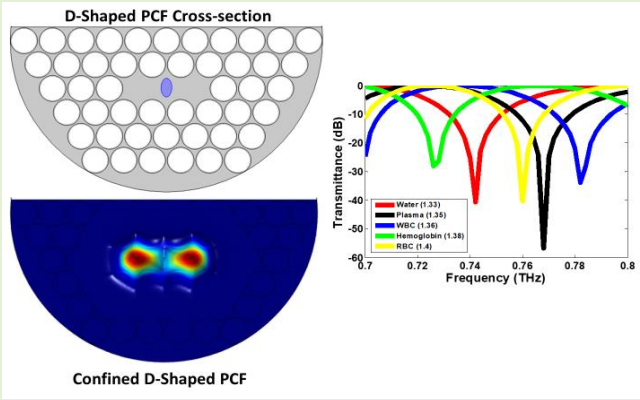
39. Van Der Burgt J, Vera SP, Wille-Haussmann B, Andersen AN, Tam- bjerg LH. "Grid impact of charging electric vehicles; study cases in Denmark, Germany and The Netherlands." IEEE. June 2015. doi:10.1109/ptc.2015.7232234.
40. Goel S, Sharma R, Rathore AK. "A review on barrier and challenges of electric vehicle in India and vehicle to grid optimisation." Transportation Engineering. 2021;4:100057. doi:10.1016/j.treng.2021.100057.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

Design and Optimization of Terahertz Based D-shaped Photonic Crystal Fiber for Blood Component Detection

Vishal Chaudhary, Sonal Singh, Vijay Shanker Chaudhary and Dharmendra Kumar, *Senior Member IEEE*

***Abstract*— In order to sense and detect significant constituents inside the blood in the terahertz (THz) frequency regime, a new kind of D-shape dual core photonic crystal fiber (DC-PCF) has been designed using COMSOL Multiphysics Software. The solid core architecture is selected for this sensor due to its capacity to transmit light effectively across a broad spectrum of wavelengths. Because blood is a necessary fluid within the human body, it is important to understand the differences between its numerous components. Red blood cells (RBC), white blood cells (WBC), water, plasma and hemoglobin are the five main blood components that must be identified in the analyte sample. The proposed sensor uses mode coupling theory to spot shifts in the transmittance curve, which makes it possible to identify blood components. The finite element method (FEM) is used for the optical examination and analysis of proposed DC-PCF properties. The proposed refractive index (RI) based sensor is designed to operate within a RI range of 1.33 to 1.4, facilitating the detection of main components in blood. Superior sensitivity has been shown in the proposed sensor, which has a sensitivity of 1487500 nm/RIU with a fiber length of 0.2 meter (m). Additionally, the analysis investigates how varying the pitch of air holes affects the sensitivity of the proposed model. This simple sensor design, which offers high sensitivity, creates opportunities for several biomedical applications.**



***Index Terms*— Blood components; Sensitivity; Finite element method; D-shape; Terahertz; Transmission**

I. Introduction

Biosensors are devices that employ biological components like antibodies, enzymes and nucleic acids to detect the presence of certain biological components or substances. RBCs, hemoglobin, WBCs, plasma, and water are among the components of human blood that are able to be detected using photonic crystal fiber (PCF) based biosensors [1]. Human blood is an intricate fluid that composed of over 4000 functional elements [2]. It is essential for supplying nutrients and oxygen to body cells while aiding in the removal of waste. Blood is thicker than water because it contains numerous cells and proteins. The average quantity of blood within a human body is 5 liters (7%) of blood. This blood volume contains 55% plasma, a liquid portion of which comprises 99% water and the remaining blood volume contains various components [3]. Protein is an essential part of plasma, playing crucial roles in promoting blood clotting and facilitating the transportation of diverse substances within the blood. Blood also contains glucose and variety of dissolved nutrients. Approximately 45% of the total blood volume is made up of blood cells, with white blood cells engaging in infection defense, red blood cells are specialized cells responsible for oxygen delivery to tissues and dedicated to transporting respiratory gases [4]–[6], and platelets aiding in the clotting process. Analyzing blood samples is an essential part of the diagnostic process to identify the presence of numerous disorders associated with blood such as thalassemia, anaemia, myeloma and haemophilia [1].

By using PCF technology, the sensor not only addresses the limitations of conventional systems but also enhances optical properties with greater diversity. Moreover, it boasts quicker detection of biological analytes compared to conventional methods. The PCF-based sensor's standout achievement is its ability to adjust optical properties by fine-tuning the arrangement, dimensions, and configurations of its air holes. Furthermore, these sensors built on PCF technology guarantee elevated sensitivity, compactness, affordability, resilience, and heightened versatility. This is accomplished by offering effective remedies for a wide array of environmental and industrial challenges [7]. In recent years, the PCF-based sensor has emerged as an intelligent component within fiber

Vishal Chaudhary and Sonal Singh are with the Department of Electronics and Communication Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India (email:vishalchaudhary_2k21phdec14@dtu.ac.in;sonalsingh@dtu.ac.in) (Corresponding author: Sonal Singh)

Vijay Shanker Chaudhary is with the Galgotias College of Engineering and Technology, Greater Noida, UP, India (e-mail: vijayshanker.chaudhary@galgotiacollege.edu)

Dharmendra Kumar is with the Department of Electronics and Communication Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh, India (e-mail: dharmendra0127@gmail.com)

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 19,2024 at 07:21:39 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

optic technology, showcasing its substantial potential across various applications. These encompass chemical sensing [8]–[10], alcohol sensing [11]–[13], genetic diagnostics [14], time domain spectroscopy [15], [16], cancer detection [17]–[20], medical diagnosis [21], RNA analysis [15], DNA analysis [22], [23] etc. particularly within the terahertz frequency range. This frequency band, spanning from 0.1 to 10 THz, occupies the space between microwave and infrared frequencies within the electromagnetic spectrum. Of particular note, terahertz-based photonic biosensors have garnered significant attention for their applications in sensing and detection, notably within the realm of biomedical sensing [24].

However, the benefits of photonic crystal fibers (PCF) provide researchers with the opportunity to investigate diverse models for detecting various analytes within the THz range. For instance, a noteworthy achievement in chemical sensing was demonstrated by Asaduzzaman et al., who attained a sensitivity of 49% at a wavelength range from 0.7 to 1.7 µm [25]. Several studies have been conducted to detect substances like water, benzene, and ethanol. Sen. et al. introduced a model with a hexagonal structure that achieved relative sensitivities of 57%, 57.27% and 57.18% for water (n = 1.33), benzene (n = 1.366) and ethanol (n = 1.354), respectively [26]. Additionally, Kanmani et al. proposed a PCF model that achieved remarkable relative sensitivities of 77.23%, 77.18% and 77.08% for water (n = 1.33), ethanol (n = 1.354) and benzene (n = 1.366), respectively [27]. In 2019, Podder et al. developed a PCF sensor that exhibited a relative sensitivity of 63.4% in detecting sulfuric acid for the maximum concentration of H_2SO_4 at a wavelength range from 0.8 to 1.8 µm [28]. Another notable example is the circular PCF designed for sensing blood components presented, where a maximum sensitivity of 66.49% was achieved at a wavelength range from 0.7 to 1.4 µm. The primary aim of this proposed work is to detect numerous human diseases and play a crucial role in medical science [1]. Further advancements were made by Ahmed et al. in 2018, achieving an even higher sensitivity of approximately 80.93% for RBCs, 80.56% for hemoglobin, 80.13% for WBCs, 79.39% for water, and 79.91% for plasma in the THz frequency range from 1.5 THz to 3.5 THz [3]. In 2020, a novel hollow rectangular core photonic crystal fiber (RCPCF) was specifically designed for detecting crucial blood components. This innovative fiber structure features six rectangular air holes arranged in a horizontal configuration, forming the cladding. Topas was chosen as the primary bulk material due to its remarkable properties in the THz range. The sensor operational range spans from 0.8 THz to 1.8 THz, offering significant advantages such as a high mode area, minimal confinement loss, and exceptional relative sensitivity. At the high-frequency end of 1.8 THz, the rectangular core PCF demonstrates exceptional relative sensitivity results of 94.38% for RBCs, 93.72% for hemoglobin, 92.94% for WBCs, 90.8% for water, and 92.14% for plasma [29]. Recently in 2023, Vijayalakshmi et al. developed a bi-core PCF sensor that exhibited a maximum sensitivity of 6930 nm/RIU in detecting blood component such as RBCs, hemoglobin, WBCs, plasma, and water, covering the refractive index range of 1.33 to 1.40 [30]. These discussions clearly indicate that there exists significant potential for enhancing the sensing properties of PCF-based models.

In this study, a DC-PCF has been proposed for detecting primary blood components such as RBC, WBC, water, plasma and hemoglobin. The proposed sensor has demonstrated superior sensitivity, achieving 1487500 nm/RIU with a fiber length of 0.2 m in the frequency range from 0.7 THz to 0.8 THz. The achieved high sensitivity of the designed sensor is attributed to a thorough analysis of diverse PCF parameters, thereby rendering it highly applicable for a wide range of applications in the field of biosensing. Operating effectively within a refractive index (RI) range spanning from 1.33 to 1.40, the sensor distinguishes itself by achieving remarkable sensitivity without resorting to the use of any plasmonic materials. A comprehensive exploration of various PCF parameters has been undertaken, encompassing factors like RI difference, birefringence, coupling length, transmission spectrum and transmittance curve. By closely monitoring spectral shifts of the transmittance curve aligned with their corresponding refractive indices, the distinctive components present in blood can be accurately discerned and identified. The proposed sensor presented here is designed using COMSOL Multiphysics software. The proposed biosensor has the potential to be a highly sensitive and selective option for detecting blood compositions.

II. Design Methodology

Figure 1(a) depicts the design configuration of the proposed THz PCF sensor, which is constructed using TOPAS material, alongside a detailed depiction of the D-shape structure. We opt for a solid core due to its ability to facilitate the operation of light across a wide range of wavelengths. The central elliptical hole of the PCF is utilized for the insertion of blood components, with the aim of their detection. Different analytes, including RBC, WBC, water, plasma, and hemoglobin have a distinct refractive index (RI), with values of 1.40 for RBC, 1.36 for WBC, 1.33 for water, 1.35 for Plasma and 1.38 for hemoglobin. To enhance sensitivity, we have removed one air hole on each side from the central elliptical hole. This results in the creation of a dual-core structure for the sensor, where light confinement occurs between these two cores. In the proposed configuration, the elliptical hole's major axis (a) is 340 µm and minor axis (b) is 190 µm are taken into consideration. The air holes, with a diameter (d) of 470 µm and spaced at a pitch (Λ) of 500 µm between any two consecutive holes, ensure a consistent air filling fraction (AFF) of 0.94 throughout the PCF. A fine mesh is employed during simulation, whereas a parametric sweep is applied to vary the operating frequency. As depicted in Figure 1(b), the geometric meshing statistics for the proposed PCF included 2,016 edge elements, 203 vertex elements, and 14,556 triangular elements. TOPAS has been selected as the background material for the proposed structure due to its remarkable optical properties, including exceptional transparency, strong resistance to acids and alkalis, nice biocompatibility, a wide range of adjustable deflection temperatures, a constant refractive index (n = 1.53) across a broad frequency range, impressive strength, excellent heat resistance, outstanding water vapor barrier characteristics, significant light transmission (91%), suitability for high humidity conditions, and minimal material dispersion [7], [31]. To date, various fabrication methods have been

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:21:38 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

documented, including capillary stacking, sol-gel, extrusion, and 3D printing. Capillary stacking and the sol-gel technique are commonly employed for manufacturing circular and elliptical holes [32], while asymmetric hole structures (rectangular) have been successfully fabricated through extrusion [33] and 3D printing [34]. Consequently, the capillary stacking and sol-gel methods are recommended for fabricating the proposed DC-PCF sensor structure. The method prototype, depicted in Figure (2), involves stacking small span capillaries of fixed diameters into a glass tube of larger diameter to achieve completion. The proposed sensor introduced here is constructed using COMSOL Multiphysics Software, which is affiliated with the numerical approach known as the finite element method (FEM).

The experimental setup for assessing the accuracy of identifying blood components using the proposed PCF biosensor is depicted in Figure (3). A broadband light source (BLS) is used to transmit optical light via a single-mode fiber (SMF). The SMF can be connected to the biosensor using the Fujikura Splicer splicing method. Various blood components are positioned in the sensing channel. Prior to measuring new samples, the PCF biosensor probes are required to undergo washing with deionized water. Changes occur in blood

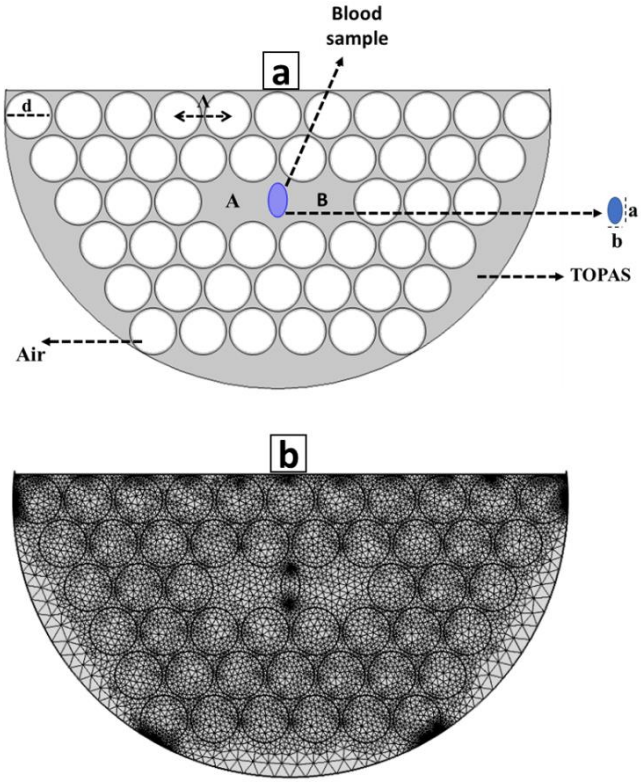


Fig. 1 Design of the proposed PCF sensor with dual solid cores: (a) 2D cross section view and (b) Fine mesh statistics.

component samples as they react with ligands inside the sensing channel. The interaction between ligands and target molecules induces changes in the optical properties (wavelength shift in the transmittance spectrum) of the guided mode in the DC-PCF sensor. These optical changes are detected and analyzed to quantify the various blood components. The optical spectrum analyzer (OSA) captures the focused light through the SMF. In the end, the computer connected to the OSA generates the final spectrum.

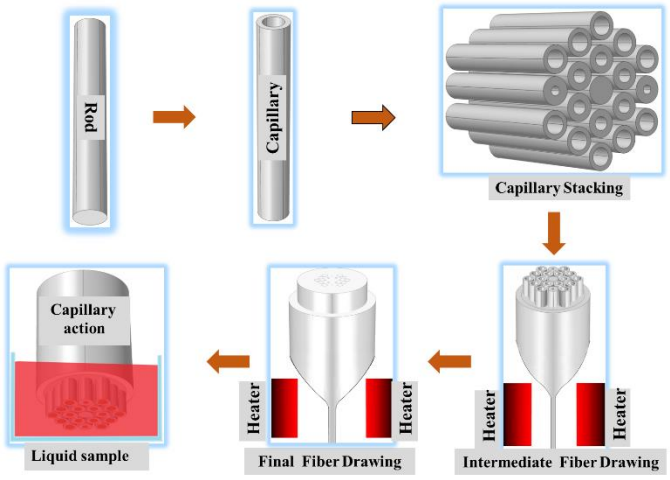


Fig. 2 Fabrication technique for proposed PCF using capillary stacking.

III. Results Analysis and Discussion

In the cross-sectional view, there are two fiber cores that establish separate waveguides, and within PCFs, mode coupling can occur. In DC-PCF sensors, mode coupling describes the interaction and transfer of optical power between the modes of the two fiber cores. This phenomenon is influenced by factors such as core separation, wavelength, and external perturbations like temperature, pressure, and strain. The strength of mode coupling, quantified by the coupling coefficient, is critical for the sensor's performance and sensitivity [35]. According to the theory of mode coupling, the interaction in a dual-core fiber can be explained through

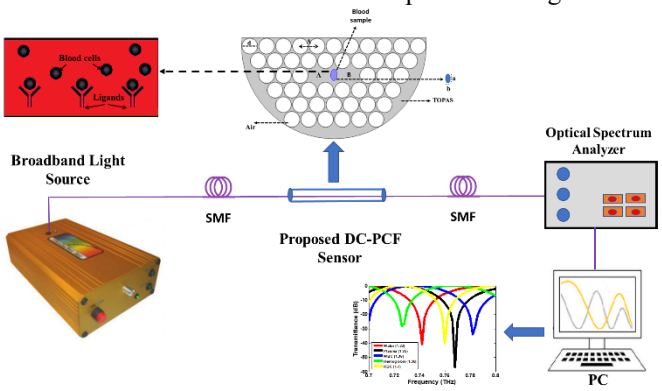


Fig. 3 Representation of experimental setup for blood component detection.

the description of four super modes, encompassing both even (x-even and y-even) and odd (x-odd and y-odd) modes. These super-modes arise from the inherent modes of the individual cores, showcasing symmetric and anti-symmetric field distributions, respectively. Figure (4) shows the distribution of the electric field among the dual core for the x and y polarized super-modes at a frequency of 0.75 THz. Figure (5) depicts a three-dimensional (3-D) visualization showcasing the electric field strength across various modes.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:21:38 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

The initial estimation of the difference in effective refractive indices between the even and odd super-modes serves as a starting point for calculating the other parameters. It is evaluated by

$$\Delta n_{eo} = |n_e - n_o| \quad (1)$$

where, n_e and n_o represent the effective refractive indices for even and odd modes, respectively. The graph in Figure (6) illustrates how the effective refractive index difference (Δn_{eo}) between even and odd modes varies for x-polarized directions. It demonstrates that as the frequency increases, the effective refractive index difference decreases for primary blood components.

Birefringent materials can induce the phenomenon of double refraction, wherein light is split into two distinct rays traveling in different directions upon polarization. The calculation of the birefringence parameter, B, specific to DC-PCF, involves employing equation (2).

$$B = \left|Re\left(n_{eff}^{x}\right) - Re\left(n_{eff}^{y}\right)\right| \quad (2)$$

where, the symbols n_{eff}^{x} and n_{eff}^{y} represent the effective refractive indices corresponding to the x and y-polarized modes, respectively [28]. Figure (7) illustrates the overall variation in birefringence with frequency changes. Within the frequency range of 0.0062 to 0.0035, there is a discernible descending trend in the maximum range of birefringence. The sensitivity and coupling length of the proposed design exhibit a direct correlation with its birefringence. The initial birefringence of the proposed sensor is 0.0062, and it undergoes a linear decrease with increasing frequency until it reaches a minimal birefringence value of 0.0035.

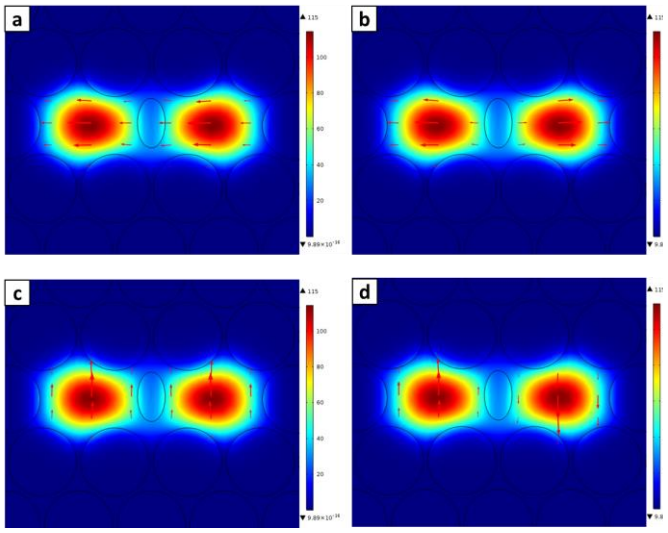
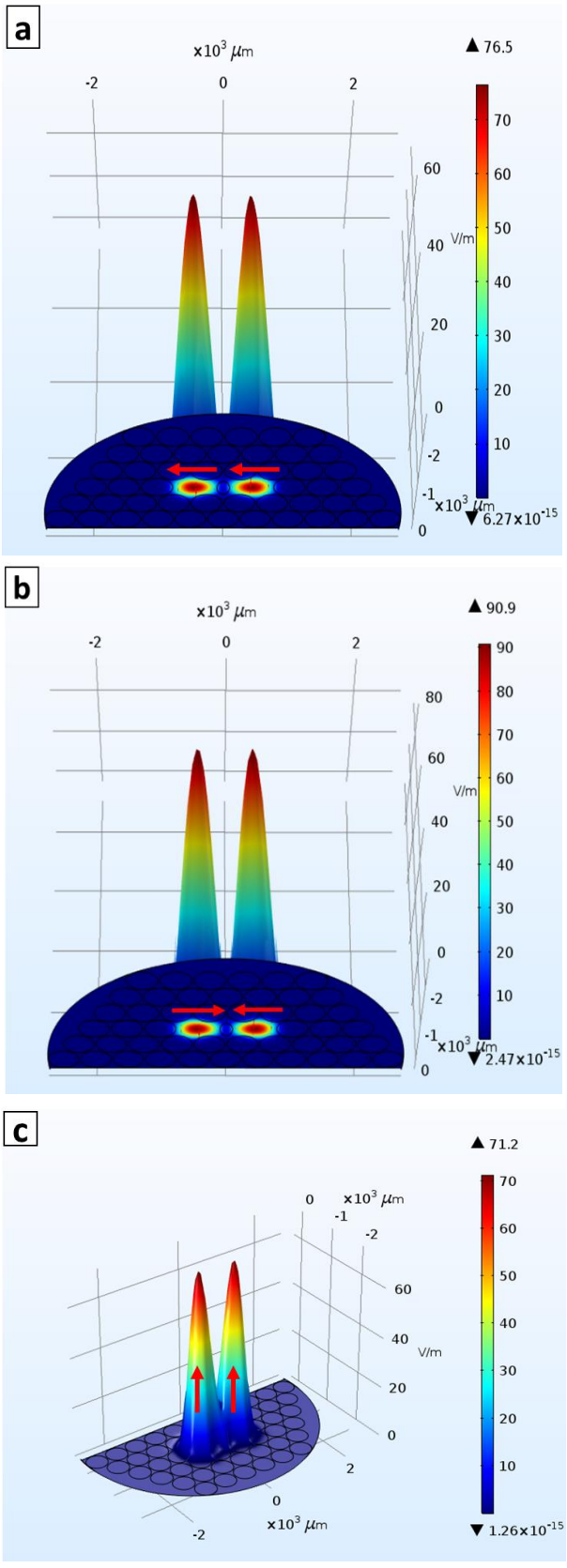


Fig. 4 Electric field distribution within a symmetric core, showcasing four distinct modes: (a) even x-polarization (b) odd x-polarization (c) even y-polarization and (d) odd y-polarization.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:21:38 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

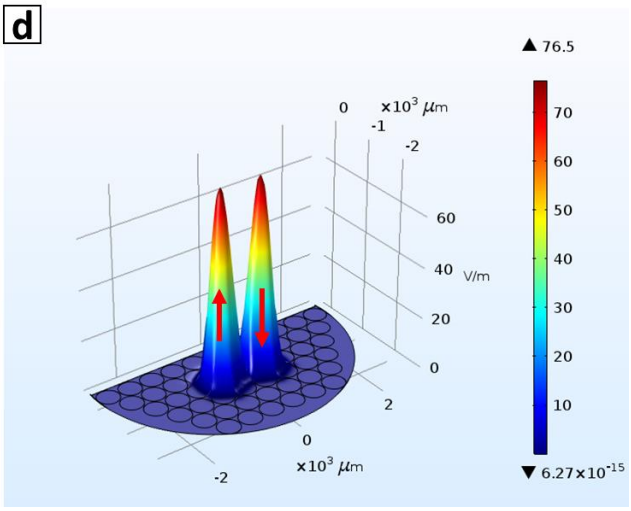


Fig. 5 3-D view of Electric field distribution for four distinct modes: (a) even x-polarization (b) odd x-polarization (c) even y-polarization and (d) odd y-polarization.

The coupling length, defined as the minimum length at which optimal light transmission occurs through the core of a PCF. It can be determined by using equation (3).

$$L_c = \frac{\pi}{|\beta_e^i - \beta_o^i|} = \frac{\lambda}{2|n_e^i - n_o^i|} \quad (3)$$

where, β_e^i and β_o^i represent the propagation constants of the even and odd supermodes with $i = x, y$ polarization, λ is the wavelength, and n_e^i and n_o^i denote the effective refractive indices for the even and odd modes, respectively [36]. The coupling length changes gradually with the frequency, as depicted in Figure (8). The graph reveals that the maximum achievable coupling length falls within the range of 0.0387 m to 0.1011 m, exhibiting an upward-curving trend.

To detect blood constituents in a sample analyte, the coupling mode theory is employed, where power is transmitted between the two cores. The power transferred from first core to another in the designed DC-PCF is computed by utilizing equation (4).

$$\mathrm{I}(\lambda) = \sin^2\left(\frac{\pi}{\lambda}\Delta n_{eo} L\right) \quad (4)$$

where, symbol L represents the length of the fiber, and λ denotes the wavelength [37]. Figure (9) depict the transmission curve achieved for the primary blood constituents for x polarized light with corresponding fiber length of 0.2 m. These graphs cover the refractive index range of 1.33 to 1.4. It is worth noting that the power transmission graph of the proposed DC-PCF follows a sinusoidal pattern.

Moreover, when taking into account the optical light entering the initial core, the transmittance can be determined by using equation (5).

$$T_r = 10 log_{10} \frac{P_{out}}{P_{in}} \quad (5)$$

here, we consider P_{in} as the maximum power of P_{out} [30].

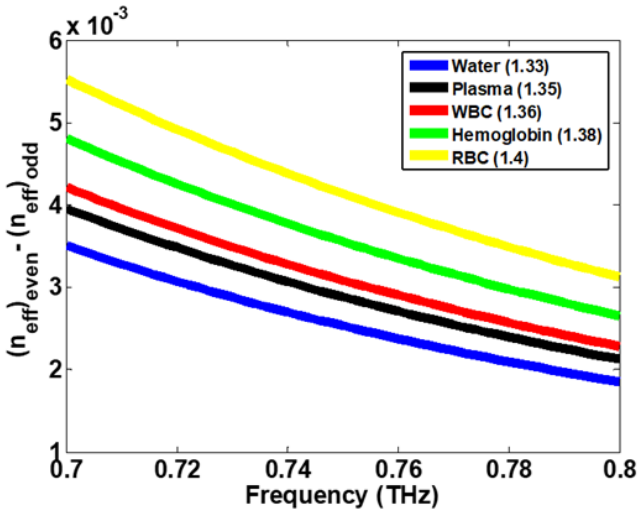


Fig. 6 Change in refractive index difference varies with different frequency.

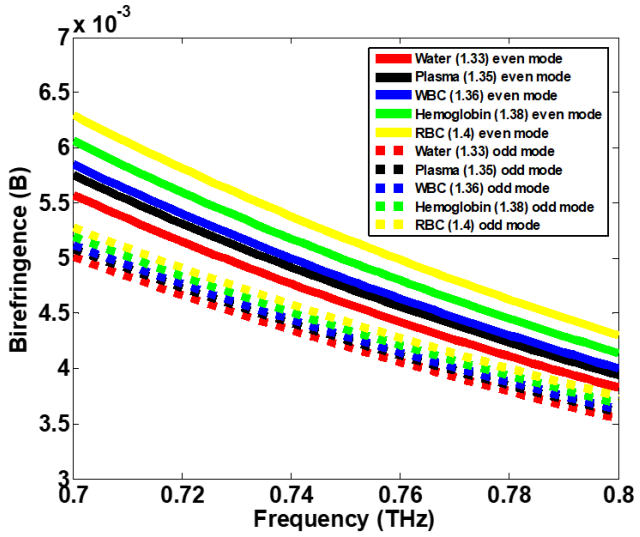


Fig. 7 Change in birefringence with respect to different frequency.

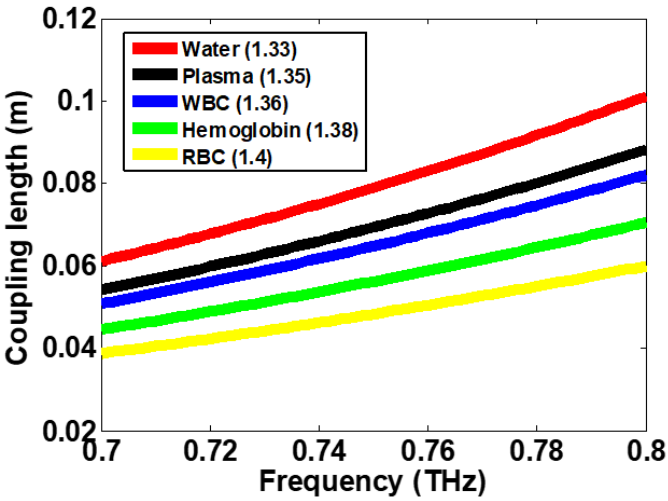


Fig. 8 Change in coupling length with respect to different frequency.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:21:38 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

Figure (10) display the transmittance curve for the five main blood constituents, covering the refractive index range of 1.33 to 1.4, for x polarized light with corresponding fiber length of 0.2 m. By monitoring the shift in refractive index (RI), it is possible to sense and detect the corresponding blood component. The frequency peaks (f_{peak}) observed for major blood constituents, like water, plasma, WBC, Hemoglobin and RBC at 0.74 THz, 0.766 THz, 0.78 THz, 0.724 THz and 0.758 THz, respectively.

Sensitivity is a crucial parameter used to evaluate the performance of PCF sensor. It is determined by examining the peak shift in response to changes in the refractive index of the blood constituent. This is articulated as follows;

$$S_\lambda(nmRIU^{-1}) = \Delta\lambda_{peak}/\Delta n_a \quad (6)$$

where, $\Delta\lambda_{peak}$ represents the shift in wavelength peak, Δn_a denotes the difference in RI of the blood samples and $\left(\lambda_{peak} = \frac{\text{speed of light (c)}}{f_{peak}}\right)$ [38]. The maximum sensitivity determined for the main blood constituents, spanning from 1.33 to 1.4, is 1487500 nm/RIU. Table I depicts a thorough examination of the anticipated outcomes of the proposed PCF biosensor study alongside a recently reported sensor with a similar sensing range. The analysis reveals that the sensitivity of the proposed PCF significantly exceeds that of the other PCF biosensor. This technique is recognized as the simple and most effective approach for blood components detection.

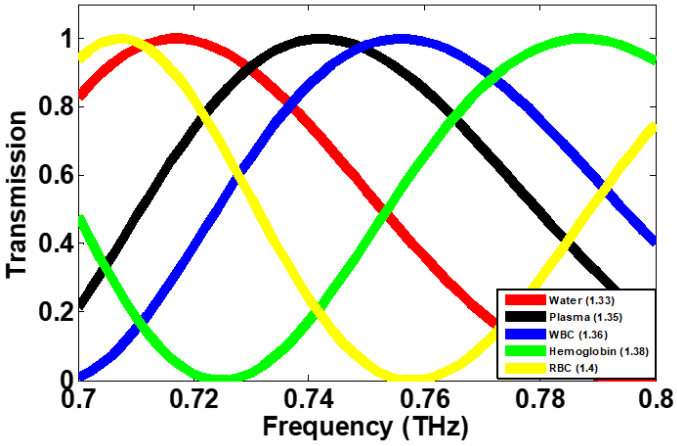


Fig. 9 Transmission curve of a DC-PCF with a length of 0.2 m.

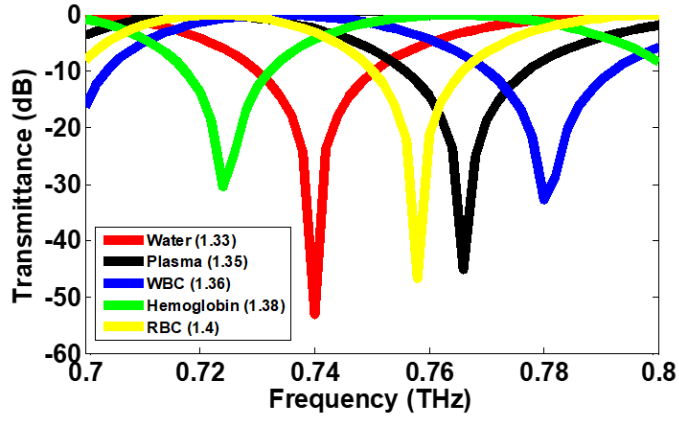


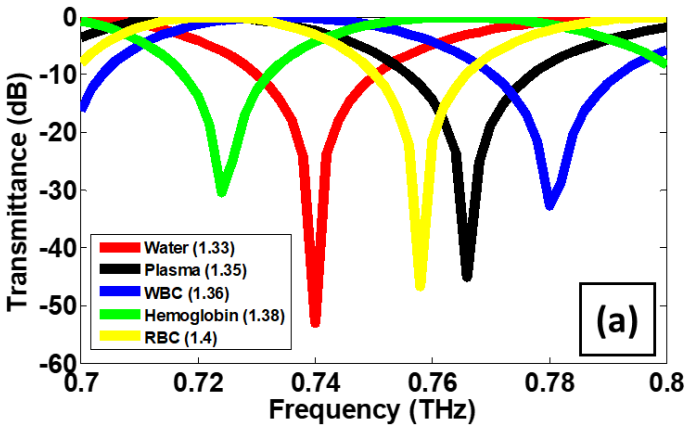
Fig. 10 Transmittance curve of DC-PCF for primary blood components.

Table I
Comparison of proposed PCF biosensor with recently reported biosensors

Reference	Design of sensor structure	Sensing range	Sensitivity (nm/RIU)
[39]	D-shape PCF with pyramid grating	1.355-1.385	5200
[40]	AZO-coated plasmonic PCF nano sensor	1.33-1.40	1950
[41]	D-shaped PCF RI sensor	1.33-1.354	294.11
[42]	D-shape open-loop PCF temperature sensor	1.35-1.40	3300
[30]	Bi-core PCF	1.33-1.40	6930
Proposed structure	THz based D-shape DC-PCF	1.33-1.40	1487500

IV. Effect of Pitch Variation on Sensitivity

Figure (11) demonstrates the pitch variation in the transmittance curve of DC-PCF for the five primary blood constituents with corresponding fiber length of 0.2 m. Figure (11-a), (11-b), and (11-c) reveal that the proposed DC-PCF achieves a sensitivity of 1479500 nm/RIU at pitch (Λ) = 498 µm, 1487500 nm/RIU at pitch (Λ) = 500 µm, and 1483500 nm/RIU at pitch (Λ) = 502 µm, respectively. Optimizing the pitch demonstrates that the proposed DC-PCF sensor achieves maximum sensitivity at Λ = 500 µm.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:21:38 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

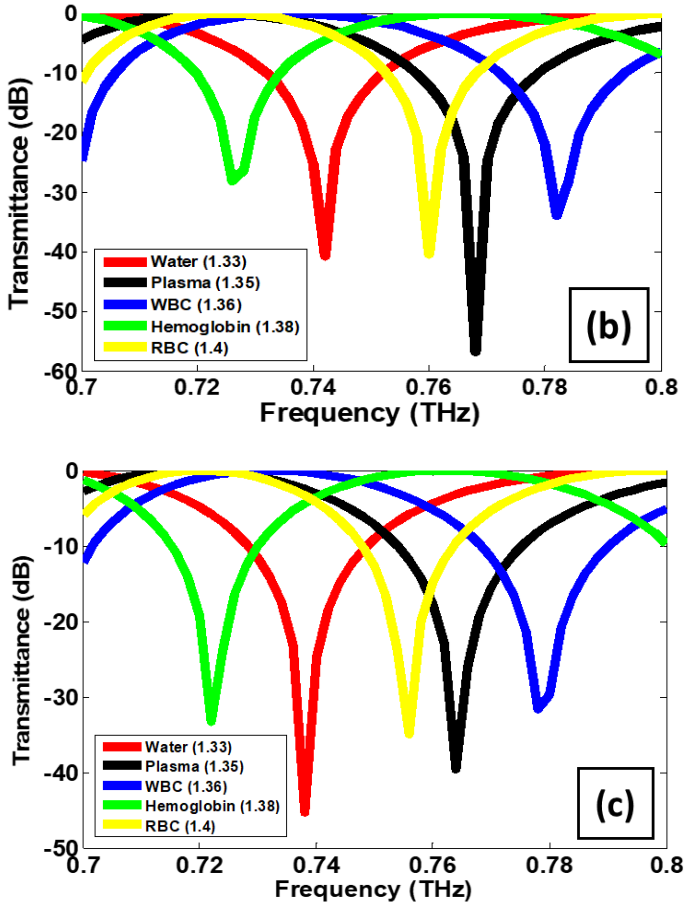


Fig. 11 Transmittance curve with different pitch (Λ), for five primary blood constituents (a) 498 µm (b) 500µm and (c) 502 µm.

V. Conclusion

In this manuscript, a highly sensitive DC-PCF based biosensor is presented using COMSOL Multiphysics Software and FEM method, encompassing a frequency range from 0.7 to 0.8 THz. The proposed sensor is designed to operate within a refractive index (RI) range of 1.33 to 1.4, facilitating the detection of main components in blood. The blood constituents analyzed include water, plasma, WBC, hemoglobin and RBC. Detection and sensing of these components are achieved through coupling mode theory involving the transfer of power between the twin cores. A comprehensive investigation of various PCF parameters has been carried out, including refractive index difference, birefringence, coupling length, transmission spectrum, and transmittance curve. Our research reveals that the sensor reaches its highest sensitivity, measuring 1487500 nm/RIU, when utilizing a 0.2 m long fiber at pitch (Λ) = 500 µm. The simplicity in structure ensures an easy fabrication process for the proposed biosensor. For the fabrication of the proposed DC-PCF sensor structure, it is recommended to utilize capillary stacking and sol-gel methods. The exceptional sensing proficiency and straightforward design of this proposed biosensor establish it as a valuable contender for utilization in both industrial and medical applications.

References

[1] S. Singh and V. Kaur, "Photonic crystal fiber sensor based on sensing ring for different blood components: Design and analysis," *Int. Conf. Ubiquitous Futur. Networks, ICUFN*, pp. 399–403, 2017, doi: 10.1109/ICUFN.2017.7993816.

[2] P. Sharma and P. Sharan, "Design of photonic crystal based ring resonator for detection of different blood constituents," *Opt. Commun.*, vol. 348, pp. 19–23, 2015, doi: 10.1016/j.optcom.2015.03.015.

[3] K. Ahmed *et al.*, "Refractive Index-Based Blood Components Sensing in Terahertz Spectrum," *IEEE Sens. J.*, vol. 19, no. 9, pp. 3368–3375, 2019, doi: 10.1109/JSEN.2019.2895166.

[4] G. Barshtein, "Biochemical and Biophysical Properties of Red Blood Cells in Disease," *Biomolecules*, vol. 12, no. 7, 2022, doi: 10.3390/biom12070923.

[5] C. Bray *et al.*, "Erythrocyte sedimentation rate and C-reactive protein measurements and their relevance in clinical medicine," *Wis. Med. J.*, vol. 115, no. 6, pp. 317–321, 2016.

[6] M. P. Spencer, "Systemic circulation.," *Annu. Rev. Physiol.*, vol. 28, pp.311–346, 1966, doi:10.1146/annurev.ph.28.030166.001523.

[7] M. S. Islam *et al.*, "A novel approach for spectroscopic chemical identification using photonic crystal fiber in the terahertz regime," *IEEE Sens. J.*, vol. 18, no. 2, pp. 575–582, 2018, doi: 10.1109/JSEN.2017.2775642.

[8] B. Fischer, M. Hoffmann, H. Helm, G. Modjesch, and P. U. Jepsen, "Chemical recognition in terahertz time-domain spectroscopy and imaging," *Semicond. Sci. Technol.*, vol. 20, no. 7, 2005, doi: 10.1088/0268-1242/20/7/015.

[9] B. M. Fischer, H. Helm, and P. U. Jepsen, "Chemical recognition with broadband THZ spectroscopy," *Proc. IEEE*, vol. 95, no. 8, pp. 1592–1604, 2007, doi: 10.1109/JPROC.2007.898904.

[10] M. Walther, B. M. Fischer, A. Ortner, A. Bitzer, A. Thoman, and H. Helm, "Chemical sensing and imaging with pulsed terahertz radiation," *Anal. Bioanal. Chem.*, vol. 397, no. 3, pp. 1009–1017, 2010, doi: 10.1007/s00216-010-3672-1.

[11] P. U. Jepsen, U. Møller, and H. Merbold, "Investigation of aqueous alcohol and sugar solutions with reflection terahertz time-domain spectroscopy," *Opt. Express*, vol. 15, no. 22, p. 14717, 2007, doi: 10.1364/oe.15.014717.

[12] P. U. Jepsen, J. K. Jensen, and U. Møller, "Characterization of aqueous alcohol solutions in bottles with THz reflection spectroscopy," *Opt. Express*, vol. 16, no. 13, p. 9318, 2008, doi: 10.1364/oe.16.009318.

[13] E. Arik, C. Koral, H. Altan, and O. Esentürk, "A new method for alcohol content determination of fuel oils by terahertz spectroscopy," *Int. Conf. Infrared, Millimeter, Terahertz Waves, IRMMW-THz*, 2013, doi: 10.1109/IRMMW-THz.2013.6665885.

[14] M. Nagel, P. Haring Bolivar, M. Brucherseifer, H. Kurz, A. Bosserhoff, and R. Büttner, "Integrated THz technology for label-free genetic diagnostics," *Appl. Phys. Lett.*, vol. 80, no. 1, pp. 154–156, 2002, doi: 10.1063/1.1428619.

[15] B. M. Fischer *et al.*, "Terahertz time-domain spectroscopy and imaging of artificial RNA," *Opt. Express*, vol. 13, no. 14, p. 5205, 2005, doi: 10.1364/opex.13.005205.

[16] J. Zhang and D. Grischkowsky, "Waveguide terahertz time-domain spectroscopy of nanometer water layers," *Opt. Lett.*, vol. 29, no. 14, p. 1617, 2004, doi: 10.1364/ol.29.001617.

[17] C. Yu, S. Fan, Y. Sun, and E. Pickwell-Macpherson, "The potential of terahertz imaging for cancer diagnosis: A review of investigations to date.," *Quant. Imaging Med. Surg.*, vol. 2, no. 1, pp. 33–45, 2012, doi: 10.3978/j.issn.2223-4292.2012.01.04.

[18] A. Rahman, A. K. Rahman, and B. Rao, "Early detection of skin cancer via terahertz spectral profiling and 3D imaging," *Biosens. Bioelectron.*, vol. 82, pp. 64–70, 2016, doi: 10.1016/j.bios.2016.03.051.

[19] D. J.S., M. B.P., F. C.I., and G. R.A., "Beyond mammography: New frontiers in breast cancer screening," *Am. J. Med.*, vol. 126, no. 6, pp. 472–479, 2013.

[20] H. Cheon, H. J. Yang, and J. H. Son, "Toward Clinical Cancer Imaging Using Terahertz Spectroscopy," *IEEE J. Sel. Top. Quantum Electron.*, vol. 23, no. 4, 2017, doi: 10.1109/JSTQE.2017.2704905.

[21] K. I. Zaytsev *et al.*, "Medical diagnostics using terahertz pulsed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:21:38 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. 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/JSEN.2024.3437245

spectroscopy," *J. Phys. Conf. Ser.*, vol. 486, no. 1, 2014, doi: 10.1088/1742-6596/486/1/012014.

[22] A. Roitberg, E. J. Heilweil, and A. G. Markelz, "Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz," *Chem. Phys. Lett.*, vol. 320, no. 1–2, pp. 42–48, 2000.

[23] B. M. Fischer, M. Walther, and P. U. Jepsen, "Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy," *Phys. Med. Biol.*, vol. 47, no. 21, pp. 3807–3814, 2002, doi: 10.1088/0031-9155/47/21/319.

[24] Y. He, P. I. Ku, J. R. Knab, J. Y. Chen, and A. G. Markelz, "Protein dynamical transition does not require protein structure," *Phys. Rev. Lett.*, vol. 101, no. 17, 2008, doi: 10.1103/PhysRevLett.101.178103.

[25] S. Asaduzzaman, K. Ahmed, T. Bhuiyan, and T. Farah, "Hybrid photonic crystal fiber in chemical sensing," *Springerplus*, vol. 5, no. 1, 2016, doi: 10.1186/s40064-016-2415-y.

[26] S. Sen, S. Chowdhury, K. Ahmed, and S. Asaduzzaman, "Design of a porous cored hexagonal photonic crystal fiber based optical sensor with high relative sensitivity for lower operating wavelength," *Photonic Sensors*, vol. 7, no. 1, pp. 55–65, 2017, doi: 10.1007/s13320-016-0384-y.

[27] R. Kanmani, K. Ahmed, S. Roy, F. Ahmed, B. Kumar Paul, and M. S. Mani Rajan, "The performance of hosting and core materials for slotted core Q-PCF in terahertz spectrum," *Optik (Stuttg).*, vol. 194, 2019, doi: 10.1016/j.ijleo.2019.163084.

[28] E. Podder, M. B. Hossain, R. H. Jibon, A. A. M. Bulbul, and H. S. Mondal, "Chemical sensing through photonic crystal fiber: sulfuric acid detection," *Front. Optoelectron.*, vol. 12, no. 4, pp. 372–381, 2019, doi: 10.1007/s12200-019-0903-8.

[29] E. Podder, M. B. Hossain, M. E. Rahaman, A. A. M. Bulbul, and K. Ahmed, "Design and optimization of terahertz blood components sensor using photonic crystal fiber," *Sens. Bio-Sensing Res.*, vol. 30, 2020, doi: 10.1016/j.sbsr.2020.100386.

[30] D. Vijayalakshmi, C. T. Manimegalai, and P. Selvakumar, "Bi-core photonic crystal fiber for blood component detection," *J. Opt.*, vol. 52, no. 1, pp. 42–49, 2023, doi: 10.1007/s12596-022-00848-6.

[31] "Topas advanced polymer. accessed: 15 th march, 2021." [online]. available:https://topas.com/markets/electronics/more electronics-apps.

[32] R. T. Bise and D. J. Trevor, "Sol-gel derived microstructured fiber: Fabrication and characterization," *Conf. Opt. Fiber Commun. Tech. Dig. Ser.*, vol. 3, pp. 269–271, 2005, doi: 10.1109/ofc.2005.192772.

[33] A. Ghazanfari, W. Li, M. C. Leu, and G. E. Hilmas, "A novel freeform extrusion fabrication process for producing solid ceramic components with uniform layered radiation drying," *Addit. Manuf.*, vol. 15, pp. 102–112, 2017, doi: 10.1016/j.addma.2017.04.001.

[34] H. Ebendorff-Heidepriem, J. Schuppich, A. Dowler, L. Lima-Marques, and T. M. Monro, "3D-printed extrusion dies: a versatile approach to optical material processing," *Opt. Mater. Express*, vol. 4, no. 8, p. 1494, 2014, doi: 10.1364/ome.4.001494.

[35] W.-P. Huang, "Coupled-mode theory for optical waveguides: an overview," *J. Opt. Soc. Am. A*, vol. 11, no. 3, p. 963, 1994, doi: 10.1364/josaa.11.000963.

[36] V. Chaudhary and S. Singh, "Highly sensitive twin core photonic crystal fiber for hazardous cancer cell detection in THz frequency regime," *Opt. Quantum Electron.*, pp. 1–17, 2023, doi: 10.1007/s11082-023-05462-9.

[37] V. Chaudhary and S. Singh, "Twin core photonic crystal fiber based temperature sensor with improved sensitivity over a wide range of temperature," *Opt. Quantum Electron.*, vol. 55, no. 1, 2023, doi: 10.1007/s11082-022-04304-4.

[38] V. Chaudhary and D. Kumar, "Design and analysis of refractive index sensor based on dual-core photonic crystal fiber (DC-PCF) with rectangular air hole lattice structure," *Lect. Notes Electr. Eng.*, vol. 546, pp. 207–213, 2020, doi: 10.1007/978-981-13-6159-3_22.

[39] N. A. Elmahdy, M. F. O. Hameed, and S. S. A. Obayya, "Refractive index sensor based on plasmonic D-shaped photonic crystal fiber with pyramid grating," *Opt. Quantum Electron.*, vol. 54, no. 11, 2022, doi: 10.1007/s11082-022-04102-y.

[40] M. R. Islam, A. N. M. Iftekher, F. Noor, M. R. H. Khan, M. T. Reza, and M. M. Nishat, "AZO-coated plasmonic PCF nanosensor for blood constituent detection in near-infrared and visible spectrum," *Appl. Phys. A Mater. Sci. Process.*, vol. 128, no. 1, 2022, doi: 10.1007/s00339-021-05220-2.

[41] M. Y. Azab, M. F. O. Hameed, G. A. Mahdiraji, F. R. M. Adikan, and S. S. A. Obayya, "Experimental and numerical characterization of a D-shaped PCF refractive index sensor," *Opt. Quantum Electron.*, vol. 54, no. 12, 2022, doi: 10.1007/s11082-022-04232-3.

[42] F. Meng, H. Wang, and D. Fang, "Research on D-Shape Open-Loop PCF Temperature Refractive Index Sensor Based on SPR Effect," *IEEE Photonics J.*, vol. 14, no. 3, 2022, doi: 10.1109/JPHOT.2022.3166822.

Vishal Chaudhary received B.Tech degree from Feroze Gandhi Institute of Engineering & Technology, Raebareli, India and M.Tech degree from Madan Mohan Malviya University of Technology Gorakhpur, India. He is currently pursuing Ph.D from Delhi Technological University, Delhi, India. His research interests include Photonic Crystal Fiber, PCF based Sensors and biosensors.

Sonal Singh received B.Tech degree in electronics and communication engineering from UPTU, India, M. Tech degree in nanotechnology from Jamia Millia Islamia, Delhi, India and Ph.D degree in electronics engineering from University of Petroleum & Energy Studies (UPES), Dehradun, India. She is currently working as an Assistant Professor in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India. Her area of interest includes Photonic Crystal Fiber, Nanoelectronics/Nanotechnology, synthesis, characterization, applications including analog circuits, PEC and photocatalysis. She has published more than 20 papers in reputed international journals, conferences, research articles and book chapters.

Vijay Shanker Chaudhary received B.Tech degree from MIT, Moradabad, India and M.Tech and Ph.D degree from Madan Mohan Malviya University of Technology, Gorakhpur. He is currently working as Associate Professor with the Galgotias College of Engineering and Technology, Greater Noida, India. His research interest's photonic crystal fiber, optical fiber sensor and terahertz sensing properties.

Dharmendra Kumar received the B.Tech degree from IET, MJP Rohilkhand University Bareilly, India, and M. Tech and Ph. D degree from IIT (ISM) Dhanbad, India. He is currently working as an Assistant Professor with the Department of ECE at Madan Mohan Malviya University of Technology, Gorakhpur, India and a senior member of IEEE Society. His research interests include photonic crystal fiber, optical fiber sensor and nano photonics.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:21:38 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Dynamic Filter Application in Graph Convolutional Networks for Enhanced Spectral Feature Analysis and Class Discrimination in Medical Imaging

ARYAN SINGH[1,2], PEPIJN VAN DE VEN[1,2], CIARÁN EISING[1,2](Senior Member, IEEE), and PATRICK DENNY[2,3](Member, IEEE)

[1]Dept. of Electronic and Computer Engineering, University of Limerick, Limerick, V94 T9PX, Ireland
[2]Data-Driven Computer Engineering Research Centre, University of Limerick, Limerick, V94 T9PX, Ireland
[3]Dept. of Computer Science and Information Systems (CSIS), University of Limerick, Limerick, V94 T9PX, Ireland

Corresponding author: Aryan Singh (e-mail: aryan.singh@ul.ie).

ABSTRACT Graph Neural Networks (GNNs) offer a promising direction for medical image analysis, particularly due to their ability to capture complex relationships and handle non-Euclidean data structures often encountered in this domain. This study aims to demonstrate the potential of GNNs in medical image classification, showcasing their comparable performance to leading CNNs while using significantly fewer parameters. We introduce a novel GNN model, Graph Convolution Neural Network Enhanced Connectivity (GCNN-EC), incorporating an enhanced connectivity technique that leverages interrelationships between RGB channel features. Evaluated on the MedMNIST dataset, a standardized benchmark for medical image analysis, GCNN-EC performs competitively against pre-trained DNNs, highlighting its potential as an efficient and effective approach for medical image analysis. Importantly, we do not claim that GCNN-EC outperforms all CNNs in every scenario, but rather aim to showcase the potential of GNNs as a valuable tool in the medical imaging toolkit that can be used on its own or in conjuction with CNNs to further advance the state of the art. Additionally, our model outperforms traditional GNNs on standard image datasets like MNIST and CIFAR-10, mitigating the common issue of over-smoothing. These findings underscore the effectiveness of GCNs with dynamic filtering and encourage further research into GNN-based approaches for medical imaging and broader computer vision tasks.

INDEX TERMS GNN, GCN, CNN, Over-smoothing

I. INTRODUCTION

In the field of image analysis, CNNs have emerged as a powerful tool, particularly within the domain of medical imaging. Their success stems partly from their ability to learn spatial features effectively from grid-like image data [1, 2], a common characteristic of many medical images. However, for certain complex medical image modalities and representations that extend beyond simple Euclidean space, CNN architectures may not be optimal. This has paved the way for the exploration of alternative models, such as GNNs, which can handle more intricate data structures.

Our research focuses on applying GCNs to medical image analysis tasks, leveraging their potential advantages in dealing with non-Euclidean data. This form of structured data frequently employs graphs for representation [3, 4]. The early GCN approaches relied on spectral domain convolutions, defined in the graph's Fourier space [5], recent advancements have led to spatial domain convolution methods, offering more flexibility and computational efficiency. Graph-based filtering techniques, increasingly applied in image processing, involve constructing graphs to define relationships between image pixels. These methods extend beyond the restriction of physical proximity. Graphs are formed using both non-local and semi-local connections, thus enabling links between pixels that consider more than just spatial closeness. As highlighted in [6, 7], these techniques excel at identifying and adapting to edges and textures in images, offering effective approaches to image analysis. Critically,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

these filtering techniques pave the way for the application of graph convolutions within GCNs. Convolution as a specific filtering operation defined on the graph allows GCNs to leverage the rich information captured in the graph structure. By aggregating information from neighboring nodes and their features, GCN convolutions can refine the representation of each pixel, considering its context within the broader image.

While GNNs have shown promise in various domains, their application to image analysis tasks, especially those involving grid-structured data like images, presents unique challenges.

Our study introduces a novel Graph Convolutional Neural Network with Enhanced Connectivity (GCNN-EC). Inspired by graph-based image segmentation techniques like normalized cuts, GCNN-EC aims to improve feature representation within GCNs for enhanced structural understanding of images. GCNN-EC leverages n-hop neighbors for both feature aggregation and edge prediction. Feature aggregation collects information from nodes within a defined neighborhood, while the edge prediction mechanism utilizes both pixel coordinates and existing node features to establish new connections that capture meaningful image structures. This optimized structural awareness mitigates the well-known issue of over-smoothing in the context of graph representation of images. Our method (GCNN-EC) centers on dynamically adapting edge features to significantly improve node interaction. We enhance graph representation by focusing on long-range feature aggregation and new edge assignment. These dynamic edge and node features, integrated into GCN, demonstrate superior performance in graph-level classification tasks.

To evaluate the robustness of our GCCN-EC model, we primarily utilize the MedMNIST dataset [8], a diverse benchmark of 2D medical images is detailed in Table 1. Additionally, we demonstrate its broader applicability with experiments on established datasets like CIFAR-10, MNIST, and MSTAR 10 classes.

The contributions of this work are as follows:

- **Dynamic Edge Feature Adaptation in GCNs**: The proposed GCNN-EC introduces a novel mechanism to dynamically adapt edge features, significantly improving the representation of relationships between image elements. This enhances the understanding of image structures and long-range dependencies.
- **Mitigation of Over-Smoothing in Graph Representations for Enhanced Image Representation**: GCNN-EC mitigates the common issue of over-smoothing in graph representations of images by emphasizing the importance of meaningful connected components. This ensures that information propagates effectively within relevant regions, allowing GNNs to capture distinct features and generate accurate image representations.
- **Robust Performance Across Diverse Image Datasets**: Evaluation on a variety of datasets, including medical images (MedMNIST) and established benchmarks (CIFAR-10, MNIST, MSTAR), demonstrates GCNN-EC's strong performance and generalizability across different image domains. This showcases its potential for real-world computer vision applications. GCNN-EC significantly outperforms well-established GNNs across most datasets in this study.
- **Comparison with popular DNNs**: We have provided comparison with the popular DNN architectures such as DenseNet121, ResNet18, and EfficientNet-B0, along with a simple CNN model. This comparsion is to demonstrate the effectiveness of the proposed method with 100 times fewer parameters.

The remainder of this paper is organized as follows: Section 2 provides an overview of CNNs, which are the most popular models in computer vision and then introduces GNNs, tracing their development from Graph Signal Processing (GSP) and highlighting their relevance to image analysis tasks. Section 3 delves into the limitations of GCNs, with a particular focus on the issue of over-smoothing. Section 4 presents our proposed GCNN-EC model, detailing its novel dynamic filter-based features and the architecture of our graph classification network. Section 5 outlines the experimental setup, including the datasets used, implementation details, and evaluation metrics. The results of our experiments are presented and analyzed in Section 6. Finally, Section 7 concludes the paper with a summary of our key findings and proposes potential avenues for future research.

II. BACKGROUND AND RELATED WORK

CNNs have achieved remarkable success in image analysis, including in the medical domain. Their ability to extract hierarchical features from structured image data has made them the dominant paradigm for numerous vision tasks. To effectively benchmark the performance and advantages of graph-based approaches like our proposed GCNN-EC model, it is essential to establish a performance baseline using leading CNN architectures. This comparison offers a reference point for understanding the potential gains offered by GNN-based methods, particularly when dealing with the complexities of medical imagery. In this section, we briefly examine several state-of-the-art CNN models recognized for their robust performance in medical image classification.

A. CNNS

CNNs, as outlined in [9], are named after the convolution operation, involving the overlay of a kernel onto an image grid, which is slid across the grid to extract local pixel information. This operation involves a dot product between the filter and the corresponding elements of the image grid, the results of which are stored in an output matrix often termed a feature map or convolved feature.

The convolution operation systematically extracts spatial hierarchies or patterns, transitioning from local pixel details to broader abstract concepts as the network depth increases. The features extracted from convolution and pooling layers are flattened into a vector, serving as the input for tasks such as classification, segmentation, or feature localization.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

CNNs have been a cornerstone in the field of computer vision [10, 11], demonstrating remarkable performance in tasks such as image classification, object detection, and semantic segmentation, including applications in medical imaging. Recent research has explored the integration of GNNs with CNN-based features and the use of standalone GNNs in medical applications, demonstrating their potential to capture complex relationships and enhance performance in tasks such as medical image classification and analysis [12, 13]. This growing body of work underscores the potential of GNNs as a valuable tool in the medical imaging domain. In our study, we have used three state-of-the-art Deep Convolutional Neural Networks (DCNNs) like ResNet (known for residual blocks) [14], DenseNet (maximizing feature reuse) [15], and EfficientNet (offering superior computational efficiency) [16], to establish performance benchmarks.

While CNNs have undoubtedly been a cornerstone in the field of computer vision, demonstrating remarkable success in tasks such as image classification, object detection, and semantic segmentation [10, 11], they are not without their limitations. Research has highlighted that CNNs can be susceptible to failure when presented with inputs that deviate from the training distribution [17], which can hinder their generalizability to new and unseen data [18]. Moreover, the lack of inherent explainability in CNNs makes it challenging to interpret their decision-making process [19], a critical concern in domains like medical imaging where understanding the rationale behind a model's predictions is paramount. Additionally, CNNs can face challenges in adapting to varying color distributions [20], potentially impacting their performance across diverse datasets [21]. These limitations underscore the ongoing need for research to enhance CNN performance and address issues related to robustness, interpretability, and adaptability in complex image analysis tasks. This has led to a growing interest in exploring alternative or complementary approaches. The unique capabilities of GNNs in capturing complex relationships between features and diverse modalities, as demonstrated in recent investigations [22], offer a promising direction for advancing image analysis capabilities. Studies have shown the potential of GNNs when used alongside CNNs [13] and as standalone models [12], highlighting their effectiveness in various computer vision domains.

The forthcoming section will explore GNNs in-depth, focusing on their fundamental components, historical development, and the challenges they face. We will also discuss various strategies developed to address these challenges [23, 24], emphasizing their applications in medical imaging [25].

B. GNNS

GNNs extend the principles behind convolutional operations in traditional CNNs to the realm of non-Euclidean data represented as graphs. A graph, denoted as $G = (N, E)$, consists of a set of nodes (or vertices) N, and a set of edges E representing connections between them. Edges are formulated as $E \subseteq \{(u, v) \mid u, v \in N\}$. For simplicity, we will focus on undirected graphs where edge ordering does not matter: (u, v) is the same as (v, u).

In image analysis, nodes often represent pixels or regions, while edges might encode spatial relationships, intensity similarities, or other relevant connections. Both nodes and edges can be characterized by features encapsulating their properties. For example, a node feature in a medical image might include pixel intensity or tissue type, while an edge feature could represent the distance between regions or the presence of a boundary. We will refer to a node feature for node n as f_n, and an edge feature between nodes u and v as $f_{(u,v)}$.

A fundamental concept in GNNs is the adjacency matrix, denoted as A. Each element A_{ij} indicates whether a direct connection (edge) exists between nodes i and j. GNNs leverage the adjacency matrix, along with node and edge features, to learn powerful representations of the graph structure, enabling them to solve various image analysis tasks[26].

GNNs have emerged from a rich lineage of research, drawing inspiration from the principles of signal processing, particularly GSP. This perspective empowers us to understand how information propagates and interacts across the intricate connections within a graph structure. Within the broad spectrum of GNNs, we will primarily focus on GCNs as a prominent example. GCNs directly adapt the convolution operation for graph-structured data and have demonstrated remarkable performance in numerous applications. To gain a comprehensive understanding, we will also explore other well-established GNN architectures like Graph Isomorphism Networks (GINs) and Graph Attention Networks (GATs). These additional models will serve as valuable points of comparison in our later discussions, particularly when evaluating their effectiveness for computer vision tasks.

1) Graph Convolutional Networks

(GCNs) enhance the capabilities of Convolutional Neural Networks (CNNs) by extending their powerful feature extraction capabilities to data represented in graph structures. This allows for the processing of non-Euclidean data, which is prevalent in various fields such as social network analysis, neuroscience (specifically in brain connectivity graphs), and computer vision applications involving meshes.

Formally, a GCN can be described as a function $f(X, A)$, where:

- **Node Feature Matrix** X: This is an $N \times F$ matrix, where N represents the number of nodes, and F is the number of features per node. Examples of node features include:
  1) **RGB Values & Coordinates:** These represent the pixel intensities and their respective locations within an image.
  2) **Haralick Features:** These are statistical texture measures that are derived from specific regions of an image.
  3) **CNN-extracted Features:** These high-level repre-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

sentations are obtained from pre-trained CNN models.

- **Adjacency Matrix** A: Previously introduced, the $N \times N$ matrix, which denotes the connections between nodes. It is often normalized to ensure the proper scaling of features, as discussed in various sources, such as [27].

GCNs update node representations through layer-wise propagation, taking into account the information from neighboring nodes. A simplified representation of this propagation rule is illustrated in (1):

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right). \tag{1}$$

where:

- $\tilde{A} = A + I$ is the adjacency matrix with added self-connections.
- $\tilde{D}$ is the degree matrix corresponding to $\tilde{A}$.
- $W^{(l)}$ is the layer-specific trainable weight matrix with dimensions $F^{(l)} \times F^{(l+1)}$.
- $H^{(l)}$ represents the node features at layer l.
- $\sigma(\cdot)$ is a non-linear activation function.

GCNs are inspired by GSP, a discipline that focuses on analyzing signals defined on the nodes or edges of graphs. A fundamental component in GSP is the graph Laplacian, with its unnormalized form defined as $L = D - A$, where D represents the degree matrix.

The Laplacian matrix is crucial for several reasons:

- **Structural Insights:** It encodes vital information about the graph's structure, with its eigenvalues and eigenvectors revealing essential characteristics such as community presence, bottlenecks, and other significant structural patterns within the graph.
- **Normalization's Role:** Techniques like the symmetric normalized Laplacian ($L_{sym} = D^{-1/2}AD^{-1/2}$) are essential for achieving a stable and comparable spectral representation. This normalization is particularly important to prevent high-degree nodes from dominating the spectrum, which could hinder the analysis of other nodes.

Further, the graph Fourier transform, derived from the eigendecomposition of the Laplacian ($L = U\Lambda U^T$), facilitates the analysis of graph signals in a spectral domain, analogous to traditional Fourier analysis. In this domain, the eigenvectors of the Laplacian form a basis for a space akin to Fourier analysis. This approach allows for the dissection and understanding of a graph signal's frequency components, enabling tasks like signal filtering, pattern identification, and comprehending signal smoothness in relation to the graph's structure[28].

Whilst graph signal processing is a powerful technique for filtering, compression, and representation learning [28], it is associated with a high computational complexity of $\mathcal{O}(n^3)$. GCNs provide an innovative solution to the primary limitation of GSP and it achieves this by efficiently approximating spectral convolutions on graphs, utilizing the mathematical properties of Chebyshev polynomials.

This methodological shift has profound implications for how we process graph-structured data. By avoiding the expensive eigendecomposition of the Laplacian matrix, GCNs facilitate a more scalable and efficient means of performing deep learning on graphs. The underlying principles and mechanics of this approximation, particularly the use of Chebyshev polynomials in redefining spectral convolutions, will be examined in greater detail in the following sections. This discussion will delve into the theoretical underpinnings of GCNs and explore how these networks manage to retain the depth and effectiveness of traditional GSP approaches while dramatically reducing their computational burden.

GCNs achieve this by adapting the approach used in ChebNet [29] by employing only the first two terms of the Chebyshev polynomial expansion, effectively setting $K = 1$. Chebyshev interpolation, as explained in [29], is a method to approximate a continuous filter function $h(\hat{\lambda})$ known at a finite number of points $\hat{\lambda}_k$ by a polynomial P_K of degree K such that $h(\hat{\lambda}_k) = P_K(\hat{\lambda}_k)$. This interpolation approach uses Chebyshev nodes rather than uniformly distributed points, leading to faster convergence and avoiding issues like the Runge Phenomenon. The Chebyshev nodes are the zeros of the (K + 1)-th Chebyshev polynomial (The zeros of a Chebyshev polynomial of a certain degree are the points at which the polynomial equals zero), providing an efficient and effective means for polynomial interpolation in spectral graph analysis.

The convolution operation in GCNs is thus represented as:

$$g_\theta * x \approx \sum_{k=0}^{K} \theta_k T_k(\tilde{\Lambda})x \tag{2}$$

where g_θ is the spectral filter function with learnable parameters θ. $*$ denotes the convolution operation. x is the graph signal (e.g., node features). $T_k(\tilde{\Lambda})$ is the Chebyshev polynomial of order k evaluated at the scaled eigenvalue matrix $\tilde{\Lambda}$ of the graph Laplacian.

However, directly computing (2) has high computational complexity due to eigendecomposition. GCNs address this by approximating spectral convolutions. Using a truncated Chebyshev expansion (setting K = 1) and sharing parameters, we arrive at:

$$g_\theta * x \approx \theta_0 T_0(\tilde{\Lambda})x + \theta_1 T_1(\tilde{\Lambda})x \tag{3}$$

Simplifying further by allowing $\theta = \theta_0 = \theta_1$ (which is a design choice in GCNs [30] aimed at reducing the number of learnable parameters and improving computational efficiency) yields:

$$g_\theta * x \approx \theta(I + D^{-\frac{1}{2}}AD^{-\frac{1}{2}})x \tag{4}$$

This demonstrates how GCN layers efficiently aggregate information from a node's neighborhood. This localized operation significantly reduces computational costs compared to full spectral convolutions.

In this section, we introduced GCNs, tracing their origins in GSP and highlighting the computational advancements

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

that have enabled their successful application to large-scale real-world graphs. This has led to their use in tasks like object localization [31], visual question answering [32], video analysis[33], and image classification [34]. While demonstrating remarkable results, GCNs face limitations similar to those in DNNs, such as vanishing and exploding gradients. These limitations can be addressed using techniques like residual connections, which facilitate the learning of long-term dependencies in deeper architectures [35]. A detailed exploration of GCN limitations is presented in Section III.

III. LIMITATIONS OF GCN

While GCNs have made significant strides in encoding graph structures, they do face limitations that researchers continue to address. Some key categories of limitations include:

- **Over-smoothing**: As GCN layers aggregate information across neighborhoods, node representations can become overly similar, especially in deeper architectures. This hinders the ability to distinguish subtle structural differences [36].
- **Limits of the Weisfeiler-Lehmann Test**: GCNs' discriminative power often aligns with the Weisfeiler-Lehmann (WL) graph isomorphism test [37]. This suggests limitations in distinguishing graph structures that the WL test can differentiate.
- **Long-range Dependencies**: Capturing complex, long-range dependencies within the graph structure can be challenging for standard GCN variants [38]. This has led to the development of attention-based mechanisms like Graph Attention Networks (GATs) [39].
- **Expressive Power and Graph Isomorphism**: While achieving significant progress, even advanced GCNs like Graph Isomorphism Networks (GINs) [40] face theoretical limits in their ability to fully capture all nuances that distinguish non-isomorphic graphs.

The predominant limitation in current GCNN models lies in their constrained ability to represent topology [41]. This limitation is further compounded by the issue of over-smoothing [42], a consequence of iterative feature aggregation from connected nodes, leading to increasingly indistinguishable node embeddings. Such over-smoothing challenges the capability of GCNNs to capture complex structural patterns, particularly those beyond the scope of the 1-WL test (where '1' denotes the consideration of 1-hop neighbors) [43]. A detailed discussion on over-smoothing and its implications on the effectiveness of GCNNs in encoding topological nuances will be further elaborated in this section. This exploration is critical for enhancing the representation learned by GCNNs.

A. THE CHALLENGE OF OVER-SMOOTHING IN GCNS

GCNs face the inherent challenge of over-smoothing, especially when many convolutional layers are employed. Over-smoothing in GCNs is primarily due to the nature of graph convolution, which essentially acts as a type of Laplacian smoothing [44]. We can define Over-smoothing as layer-wise

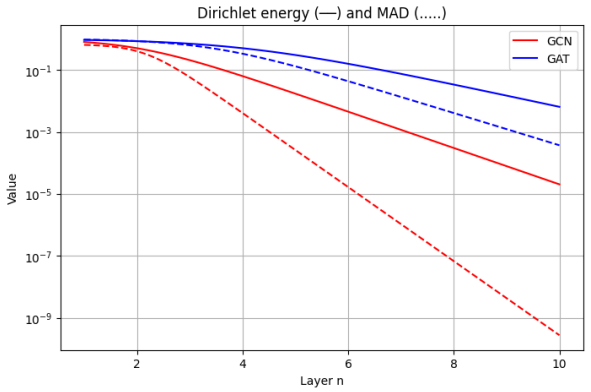


FIGURE 1: Comparison of Dirichlet Energy and MAD across layers for GCN, GAT on synthetic graph data generated using SBM, illustrating the over-smoothing, faster damping of Dirichlet energy means more over-smoothing

exponential convergence of the node-similarity measure μ to zero. As more convolutional layers are added, the distinctive features of vertices from different clusters may blend, leading to indistinguishable node representations [36]. Node similarity measures μ can be estimated in two ways [36]:

- Using Dirichlet energy which is obtained by summing the squared differences of features between adjacent nodes.
- Using Mean Average Distance (MAD) which measures feature similarity by averaging the cosine dissimilarity (one minus the normalized dot product) of features between each node and its neighbors.

In graph-level classification, the effect of smoothing is most pronounced. This is because the final graph-level embeddings are derived by aggregating, or pooling, the features of all nodes. Such pooling inherently smooths out the distinctive features of individual nodes across the graph.

B. EMPIRICAL OBSERVATION WITH BREASTMNIST GRAPH DATASET AND SYNTHETIC DATASET

To investigate the over-smoothing phenomenon, we created a synthetic dataset with two distinct blocks of five nodes each. Nodes within a block are connected, but there are no connections between blocks. Each node has five unique features, ensuring clear differentiation between blocks. Our findings confirm that GCNs' message-passing and aggregation mechanisms inherently lead to over-smoothing of node features. The rate of this feature mixing is influenced by both connectivity and the features themselves as shown in Figure 2. Figure 1 demonstrates this effect using two established over-smoothing metrics – Dirichlet Energy and MAD – on a synthetic graph. Over-smoothing leads to a faster decrease in Dirichlet Energy, indicating a loss of feature distinctiveness.

This adversely affects the downstream task, To illustrate the negative impact of over-smoothing on graph classification, we turn to a real-world example. We see this phenomenon in action with the BreastMNIST dataset. This

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

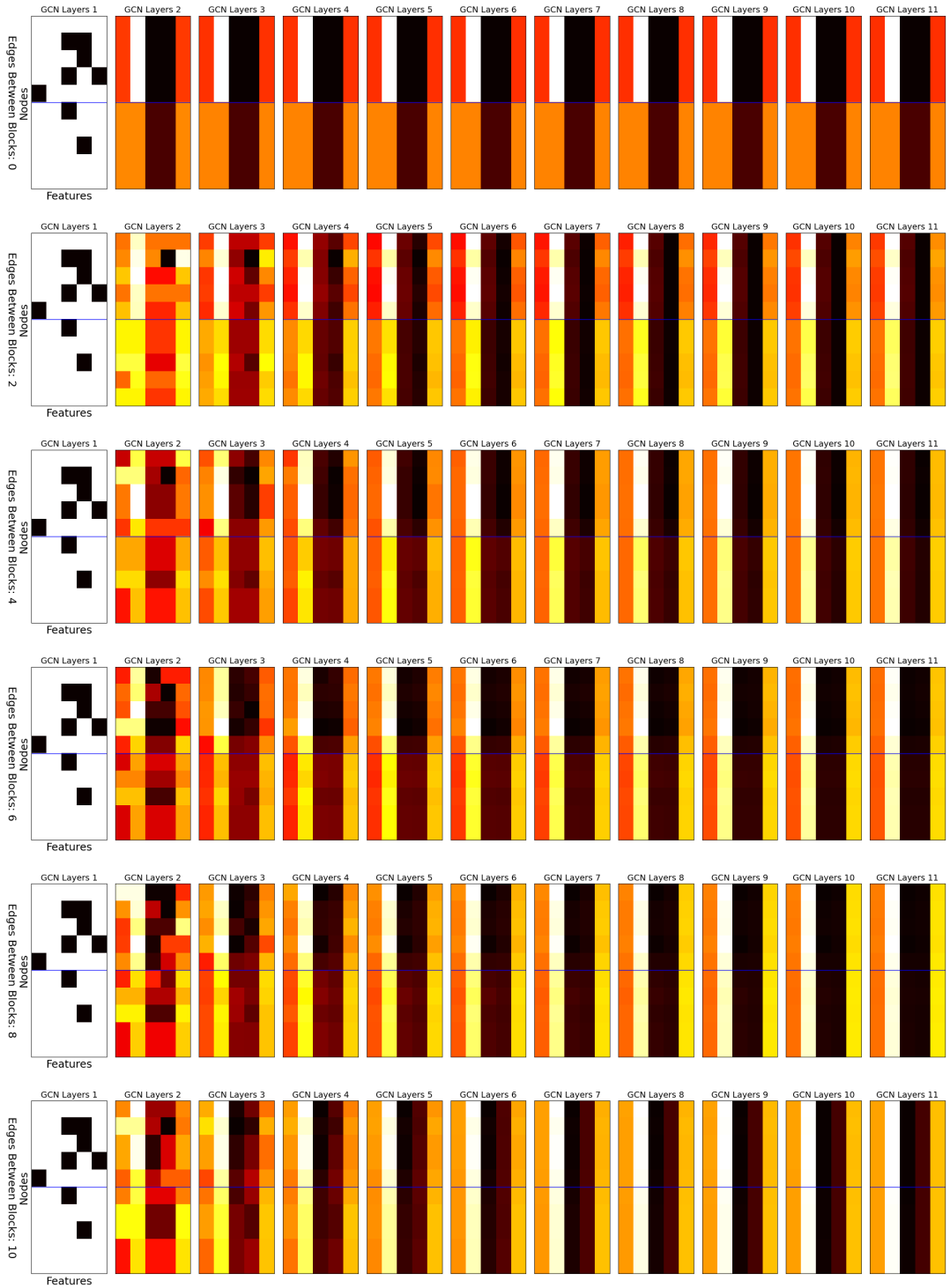


FIGURE 2: **Heatmap visualization of GCN node features over multiple layers.** The y-axis represents nodes (separated by a blue line between blocks), and the x-axis shows the number of GCN layers. Note the lack of connection between blocks, and the progressive blurring of features as layer count and edges between blocks increase, demonstrating over-smoothing.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

dataset consists of 780 graphs representing images, where each pixel serves as a vertex connected to its four neighbors. This creates graphs with 784 vertices and 6,724 edges, divided into two distinct classes. As GCN layers are applied, graph-level features rapidly converge, rendering them indistinguishable. This convergence, driven by the direct correlation between node features and pixel values, hampers graph classification performance (see Figure 3).

C. THEORETICAL INSIGHTS INTO FEATURE CONVERGENCE

Oversmoothing in GCNs has both empirical and theoretical underpinnings. Images transformed into graphs, like the BreastMNIST dataset, often lack structurally significant edges compared to other types of graph data. This leads to the formation of connected components with relatively weak distinctions. Repeated applications of GCNs propagate features within these components, ultimately leading to oversmoothing and hindering classification tasks.

Theoretical analysis reveals that with repeated application of Laplacian smoothing, the features of vertices within each connected component of a graph tend to converge to the same values [44]. For symmetric Laplacian smoothing, this convergence aligns with the square root of the vertex degree. Mathematically, this is expressed as follows:

Suppose that a graph G has k connected components $\{C_i\}_{i=1}^{k}$, and the indication vector for the i-th component is denoted by $1^{(i)} \in \mathbb{R}^n$, which serves to identify the membership of vertices in the component. This vector indicates whether a vertex is in the component C_i, i.e.,

$$1_j^{(i)} = \begin{cases} 1, & \text{if } v_j \in C_i \\ 0, & \text{if } v_j \notin C_i \end{cases}$$

Mathematically, applying a GCN layer can be represented by multiplying the feature matrix X by a specific transformation matrix. This transformation depends on the Laplacian matrix L of the graph, which encodes the connectivity information, and a hyperparameter α that controls the influence of the Laplacian. Normalization of the Laplacian matrix is often employed to stabilize the numerical computations and control the scale of the feature transformations. The two most common types of normalization used are the random walk Laplacian L_{rw} and the symmetric normalized Laplacian L_{sym}.

After repeated applications of these GCN layers (denoted by m iterations), a fascinating property arises: the node features start converging towards a specific combination of the eigenvectors of the Laplacian matrix being used. We can express this mathematically using the following equations:

$$\lim_{m\to+\infty}(I-\alpha L_{rw})^m X = [1^{(1)}, 1^{(2)}, \ldots, 1^{(k)}]\theta_1,$$

$$\lim_{m\to+\infty}(I-\alpha L_{sym})^m X = D^{-\frac{1}{2}}[1^{(1)}, 1^{(2)}, \ldots, 1^{(k)}]\theta_2,$$

Here, I is the identity matrix, and D is a diagonal matrix containing the degrees of each node in the graph. The key elements are:

- **Eigenvectors:** These are special directions associated with the Laplacian matrix. Their properties influence how information propagates through the graph during GCN application.
- **Indicator vectors:** $1^{(i)} \in \mathbb{R}^n$ represents a vector where all elements are zero except for the entries corresponding to nodes in the i^{th} connected component of the graph. These indicate the membership of nodes within connected components.
- **Weight vectors:** θ_1 and θ_2 are vectors in $\mathbb{R}^k$ (where k is the number of connected components) that determine the specific contribution of each indicator vector to the final combination.

While convergence seems like a natural consequence of information propagation, it becomes problematic when the number of GCN layers (iterations, m) becomes very large [42]. In this scenario, the features of nodes within the same connected component become increasingly similar and ultimately converge to the same linear combination of the Laplacian's eigenvectors. This phenomenon is called oversmoothing. Consider a graph representation of an image where nodes correspond to pixels or superpixels, and edges denote neighboring relationships. Over-smoothing in this context leads to a blurring of features within distinct image regions or objects. In the following section, we will explore the implications of these limitations for GCN design.

D. IMPLICATIONS FOR GCN DESIGN

The phenomenon of over-smoothing highlights limitations in conventional GCN designs. This underscores the importance of rethinking how graphs are represented and utilized in GNNs. To mitigate over-smoothing, strategies such as skip connections, have been employed:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) + H^{(l)} \tag{5}$$

This allows the network to retain original feature information at each layer, thereby preserving the uniqueness of node embeddings and combating the over-smoothing effect.

Building on this, we explore additional avenues for enhancing GCN architectures:

- **Beyond Local Neighborhoods**: While traditional message-passing schemes often aggregate features from immediate neighbors, we will explore techniques to incorporate information from extended neighborhoods. This allows for the extraction of richer features that capture both local and long-range dependencies within the graph.
- **Structural Refinement**: We will investigate methods to use the learned features and topological insights to modify or create new edges within the graph. This dynamic graph refinement enables the GCN to better represent complex structural and spatial relationships within the image data.

While skip connections help preserve feature distinctiveness, they have limitations when dealing with node attributes

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

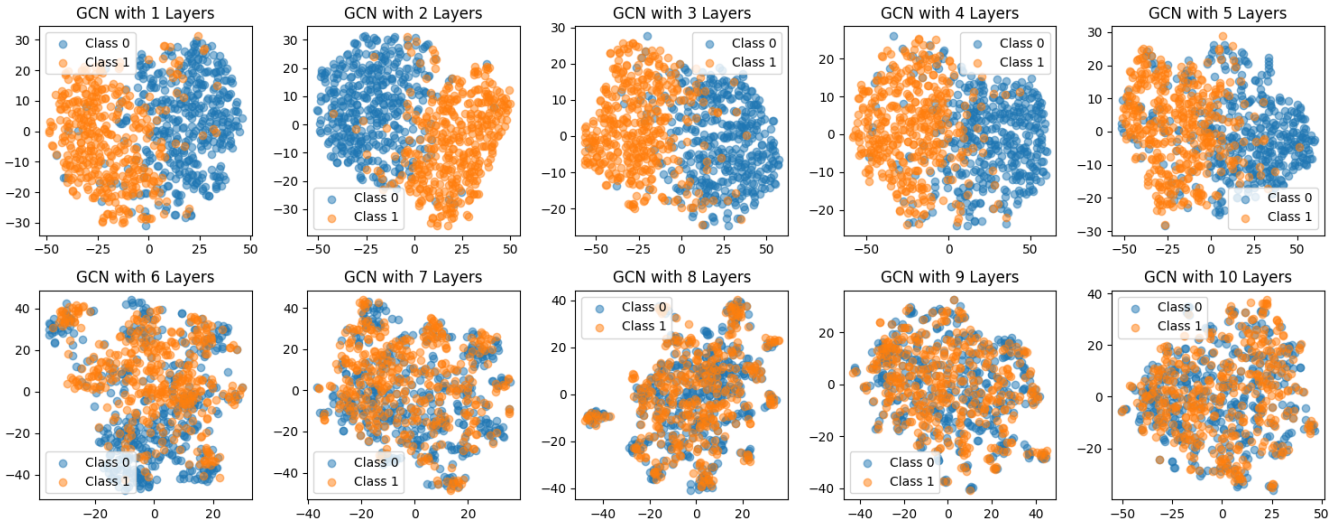


FIGURE 3: **GCN graph level embeddings of BreastMNIST binary classes.** We can observe that increasing the number of layers leads to feature mixing. This mixing results in graph-level embeddings that become less distinguishable between the two classes.

like RGB values, spatial coordinates, or Haralick features. These attributes often lack the richness needed to fully prevent over-smoothing. This underscores the need for a new approach that can better capture and utilize local image patterns. To address this challenge, we introduce an innovative method that dynamically integrates edge features created by applying filters to adjacent nodes. This technique strengthens the model's ability to learn complex patterns around each node, enhancing both connectivity and feature distinction.

IV. METHOD

This section provides a detailed overview of our innovative approach designed to address the challenges faced by conventional GCNs. Central to our method is the dynamic adaptation of edge features, which significantly enhances the interaction between nodes.

Our proposed method harnesses the best of the node features and graph topology by advancing graph connectivity and node interaction. This is achieved by focusing on two ways to enhance graph representation:

1) **Long-Range Feature Aggregation**: This function facilitates the gathering of features from distant nodes, beyond the immediate neighbors typically considered in traditional GCN architectures. By aggregating long-range features, the dynamic filter layer enables a broader view of the graph's topology, thereby enriching the node feature set with diverse and contextually relevant information.
2) **New Edge Assignment**: Utilizing the aggregated long-range features, this layer dynamically assigns new edges within the graph. This process is geared towards enhancing the shape and structure of the connected components.

We have used a *Dynamic Filter Layer* which is responsible for feature aggregation and edge prediction.

Following this, we have *Dynamic Filter Based Features*, where these dynamically computed edge features are integrated to refine node representations. This integration not only improves the individual node characteristics but also the overall structure of the graph.

Subsequently, these elements are incorporated into the *Graph Classification Network*. This integration demonstrates the combined effectiveness of dynamic edge and node features in graph-level classification tasks.

Our method's pipeline is depicted in Figure 4, which illustrates the sequential processing and integration of these innovative components within our enhanced GCN framework.

A. DYNAMIC FILTER LAYER

Our Dynamic Filter Layer draws inspiration from graph partitioning and spectral clustering, particularly the n-cut method. This layer learns filters that capture multi-scale interactions and predict optimal connectivity, leading to refined graph structures.

We begin with the graph Laplacian, denoted as L. To capture interactions between a node and its n-hop neighbors, we apply a polynomial of the Laplacian with learnable coefficients:

$$p(\mathbf{L}) = \sum_{k=0}^{K} \alpha_k \mathbf{L}^k \tag{6}$$

This polynomial is applied to the node feature vectors x:

$$\mathbf{X}_{ij}^{e} = \text{Filter}\left(p(\mathbf{L})\mathbf{x}_i, \frac{1}{|\mathcal{N}i^{(n)}|}\sum u \in \mathcal{N}_i^{(n)} p(\mathbf{L})\mathbf{x}_u\right) \tag{7}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

TABLE 1: Summary of MedMNIST-2D datasets including OR: Ordinal Regression. MC: Multi-Class. ML: Multi-Label. BC: Binary-Class.

Name	Data Modality	Task (Classes / Labels)	Samples	Training / Validation / Test
ChestMNIST	Chest X-Ray	ML (14), BC (2)	112,120	78,468 / 11,219 / 22,433
PathMNIST	Colon Pathology	MC (9)	107,180	89,996 / 10,004 / 7,180
OCTMNIST	Retinal OCT	MC (4)	109,309	97,477 / 10,832 / 1,000
DermaMNIST	Dermatoscope	MC (7)	10,015	7,007 / 1,003 / 2,005
RetinaMNIST	Fundus Camera	OR (5)	1,600	1,080 / 120 / 400
PneumoniaMNIST	Chest X-Ray	BC (2)	5,856	4,708 / 524 / 624
BreastMNIST	Breast Ultrasound	BC (2)	780	546 / 78 / 156
TissueMNIST	Kidney Cortex Microscope	MC (8)	236,386	165,466 / 23,640 / 47,280
BloodMNIST	Blood Cell Microscope	MC (8)	17,092	11,959 / 1,712 / 3,421
OrganAMNIST	Abdominal CT	MC (11)	58,850	34,581 / 6,491 / 17,778
OrganCMNIST	Abdominal CT	MC (11)	23,660	13,000 / 2,392 / 8,268
OrganSMNIST	Abdominal CT	MC (11)	25,221	13,940 / 2,452 / 8,829

In the above equation, Filter denotes the operation of applying $p(\mathbf{L})$ to $\mathbf{x}_i$, i.e., the node feature, along with the n-hop neighbors of $\mathbf{x}_i$ denoted by $\mathcal{N}_i^{(n)}$.

(6) and (7) start with the initial graph and then learn new coefficients for Laplacian as in Algorithm 1.

The edge features generated by Filter, denoted as $\mathbf{Xe}_{ij}$, encapsulate the connectivity potential between nodes i and j, based on their similarity and affinity. Incorporating features from n-hop neighbors enriches the generation of optimally connected components, laying the groundwork for the graph's initial structure.

We extend the utility of these features by integrating a Multi-Layer Perceptron (MLP) to project them into a 16-dimensional space for each edge:

$$\mathbf{Xe}'_{ij} = \text{MLP}(\mathbf{Xe}_{ij}) \tag{8}$$

Subsequently, as outlined in Algorithm 2, the transformed edge feature vector $\mathbf{Xe}'_{ij}$ is processed through an activation function, as demonstrated in Algorithm 3, to predict the presence of edges, thus optimizing the graph structure:

$$\mathbf{e}''_{ij} = \text{Activation(Tanh)}(\mathbf{Xe}'_{ij}) \tag{9}$$

Here, $\mathbf{e}''_{ij}$ denotes the output representing the existence of an edge between nodes i and j, essential for refining the graph's structure.

Through the adoption of dynamic filtering concepts from image processing to GNNs, our approach enables filters to dynamically interact with node and edge features. This dynamic interaction facilitates position-specific transformations and adaptability, closely modeling significant node and edge relationships and aligning with the foundational principles of the n-graph cut method and spectral clustering in a machine learning framework.

B. DYNAMIC FILTER BASED FEATURES

The dynamic filter layer enhances the graph by refining edge connections and integrating edge features. These edge features encapsulate long-range regional representations, which are then merged with node features. This fusion serves a

Algorithm 1 Learned Laplacian Layer

Require: in_ch, out_ch, deg (default=2)
  $coeff \leftarrow \mathbb{R}^{deg+1}$
  $linear \leftarrow$ Linear(in_ch, out_ch)
  Function Forward(x, $edge$):
    $edge, norm \leftarrow$ GetNormalizedLaplacian$(edge)$
    $tx \leftarrow x$
    $out \leftarrow coeff[0] \times tx$
  for $i = 1$ **to** deg **do**
    $tx \leftarrow$ Matmul$(edge, tx, norm)$
    $out \leftarrow out + coeff[i] \times tx$
  end for

  return Linear(out)

Algorithm 2 Dynamic Edge Feature Layer

Require: in_channels, out_channels
  MLP $\leftarrow$ Sequential(Linear(2 $\times$ in_channels, out_channels), ReLU, Linear(out_channels, in_channels))
  Function Forward(x, $edge_index$):
    $row, col \leftarrow$ Decompose $edge_index$
    $edge_features \leftarrow$ Concat$(x[row], x[col])$
    $edge_features \leftarrow$ MLP$(edge_features)$
    $agg_edge_features \leftarrow$ ScatterAdd$(edge_features, row)$

  return $agg_edge_features$

Algorithm 3 Edge Prediction Layer

Require: in_ch
  $MLP \leftarrow$ Sequential(Linear$(in_ch, 1)$, Tanh)
  Function Forward(x, $edge_idx$):
    $r, c \leftarrow$ Decompose$(edge_idx)$
    $pred \leftarrow MLP(x[r] - x[c])$
    $filtered_idx \leftarrow$ Filter based on $(pred > 0)$

  return $filtered_idx, pred$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

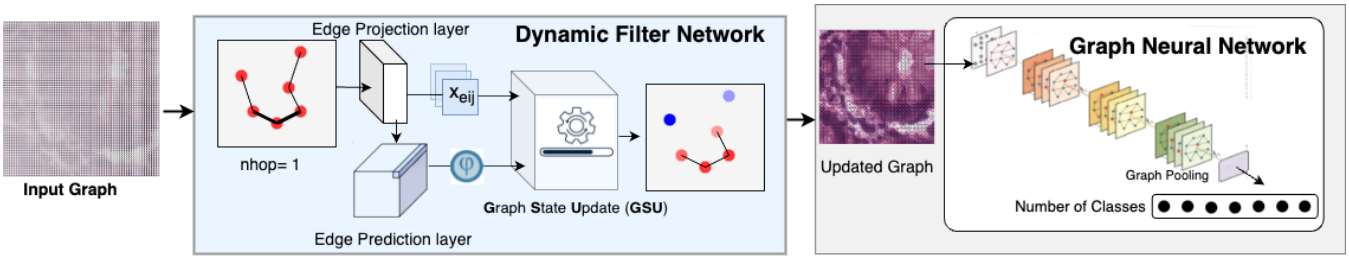


FIGURE 4: **Depiction of the *GraphClassificationNetwork* within our model architecture**, highlighting the flow from dynamic layers through GCN layers to the global mean pooling layer, and ultimately to the classifier for final predictions.

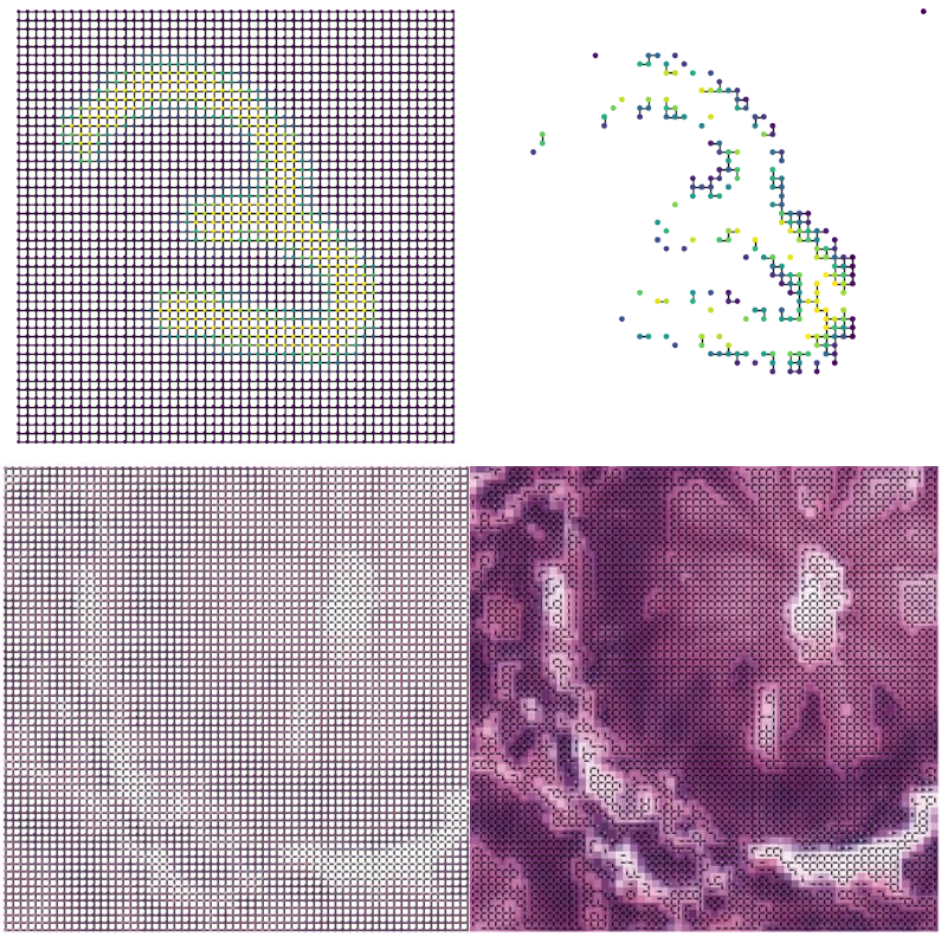


FIGURE 5: **Result of Dynamic filter applied to pixels of MNIST (1 channel) top and MedMNIST (3 channel) bottom.** The final graph representation, which holds the optimal graph structure, is further passed to GCN layers.

dual purpose: while node features capture local characteristics, edge features introduce global or regional context, enriching the detection of distinct communities. Nodes are subsequently updated with this enriched feature set, forming a new graph representation for graph-level classification in our study.

Update Mechanism: Node feature vectors are augmented by concatenating them with the features of their eight adjacent edges. This update is mathematically represented as:

$$\mathbf{X}_i' = [\mathbf{X}_i; \mathbf{X}\mathbf{e}_{i1}; \ldots; \mathbf{X}\mathbf{e}_{i4}] \tag{10}$$

C. GRAPH CLASSIFICATION NETWORK

The *GraphClassificationNetwork* (Figure 4) in our model architecture plays a pivotal role in graph-level classification. This network is an extension of the baseline model [27], which consists of bare-bone GCN layers. Structurally, it combines the Dynamic Filter Layer and Dynamic Filter Features aggregation to enhance the initial graph representation and subsequent GCN layers are responsible for performing deep graph convolution operations. To aggregate node features into a unified graph-level representation, a global mean

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

pooling layer is employed. Finally, the network features a classifier, tasked with generating the final predictions based on these graph-level embeddings.

V. EXPERIMENTAL SETUP

To comprehensively evaluate the effectiveness of our dynamic filtering technique in GCNs for image classification, we conducted experiments on four carefully selected benchmark datasets: MNIST, CIFAR-10, MedMNIST and MSTAR-10. Here's a breakdown of our choices:

- **MNIST:** This widely used dataset provides a well-established baseline for evaluating image classification algorithms. While considered a simpler dataset, MNIST allows us to clearly illustrate the core functionalities of our approach and serves as a foundational test case.
- **CIFAR-10:** Building upon MNIST, CIFAR-10 offers a more challenging dataset with higher image resolution and increased class complexity (10 object classes). This selection allows us to assess how our method scales to more intricate image classification tasks.
- **MedMNIST:** MedMNIST includes different types of medical images like X-rays and MRIs, similar to real-world medical situations. It helps test how well our method works on various complex tasks. The details of the 12 2D image datasets are shown in Table 1 MedMNIST is a powerful tool that has been actively used as a benchmark dataset in numerous studies [45, 46, 47, 48, 49, 50, 51, 52], proving to be highly valuable for evaluating model performance and advancing research in medical image analysis. Its utility stems from its diverse collection of standardized medical image datasets, ranging from small to large scales and covering a variety of modalities. The availability of both 28×28 and 224×224 image sizes further enhances its flexibility, allowing for fast benchmarking and comprehensive evaluation of different models.
- **MSTAR-10:** This dataset focuses on the classification of military vehicles in Synthetic Aperture Radar (SAR) imagery. The unique challenges of SAR data, such as speckle noise and viewpoint variations, provide a distinct test case for our method. Its inclusion expands the scope of our evaluation to a different imaging modality and a domain-specific application.

By using this diverse array of datasets, we aimed to compare the effectiveness of our GCNN-EC model across different domains, varying levels of difficulty, and color schemes. This strategic choice enhances our evaluation, showcasing the model's adaptability and effectiveness in a broad spectrum of image classification challenges, including tasks within the medical field.

A. IMPLEMENTATION DETAILS

In our research, we established a basic model using a simple GCN structure, which includes two layers of GCN convolution (GCNConv). We implemented this model in the PyTorch framework, specifically leveraging the PyTorch Geometric (PyG) library, recognized for its effectiveness in geometric deep learning applications. We established a GCN model using a four-layer GCN convolution structure (GCNConv). As with our Dynamic Filter + GCN model, this model is implemented within the PyTorch framework, leveraging the PyTorch Geometric (PyG) library. Both models feature four layers, allowing us to directly isolate and evaluate the impact of the Dynamic Filter layer. To construct graphs from images, we represent each pixel as a node. Nodes are connected using a 4-neighbor connectivity pattern. Each node's feature vector incorporates the pixel's RGB values along with its spatial coordinates (x, y) within the 2D image. These graph representations were formatted for compatibility with the PyG dataloader, facilitating both training and testing phases.

B. EXPERIMENT

In our experiment, the model was trained using a batch size of 64 on an NVIDIA T4 with 12 GB of GPU memory and 32 GB of RAM. We implemented early stopping to prevent overfitting. The learning rate was set at 0.001, coupled with a weight decay of 0.001. Dropout rates of 0.3 and 0.5 were applied across the two layers to enhance generalization. The training utilized cross-entropy loss and the Adam optimizer. Our model underwent training for 100 epochs, with early stopping criteria. The optimal learning rate was determined through fine-tuning using the Optuna hyperparameter optimization framework [53], ensuring the best possible model performance.

VI. RESULTS

Our GCN-EC model implements dynamic filtering to optimize graph configurations, notably enhancing the formation of distinct connected components. This effect is directly observable in Figure 5, where connected components enriched by node features from their k-hop neighbors are clearly distinguishable.

Further validating our model's performance, we present a comparative analysis of the MedMNIST dataset in Table 2. Our approach achieves performance comparable to that of advanced DNNs, yet it requires significantly fewer parameters, a comparison detailed in Table 4. This demonstrates not only the efficiency of our model but also its ability to maintain high accuracy with a more compact model structure.

Figure 2 demonstrates how features from two connected blocks become indistinguishable as more GCN layers are added, reflecting a typical over-smoothing issue. In contrast, Figure 8 shows that feature mixing is effectively restricted through continuous changes in graph topology, facilitated by the dynamic filter layers. This is further substantiated by Figure 7 highlights how GCNN-EC generates clearly distinguishable graph embeddings, contrasting with the less discriminative embeddings produced by traditional GCNs (Figure 3). This separation visually reinforces our model's ability to mitigate the over-smoothing problem. Furthermore, Figure 6 directly reveals this difference in over-smoothing behavior. We observe that the cosine similarity between

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

TABLE 2: **Comparison of various methods on MedMNIST datasets.** The best-performing model for each of the MedMNIST datasets is highlighted (ACC = accuracy, AUC = area under the ROC curve).

Methods	PathMNIST		ChestMNIST		DermaMNIST		OCTMNIST		PneumoniaMNIST		RetinaMNIST	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
DenseNet121	0.983	0.907	0.768	0.947	**0.917**	0.735	0.943	0.743	0.944	0.854	0.717	0.524
EfficientNetB0	**0.990**	**0.911**	0.769	**0.947**	0.913	0.735	0.952	0.762	0.948	0.854	**0.726**	**0.528**
ResNet-18	0.959	0.834	0.742	0.937	0.915	0.749	**0.955**	**0.763**	0.947	0.878	0.719	0.503
CNN (2 conv)	0.953	0.739	0.540	0.947	0.816	0.669	0.895	0.566	0.900	0.808	0.665	0.435
GCN	0.641	0.450	0.811	0.630	0.778	0.425	0.488	0.256	0.460	0.625	0.460	0.267
GIN	0.870	0.703	0.811	0.781	0.793	0.661	0.744	0.694	0.451	0.625	0.661	0.437
GAT	0.558	0.410	0.838	0.761	0.661	0.477	0.610	0.400	0.460	0.625	0.511	0.305
GCNN-EC	0.944	0.791	**0.937**	0.874	0.892	**0.755**	0.903	0.761	**0.954**	0.877	0.712	0.430

Methods	BreastMNIST		BloodMNIST		TissueMNIST		OrganAMNIST		OrganCMNIST		OrganSMNIST	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
DenseNet121	**0.901**	**0.863**	0.998	0.958	0.930	0.676	**0.997**	**0.935**	0.992	0.900	0.972	0.782
EfficientNetB0	0.857	0.812	0.997	0.956	0.931	0.680	0.997	0.935	**0.992**	**0.905**	0.972	0.770
ResNet-18	0.871	0.831	**0.998**	**0.961**	**0.941**	**0.703**	0.994	0.905	0.990	0.879	**0.974**	**0.813**
CNN (2 conv)	0.633	0.731	0.954	0.750	0.866	0.559	0.821	0.780	0.947	0.818	0.771	0.553
GCN	0.688	0.419	0.373	0.119	0.377	0.230	0.701	0.435	0.635	0.485	0.741	0.435
GIN	0.713	0.670	0.738	0.504	0.716	0.677	0.594	0.605	0.667	0.451	0.249	0.326
GAT	0.769	0.668	0.511	0.384	0.738	0.529	0.601	0.472	0.260	0.315	0.301	0.228
GCNN-EC	0.865	0.795	0.909	0.869	0.771	0.599	0.937	0.834	0.929	0.751	0.904	0.674

the final and pre-final layer features increases more slowly in GCNN-EC compared to standard GCN and GAT models. This slower homogenization of features, coupled with a less drastic decline in accuracy, further underscores the effectiveness of our dynamic filtering approach in creating informative graph-level embeddings.

Our model demonstrates competitive performance against leading CNNs on the MedMNIST dataset, particularly excelling in specific classes, detailed in Table 3 while operating with significantly fewer parameters ranging from 10 to 100 times less as shown in Table 3, it surpasses simple CNN with similar parameter size. Additionally, it surpasses all other GNNs on the MNIST, CIFAR-10, and MSTAR-10 datasets, effectively addressing the critical challenge of over-smoothing and underscoring the robustness of our dynamic filtering approach.

A. DISCUSSION

The proposed GCNN-EC model, incorporating the dynamic filter layer, can be seamlessly integrated with existing GCN layers within the PyTorch Geometric (PyG) package. Other models can also leverage the dynamic filter layer, which can utilize input node features and generate refined features that contribute to a more effective graph representation of images by applying filters over connected nodes and their associated features. In our experiments, we found that an input node feature size of 5 (r, g, b, x, y), subsequently transformed to feature sizes of 16 and 32, yielded optimal results in terms of creating distinct connected components. The output of the layer can be further processed by additional GCN or other neural network layers to extract higher-level features, demonstrating its versatility and adaptability within existing architectures. This flexibility could be particularly beneficial for medical imaging applications, where there is often a need for low-power and efficient models that can be deployed on edge devices or battery-powered devices (e.g., portable ultrasound machines) [54]. The significantly fewer parameters required by GCNN-EC compared to popular DNNs, as illustrated in Table 4, highlight the potential of GNNs for developing computationally efficient models. This could be particularly advantageous in medical imaging, where computational resources may be limited. Furthermore, the explainability of GNNs, as discussed in [55], is a significant advantage over traditional CNNs, especially in the medical domain where understanding the reasoning behind a model's decisions is crucial for building trust and ensuring safety [56].

TABLE 3: **Graph Classification Results (Accuracy)**

METHOD	MNIST	CIFAR-10	MSTAR10
GCN-4	0.901	0.541	0.568
GIN-4	0.964	0.552	0.731
GAT-4	0.955	0.642	0.580
GCNN-EC	**0.971**	**0.661**	**0.951**

TABLE 4: **Parameter counts of various models**

METHOD	Feature size	Layers	Parameter (for RGB dataset)
DenseNet121	3*224*224	_	7,978,856
ResNet18	3*224*224	_	11,689,512
EfficientNetB0	3*224*224	_	4,014,658
CNN	3*32*32	conv(2)	155,097
GCN	5	GCNConv(4)	44,194
GIN	5	GINConv(4)	64,198
GAT	5	GATConv(4)	671,586
GCNN-EC	5	GCNN-EC(4)	95,210

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

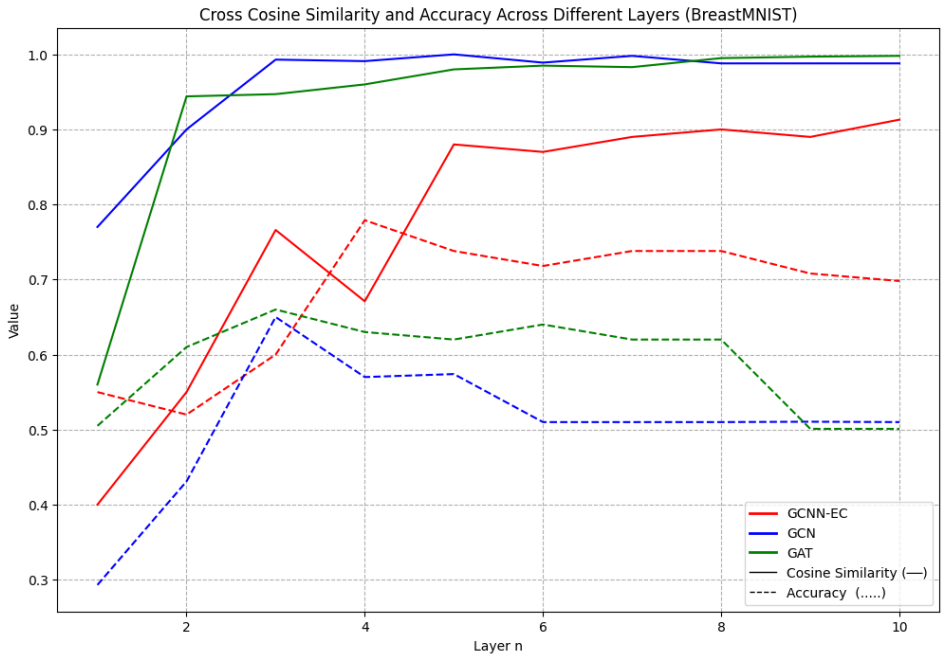


FIGURE 6: **Feature Evolution in Layered GNNs for BreastMNIST Graphs:** The plot illustrates the cosine similarity across 10 layers in GCN models, contrasting the quick convergence in standard GCN and GAT with the gradual increase observed in GCN-EC. This suggests the potential of GCN-EC to maintain diverse feature learning across layers, avoiding over-smoothing.

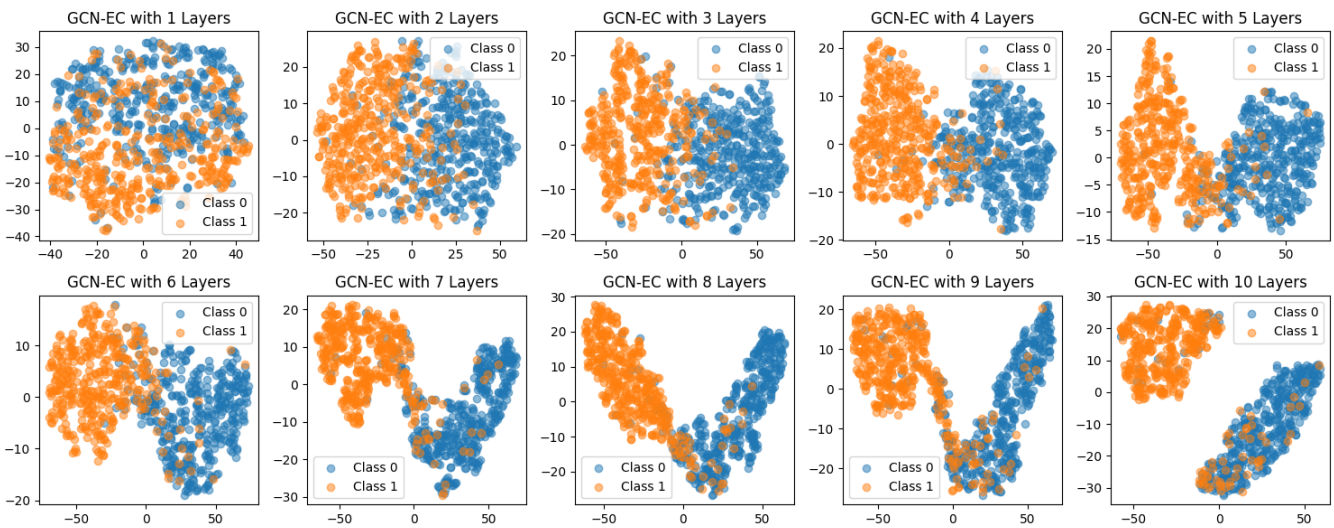


FIGURE 7: **GCNN-EC graph level embeddings of BreastMNIST binary classes visualized.** The overmixing of features is mitigated, resulting in graph-level embeddings that are more distinguishable between the two classes.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042



FIGURE 8: **Heatmap visualization of GCNN-EC node features across various layers shows clear differentiation.** The y-axis categorizes nodes, divided by a blue line between blocks, while the x-axis represents the progression through GCNN-EC layers. Notably, there is minimal feature mixing between blocks as the layers advance, and the connected components become more distinct. This visualization highlights the model's ability to limit over-smoothing, maintaining clearer separations and more defined features across layers, contrary to typical GCN models where increased layer depth blurs feature distinctiveness.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

VII. CONCLUSION

In this paper, we introduced an innovative dynamic filter layer for GNNs, inspired by spectral graph clustering techniques. This layer significantly improves graph-level classification by focusing on long-range feature aggregation and dynamic edge assignment. Our method is computationally efficient, adaptable to various image types, and demonstrates a clear advantage in creating more discriminative graph representations. Our results convincingly demonstrate that GCNN-EC optimizes graph configurations, leading to the formation of distinct connected components. This not only translates to comparable performance to advanced DNNs on the MedMNIST dataset while requiring significantly fewer parameters, underscoring our model's efficiency and power but also highlights the improved ability of the model to distill information from the images. The fewer parameters and the distinct connected components produced by the model showcase its ability to not only extract information effectively but also maintain the separation of distinct information within the image, contributing to enhanced performance and interpretability.

Crucially, GCNN-EC effectively mitigates the common over-smoothing problem in GCNs. Through dynamic changes in graph topology, our model maintains distinguishable features and generates superior graph-level embeddings. This advantage is further solidified by outperforming leading GNNs across three benchmark datasets.

For future work, we aim to further optimize performance by exploring more sophisticated node representations. Instead of solely relying on individual pixels, we will investigate the use of groups of homogeneous pixels, bounding boxes, and feature-based connections as node candidates. This will involve incorporating feature representations from pre-trained DNNs, or Vision Transformers (ViTs) to leverage both local and global image characteristics. Additionally, we will integrate the dynamic filter layer with more complex node representations and broaden its compatibility within the PyTorch Geometric framework. This will facilitate extensive testing and validation of the dynamic filter concept's applicability across a wider range of tasks and datasets in the medical imaging domain.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in Lecture Notes in Computer Science, vol. 2014. Springer International Publishing, 2014, pp. 346–361. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10578-9_23

[2] S. Gulshad, T. Long, and N. van Noord, "Hierarchical explanations for video action recognition," arXiv, vol. abs/2301.00436, 2023. [Online]. Available: https://arxiv.org/abs/2301.00436

[3] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," 2017. [Online]. Available: https://arxiv.org/abs/1711.05225

[4] H. Lee, H. Park, and K. Yoon, "Towards better generalization with flexible representation of multi-module graph neural networks," 2022. [Online]. Available: https://arxiv.org/abs/2209.06589

[5] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013. [Online]. Available: https://arxiv.org/abs/1312.6203

[6] J. Galon, A. Costes, F. Sanchez-Cabo, A. Kirilovsky, B. Mlecnik, C. Lagorce-Pagès, M. Tosolini, M. Camus, A. Berger, P. Wind, F. Zinzindohoué, P. Bruneval, P. H. Cugnenc, Z. Trajanoski, W. H. Fridman, and F. Pagès, "Type, density, and location of immune cells within human colorectal tumors predict clinical outcome," Science, 2006. [Online]. Available: https://doi.org/10.1126/science.1129139

[7] Y. Shen, B. Zhou, X. Xiong, R. Gao, and Y. G. Wang, "How gnns facilitate cnns in mining geometric information from large-scale medical images," arXiv, 2022. [Online]. Available: https://arxiv.org/abs/2206.07599

[8] J. Yang, R. Shi, and B. Ni, "MedMNIST classification decathlon: A lightweight AutoML benchmark for medical image analysis," in 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, apr 2021. [Online]. Available: https://doi.org/10.1109%2Fisbi48211.2021.9434062

[9] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[10] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. V. Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," Proceedings of the IEEE, 2021. [Online]. Available: https://doi.org/10.1109%2Fjproc.2021.3054390

[11] M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," 2018. [Online]. Available: https://arxiv.org/abs/1811.04241

[12] Q. Liu, M. Kampffmeyer, R. Jenssen, and A.-B. Salberg, "Scg-net: Self-constructing graph neural networks for semantic segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2009.01599

[13] A. Singh, P. V. de Ven, C. Eising, and P. Denny, "Connecting the dots: Graph neural network powered ensemble and classification of medical images," 2023. [Online]. Available: https://arxiv.org/abs/2311.07321

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016. [Online]. Available: https://arxiv.org/abs/1608.06993

[16] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2019. [Online]. Available: https://arxiv.org/abs/1905.11946

[17] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," Journal of Big Data, vol. 8, no. 1, p. 53, 2021. [Online]. Available: https://doi.org/10.1186/s40537-021-00444-8

[18] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, "Reducing overfitting and improving generalization in training convolutional neural network (cnn) under limited sample sizes in image recognition," in 2020 - 5th International Conference on Information Technology (InCIT), 2020, pp. 300–305.

[19] J. Gu, "Explainability and robustness of deep visual classification models," 2023. [Online]. Available: https://arxiv.org/abs/2301.01343

[20] P. Elavarthi, J. Lee, and A. Ralescu, "On the ability of cnns to extract color invariant intensity based features for image classification," 2023. [Online]. Available: https://arxiv.org/abs/2307.06500

[21] A. Hashemi, Y. Feng, and H. Sabet, "Spherical cnn for medical imaging applications: Importance of equivariance in image reconstruction and denoising," 2023.

[22] P. Celard, E. L. Iglesias, J. M. Sorribes-Fdez et al., "A survey on deep learning applied to medical images: From simple artificial neural networks to generative models," Neural Computing and Applications, vol. 35, p. 2291–2323, 2023. [Online]. Available: https://doi.org/10.1007/s00521-022-07953-4

[23] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Distance metric learning using graph convolutional networks: Application to functional brain networks," arXiv, 2017. [Online]. Available: https://arxiv.org/abs/1703.02161

[24] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," IEEE Signal Processing Magazine, 2017. [Online]. Available: https://doi.org/10.1109/MSP.2017.2693418

[25] K. Ding, M. Zhou, Z. Wang, Q. Liu, C. W. Arnold, S. Zhang, and D. N. Metaxas, "Graph convolutional networks for multi-modality medical imaging: Methods, architectures, and clinical applications," 2022.

[26] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," IEEE Transactions on Neural Networks, vol. 20, no. 1, pp. 61–80, 2009.

[27] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," CoRR, vol. abs/1609.02907, 2016. [Online]. Available: http://arxiv.org/abs/1609.02907

[28] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges and applications," arXiv, 2018. [Online]. Available: https://arxiv.org/abs/1712.00468

[29] Author(s), "Chebyshev polynomial approximates is one such approach," Source/URL, Year.

[30] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," CoRR, vol. abs/1609.02907, 2016. [Online]. Available: http://arxiv.org/abs/1609.02907

[31] R. Vishwakarma, R. B. Joshi, and S. Mishra, "Indoorgnn: A graph neural network based approach for indoor localization using wifi rssi," in Lecture Notes in Computer Science. Springer Nature Switzerland, 2023, pp. 150–165.

[32] W. Norcliffe-Brown, E. Vafeias, and S. Parisot, "Learning conditioned graph structures for interpretable visual question answering," arXiv, 2018. [Online]. Available: https://arxiv.org/abs/1806.07243

[33] ——, "Learning conditioned graph structures for interpretable visual question answering," arXiv, 2018. [Online]. Available: https://arxiv.org/abs/1806.07243

[34] A. Singh, P. V. de Ven, C. Eising, and P. Denny, "Connecting the dots: Graph neural network powered ensemble and classification of medical images," arXiv, 2023. [Online]. Available: https://arxiv.org/abs/2311.07321

[35] B. Huang and K. M. Carley, "Residual or gate? towards deeper graph neural networks for inductive graph representation learning," arXiv, 2019. [Online]. Available: https://arxiv.org/abs/1904.08035

[36] T. K. Rusch, M. M. Bronstein, and S. Mishra, "A survey on oversmoothing in graph neural networks," arXiv, 2023. [Online]. Available: https://arxiv.org/abs/2303.10993

[37] B. Y. Weisfeiler and A. A. Leman, "The reduction of a graph to canonical form and the algebra which appears therein," 1968.

[38] Z. Wu, P. Jain, M. A. Wright, A. Mirhoseini, J. E. Gonzalez, and I. Stoica, "Representing long-range context for graph neural networks with global attention," 2022.

[39] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017. [Online]. Available: https://arxiv.org/abs/1710.10903

[40] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" arXiv, 2019. [Online]. Available: https://arxiv.org/abs/1810.00826

[41] D. Jin, X. Song, Z. Yu, Z. Liu, H. Zhang, Z. Cheng, and J. Han, "Bite-gcn: A new gcn architecture via bidirectional convolution of topology and features on text-rich networks," arXiv, 2021. [Online]. Available:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042

https://arxiv.org/abs/2010.12157

[42] U. Alon and E. Yahav, "On the bottleneck of graph neural networks and its practical implications," arXiv, 2021. [Online]. Available: https://arxiv.org/abs/2006.05205

[43] D. Chen, Y. Lin, W. Li, P. Li, J. Zhou, and X. Sun, "Measuring and relieving the over-smoothing problem for graph neural networks from the topological view," arXiv, 2019. [Online]. Available: https://arxiv.org/abs/1909.03211

[44] Q. Li, Z. Han, and X.-M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," arXiv, 2018. [Online]. Available: https://arxiv.org/abs/1801.07606

[45] M. M. Derakhshani, I. Najdenkoska, T. V. Sonsbeek, X. Zhen, D. Mahapatra, M. Worring, and C. G. M. Snoek, "Lifelonger: A benchmark for continual disease classification," International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022.

[46] W. Lin, Z. Zhao, X. Zhang, C. Wu, Y. Zhang, Y. Wang, and W. Xie, "Pmc-clip: Contrastive language-image pre-training using biomedical documents," International Conference on Medical Image Computing and Computer-Assisted Intervention, 2023.

[47] J. Liu, Y. Li, G. Cao, Y. Liu, and W. Cao, "Feature pyramid vision transformer for medmnist classification decathlon," IEEE International Joint Conference on Neural Network, 2022.

[48] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, "Medmnist v2 - a large-scale lightweight benchmark for 2d and 3d biomedical image classification," Scientific Data, 2023.

[49] Z. Zheng and X. Jia, "Complex mixer for medmnist classification decathlon," arXiv.org, 2023.

[50] F. D. Salvo, S. Doerrich, and C. Ledig, "Medmnist-c: Comprehensive benchmark and improved classifier robustness by simulating realistic image corruptions," null, 2024.

[51] S. Doerrich, F. D. Salvo, J. Brockmann, and C. Ledig, "Rethinking model prototyping through the medmnist+ dataset collection," 2024. [Online]. Available: https://arxiv.org/abs/2404.15786

[52] O. N. Manzari, H. Ahmadabadi, H. Kashiani, S. B. Shokouhi, and A. Ayatollahi, "Medvit: A robust vision transformer for generalized medical image classification," Computers in Biology and Medicine, vol. 157, p. 106791, May 2023. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2023.106791

[53] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," 2019.

[54] P. William Davros, Medical Imaging: Principles, Detectors, and Electronics. Wiley, 2009, first published: 27 October 2009. [Online]. Available: https://doi.org/10.1118/1.3246610

[55] C. Agarwal, O. Queen, H. Lakkaraju et al., "Evaluating explainability for graph neural networks," Scientific Data, vol. 10, p. 144, 2023. [Online]. Available: https://doi.org/10.1038/s41597-023-01974-x

[56] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of explainable ai techniques in healthcare," Sensors (Basel), vol. 23, no. 2, p. 634, 2023. [Online]. Available: https://doi.org/10.3390/s23020634



ARYAN SINGH Aryan Singh holds a B.Tech in Computer Science from Delhi Technological University and is currently pursuing his PhD at the University of Limerick. His professional journey includes brief stints as a Java Developer at Mobikwik and a Backend Developer at Netcore Cloud Solutions, focusing on data analytics. His PhD research is concentrated on the application of deep learning in medical image analysis and graph-based data representation. Aryan aims to contribute to advancements in data science that benefit medical diagnostics and understanding of image data.



CIARÁN EISING (SENIOR MEMBER, IEEE) got his BE in 2003 and PhD in 2010 from NUI Galway. After, he worked in autonomous driving perception at Valeo. He joined the University of Limerick as an Associate Professor in 2020.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444042



PATRICK DENNY earned his B.Sc. in Experimental Physics and Mathematics from NUI Maynooth, Ireland, in 1993, and his M.Sc. in Mathematics from the University of Galway, Ireland, in 1994. He completed his Ph.D. in Physics at the University of Galway in 2000, conducting part of his research at GFZ Potsdam, Germany. He has been an IEEE member since 2010. From 1999 to 2001, Denny was an RF Engineer at AVM GmbH, Germany, where he contributed to developing the RF hardware for the first integrated GSM/ISDN/USB modem. He then worked in supercomputing at Compaq-HP between 2001 and 2002, before joining Connaught Electronics Ltd. (later Valeo) in Galway, Ireland as the Team Leader of RF Design. Over the following two decades, he led efforts in developing innovative RF and imaging systems, including the first mass-production HDR automotive cameras for prominent car manufacturers such as Jaguar Land Rover, BMW, and Daimler. In 2010, he was appointed Adjunct Professor of Engineering and Informatics at the University of Galway. In 2022, he became a Lecturer in Artificial Intelligence in the Department of Electronic and Computer Engineering at the University of Limerick, and in 2024, he advanced to Associate Professor of Artificial Intelligence and Imaging in the Department of Computer Science and Information Systems (CSIS) at the University of Limerick. He is also a co-founder and committee member of the IEEE P2020 Automotive Imaging Standards Group, the AutoSens Conference on Automotive Imaging, and the IS&T Electronic Imaging Autonomous Vehicles and Machines (AVM) conference.



PEPIJN VAN DE VEN obtained his MSc in Electronic Engineering from the Eindhoven University of Technology, Netherlands, in 2000 and completed his PhD in Artificial Intelligence for autonomous underwater vehicles at the University of Limerick in 2005. He is currently serving as a Professor of Artificial Intelligence at the University of Limerick. His research primarily focuses on artificial intelligence and machine learning, with a special emphasis on applications in the medical field.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

Dynamic Modelling and PI Control Design for Bidirectional SEPIC/ZETA DC-DC Converter in Battery Applications

Prashant Singh
Electrical Engineering Department
Delhi Technological University
prashantsingh0719@gmail.com

Alka Singh
Electrical Engineering Department
Delhi Technological University
alkasingh.eed@gmail.com

Ankita Arora
Electrical Engineering Department
Delhi Technological University
aroraankita@dtu.ac.in

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10626991

***Abstract*: The effective management of energy flow is essential in modern power systems, especially when it comes to applications involving Battery Energy Storage System (BESS). The bidirectional SEPIC/ZETA DC-DC Converter offers a promising solution that enables battery charging and discharging operations between the battery and the grid or other power sources. In this work, we describe the simulation of a proposed Bidirectional DC-DC Converter employing a state-space circuit averaged modelling technique to determine the transfer function for managing the switching pulses of the converter via a PID controller in both SEPIC and ZETA modes. The tuning process for the controller leverages the classic Ziegler-Nichols technique. The converter utilizes a dual loop control strategy implemented in MATLAB/Simulink, for constant current charging and discharging of the battery in which outer voltage control loop as well as inner current control loop is involved. The functionality of the bidirectional power flow control system has been thoroughly examined using MATLAB/Simulink simulations, ensuring its efficacy during both battery charging and discharging phases across diverse conditions.**

Keywords — Bidirectional SEPIC/ZETA DC-DC Converter, State Space Modelling, Constant Current Charging and Discharging, PID Controller, Ziegler-Nichol's Technique.

I. INTRODUCTION

The importance of clean energy sources has surged due to the detrimental effects of fossil fuel usage on the environment. In recent times, there has been rapid development in PV system, fuel-cell, and wind-power generating systems. However, these systems often face challenges in providing stable power output. To address this issue, hybrid power systems combining renewable energy sources with batteries have emerged. In such systems, when renewable sources fail to meet the power demand, batteries step in to supply the shortfall. Additionally, these batteries can be used to store excess energy generated. Battery Energy Storage System (BEES) is generally performed as a storage system for energy. Bidirectional DC-DC converters play a crucial role in facilitating power transfer between different DC sources, making them essential components in renewable energy hybrid power systems, hybrid electric vehicles, and uninterruptible power supplies.[1]-[6].

Bidirectional DC-DC converters serve a dual purpose in battery charging and discharging systems. They not only manage the flow of current during charging and discharging but also ensure that the output voltage of the discharger remains steady at a set level when utilizing stored battery energy again. In scenarios where unidirectional DC-DC converters are used instead, two separate converters would be necessary to handle each direction of energy flow.[7]-[10]

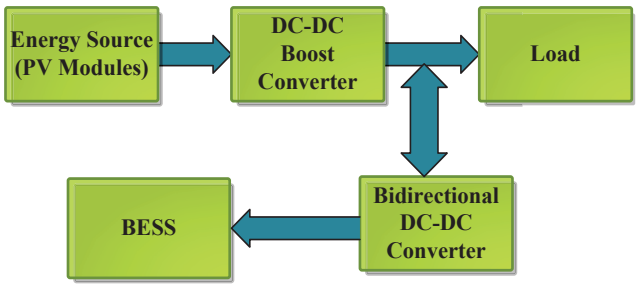


Fig. 1. System Block Diagram

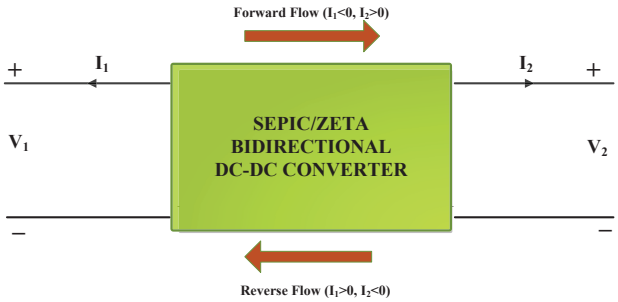


Fig. 2. Bidirectional SEPIC/ZETA DC-DC Converter

Fig. 2 shows the block diagram of the Bidirectional SEPIC/ZETA DC-DC Converter in which forward and reverse power flow condition have been shown. [11]

The proposal suggests utilizing a Bidirectional SEPIC/Zeta converter due to its advantageous feature of lower output voltage ripple when compared to other converters. This converter operates in two modes. In SEPIC mode, the battery acts as a source, supplying power to the DC grid with a voltage boost (Battery Discharging Mode). Conversely, Zeta mode allows the converter to draw power from the grid and charge the battery at a lower voltage (Battery Charging Mode). While an energy management system would typically handle mode selection, this chapter focuses on exploring SEPIC and ZETA modes independently.

The Bidirectional SEPIC/ZETA converter is used in various applications like EV battery management system, UPS, HVDC transmission, smart grid, etc [12]. This paper explores the transfer function by performing converter's mathematical analysis relating the converter's output voltage to the duty cycle of its switches. The focus is on battery charging and discharging operations, with a deep dive into the different modes of the SEPIC/ZETA converter so that battery is charging and discharging with constant current. To optimize the performance of the PI controllers, the Ziegler-Nichols method is employed to determine the appropriate values for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:54:54 UTC from IEEE Xplore. Restrictions apply.

the proportional (Kp), integral (Ki), gains. The proportional component of controllers enhances transient response speed, while the integral component improves steady-state response and diminishes steady-state error. [13]

This paper includes: Section II of the paper contains the system under consideration in which battery specification table and proposed SEPIC/ZETA Bidirectional circuit are mentioned in detail. Section III of this study delves the operating modes of Bidirectional converter. Section IV of the work presents the State Space modelling of the bidirectional converter. Section V presents a Converter parameter and transfer function. Section VI provides control strategy for the Bidirectional converter. Section VII delves with the simulation results and discussion. Section VIII of the paper offers an extensive conclusion and explores potential applications and future research avenues concerning SEPIC/ZETA bidirectional converter control and optimization.

II. SYSTEM UNDER CONSIDERATION

Under this system there is an integration of Battery and Bidirectional DC-DC Converter. This integrated approach not only ensures seamless energy flow but also underscores the system's adaptability and efficiency in managing power resources. Battery and Bidirectional SEPIC/ZETA DC-DC Converter specification consideration have been discussed below:

A. *Specifications Of Battery*

Storage of the excess power is performed using a rechargeable battery for the DC bus and release it when needed, ensuring a stable voltage. It plays a crucial role in managing the charging **and** discharging processes to maintain optimal performance and prolong battery life. Monitoring the state of charge (SOC) helps determine when to charge or discharge the battery, contributing to its longevity. Table I outlines the battery specifications.

TABLE I. BATTERY PARAMETERS

Parameters	Values
Type	Lead Acid
Nominal voltage	48V
Rated capacity	25Ah
Internal resistance	0.0192Ω

B. *Bidirectional Sepic/Zeta Dc-Dc Converter*

This bidirectional converter, like regular SEPIC/ZETA converters, can both increase (boost) and decrease (buck) the voltage regardless of which direction power is flowing. The design uses two inductors (L_A and L_B), three capacitors (C_A, a coupling capacitor C_C, and C_B), and two switches (SW_A and SW_B) that can handle current in both directions, as illustrated in Fig.2. The system has a DC bus or Grid on one side and a battery on the other side. Table II summarizes how the converter operates under various conditions.

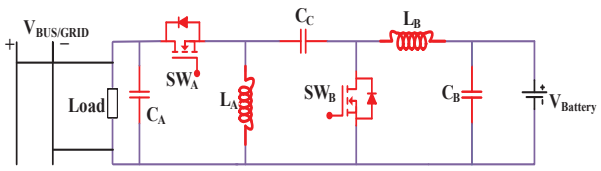


Fig. 3. Bidirectional SEPIC/ZETA DC-DC Converter Systematic Diagram

TABLE II. OPERATIONAL MODES

Conditions	Operating Mode	Current Flow
Grid Voltage < load voltage	SEPIC	Battery to Grid
Grid Voltage > load voltage	ZETA	Grid to Battery
Grid voltage = load voltage	-	Battery in standby mode

III. OPERATIONAL MODES

A. *Mode 1: Charging (Zeta)*

This mode charges the battery. Power flows from the DC bus (VBUS), which is typically 80volts, to the 48volt battery through the Bidirectional Converter (BDC). In this scenario, the BDC acts like a Zeta converter, and switch SW_A turns on. To achieve the necessary voltage step-down for charging, the switch typically operates at a duty cycle below 50%.

B. *MODE 2: DISCHARGING (SEPIC)*

This mode handles battery discharge. Power flows from the battery ($V_{Battery}$), which is typically 48 volts, to the 80volt DC bus through the BDC. In this scenario, the BDC acts like a SEPIC converter, and switch SW_B takes over. To achieve the necessary voltage step-up for discharge, the switch generally operates at a duty cycle above 50%.

IV. STATE SPACE MODELLING OF THE CONVERTER

In examining the state space modeling of ZETA and SEPIC converters, we delve into their dynamic behaviours and control mechanisms. By constructing state space equations, we uncover the intricate relationships among crucial parameters, offering insights into both transient and steady-state responses. This analytical approach not only allows for accurate system characterization but also facilitates the development of tailored control techniques, optimizing the performance and effectiveness of ZETA and SEPIC converters across various applications.

A. *Zeta Mode (Charging)*

Fig. 4 depicts the circuitry of the bidirectional converter configured in the ZETA arrangement during battery charging mode. The mathematical equations describing the ZETA circuit's behaviour in both on and off modes, necessary for deriving the average state space model, are presented in equations (4)-(17).

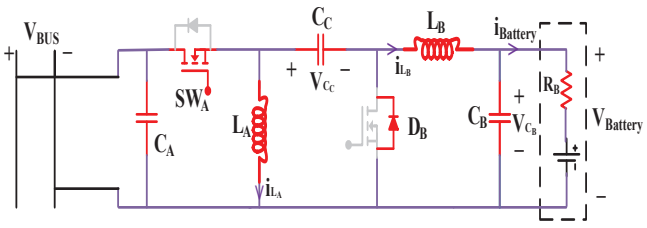


Fig. 4. ZETA mode configuration

Based on the state space representation obtained from the circuit equations, a small-signal model is built to examine the converter's small-signal behaviour. The currents (i_{L_A} and i_{L_B}) passing through the inductors and the voltages (V_{C_C} and V_{C_B}) across **the** capacitors are identified in this research paper as the state variables of the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:54:54 UTC from IEEE Xplore. Restrictions apply.

Zeta converter. The output voltage functions as the output variable, and the input voltage is regulated as the controlled input variable. The equations listed below describe the Zeta converter.

$$x = \begin{bmatrix} i_{L_A} \\ i_{L_B} \\ V_{C_C} \\ V_{C_B} \end{bmatrix}, u = [V_{BUS}], y = [V_{Battery}] \quad (1)$$

This system is represented in state space model as:

$$\dot{x} = AX + BU \quad (2)$$

$$y = CX + EU \quad (3)$$

In mode I ZETA converter's state space equations are as follows:

$$\frac{di_{L_A}}{dt} = \frac{V_{BUS}}{L_A} \quad (4)$$

$$\frac{di_{L_B}}{dt} = \frac{V_{BUS}+V_{C_C}-V_{C_B}}{L_B} \quad (5)$$

$$\frac{dV_{C_C}}{dt} = -\frac{i_{L_B}}{C_C} \quad (6)$$

$$\frac{dV_{C_B}}{dt} = \frac{i_{L_B}}{C_B} - \frac{V_{C_B}}{R_B C_B} \quad (7)$$

$$V_{Battery} = V_{C_B} \quad (8)$$

$$\begin{bmatrix} \frac{di_{L_A}}{dt} \\ \frac{di_{L_B}}{dt} \\ \frac{dV_{C_C}}{dt} \\ \frac{dV_{C_B}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{L_B} & -\frac{1}{L_B} \\ 0 & -\frac{1}{C_C} & 0 & 0 \\ 0 & \frac{1}{C_B} & 0 & -\frac{1}{RC_B} \end{bmatrix} \begin{bmatrix} i_{L_A} \\ i_{L_B} \\ V_{C_C} \\ V_{C_B} \end{bmatrix} + \begin{bmatrix} \frac{1}{L_A} \\ \frac{1}{L_B} \\ 0 \\ 0 \end{bmatrix} [V_{BUS}] \quad (9)$$

$$[V_{Battery}] = [0 \quad 0 \quad 0 \quad 1] \begin{bmatrix} i_{L_A} \\ i_{L_B} \\ V_{C_C} \\ V_{C_B} \end{bmatrix} + [0][V_{BUS}] \quad (10)$$

In mode II ZETA converter's state space equations are as follows:

$$\frac{di_{L_A}}{dt} = -\frac{V_{C_C}}{L_A} \quad (11)$$

$$\frac{di_{L_B}}{dt} = -\frac{V_{C_B}}{L_B} \quad (12)$$

$$\frac{dV_{C_C}}{dt} = \frac{i_{L_A}}{C_C} \quad (13)$$

$$\frac{dV_{C_B}}{dt} = \frac{i_{L_B}}{C_B} - \frac{V_{C_B}}{R_B C_B} \quad (14)$$

$$V_{Battery} = V_{C_B} \quad (15)$$

$$\begin{bmatrix} \frac{di_{L_A}}{dt} \\ \frac{di_{L_B}}{dt} \\ \frac{dV_{C_C}}{dt} \\ \frac{dV_{C_B}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\frac{1}{L_A} & 0 \\ 0 & 0 & 0 & -\frac{1}{L_B} \\ \frac{1}{C_C} & 0 & 0 & 0 \\ 0 & \frac{1}{C_B} & 0 & -\frac{1}{R_B C_B} \end{bmatrix} \begin{bmatrix} i_{L_A} \\ i_{L_B} \\ V_{C_C} \\ V_{C_B} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} [V_{BUS}] \quad (16)$$

$$[V_{Battery}] = [0 \quad 0 \quad 0 \quad 1] \begin{bmatrix} i_{L_A} \\ i_{L_B} \\ V_{C_C} \\ V_{C_B} \end{bmatrix} + [0][V_{BUS}] \quad (17)$$

A. SEPIC MODE (DISCHARGING)

Fig. 5 depicts the circuitry of the bidirectional converter configured in the SEPIC arrangement during battery discharging mode. The mathematical equations describing the SEPIC circuit's behaviour in both on and off modes, necessary for deriving the average state space model, are presented in equations (18)-(31).

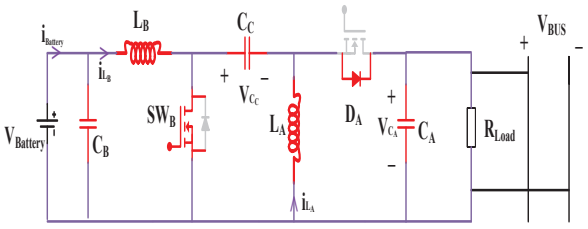


Fig. 5. SEPIC Mode configuration

In mode I SEPIC converter's state space equations are as follows:

$$\frac{di_{L_B}}{dt} = \frac{V_{Battery}}{L_B} \quad (18)$$

$$\frac{di_{L_A}}{dt} = \frac{V_{C_C}}{L_A} \quad (19)$$

$$\frac{dV_{C_C}}{dt} = -\frac{i_{L_A}}{C_C} \quad (20)$$

$$\frac{dV_{C_A}}{dt} = -\frac{V_{C_A}}{R_{Load}C_A} \quad (21)$$

$$V_0 = V_{C_A} \quad (22)$$

$$\begin{bmatrix} \frac{di_{L_B}}{dt} \\ \frac{di_{L_A}}{dt} \\ \frac{dV_{C_C}}{dt} \\ \frac{dV_{C_A}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{L_A} & 0 \\ 0 & -\frac{1}{C_C} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{R_{Load}C_A} \end{bmatrix} \begin{bmatrix} i_{L_B} \\ i_{L_A} \\ V_{C_C} \\ V_{C_A} \end{bmatrix} + \begin{bmatrix} \frac{1}{L_B} \\ 0 \\ 0 \\ 0 \end{bmatrix} [V_{Battery}] \quad (23)$$

$$[V_{Bus}] = [0 \quad 0 \quad 0 \quad 1] \begin{bmatrix} i_{L_B} \\ i_{L_A} \\ V_{C_C} \\ V_{C_A} \end{bmatrix} + [0][V_{Battery}] \quad (24)$$

In mode II SEPIC converter's state space equations are as follows:

$$\frac{di_{L_B}}{dt} = \frac{V_{Battery}-V_{C_C}-V_{C_A}}{L_B} \quad (25)$$

$$\frac{di_{L_A}}{dt} = -\frac{V_{C_A}}{L_A} \quad (26)$$

$$\frac{dV_{C_C}}{dt} = \frac{i_{L_B}}{C_C} \quad (27)$$

$$\frac{dV_{C_A}}{dt} = \frac{i_{L_B}}{C_A} + \frac{i_{L_A}}{C_A} - \frac{V_{C_A}}{R_{Load}C_A} \quad (28)$$

$$V_{Bus} = V_{C_A} \quad (29)$$

$$\begin{bmatrix} \frac{di_{L_B}}{dt} \\ \frac{di_{L_A}}{dt} \\ \frac{dV_{C_C}}{dt} \\ \frac{dV_{C_A}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\frac{1}{L_B} & -\frac{1}{L_B} \\ 0 & 0 & 0 & -\frac{1}{L_A} \\ \frac{1}{C_C} & 0 & 0 & 0 \\ \frac{1}{C_A} & \frac{1}{C_A} & 0 & -\frac{1}{R_{Load}C_A} \end{bmatrix} \begin{bmatrix} i_{L_B} \\ i_{L_A} \\ V_{C_C} \\ V_{C_A} \end{bmatrix} + \begin{bmatrix} \frac{1}{L_B} \\ 0 \\ 0 \\ 0 \end{bmatrix} [V_{Battery}] \quad (30)$$

$$[V_{Bus}] = [0 \quad 0 \quad 0 \quad 1] \begin{bmatrix} i_{L_B} \\ i_{L_A} \\ V_{C_C} \\ V_{C_A} \end{bmatrix} + [0][V_{Battery}] \quad (31)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:54:54 UTC from IEEE Xplore. Restrictions apply.

V. PARAMETERS OF CONVERTER

Table III shows detailed parameters related to SEPIC/ZETA Bidirectional (DC-DC) Converter.

TABLE III. CONVERTER PARAMETERS:

PARAMETERS	RATINGS
V_{BUS}	80V
$V_{Battery}$	48V
$I_{Battery}$	5A
L_A	6.7mH
L_B	4mH
C_A	50µF
C_B	12.6µF
Load	26.667Ω
C_C **(Coupling capacitor)**	26.042µF
f_{SW} **(Switching frequency)**	50kHz

The circuit **calculations** eliminate the switching ripple component by averaging across one switching period, resulting in the creation of the averaged switching model.

$$A = A_1 d + A_2(1-d) \quad (32)$$
$$B = B_1 d + B_2(1-d) \quad (33)$$
$$C = C_1 d + C_2(1-d) \quad (34)$$
$$E = E_1 d + E_2(1-d) \quad (35)$$

Before obtaining the small-signal model of the converter, it's necessary to linearize the nonlinear averaged state-space model around a **specific** operating point. This involves introducing small disturbances, termed perturbations, to the state **variables** of the system and its parameter, also to its input variables as well.

$$d = D + \hat{d},\ x = X + \hat{x},\ u = U + \hat{u},\ y = Y + \hat{y} \quad (36)$$

By replacing the averaged state space model (eq. 32-35) from the **perturbed** variables and neglecting steady-state terms and higher-order influences, the linearized small-signal model is obtained in **the** following manner:

$$s\hat{x}(s) = A\hat{x}(s) + B\hat{u}(s) + \big((A_1 - A_2)X(s) + (B_1 - B_2)U(s)\big)\hat{d}(s) \quad (37)$$

$$\hat{y}(s) = C\hat{x}(s) + E\hat{u}(s) + ((A_1 - A_2)X(s) + (E_1 - E_2)U(s))\hat{d}(s) \quad (38)$$

$$\frac{\hat{x}(s)}{\hat{d}(s)} = (sI - A)^{-1}\big((A_1 - A_2)X(s) + (B_1 - B_2)U(s)\big) \quad (39)$$

The transfer function for the ZETA mode, determined through **small** signal analysis and averaging while incorporating circuit parameters in terms of duty cycle, can be represented as:

$$\frac{\widehat{v_0}(s)}{\hat{d}(s)} = \frac{1.069e08\ s^3 + 2.086e11\ s^2 + 1.288e13\ s + 1.178e18}{s^5 + 4e05\ s^4 + 1.07e09\ s^3 + 1.644e12\ s^2 + 2.162e15\ s + 1.178e18} \quad (40)$$

Likewise, for the SEPIC mode, the transfer function can be formulated as follows:

$$\frac{\widehat{v_0}(s)}{\hat{d}(s)} = \frac{-5.12e05\ s\text{^}3 + 1.229e09\ s\text{^}2 - 1.842e12\ s + 4.422e15}{s\text{^}4 + 2400\ s\text{^}3 + 7.198e06\ s\text{^}2 + 8.637e09\ s + 1.295e13} \quad (41)$$

VI. CONTROL STRATEGY FOR ZETA/SEPIC BIDIRECTIONAL CONVERTER

Fig. 6 shows the basic idea of the control strategy (PID) which is proposed for the system. Using the ZETA mode for charging and the SEPIC mode for discharging, a PID controller is used in this configuration to control both charging and discharging operations. The reference values (V_{Refb}, V_{RefL}), feedback, and control output of the PID system are modified in accordance with whether the system is in ZETA or SEPIC mode.

The tuning of the PI controller for the system is carried out using the **classical** Ziegler-Nichols (ZN) method. This involves determining the parameters by performing the averaging of the state space model for the transfer function. As a result, the PI gains obtained are presented in Table IV.

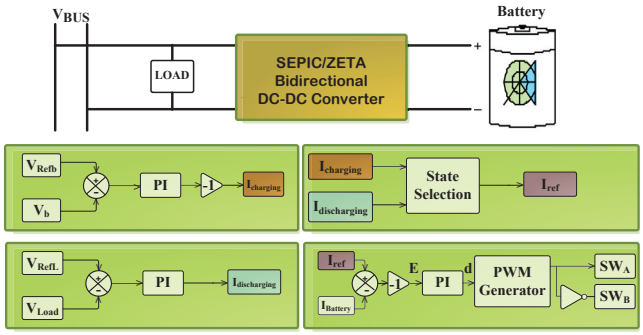


Fig. 6. Fig.6 Proposed Control Strategy

TABLE IV. PARAMETERS OF CONTROLLER

Loop	Mode	K_p	K_i
Voltage Loop	Charging	50	10e3
	Discharging	0.5	50
Current Loop	Charging/ Discharging	0.05	10

VII. RESULTS AND ANALYSIS

The output results for the converter are evaluated in MATLAB/SIMULINK for **scenarios** involving both charging and discharging of the battery.

A. Charging Mode (Zeta Converter)

During the charging phase, the state of charge (SOC%) of the battery **gradually** increases, accompanied by a rise in battery voltage. The corresponding waveforms depicting the battery current are illustrated in Fig.7, 8, and 9.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:54:54 UTC from IEEE Xplore. Restrictions apply.

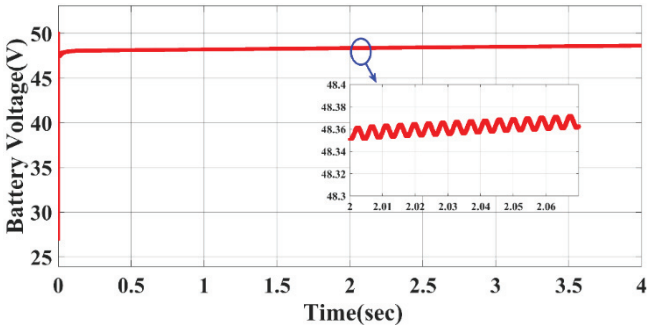


Fig. 7. Battery Voltage

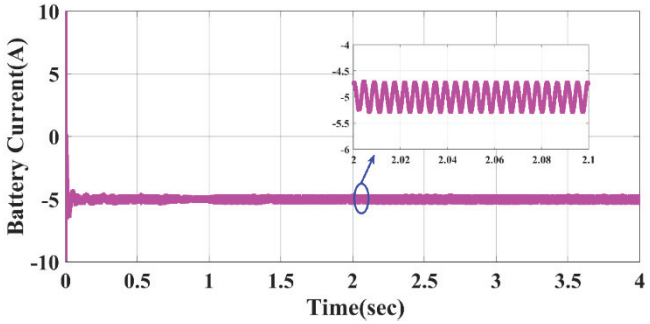


Fig. 8. Battery Current

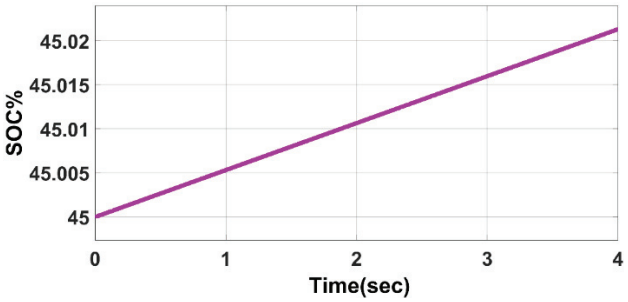


Fig. 9. SOC%

A. *Discharging Mode (Sepic Converter*

During the charging phase, the state of charge (SOC%) of the battery gradually **decreases**, accompanied by a fall in battery voltage. The corresponding waveforms depicting the battery current are illustrated in Fig.10,11, and 12.

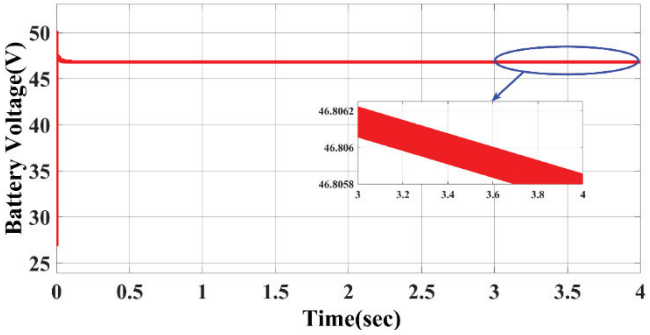


Fig. 10. Battery Voltage

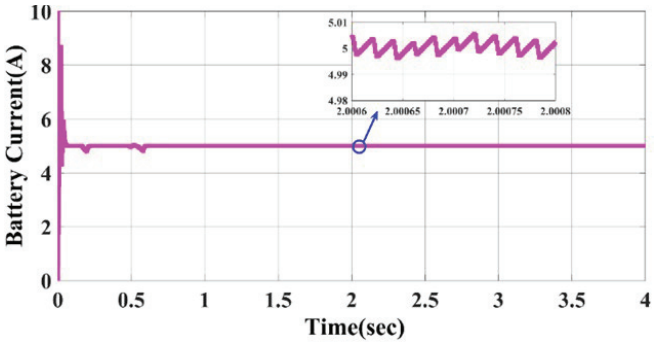


Fig. 11. Battery Current

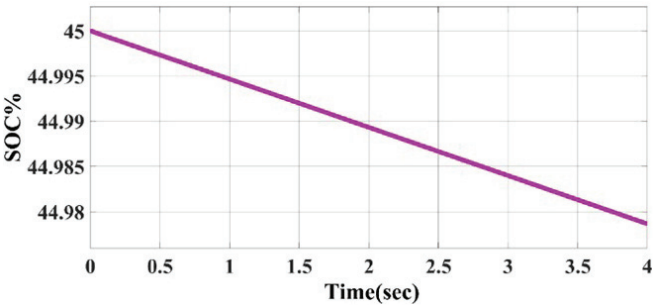


Fig. 12. SOC%

B. *Charging And Discharging Mode*

The bidirectional converter showcases its effectiveness in both charging and discharging modes, as evidenced by the comprehensive depiction in Fig. 13, 14, and 15. These figures vividly **illustrate** the converter's adeptness in seamlessly handling operations in both modes, underscoring its efficiency and versatility.

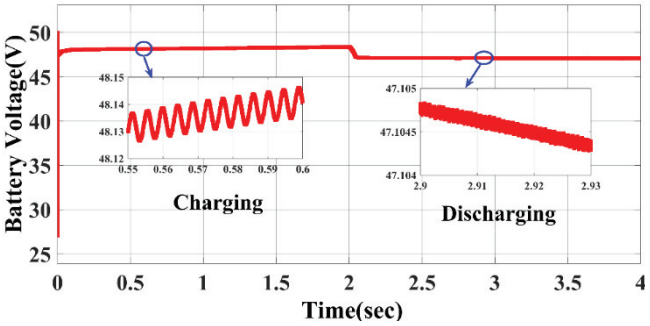


Fig.13 Battery Voltage

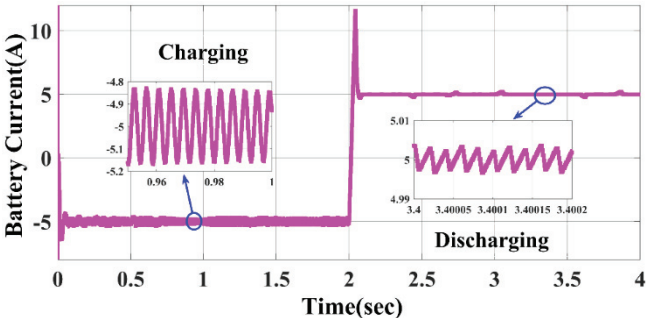


Fig.14 Battery Current

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:54:54 UTC from IEEE Xplore. Restrictions apply.

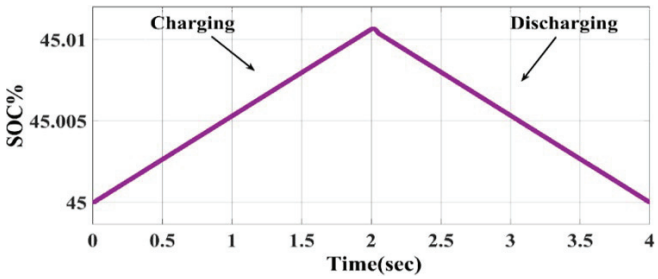


Fig. 15 SOC%

VIII. CONCLUSION AND FUTURE SCOPE

This study demonstrates a comprehensive modeling of a bidirectional DC-DC converter using SEPIC/ZETA circuitry, that is designed for applications related to battery. The battery charging and discharging processes is effectively managed by implementing PI control system, resulting in satisfactory state of charge (SOC) as validated by MATLAB/SIMULINK. The Ziegler-Nichols method has been employed to ascertain the PI parameters (Kp, and Ki), enhancing the converter's control performance. The analysis shows that, whether the converter is operating in the charging or discharging mode, the battery has a constant current. Future work will focus on comparative study of Ziegler-Nichols with any optimisation method like Particle Swarn Optimisation, Fuzzy Logic and validating the proposed system through hardware implementation and comparing its performance with simulation results.

REFERENCESS

[1] Lin, C-C., L-S. Yang, and G. W. Wu. "Study of a non-isolated bidirectional DC–DC converter." IET Power Electronics (2013)

[2] Sebaje, Alex Sander, Mário Lúcio da Silva Martins, and Carlos. "A hybrid bidirectional DC-DC converter based on a SEPIC/Zeta converter with a modified switched capacitor cell." 2021 Brazilian IEEE, 2021.

[3] Sunarno, Epyk, et al. "Design and implementation bidirectional SEPIC/ZETA converter using Fuzzy Logic Controller in DC microgrid application." Journal of Physics: Conference Series 2019.

[4] Mittal, Sudhanshu, Alka Singh, and Prakash Chittora. "Solar PV Array based Grid-Connected Bi-Directional EV Charger Controlled using NARLMMN Algorithm." 2023 9th IEEE India International Conference on Power Electronics (IICPE). IEEE, 2023.

[5] Mittal, S., Singh, A. & Chittora, P. Design and development of Reduced Switch Five Level Inverter for interfacing renewable energy sources and EV charger. Electr Eng (2024).

[6] Ravi, Deepak, et al. "Bidirectional DC to DC converters: an overview of various topologies, switching schemes and control techniques." International Journal of Engineering & Technology 7.4.5 (2018).

[7] Kim, In-Dong. "Design of bidirectional PWM Sepic/Zeta DC-DC converter." 2007 7th IEEE.

[8] V. Kumar, A. Singh and P. Chittora, "Control and Analysis of a Grid Following Converter using Adaptive Log Family based SOGI," 2023 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA), Roorkee, India, 2023

[9] Kumar, Vineet, Haroon Ashfaq, and Rajveer Singh. "Coordinated Control and Energy Management Strategy of Battery and Super-capacitor of Micro-grid." Intelligent Systems and Smart Infrastructure. CRC Press, 2023.

[10] Mittal, Sudhanshu, Alka Singh, and Prakash Chittora. "EV Control in G2V and V2G modes using SOGI Controller." 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT). IEEE, 2022.

[11] Gorji, Saman A., et al. "Topologies and control schemes of bidirectional DC–DC power converters: An overview." IEEE (2019).

[12] Lithesh, Gottapu, Bekkam Krishna "Review and comparative study of bi-directional DC-DC converters." 2021 IEEE International Power and Renewable Energy Conference (IPRECON).

[13] Sharma, Vasudha, Narendra Kumar, and Ram Bhagat. "Tuning of PID Controller in Magnetic Levitation System Applying Honey Badger Algorithm." IEEE 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:54:54 UTC from IEEE Xplore. Restrictions apply.

AIP Conference Proceedings

RESEARCH ARTICLE | AUGUST 03 2024

Engineering optoelectronic properties of Cs_2SnI_6 using pressure

Parmeet Singh Chani; Mehak Gole; Yashjeet Singh; Mukhtiyar Singh

AIP Conf. Proc. 3044, 050001 (2024)
https://doi.org/10.1063/5.0209092

View Online
Export Citation

13 August 2024 07:14:15

Nanotechnology & Materials Science
Optics & Photonics
Impedance Analysis
Scanning Probe Microscopy
Sensors
Failure Analysis & Semiconductors



Unlock the Full Spectrum.
From DC to 8.5 GHz.
Your Application. Measured.
Find out more
Zurich Instruments

AIP Publishing

Engineering Optoelectronic properties of Cs_2SnI_6 using Pressure

Parmeet Singh Chani[a)], Mehak Gole[b)], Yashjeet Singh[c)], Mukhtiyar Singh[d)]

Department of Applied Physics, Delhi Technological University, Delhi 110042, India

[d)]*Corresponding author: msphysik09@gmail.com*
[a)]*parmeetsinghchani_2k19ep066@dtu.ac.in*
[b)]*mehakgole@gmail.com*
[c)]*syashjeet08@gmail.com*

Abstract. Lead free perovskites such as Cs_2SnI_6 have been given significant attention due to their extravagant optoelectronic properties. In this study, we use first principles calculations to explore the evolution of structural, optical and electronic properties of Cs_2SnI_6 when it is exposed to external hydrostatic pressure. Our results indicate that Cs_2SnI_6 is far more compressible as compared to other perovskites, mainly $Cs_2AgInCl_6$ and $CsPbI_3$. The compression of the lattice leads to a lowering of bond length which in turn results in a higher orbital overlap. The reduction of I-I bond length results in a drastic decrease in the band gap of the material. Similar to electronic properties, optical properties also show a high dependency on pressure. With the increase in pressure both the absorptivity coefficient and refractive index increase. These findings show that the application of external hydrostatic pressure can act as an effective strategy to modify the opto-electronic properties of such materials.

13 August 2024 07:14:15

INTRODUCTION

The field of optoelectronics is rapidly evolving with new materials being at the heart of this evolution. Lead based halide perovskite materials owing to their good stability, excellent photoelectric properties, high power conversion efficiency and tunable opto-electronic properties have gained a vital role in this evolution. Nonetheless, the presence of lead makes these materials bio-hazardous [1] due to which commercialization of these perovskite-based devices has not been possible. Due to this, much attention has been paid to the discovery and analysis of alternate materials which possess low toxicity along with similar opto-electronic properties as compared to the Lead-based perovskites. One prospective material for this is Cs2SnI6[2] as it possesses good stability [3], near ideal band gap of 1.26 to 1.6 eV (according to Shockley Queisser Limit) [4-6], and high absorption coefficient (over 10^5 cm^{-1} from 1.7 eV [5]).

Prior to the use of any material in an optoelectronic device, it should be optimized to be properly integrated within the device. This optimization can be acquired by several methods. One of the most prominent methods is the usage of doping and compositional tuning, however, this method has often shown a decrease in carrier lifetime, an increase in chemical impurities and a deterioration in structural stability [7,8]. Another comparatively less explored strategy is the use of high-pressure techniques which has shown promising results due to soft lattices of perovskite materials. Experimentally, diamond anvil cells in combination with pressure-transmitting media such as silicone oil are used in order to create static compression. However, one can also explore the effects of pressure on crystalline structures using theoretical calculations such as density functional theory. Bounos et. al [9] have used synchrotron X-ray diffraction and Raman spectroscopy measurements in order to analyse the structural evolution of Cs_2SnI_6 under Hydrostatic pressure. Here, we build upon their work and explore the effect of pressure on optical and electronic properties of Cs_2SnI_6 in cubic phase using first principles calculations. This study contributes to the ongoing effort to optimize the optoelectronic properties of perovskite materials. Our results indicate novel high compressibility and demonstrate the potential of high-pressure treatment as a tool for manipulating the properties of Cs2SnI6.

Smart Grid & Electric Vehicle
AIP Conf. Proc. 3044, 050001-1–050001-8; https://doi.org/10.1063/5.0209092
Published under an exclusive license by AIP Publishing. 978-0-7354-4972-5/$30.00

COMPUTATIONAL DETAILS

We have used the first-principles method as employed via density functional theory (DFT). The projector augmented wave (PAW) [10] method along with Perdew–Burke–Ernzerhof (PBE) version of the generalized gradient approximation (GGA) was used as has been implemented in Vienna Ab Initio Simulation (VASP) Package [11]. Hydrostatic Pressure was modeled through ionic minimization using the PSTRESS tag as employed in VASP. Cut-off energy and K point grid were respectively, kept to 360 eV and 4x4x4 for conventional cell relaxation. For subsequent calculations, using primitive cells, a k-mesh grid of 6x6x6 was used and a much denser grid of 16x16x16 was used for DOS calculations. The Monkhorst-pack scheme was used for sampling the Brillouin zone. A threshold of 10^{-6} eV was used for energy convergence and ionic minimization was carried out until the forces were below 0.025 eV A^{-1}. VASPKIT [12] was used for various pre-processing tasks such as obtaining the K-path, generating the primitive unit cell, and for post-processing to obtain the Electronic Band Structure, Density of States, and Optical Properties.

RESULTS AND DISCUSSIONS

Structural Transformation Under Pressure

Cs_2SnI_6 crystalizes in FCC structure where the Cs_2^+ occupies the tetragonal interstitial space and $[SnI_6]^{-2}$ octahedra are located at corners and face centers. Ionic relaxation is performed with the hydrostatic pressure of 0, 0.5, 1.7, 3.1, 5.4 and 7.4 GPa, Fig. 1b compares the experimental lattice constant [9] with the relaxed structure. The results obtained from DFT calculations are able to mimic the reduction in cell size. A slightly higher rate of decrease in the theoretical lattice constant leads to a small average error of 1.15%.

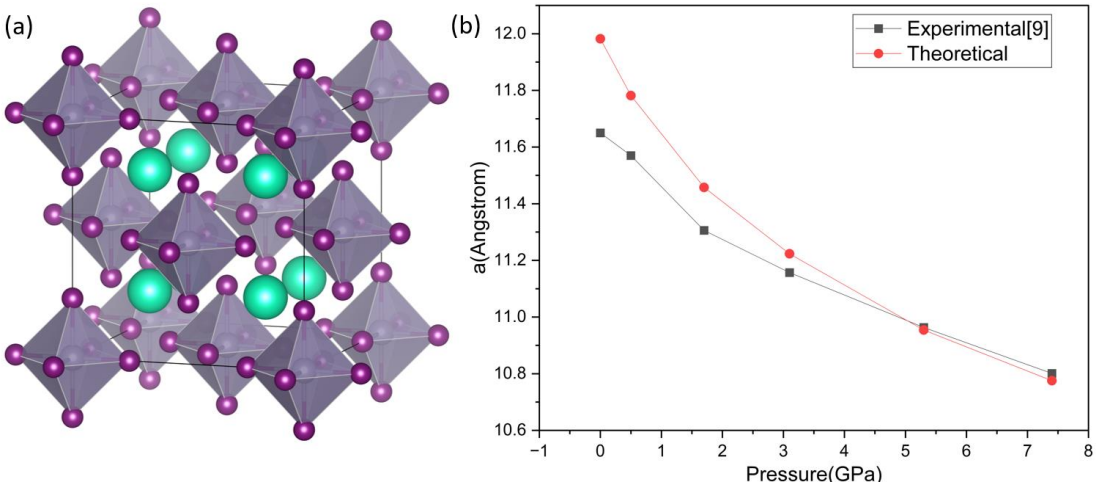


FIGURE 1. (a) Cs_2SnI_6 conventional Cell ;(b) Variation of Experimental [9] and DFT-Based Lattice Constant.

Cs_2SnI_6 can be considered as a defect variant of $CsSnI_3$ wherein alternate $[SnI_6]^{-2}$ octahedra have been removed. This leaves a 3D array of isolated octahedra, the free space makes the structure of Cs_2SnI_6 much more sensitive to pressure as compared to other perovskites. In Fig. 2, the pressure dependence of V_0/V is shown, the slope of which gives us isothermal compressibility(β_T). Subsequently the isothermal compressibility of two other famous perovskites $CsPbI_3$ and $Cs_2AgInCl_6$ have been evaluated in order to create a comparison. For Cs_2SnI_6 linear dependence was observed from 0 to 1.7 GPa and 3.1 to 7.4 GPa, and linear fitting was performed between the same extremities for all perovskites. One can observe that Cs_2SnI_6 is more compressible, especially from 0 to 1.7 GPa wherein β_T of Cs2SnI6 is ~2 times as compared to that of other perovskites. This can be attributed to the alternate octahedral structure.

Bounos et. al [9] observed that upon compression beyond 3.3 GPa, certain Raman modes vanished whereas SXRD patterns remained the same. This was a result of the structural phase transition from FCC Phase-1(Pressure < 3.3 GPa) to FCC Phase-2(Pressure > 3.3 GPa). The variation of Raman Spectroscopy measurements in Phase-2 can be attributed

13 August 2024 07:14:15

to the freezing of a few Raman vibrational modes due to the low availability of space. The primary distinction between both phases is compressibility, as the structure goes from Phase-1 to Phase-2. XRD data detects a decrease in compressibility by 44%. Similar to this, our calculations are able to detect a sudden decrease of 55.67% in β_T. No decrease of similar magnitude can be observed for other perovskites. For $CsPbI_3$ and $Cs_2AgInCl_6$ we observed a decrease of 18.10% and 36.84% respectively. This decrease in compressibility as pressure increases happens as a result of increasing repulsion between the atoms.

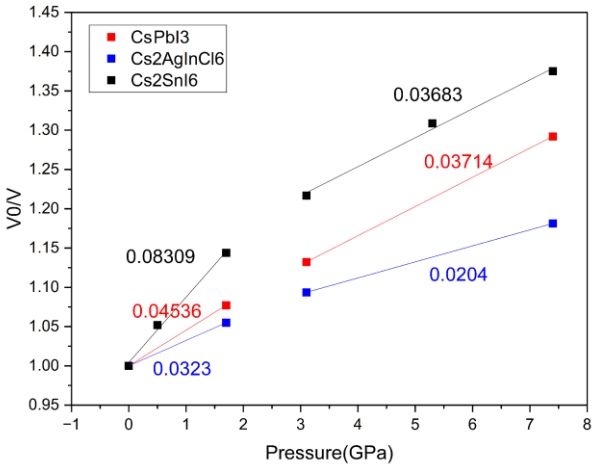


FIGURE 2. (a) Variation of V_0/V with Pressure

Electronic properties

The high compressibility of Cs_2SnI_6 is evident in Fig. 2, in order to evaluate the effect of this compression on electronic properties one must study the variation of electronic band structure and DOS with pressure. Fig. 3 shows that upon compression the electronic band gap reduces considerably. It is well known that PBE underestimates the band gap which can be seen in our results. Furthermore, our calculated band gaps are even lower than previous reports which used PBE[13,14], this can be attributed to less accurate lattice parameters used in said reports. Even though the band gaps are severely underestimated, one can still observe a decrease of a very significant proportion. Nurul et. al [15] performed similar calculations for $Cs_2AgBiBr_6$ and only a variation of ~0.08 eV in our pressure range whereas we observed a variation of ~0.18 eV. As the compressibility of Cs_2SnI_6 decreases with an increase in pressure one can observe that rate of change of the Band Gap also decreases.

13 August 2024 07:14:15

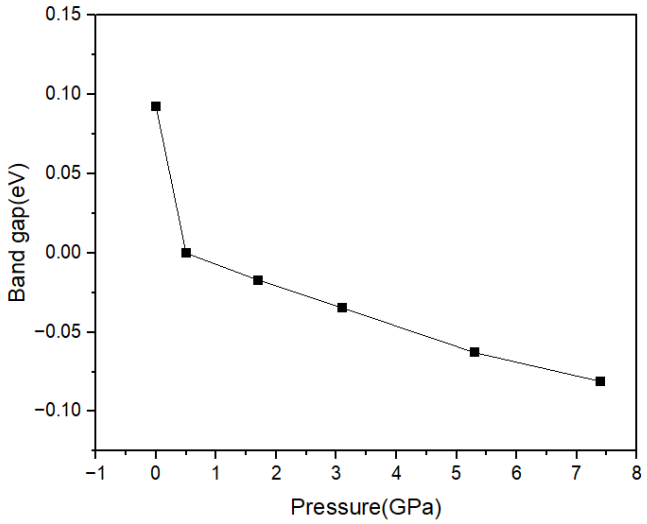


FIGURE 3. Variation of Band Gap with Pressure

Fig. 4 shows that the VBM is mainly composed of Iodine's p orbital whereas the CBM is formed by a combination of Iodine's p orbital and Tin's s orbital. The decrease in lattice size leads to a reduction in bond lengths, which in turn leads to more overlap of orbitals. The Interoctahedral I-I bonds tend to compress more as compared to other bonds which leads to a higher increase in their covalent character. As a result, the band gap reduces drastically. From Figures 4b and 4d it can also be observed that intermediate bands within the valence band also start to overlap, this could lead to interesting results as Cs_2SnI_6 can be used for intermediate band solar cells [16].

13 August 2024 07:14:15

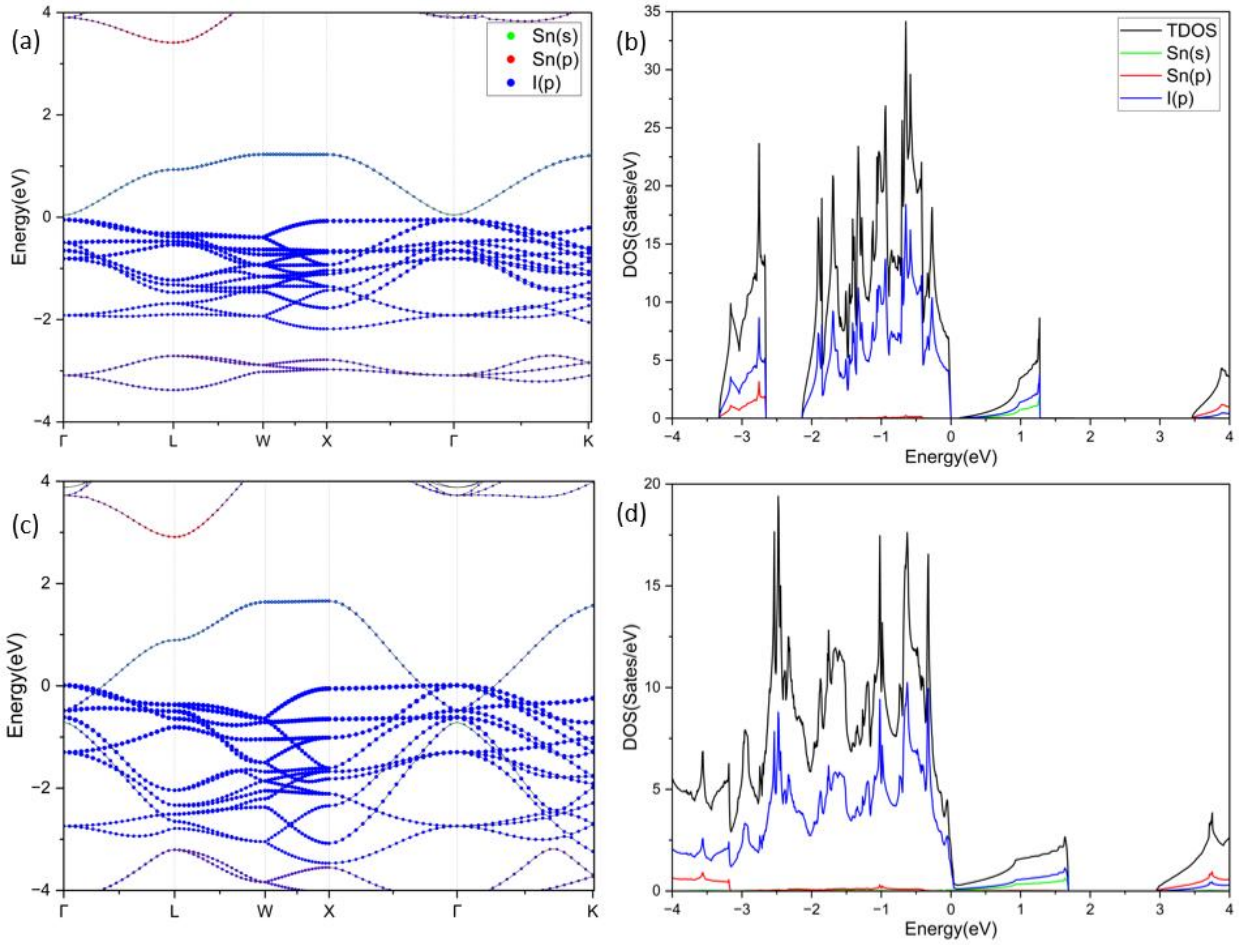


FIGURE 4. Projected Band Structures and Partial Density of States at (a),(b) 0 GPa and (c),(d) 7.4 GPa

Optical Properties

To observe the effect of pressure on optical properties we calculated the real and imaginary parts of the optical dielectric function [17] which were further used to calculate the absorption coefficients($\alpha(\omega)$), refractive index($\eta(\omega)$), extinction coefficient($\kappa(\omega)$) and Reflectivity($R(\omega)$). Fig. 5 shows these linear optical properties at 0, 1.7, 3.1 and 7.4 GPa. Absorption coefficient of the unpressurized system shows the first peak at 1.49 eV which goes along with experimental studies wherein absorption peaks are observed around 1.50 eV [5]. Calculations performed by Rasukkannu et al.[16] provide near identical values confirming the accuracy of our results. It can be observed that as pressure increases there is a significant increase in the absorption coefficient, especially in the UV region. This can be attributed to lower excitation requirements due to more orbital overlap. The increase in absorption upon compression also indicates possible applications in UV sensors. Similar observations can be made from the extinction coefficient (Fig. 5c).

The refractive index gives a measure of the interaction of incident photons with the material and can be used to analyze the type of bonding. Higher refractive indices indicate more covalent character as strong interactions lead to a lowering of photons velocity. Fig. 5b shows that the refractive index increases with the pressure which confirms our theory of increased interaction being the cause of band gap lowering.

For solar cell applications, the reflectivity of the material plays an important role as an increase in reflectivity could lead to harmful effects as a higher proportion of light will be reflected back. Fig. 5d shows that along with an increase in pressure, the Reflectivity also increases. The increase of absorption coefficient and reflectivity leads to a controversial standpoint for application in photovoltaic devices. However, one can locate the sweet spot by fine-tuning the pressure to a point where the gain from the increase in absorption coefficient is higher in comparison to the loss from increased reflectivity.

13 August 2024 07:14:15

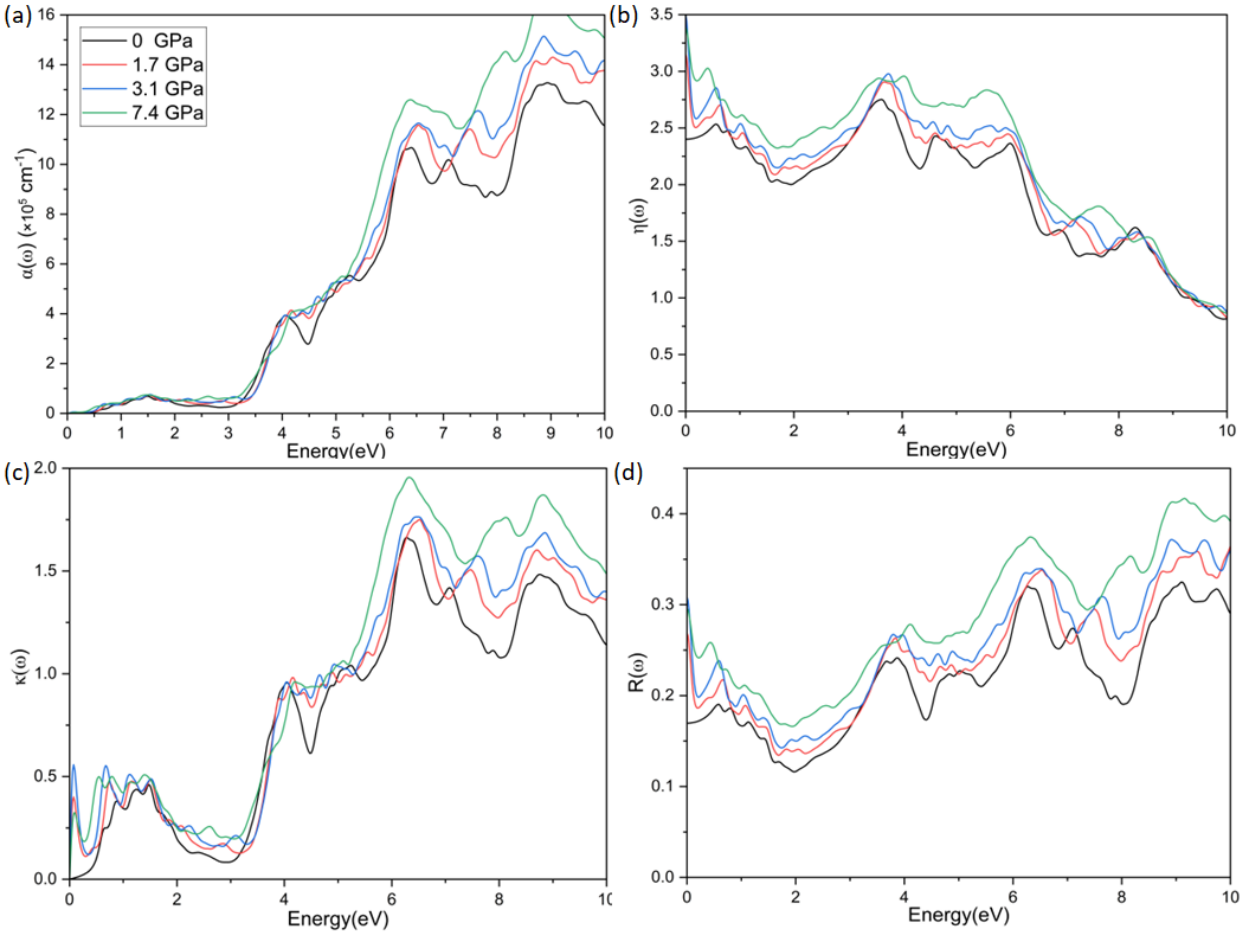


FIGURE 5. (a) Absorption Coefficient ;(b) Refractive Index;(c) Extinction coefficient; (d) Reflectivity

CONCLUSION

In the current study, we have employed DFT calculations to analyse the variation of structural, electronic and optical properties of Cs_2SnI_6 under Pressure. We observed that owing to alternative vacancy, Cs_2SnI_6 is far more compressible in comparison to other perovskite materials. Using isothermal compressibility, we were also able to validate experimental results [9] by detecting the phase transformation of crystal structure beyond 3.3 GPa from Phase-1 to Phase-2. Calculated electronic and optical properties indicate a huge dependence upon pressure which can be attributed to an increase in orbital overlap with the compression of the crystal lattice. The monumental decrease in band gap happens as inter-octahedral I-I bonds compress with ease and the band edges are mainly composed of iodine's p orbital. We also observe with pressure, the absorption coefficient and reflectivity both increase. The complete analysis indicates that hydrostatic pressure can be used to fine-tune the properties of Cs_2SnI_6 based on individual applications. Future work should include using hybrid DFT calculations in order to analyze the structural evolution at higher pressure ranges and its effect on various optical and electronic properties.

ACKNOWLEDGMENTS

We thank Professor Shyam Kishor for his valuable suggestions and guidance. This work was carried out at the HPC facility of IIT-Delhi.

13 August 2024 07:14:15

REFERENCES

1. Fabini, Douglas. "Quantifying the potential for lead pollution from halide perovskite photovoltaics." *The Journal of Physical Chemistry Letters* 6, no. 18 (2015): 3546-3548. DOI: https://doi.org/10.1021/acs.jpclett.5b01747
2. Cai, Yao, Wei Xie, Hong Ding, Yan Chen, Krishnamoorthy Thirumal, Lydia H. Wong, Nripan Mathews, Subodh G. Mhaisalkar, Matthew Sherburne, and Mark Asta. "Computational study of halide perovskite-derived A2BX6 inorganic compounds: chemical trends in electronic structure and structural stability." *Chemistry of Materials* 29, no. 18 (2017): 7740-7749. DOI: https://doi.org/10.1021/acs.chemmater.7b02013
3. Wang, Aifei, Xingxu Yan, Mian Zhang, Shibin Sun, Ming Yang, Wei Shen, Xiaoqing Pan, Peng Wang, and Zhengtao Deng. "Controlled synthesis of lead-free and stable perovskite derivative Cs2SnI6 nanocrystals via a facile hot-injection process." *Chemistry of Materials* 28, no. 22 (2016): 8132-8140. DOI: https://doi.org/10.1021/acs.chemmater.6b01329
4. Xiao, Zewen, Hechang Lei, Xiao Zhang, Yuanyuan Zhou, Hideo Hosono, and Toshio Kamiya. "Ligand-hole in [SnI6] unit and origin of band gap in photovoltaic perovskite variant Cs2SnI6." *Bulletin of the Chemical Society of Japan* 88, no. 9 (2015): 1250-1255. DOI: https://doi.org/10.1246/bcsj.20150110
5. Qiu, Xiaofeng, Bingqiang Cao, Shuai Yuan, Xiangfeng Chen, Zhiwen Qiu, Yanan Jiang, Qian Ye et al. "From unstable CsSnI3 to air-stable Cs2SnI6: A lead-free perovskite solar cell light absorber with bandgap of 1.48 eV and high absorption coefficient." *Solar Energy Materials and Solar Cells* 159 (2017): 227-234. DOI: https://doi.org/10.1016/j.solmat.2016.09.022
6. Saparov, Bayrammurad, Jon-Paul Sun, Weiwei Meng, Zewen Xiao, Hsin-Sheng Duan, Oki Gunawan, Donghyeop Shin, Ian G. Hill, Yanfa Yan, and David B. Mitzi. "Thin-film deposition and characterization of a Sn-deficient perovskite derivative Cs2SnI6." *Chemistry of Materials* 28, no. 7 (2016): 2315-2322. DOI: https://doi.org/10.1021/acs.chemmater.6b00433
7. Li, Mei, Tianbiao Liu, Yonggang Wang, Wenge Yang, and Xujie Lü. "Pressure responses of halide perovskites with various compositions, dimensionalities, and morphologies." *Matter and Radiation at Extremes* 5, no. 1 (2020): 018201. DOI: https://doi.org/10.1063/1.5133653
8. Wang, Lingrui, Kai Wang, and Bo Zou. "Pressure-induced structural and optical properties of organometal halide perovskite-based formamidinium lead bromide." *The journal of physical chemistry letters* 7, no. 13 (2016): 2556-2562. DOI: https://doi.org/10.1021/acs.jpclett.6b00999
9. Bounos, Giannis, Maria Karnachoriti, Athanassios G. Kontos, Constantinos C. Stoumpos, Leonidas Tsetseris, Andreas Kaltzoglou, Xiaofeng Guo et al. "Defect perovskites under pressure: structural evolution of Cs2SnX6 (X= Cl, Br, I)." *The Journal of Physical Chemistry C* 122, no. 42 (2018): 24004-24013. DOI: https://doi.org/10.1021/acs.jpcc.8b08449
10. Blöchl, Peter E. "Projector augmented-wave method." *Physical review B* 50, no. 24 (1994): 17953. DOI: https://doi.org/10.1103/PhysRevB.50.17953
11. Kresse, Georg, and Jürgen Furthmüller. "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set." *Physical review B* 54, no. 16 (1996): 11169. DOI: https://doi.org/10.1103/PhysRevB.54.11169
12. Wang, Vei, Nan Xu, Jin-Cheng Liu, Gang Tang, and Wen-Tong Geng. "VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code." *Computer Physics Communications* 267 (2021): 108033. DOI: https://doi.org/10.1016/j.cpc.2021.108033
13. Kaltzoglou, Andreas, Maria Antoniadou, Athanassios G. Kontos, Constantinos C. Stoumpos, Dorothea Perganti, Eirini Siranidi, Vasilios Raptis et al. "Optical-vibrational properties of the Cs2SnX6 (X= Cl, Br, I) defect perovskites and hole-transport efficiency in dye-sensitized solar cells." *The Journal of Physical Chemistry C* 120, no. 22 (2016): 11777-11785. DOI: https://doi.org/10.1021/acs.jpcc.6b02175
14. 14.Handayani, Yolla Sukma, Priastuti Wulandari, and Rahmat Hidayat. "Electronic Structure of Lead-Free Halide Perovskite Material: Cs2SnI6, for Perovskite Solar Cells Applications." In *Materials Science Forum*, vol. 1028, pp. 210-214. Trans Tech Publications Ltd, 2021. DOI: https://doi.org/10.4028/www.scientific.net/MSF.1028.210
15. Islam, Md Nurul, Jiban Podder, Tusar Saha, and Protima Rani. "Semiconductor to metallic transition under induced pressure in Cs 2 AgBiBr 6 double halide perovskite: a theoretical DFT study for photovoltaic and optoelectronic applications." *RSC advances* 11, no. 39 (2021): 24001-24012. DOI: https://doi.org/10.1039/D1RA03161A
16. Rasukkannu, Murugesan, Dhayalan Velauthapillai, and Ponniah Vajeeston. "A first-principle study of the

13 August 2024 07:14:15

electronic, mechanical and optical properties of inorganic perovskite Cs2SnI6 for intermediate-band solar cells." *Materials Letters* 218 (2018): 233-236. DOI: https://doi.org/10.1016/j.matlet.2018.02.034
17. Gajdoš, M., K. Hummer, G. Kresse, J. Furthmüller, and F. J. P. R. B. Bechstedt. "Linear optical properties in the projector-augmented wave methodology." *Physical Review B* 73, no. 4 (2006): 045112. DOI: https://doi.org/10.1103/PhysRevB.73.045112

13 August 2024 07:14:15

2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

Enhanced Credit Card Fraud Detection: A Machine Learning Approach

Priti Meena
Computer Science Engineering
Delhi Technological University
New Delhi, India
pritimeena_co20b5_73@dtu.ac.in

Mukul Raj
Computer Science Engineering
Delhi Technological University
New Delhi, India
mukulraj_co20b5_19@dtu.ac.in

Rajni Jindal
Computer Science Engineering
Delhi Technological University
New Delhi, India
rajnijindal@dce.ac.in

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10626378

***Abstract*—With the swift advancement in technology, credit cards are increasingly prevalent as substitutes for cash in daily transactions. This presents several opportunities for unscrupulous individuals to use credit cards fraudulently. 425,977 reports of credit card fraud were made in 2023, according to a Federal Trade Commission report. The credit card issuer ought to offer a service that shields customers from potential risks to guarantee their safety when using these cards. To address the problem of class imbalance, we resampled the dataset using the Synthetic Minority Oversampling Technique (SMOTE). The following variety of machine learning methods were applied to assess this framework using a soft voting mechanism: Combination 1 consists of KNN, Bagging, AdaBoost, and Logistic Regression, whereas Combination 2 consists of Random Forest, KNN, Bagging, and XGBoost. Furthermore, the usefulness of Logistic Regression, KNN, and AdaBoosting individually was evaluated to determine their stand-alone performance. This comparison served as a benchmark against which the ensemble models were evaluated. The models' accuracy, F1 score, and Area Under the Curve (AUC) were used to evaluate them. Additionally, the outcomes of the machine learning models combined produced better results, with a total accuracy of 99.81% for combination 1 and 99.86% for combination 2 compared to the individual predictions of the models. However, Combination 2 outperformed Combination 1 in terms of F1 score, suggesting enhanced overall recall and precision performance.**

Keywords*—Credit Cards, Fraud Detection, Federal Trade Commission, Machine learning algorithms, Logistic Regression, KNN, Random Forest, AdaBoost, XGBoost, Classification algorithms, Bagging, Individual Models, ROC, and AUC

I. INTRODUCTION

In the modern world, with technological advancement, major financial institutions make their services available to the public through credit and debit cards, mobile banking, etc. Credit cards are among the services that have made life easier and have shown to be beneficial for online transactions. Financial institutions provide their customers with credit cards, which allow them to borrow money to make purchases. Freedom from having to always carry cash is one of the key advantages of credit cards for users. These days, transactions involving cash have supplanted those involving credit cards. As is the case with every service, credit cards have their share of drawbacks, particularly in relation to security and fraud.

Credit card fraud is a serious threat as it involves an unauthorized person using a genuine card to make purchases without the owner's knowledge or agreement, potentially resulting in financial loss for the cardholder.[1] The financial industry is witnessing an increase in fraudulent activity, particularly credit card fraud. Criminal deception committed with the intention of making money is called fraud. As credit cards are used more often daily, credit card fraud is becoming more common [2]. The gravity of the problem is highlighted in a Federal Trade Commission (FTC) report [3], which notes that 425,977 credit card fraud claims were filed in 2023 as shown in Fig. 1. It is important to remember that a large proportion of identity theft instances go undetected, indicating that the true number may be higher than those that are publicly reported.

Year	Credit card fraud reports
2019	277,739
2020	399,721
2021	395,391
2022	448,459
2023	425,977

Fig. 1. Credit card theft within the US [3]

The unauthorized acquisition of credit card credentials for purchases is known as credit card fraud. Digital or physical methods can be used to complete credit card transactions [4]. Usually, cardholders give the credit card number, expiration date, CVV, and verification number via phone or online. Card not-present transactions, or online purchases made without a physical credit card, are becoming more common [2]. Web payment gateways manage every credit card transaction; in reality, the card issuing banks are not even informed that the card is being used.

The bulk of credit card fraud is caused by external card fraud, which has been the subject of several studies. The introduction of big data has rendered manual approaches even more unfeasible in the face of time consuming and ineffective traditional methods of fraud detection. Identification of fraudulent activities is seen as a problem for machine learning for several reasons, including the fact that very few transactions made with credit cards are fraudulent and that the distribution of data changes over time because of new attack techniques and seasonality [5]. This study is innovative in that

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:53:44 UTC from IEEE Xplore. Restrictions apply.

it suggests an ensemble learning based method based on soft voting to identify fraudulent transactions with credit cards in an unbalanced dataset. Utilizing the SMOTE technique for resampling, we explore two distinct combinations of machine learning techniques: Combination 1 (KNN, Bagging, AdaBoost, Logistic Regression) and Combination 2 (Random Forest, KNN, Bagging, XGBoost). Furthermore, we evaluate the individual performance of Logistic Regression, KNN, and AdaBoosting to establish a comparative baseline for assessing the ensemble models.

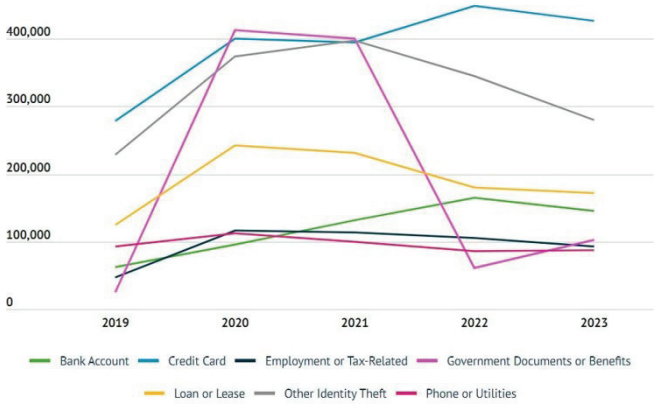


Fig. 2. Identify theft reports by type [3]

The remaining portion of the document is structured in this manner. The pertinent research on detecting fraudulent credit card transactions is covered in Section II. Section III provides an explanation of the data set and our approach hypothesis. Section IV presents a summary of the observations and results, and Section V presents the conclusion.

II. RELATED WORK

A robust understanding of credit card fraud detection technologies is important in addressing the persistent challenges associated with fraudulent activities. Mohammad et al. [14] offers a comprehensive review of prevalent credit card fraud types, highlighting application and behavior fraud as prominent categories. Application fraud involves the acquisition of credit cards through false information or legitimate cardholder data, while behavior fraud entails the unauthorized utilization of cardholder credentials by malicious actors. Shalini Gupta and Rahul Johari [8] propose a novel framework focusing on secure credit card transactions, emphasizing mutual authentication between cardholders and merchants. Their framework integrates one time password (OTP) systems and digital signatures to mitigate vulnerabilities in both onsite and online transactions. However, empirical validation and comparative analysis with existing authentication systems still need to be improved.

Mittal and Tyagi [9] evaluate various machine learning algorithms using imbalanced datasets. Their study covers supervised, hybrid, and unsupervised algorithms but falls short in discussing the potential impact of feature engineering on model performance. Thennakoon et al. [10] conducted extensive experiments using four different fraud detection models and evaluated their performance. However, a limitation of their study lies in the absence of relative efficacy of their proposed system compared to existing methods.

Xuan et al. [7] compare two variations of the random forest algorithm using data from a Chinese e-commerce company. They test these models with different ratios of legal and fraudulent transactions to assess their performance. While Random Forest II (based on CART) exhibits superior overall performance, including accuracy and F-measure, Random Forest I (based on random trees) demonstrates slightly higher precision. The study acknowledges the challenge of imbalanced data in fraud detection and suggests under sampling techniques to mitigate it. Furthermore, M. Zareapoor et al. [12] propose a comprehensive survey of fraud detection methodologies, aiming to guide the development of hybrid algorithms for more effective fraud detection in evolving transactional landscapes. Vaishnavi Nath Dornadula et al. [11] seek to solve issues with fraud detection, particularly concept drift and class imbalance. They suggested a technique in which clients are clustered according to their transactional patterns and various classifiers are trained on these groups.

The results demonstrate improvements in detection accuracy after applying techniques like SMOTE for data balancing. However, the paper lacks a comprehensive comparison with existing methods and does not provide insights into the generalizability of the proposed approach across different datasets or scenarios. Decision trees and support vector machines (SVM) are compared by Sahin and Duman [15] in terms of their ability to identify credit card fraud. Experiment data shows that the former model works better than the latter. However, it can outperform in terms of accuracy as the training dataset gets bigger.

Ganji and Mannem [13] used the SODRNN algorithm, which is based on reverse k-nearest neighbors. It's designed to efficiently detect outliers in data streams with just a single pass, unlike traditional methods that require multiple scans. The absence of a comparative analysis with other fraud detection methods, which may shed light on the relative efficacy, is a shortcoming of this work. Hassan Najadat et al. [16] utilize a dataset provided by Kaggle and apply a model along with six classifiers. They experiment with different sampling techniques to handle the highly imbalanced dataset. With random oversampling, the deep learning model achieves an AUC of 91.37%, indicating that it performs better than the machine learning classifiers. The outcomes of machine learning classifiers, on the other hand, were less promising; the highest AUC of 81% was obtained through hard voting combined with oversampling.

Randhawa et al. [17] investigate credit card fraud detection using machine learning algorithms, particularly AdaBoost and majority voting. They utilize a variety of standard models and hybrid methods, analyzing both benchmark and real-world credit card datasets. The study's main contribution lies in evaluating these methods with a real-world dataset, a departure from previous studies primarily using publicly available data. The paper's contextualization within the larger research environment may be limited, though, as it does not provide a thorough comparison with previous research using comparable algorithms.

Correa Bahnsen et al. [20] explore the crucial role of feature engineering in enhancing credit card fraud detection systems. Conventional approaches mostly rely on raw transactional data, neglecting important spending behavioral patterns. They propose a comprehensive feature engineering strategy, aggregating transactions to reveal customer spending behavior. It incorporates periodic features based on the von Mises distribution to capture transaction timing patterns. By expanding feature sets and leveraging periodic behavior

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:53:44 UTC from IEEE Xplore. Restrictions apply.

analysis, the proposed method achieves a substantial 13% increase in fraud detection savings, highlighting the importance of feature engineering in enhancing fraud detection and providing a more detailed comprehension of fraudulent activities.

III. METHODOLOGY

A. Dataset Acquisition and Preprocessing

The dataset that was used in this study was acquired from Kaggle and included 284,807 transactions throughout the period of two days. [6] There were 492 instances of fraudulence found in this dataset, demonstrating how unbalanced it is and how the fraudulent class barely accounts for 0.172% of the overall transactions. Principal Component Analysis (PCA), a commonly used technique to reduce dimensionality, produced 28 of the 31 columns that make up the total. Unfortunately, confidentiality constraints hindered the public dissemination of the dataset's initial attributes and further contextual details [18]. To guarantee data quality and compatibility, several preparation procedures were used before starting the model training process. These included filling in missing values, scaling features to ensure consistency and comparability across different variables, and encoding categorical variables when they were present. [19] Overall, the preprocessing procedures ensured that the dataset was suitably ready for further model training and analysis, even without comprehensive information on the original features and data.

B. Exploratory Data Analysis (EDA)

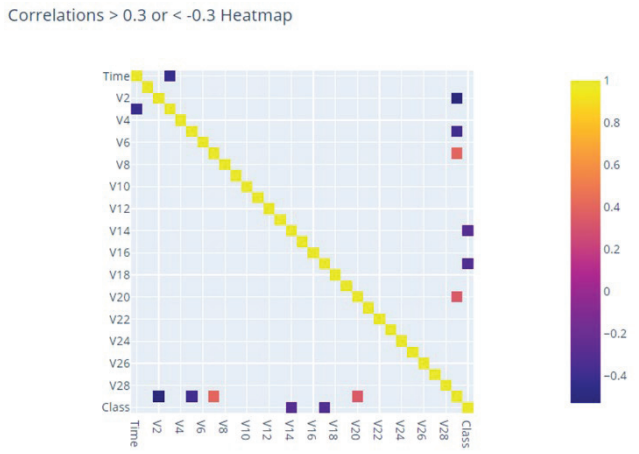


Fig. 3. Correlation matrix illustrating relationships between columns

The dataset was examined using Exploratory Data Analysis (EDA), which used visualizations such as correlation matrices (Fig. 3.), box plots, and histograms. In particular, the correlation matrix graph provided insightful information about the connections between the different PCA components and transaction amounts. Potential relationships and dependencies in the dataset were highlighted by this research, which influenced further modelling choices. The utilization of histograms and box plots facilitated a more profound comprehension of the dataset's attributes by aiding in the identification of feature distributions and variability.

C. Model Selection and Ensemble Configuration

In this investigation, two different ensemble learning model combinations were used. Combination 1 as shown in Fig. 4. used AdaBoosting, Bagging, K-Nearest Neighbors (KNN), and as shown in Fig. 5. Logistic Regression. K-Nearest Neighbors (KNN), Random Forest, Bagging, and XGBoosting were included in Combination 2. These ensembles increase prediction accuracy by strengthening the model's ability to recognize complex patterns in the data and offering robustness against overfitting by integrating the strengths of multiple algorithms. The study attempted to take advantage of the complementary qualities of each algorithm by mixing various model types and ensemble methodologies, which could result in better performance than individual models.

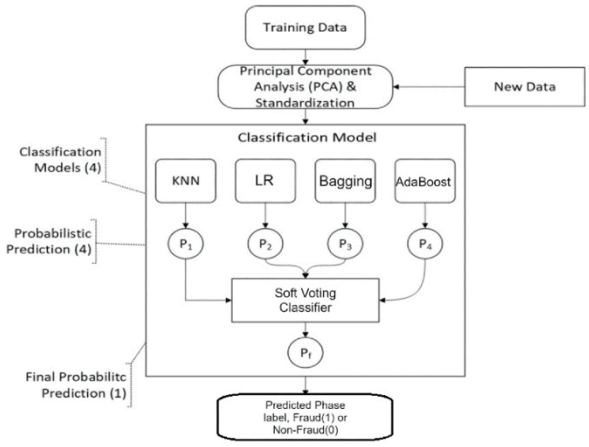


Fig. 4. Combination 1 (KNN + LR + Bagging + AdaBoost)

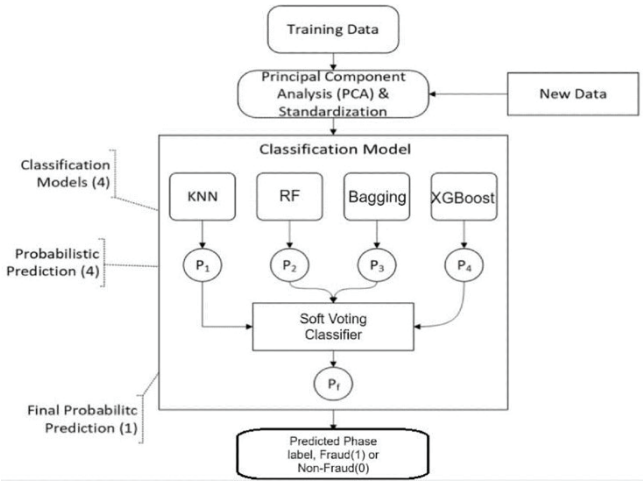


Fig. 5. Combination 2 (KNN + RF + Bagging + XGBoost)

1) KNN: KNN classifies objects based on their 'K' nearest neighbors using the majority class approach. It may identify which class label the new instance most closely matches by computing the distances between the key points of labelled instances. K indicates how many neighbors we must consider while classifying a new instance in KNN. In the case of KNN, no explicit hyperparameters were given; instead, default values were used.

2) Logistic Regression: It is an algorithm for performing binary classification. It uses a logistic function to describe the probability of an outcome and ensures predictions are between 0 and 1. It analyses the association between the dependent variable (fraud or genuine) and independent variables by fitting a sigmoid curve to the credit card data. The random state hyperparameter, which guarantees results repeatability, was used in logistic regression.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:53:44 UTC from IEEE Xplore. Restrictions apply.

3) Random Forest: The multiple trees present in the algorithm then provide us with classification using the mode of identification for each tree. The individual trees "vote" on the result during prediction about transactions, and the forecast with the largest number wins. Depending on the quality of the features found, the random forest may result in a robust classification of credit card transactions, genuine or fraudulent. The hyperparameter random state was utilized to ensure the reproducibility of random forest generation.

4) Bagging: It is the procedure for combining the output of numerous classifiers that have been trained on various subsets of the training data. Several bootstrap samples are created from the training dataset to create these subgroups. By integrating the viewpoints of several models, the combination—which is accomplished by voting or averaging—helps lower overfitting and variation. The K-Nearest Neighbors Classifier was used as the base estimator, which affected the model's performance. To make sure that the bagging procedure could be repeated, the hyperparameter random state was used.

5) AdaBoost: It is a supervised learning method that creates a strong learner from a combination of several weak or base learners (such as decision trees) to classify data. It trains a sequence of base models one after the other, changing the weights for each instance according to how accurate the models were before. Higher weights are applied to instances that past models incorrectly identified, highlighting their significance in later iterations. To get a final prediction, AdaBoost weighted sums the predictions of these weak learners. AdaBoost continuously enhances the model's generalization capacity by concentrating on cases that are challenging to categorize. The n estimators' hyperparameter was set to 50, determining how many weak learners (decision trees) would be repeatedly trained. A suitable value is essential to avoid overfitting or underfitting. random state was used to ensure the AdaBoost algorithm's reproducibility.

6) XgBoost: It uses an ensemble method for decision trees, adding trees one after the other to fix mistakes in earlier models. To increase effectiveness and performance in identifying fraudulent credit card transactions, XGBoost integrates several advances, including parallelized computing and a regularized objective function. Similarly, the n estimators hyperparameter was set to 50 to determine how many boosting stages would be conducted, and a random state was used to guarantee that the gradient boosting method was reproducible.

D. Implementation Details

Using the scikitlearn package, the selected algorithms were included in two different model combinations during the implementation stage. Combination 1 used AdaBoost, Bagging, K-Nearest Neighbors (KNN), and logistic regression classifiers wrapped in a Voting Classifier. Using a soft voting mechanism to take probabilities into account, this ensemble technique used the decisions made by each of its component models together. In Combination 2, the combination had Gradient Boosting, Bagging, KNN, and Random Forest classifiers inside of another Voting Classifier instance. To merge the predictions of individual classifiers, soft voting was once again used. The model building and group learning features of scikitlearn were heavily utilized in the implementation. A strong framework for model development and evaluation was established by utilizing the functions of scikitlearn and other pertinent libraries, such as plotly and seaborn for graph plotting and imblearn for SMOTE sampling. This allowed for the comparison of various algorithmic combinations to determine the best predictive performance.

E. Model Evaluation Metrics

This study used two primary criteria to assess the ensemble models: accuracy and ROC curve analysis. As a basic metric, accuracy shows the proportion of instances from all the model's predictions that it correctly identified. It offers a comprehensive overview of the model's overall prediction capability. The discriminatory power of the model was also evaluated using the ROC curve and its associated statistic, Area Under the Curve (AUC). For different classification thresholds, the trade-off between the true positive rate (sensitivity) and the false positive rate (specificity) is displayed by the ROC curve. Better discriminating performance, or the model's ability to reliably distinguish between positive and negative classifications, is indicated by a higher AUC value.

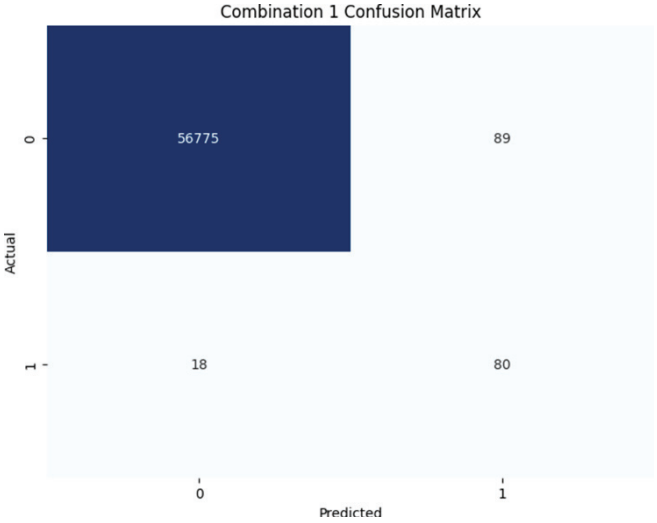


Fig. 6. Combination 1 Confusion Matrix

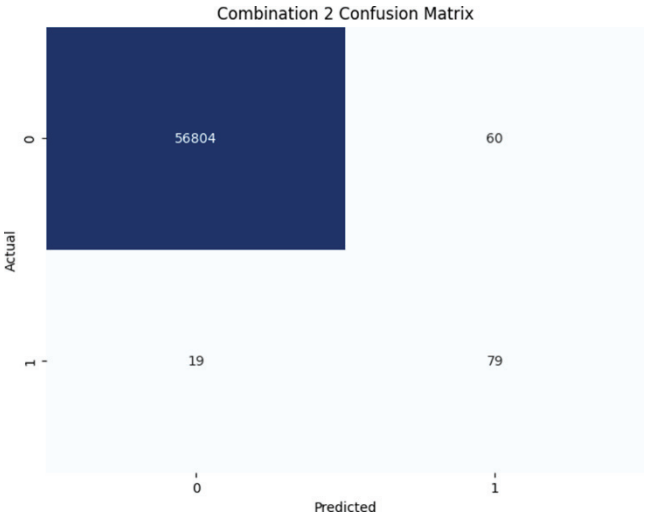


Fig. 7. Combination 2 Confusion Matrix

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:53:44 UTC from IEEE Xplore. Restrictions apply.

F. Individual Model Comparisons

Furthermore, to evaluate the effectiveness in the determination of fraud transactions, we have analyzed the individual execution of AdaBoosting, K-Nearest Neighbors (KNN), and Logistic Regression independently. The ensemble models were then assessed using this comparison as a benchmark. By separately examining the performance of each base model, insights were gained into their relative strengths and weaknesses in solving the classification problem. This approach allowed a deeper understanding of how each base learner contributed to the overall ensemble performance and whether the ensemble models outperformed the individual algorithms.

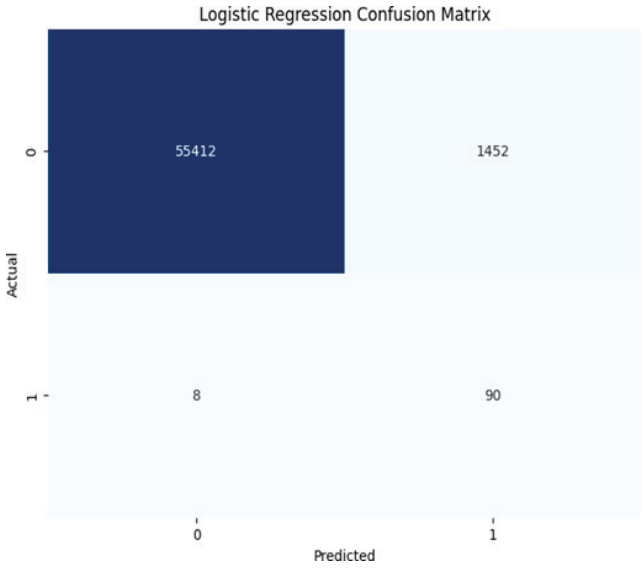


Fig. 8. Logistic Regression Confusion Matrix

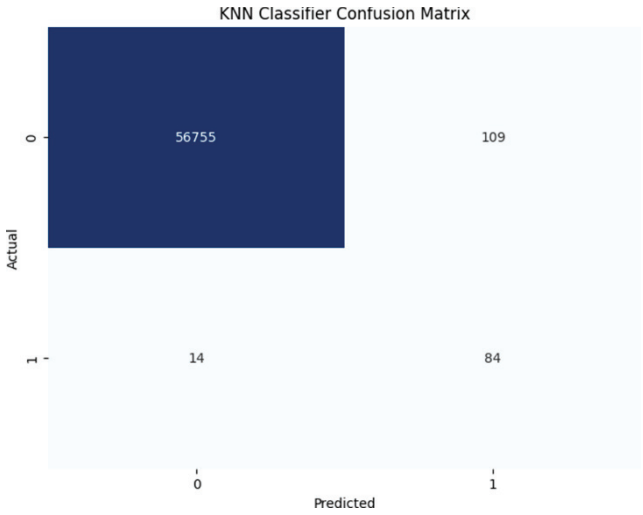


Fig. 9. KNN Classifier Confusion Matrix

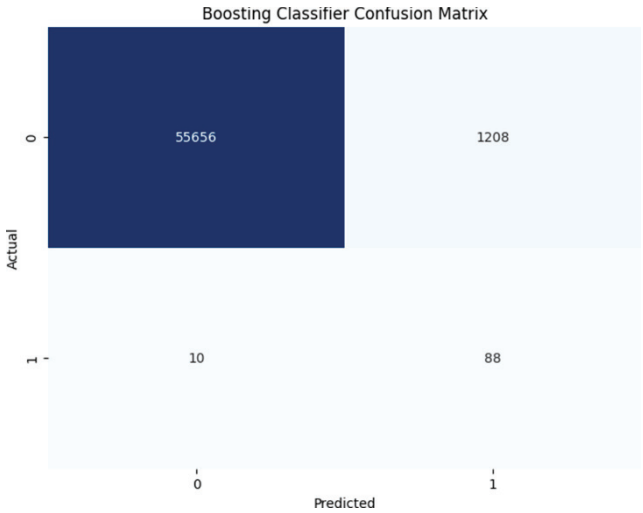


Fig. 10. Boosting Classifier Confusion Matrix

G. Ethical Considerations

Throughout the research process, ethical concerns were given top priority to maintain responsible data usage and reduce the likelihood that bias or discrimination would persist. The collection and management of the data were handled with great care. The dataset used was guaranteed to have been collected lawfully, ethically, and with the necessary permits. Furthermore, measures were implemented to reduce the possibility of bias in preprocessing the data. Among these were handling uneven class distributions and taking extra caution when encoding categorical data. Through careful evaluation of ethical issues, the research aimed to uphold integrity, accountability, and social responsibility in using data and technology.

In summary, the procedure was made up of several crucial parts. The dataset was first acquired and preprocessed to guarantee data quality and suitability for further investigation. Subsequently, models suitable for the task at hand were chosen, including both individual algorithms and ensemble approaches, depending on their attributes. Performance was assessed using pertinent metrics, such as accuracy and the ROC/AUC curve, to ascertain discrimination power and prediction ability. To assess the efficacy of ensemble models, their performance was also contrasted with that of standalone algorithms. This all-encompassing strategy addressed potential biases and ethical issues while utilizing the advantages of several approaches to create reliable predictive models. Scalability issues in this work were mostly caused by the enormous dataset size, demanding an emphasis on computing efficiency. Algorithms such as K-Nearest Neighbors (KNN) took longer to execute during inference since they relied on distance estimations and the whole training dataset. To address such computational constraints, new algorithms or optimization methodologies, including testing of different types of combinations, were investigated to improve efficiency and speed up model deployment.

IV. RESULTS

Combination 1 and Combination 2 both scored extremely high accuracy rates of over 99%, indicating high accuracy rates. Upon closer examination, however, Combination 2 performed better than Combination 1 overall when it comes to accuracy and recall as measured by the F1 score metrics. With

TABLE I. RESULT COMPARISON OF THE IMPLEMENTED MODELS

Models	Accuracy	F1-score	ROC AUC
Combination 1	99.81%	0.5993	0.9074
Combination 2	99.86%	0.6667	0.9025
Logistic Regression	97.44%	0.1098	0.8511
KNN	99.78%	0.5773	0.8871
Boosting (AdaBoost)	97.86%	0.1263	0.8752

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:53:44 UTC from IEEE Xplore. Restrictions apply.

the highest accuracy rate of 99.78% plus a respectably good F1 score of 0.5773, K-Nearest Neighbors (KNN) proved to be the most successful individual model. However, in terms of accuracy and F1 score, AdaBoost and Logistic Regression (LR) showed comparatively worse performance metrics.

While accuracy rates are a useful indicator of overall correctness, other metrics like the F1 score and ROC AUC need to be included to assess the model's effectiveness in its whole. Overall, the ensemble learning approach showed promising results, emphasizing the need to merge many models to increase prediction accuracy. The systematic evaluation of many metrics clarifies the nuances of each model's benefits and drawbacks, facilitating informed decision making and providing opportunities for continuous enhancement and change of predictive models in future projects.

V. CONCLUSION

In this study, we use credit card transaction data from Kaggle to examine the ability to predict ensemble learning techniques. After carrying out an extensive testing procedure with several ensemble models and individual algorithm combinations, we discovered favorable outcomes. The ensemble combinations showed strong predictive power, with Combination 2 showing the highest accuracy rates at 99.86%. Additionally, Combination 2 fared better on the F1 score than Combination 1, showing better recall and overall precision. The top individual model performance was K-Nearest Neighbors (KNN), It has a 99.78% accuracy rate and an F1 score of 0.5773. Logistic Regression (LR) and AdaBoost scored inferiorly when it came to the F1 score and ROC AUC, although they had comparable accuracy rates.

The F1 score and ROC AUC offer more insightful information about model performance, which emphasizes the significance of evaluation metrics beyond correctness in the outcomes. The high accuracy rates of ensemble combinations show how ensemble learning techniques, such as soft voting classification, can use the capabilities of numerous models to improve predictive performance. Future research may concentrate on adjusting hyperparameters, analyzing feature importance, ensuring model interpretability, and tackling biases or ethical issues.

REFERENCES

[1] N. Shirodkar, P. Mandrekar, R. S. Mandrekar, R. Sakhalkar, K.M Kumar, S. Aswale, "Credit Card Fraud Detection Techniques – A Survey", International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), April 2020

[2] Sahithi, G.L.; Roshmi, V.; Sameera, Y.V.; Pradeepini, G. Credit Card Fraud Detection using Ensemble Methods in Machine Learning. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 28–30 April 2022; pp. 1237–1241.

[3] Federal Trade Commission. CSN-Data-Book-2022. no. February 2023. Available online: https://www.fool.com/the-ascent/research/identity-theft-credit-card-fraud-statistics/

[4] A. O. Adewumi and A. A. Akinyelu," A survey of machine-learning and nature-inspired based credit card fraud detection techniques", Int. J. Syst. Assurance Eng. Manage., vol. 8, no. 2, pp. 937-953, 2017.

[5] P. A. Dal, G. Boracchi, O. Caelen, C. Alippi and G. Bontempi," Credit card fraud detection: A realistic modeling and a novel learning strategy", IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 8, pp. 3784-3797, Sep. 2017

[6] Kaggle, "Credit Card Fraud Detection," www.kaggle.com, 2018. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

[7] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random Forest for credit card fraud detection - IEEE Conference Publication," Ieee.org, 2018. https://ieeexplore.ieee.org/abstract/document/8361343

[8] Gupta, Shalini, and R. Johari." A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant." International Conference on Communication Systems and Network Technologies IEEE, 2011:22-26.

[9] S. Mittal and S. Tyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection," IEEE Xplore, Jan. 01, 2019. https://ieeexplore.ieee.org/document/8776925

[10] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," IEEE Xplore, Jan. 01, 2019. https://ieeexplore.ieee.org/document/8776942

[11] V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," Procedia Computer Science, vol. 165, pp. 631–641, 2019, doi: https://doi.org/10.1016/j.procs.2020.01.057.

[12] M. Zareapoor, Seeja. K. R. Seeja.K.R, and M. Afshar Alam, "Analysis on Credit Card Fraud Detection Techniques: Based on Certain Design Criteria," International Journal of Computer Applications, vol. 52, no. 3, pp. 35–42, Aug. 2012, doi: https://doi.org/10.5120/8184-1538.

[13] Ratnam, Venkata. (2012). credit card fraud detection using anti k-nearest algorithm. IJCSE. volume 4. 1035.

[14] Behdad, M., Barone, L., Bennamoun, M., and French, T. (2012). Nature inspired techniques in the context of fraud detection. IEEE Transactions on Systems Man and Cybernetics Part C, 42(6), 1273-1290.

[15] Sahin, Y., and Duman, E. (2011). Detecting credit card fraud by decision trees and support vector machines. Lecture Notes in Engineering and Computer Science, 2188(1).

[16] H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, "Credit Card Fraud Detection Based on Machine and Deep Learning," IEEE Xplore, Apr. 01, 2020. https://ieeexplore.ieee.org/document/9078935

[17] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," IEEE Access, vol. 6, pp. 14277–14284, 2018, doi: https://doi.org/10.1109/access.2018.2806420.

[18] A. Saputra, "Fraud Detection using Machine Learning in e-Commerce," International Journal of Advanced Computer Science and Applications, vol. 10, no. 9, 2019, doi: https://doi.org/10.14569/ijacsa.2019.0100943.

[19] Seiffert, C.; Khoshgoftaar, T.M.; Van Hulse, J.; Napolitano, A. RUSBoost: Improving classification performance when training data is skewed. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.

[20] Correa Bahnsen, Alejandro, Aouada, Djamila, Stojanovic, Aleksandar, Ottersten, Bjo¨rn. (2016). Feature Engineering Strategies for Credit Card Fraud Detection. Expert Systems with Applications. 51. 10.1016/j.eswa.2015.12.030

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:53:44 UTC from IEEE Xplore. Restrictions apply.

Applied Physics B (2024) 130:157
https://doi.org/10.1007/s00340-024-08294-x

Applied Physics B
Lasers and Optics

RESEARCH

Enhanced emission in the NIR range by Nd^{3+} doped borosilicate glass for lasing applications

Yasha Tayal[1] · Rupesh A. Talewar[2] · Sk. Mahamuda[3] · Kartika Maheshwari[1] · Sheetal Kumari[4] · Mohit Kumar[4] · Ravita Pilania[5] · Aman Prasad[6] · A. S. Rao[4]

Received: 4 April 2024 / Accepted: 26 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Borosilicate glasses were prepared via melt quench technique with different concentration of Nd^{3+} ions. X-ray diffraction (XRD) was performed for the prepared glass to study the amorphous structure. The glass stability was studied by Thermogravimetric Analysis (TGA). Spectroscopic analysis of the prepared glasses was done through absorption, excitation, emission and decay measurements. Absorption spectra was used to evaluate the J-O parameters. Radiative parameters are evaluated with the help of application of J-O theory. Emission spectra show three prominent bands in NIR region arising from $^4F_{3/2} \rightarrow {}^4I_{9/2}$, $^4I_{11/2}$ and $^4I_{13/2}$ centered around 910, 1068 and 1339 nm accordingly, with the most intense being the one at 1068 nm. Stimulated emission cross section, branching ratio and quantum efficiency for $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition is very high and is suitable for lasing applications. Dipole–dipole interaction is observed among activator ions on the basis of Dexter theory. The optimized Nd^{3+} doped glass (NBS10) possess maximum value for all radiative parameters and hence can be best suited for applications in the area of lasers and fiber amplifiers.

1 Introduction

Hyaline substances doped with a rare earth (RE) ions find vast usage in numerous optoelectronic devices like light emitting diodes (LEDs), fiber amplifiers, solid state lasers, optical waveguides, temperature sensors, etc. [1–5]. Intense lasing emissions in NIR range for Nd^{3+} places it as most famous RE ion for luminescence based characteristics [6, 7]. The transition at 1068 nm is sharpest and hence, is vastly explored for usage in high power lasers. Also, it can be observed easily at room temperature [8, 9]. While the one associated with 1340 nm finds huge utility in enhancing the O-band optical signal and provides a window in telecommunication transition, the emission centered around 900 nm provides a window for forging a device that could produce blue emission via monolithic integration of emission from a powerful diode laser [10, 11].

Various glassy systems like silicate, borates, phosphates, tellurites, chalcogenides and fluorides have been explored for the NIR range luminescence via Nd^{3+} ions [7, 10, 12–15]. Among these, borosilicate, a combination of borate and silicate glass systems, provides fine RE ion solubility, good mechanical strength, thermal and chemical durability [16]. Further, luminescence characteristics can be enhanced by reducing the phonon energy in the glass host, which in turn produces further utility for fiber amplifiers and high efficiency lasers [16, 17]. Phonon energy in this system can be lowered upon introduction of heavy metal oxide (PbO) which also increases the refractive index for the glass system. It further helps in reduction of transition temperature and allows for a wider spectral transmission window. Incorporation of ZnO, a non-toxic and non-hygroscopic oxide in the host matrix paves the way for better optical, magnetic and electrical characteristics.

✉ Yasha Tayal
yashaphy10nov@gmail.com

1 ABES Engineering College, Ghaziabad 201009, India

2 Physics Department, Shri Ramdeobaba College of Engineering and Management, Katol Road, Nagpur 440013, India

3 Department of Engineering Physics, College of Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522302, India

4 Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110042, India

5 Department of Physics, Chaudhary Bansilal University, Bhiwani 127021, India

6 Department of Physics and Computer Science, Dayalbagh Educational Institute (DEI), Deemed University, Agra 282005, India

Published online: 12 August 2024

ZnO also possess some unique properties like direct wide band gap, large intrinsic emitting and higher exciting binding energy which in turn helps the glass system to find applications in solar energy converters, optoelectronic devices and gas sensors. By taking up the tetrahedral and octahedral positions in the host matrix, the addition of Al_2O_3 enhances its chemical and thermal stability as well as its mechanical strength, emission characteristics, and solubility of RE ions. It also decreases crystallization. The inclusion of alkali oxides (Li_2O, Na_2O, etc.) in the borate matrix can change the three coordinated boron atoms in the system into four coordinated tetrahedral structure, which improves chemical stability and causes non-bridging oxygens (NBO) to be produced [18–22]. As compared to phosphors, glasses possess easier fabricating technique and lower melting temperature leading to better incorporation in the optical devices [10, 23].

Nd^{3+} ions populate the $^4F_{5/2}$ level via pumping and later decay to $^4F_{3/2}$ state via non-radiative transition (multi phonon relaxation). Subsequently, the ions decay to $^4I_{13/2}$, $^4I_{11/2}$ and $^4I_{9/2}$ levels from $^4F_{3/2}$ emitting radiation in NIR range at 900, 1068 and 1340 nm respectively. These NIR range emissions have two major setbacks. Firstly, nearby Nd^{3+} ions tend to suffer cross relaxation mechanism, leading to quenching thereby mandating an optimized concentration of Nd^{3+} ions [24, 25]. Secondly, to acquire a rapid multi phonon relaxation rate of Nd^{3+} ions from the $^4F_{5/2}$ ($^2H_{9/2}$) to $^4F_{3/2}$ level when excited by 585 nm, the ambient phonon energy possessed by glass host should be close to the bridge energy gap (~1000 cm^{-1}). Minimum non radiative transition produced by multiphonon relaxation provides a good window for quantum yield and hence can be better used as laser amplifiers.

Great interest by numerous researchers has been observed in studying the physical, absorption and fluorescence properties of Nd^{3+} doped glass matrix and it can be improvised by selecting appropriate glassy system [26–28]. In the present study, we aim to prepare Nd^{3+} ions doped borosilicate glass and hence to identify the improvised NIR lasing band obtained upon the introduction of optimized Nd^{3+} content. On account of above scientific parameters, we prepared a borosilicate glass with Li_2O, PbO, ZnO and Al_2O_3 as intermediates for enhanced thermal and mechanical strength. Dwelling upon the aforementioned utilities, we intend to state a pertinent examination of lithium lead zinc alumino borosilicate glass doped via various concentrations of Nd^{3+} ions figuring out optimized content of the RE ion for intense laser emission at 1068 nm.

2 Experimental procedure

With the help of melt quench technique Nd^{3+} doped glasses were prepared according to the following glass composition: $10Li_2O–5PbO–(5–x)ZnO–10Al_2O_3–10SiO_2–60B_2O_3–xNd_2O_3$, where, $x = 0.1$, 0.5, 1.0, 1.5, 2.0 and 2.5 mol%. For the ease of recognition, these samples were tagged as NBS01, NBS05, NBS10, NBS15, NBS20 and NBS25 accordingly. The starting reagents i.e. H_3BO_3, SiO_2, PbO, ZnO, Al_2O_3, Li_2O and Nd_2O_3 of high purity (98–99.99%) were taken. A batch of 7.0 g of ingredient chemicals were measured and then grounded finely via agate mortar so as to obtain homogeneous mixture. This mixture was transferred to an alumina crucible and then placed in an electric furnace at 1100 °C for 3 h. The melt so obtained was suddenly quenched by decanting it on a preheated brass plate and pressing with a similar one. The obtained glass samples were further annealed in a muffle furnace at 350 °C for about 4 h. This process helps in releasing internal mechanical stress which avoids cracking and bubbles formation in the glass samples. Figure 1 shows the photograph of Nd^{3+} ions doped NBS glasses. These samples were then characterized at room temperature (RT). Archimedes' principle and Brewster's method was used to estimate the densities and refractive indices respectively of NBS glasses using xylene as the immersion liquid. Using an XPERT-PRO diffractometer, the XRD spectral recordings were made with a step size of 0.02 and a 2θ angle that ranged from 10° to 75°. JASCO V-670 model UV–vis–NIR spectrophotometer was used to record the optical absorption spectra of the samples. Hitachi-F7000 fluorescence

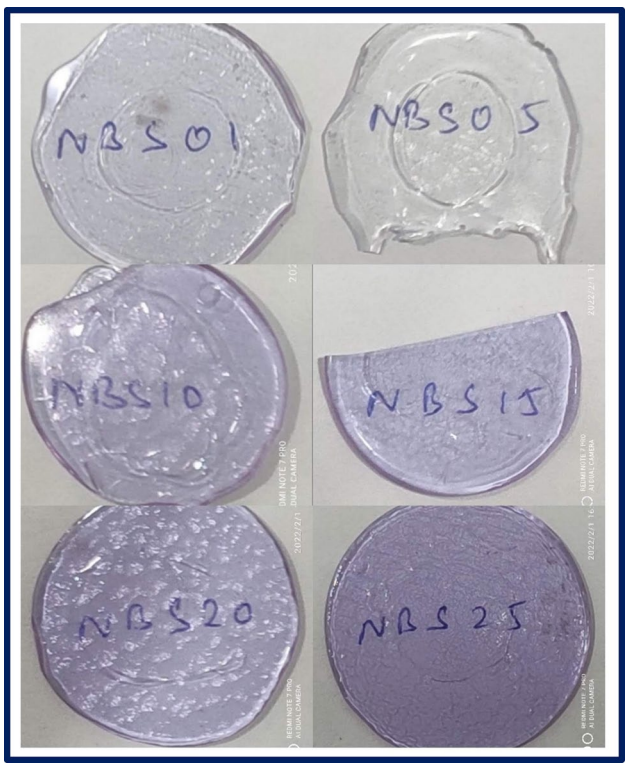


Fig. 1 Nd^{3+} ions doped NBS glass samples

spectrofluorometer was used to record photoluminescence (PL) excitation, emission spectra and fluorescence decay.

3 Results and discussion

3.1 Physical properties and XRD

Density is a useful tool for examining changes in coordination, the degree of structural compactness, the geometrical topologies of the glass network, and the dimensions of interstitial holes. The values of density measured for the present glass system are 2.64, 2.66, 2.67, 2.69, 2.73 and 2.76 g/cm^3 as the mol% of Nd^{3+} ions increased from 0.1 to 2.5 mol% respectively. The refractive index of optical materials is another crucial parameter. The values of refractive index measured for the 0.1 to 2.5 mol% of Nd^{3+} ions doped NBS glasses are 2.42, 2.47, 2.51, 2.54, 2.59 and 2.62 respectively. The calculated values of molar volume for the 0.1 to 2.5 mol% of Nd^{3+} ions doped NBS glasses are 3062, 3077, 3114, 3138, 3139 and 3151 cm^3 respectively. Figure 2 shows that the refractive index, density and molar volume of the glasses which increase as the concentration of Nd^{3+} ions rises, signifying the addition of Nd_2O_3 in place of B_2O_3. These increasing values of density and molar volume suggest that there are more non-bridging oxygen molecules (NBOs) in the glass network. Since Nd_2O_3 enters the glass network as a network modifier, occupies an interstitial space within the network and ultimately produces NBOs in the glass structure, the glass network grows with the addition of Nd_2O_3.

Figure 3 depicts the XRD pattern of the un-doped NBS glass. It shows the amorphous behaviour of the as-synthesized glass. The broad hump signifies absence of long range order in the glass lattice.

3.2 Thermal analysis

The TGA curve was utilized to calculate the weight loss and identify the weight loss mechanisms of the powder precursors to form pure NBS glass. As seen in Fig. 4, the TGA curve illustrates weight loss across three separate stages

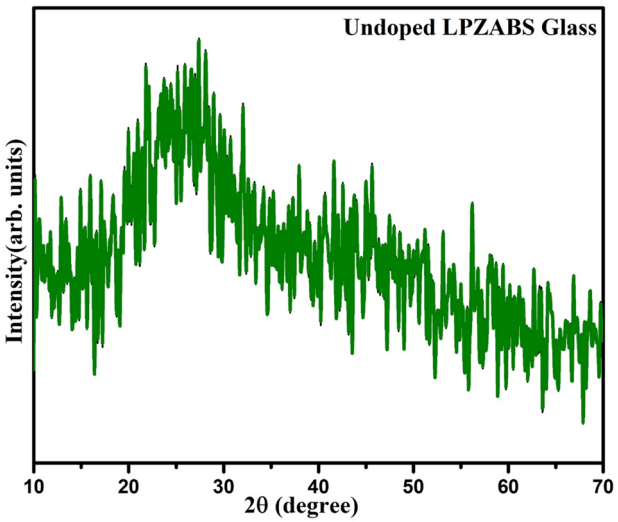


Fig. 3 XRD pattern of undoped NBS glass

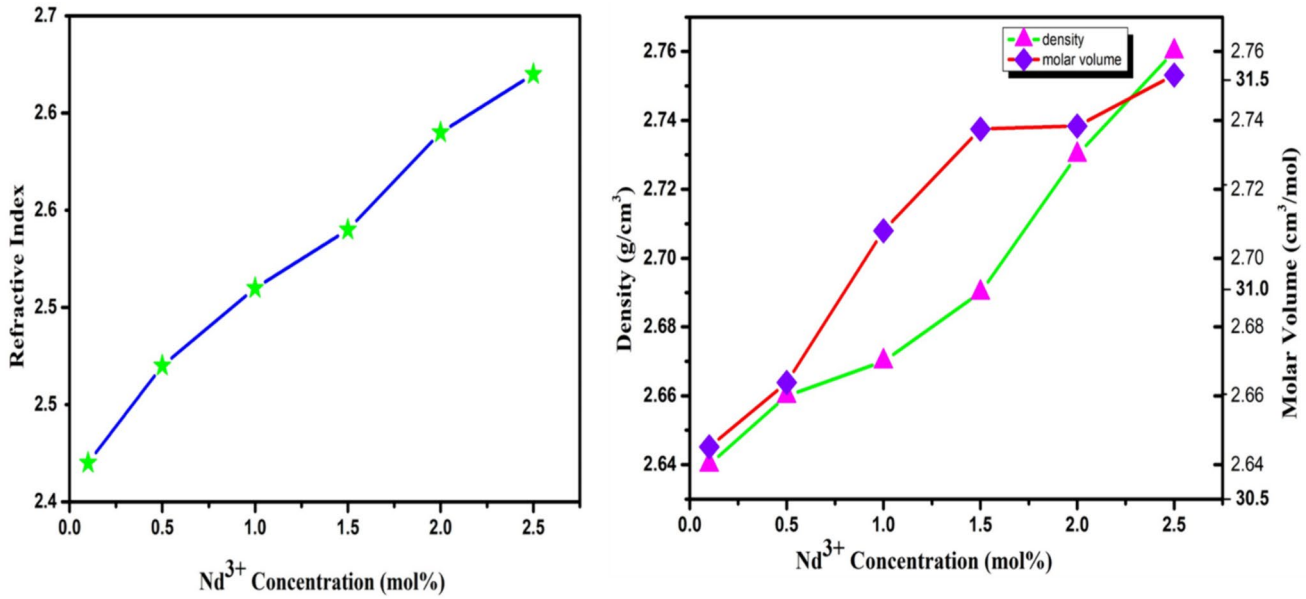


Fig. 2 Variation of refractive index, density and molar volume with the concentration of Nd^{3+} ions in NBS glass samples

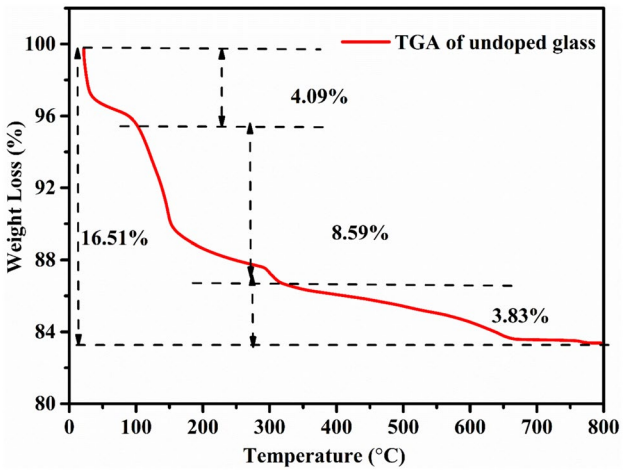


Fig. 4 TGA curve for the undoped sample

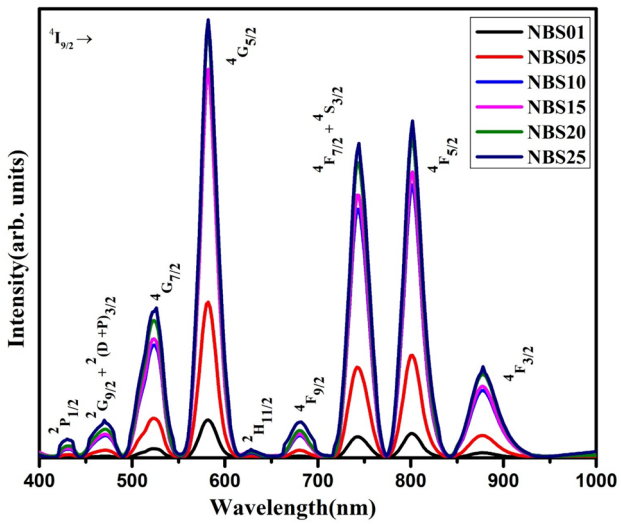


Fig. 5 Absorption spectra of Nd^{3+} ions in NBS glasses

from 21 to 800 °C. Evaporation of the absorbed water from the sample's surface causes a weight loss of 4.09% during the first stage, which is observed between 21 and 102 °C. The exhaustion and breakdown of remaining oxides in the sample is represented by the second stage (102–334 °C), which has a weight loss of 8.59%. The presence of combustibles in the sample is the reason of the third stage, which is between 334 and 768 °C and has a weight loss of 3.83%. The evaporation of various gases and the exhaustion of organic species cause a total weight loss of around 16.51%. Therefore, it is evident from Fig. 4 that the glass sample have a lesser mass loss at high temperatures and is thermally more stable.

3.3 Absorption spectrum and Judd–Ofelt parameters

Absorption spectra for Nd^{3+} doped NBS glass is shown in Fig. 5 for the wavelength range of 300–1100 nm. The figure comprises of various peaks of f–f electronic transitions belonging to Nd^{3+} ions from their ground state. The transitions found in the UV–vis–NIR range are from $^4I_{9/2} \rightarrow {}^4F_{3/2}$, $^4F_{5/2}$, $^4F_{7/2}+{}^4S_{3/2}$, $^4F_{9/2}$, $^2H_{11/2}$, $^4G_{5/2}$, $^4G_{7/2}$, $^4G_{9/2}+{}^2(D+P)_{3/2}$ and $^2P_{1/2}$ which are centered at 876, 800, 743, 680, 628, 582, 525, 470 and 429 nm respectively. These transitions are quite identical to other reported Nd^{3+} doped glasses excluding the minor changes in band positions and respective intensities, on account of the present glass system possessing diverse ligand fields having corresponding nature [7, 29]. As the intensity for all bands increases on increasing the Nd^{3+} content in the glass, it can be concluded that the Nd^{3+} ions are very well homogeneously dissolved in the glass matrix.

Nephelauxetic ratio (β) and bonding parameter (δ%) help in determination of the type of Nd^{3+} ligand bond [16].

$$\beta = \frac{\Upsilon_a}{\Upsilon_b} * 100 \tag{1}$$

here, Υ_a, Υ_b represents the energies for the transitions in complex and free ion or aqua ion respectively as given in Carnall paper [30]. From all calculated β values respective to the transitions occurring in absorption spectra, $\bar{\beta}$ the average value was calculated, which further helps in determination of the bonding parameters by using following expression:

$$\delta\% = \left[\frac{1-\bar{\beta}}{\bar{\beta}}\right] \times 100 \tag{2}$$

Bonding parameter value is either positive or negative corresponding to covalent or ionic bonding of the RE ions with the local host in glassy system. Hence, as the values obtained for the bonding parameters are positive for all Nd^{3+} doped NBS glasses, we can conclude that the bonding between the ions is covalent in nature. The values of nephelauxetic ratios and bonding parameters of all the samples are tabulated in Table 1.

Various important parameters like experimental oscillator strength, Judd–Ofelt parameters, various radiative properties such as transition probability, radiative lifetime and fluorescence branching ratio rely upon the data from the absorption spectra. The area under the curve for each transition occurring in absorption band is used to estimate the experimental oscillator strength (f_{exp}) of f–f induced electric -dipole transitions by the following expression [1, 8]:

$$f_{exp} = 4.32 \times 10^{-9} \int_{\nu 1}^{\nu 2} \varepsilon(\nu) d\nu \tag{3}$$

Table 1 Experimental (f_{exp}) ($\times 10^{-6}$), calculated (f_{cal}) ($\times 10^{-6}$) oscillator strengths, r.m.s deviation (δ_{rms}), nephelauxetic ratio ($\bar{\beta}$) and bonding parameters (δ %) of Nd^{3+} ions in NBS glasses

Transitions $^4I_{9/2}\rightarrow$	NBS01		NBS05		NBS10		NBS15		NBS20		NBS25	
	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}
$^4F_{3/2}$	3.32	2.72	3.81	2.92	4.07	3.13	2.82	2.41	2.37	2.03	1.97	1.76
$^4F_{5/2}$	11.94	12.52	12.04	13.15	14.91	15.98	10.34	11.35	8.74	9.61	7.28	8.01
$^4F_{7/2}+{}^4S_{3/2}$	10.74	9.98	11.75	10.42	14.44	13.11	10.27	9.12	8.69	7.72	7.17	6.37
$^4F_{9/2}$	0.92	1.45	0.93	1.52	1.06	1.89	0.95	1.32	0.85	1.12	0.74	.92
$^2H_{11/2}$	0.18	.39	0.19	.41	0.20	.50	0.13	.35	0.11	.30	0.10	.25
$^4G_{5/2}$	32.67	32.65	33.64	33.65	37.99	37.97	25.58	25.64	19.65	19.71	15.80	15.86
$^4G_{7/2}$	5.45	5.72	6.07	6.02	6.54	6.76	6.09	4.87	5.07	3.98	4.48	3.34
$^2G_{9/2}+{}^2(D+P)_{3/2}$	0.48	.54	1.04	.57	1.15	.65	0.99	.48	0.87	.40	0.63	.34
$^2P_{1/2}$	0.65	.42	0.91	.47	1.05	.42	0.79	.36	0.66	.31	0.61	.28
$\delta_{rms}(\times 10^{-6})$	0.439		0.715		0.768		0.720		0.617		0.555	
$\bar{\beta}$	0.9954		0.9955		0.9957		0.9958		0.9965		0.9969	
δ%	0.4591		0.4495		0.4239		0.4164		0.3504		0.3084	

Here, for the wavenumber (cm^{-1}), $\varepsilon(\nu)$ represents the molar extinction coefficient while $d\nu$ represents the half-bandwidth corresponding to the absorption band. If Gaussian shapes do occur for the absorption bands, then for such circumstances oscillator strength could be easily determined by half width technique according to the relation [1, 8]:

$$f_{exp} = 4.32 \times 10^{-9} \sum \Delta\nu \tag{4}$$

Here, $\Delta\nu$ (cm^{-1}) represents the width of the band at half the intensity of the peak. However Gaussian shape does not occur for glasses generally and bands suffer from inhomogeneous broadening which forces us to not use the latter technique, as it would give inaccurate results. For the present study, intensities for all absorption bands have been determined via area method. Calculated oscillator strength (f_{cal}) are evaluated in accordance with the Judd–Ofelt (J–O) theory, for f–f intensity model with the help of least square fit analysis from initial to final state [31]. Proper relation for estimating the f_{cal} was used from literature [32, 33]. Root mean square deviation (δ_{rms}) is used to estimate the goodness of fit among f_{exp} and f_{cal} by the equation:

$$\delta\text{rms} = \left[\frac{\sum \left(f_{exp} - f_{cal}\right)^2}{p}\right]^{1/2} \tag{5}$$

The number of energy levels in totality for present fitting technique are represented by p. The experimental and calculated oscillator strengths along with δ_{rms} are shown in Table 1 for all Nd^{3+} doped NBS glasses. Smaller values of δ_{rms} signifies the precision of estimation used. Applying least square fitting procedure provided in literature [34], the three J–O intensity parameters Ω_t (t = 2, 4, 6) for Nd^{3+} ions doped NBS glasses have been determined and are listed in Table 2 with other reported JO parameters [35–38]. The trend followed by JO parameters for Nd^{3+} doped NBS glasses is $\Omega_6 > \Omega_2 > \Omega_4$ and these values increase upto 1 mol%. Beyond this concentration, the values for these three parameters decrease. As similar dopant sites are occupied, the JO intensity characteristics ideally do not need to vary with concentration in the same host. However, if the clustering of dopant ions occurs, then variation in JO intensity parameters is observable, like the one noticed in the present work [39]. These three parameters are concerned with numerous physical and structural characteristics of glasses. Ω_2 is highly sensitive to the RE ions surroundings and is associated with the lack of symmetry with respect to coordination structure, type of bonding and polarizability of ligand ions. Since Ω_2 is correlated with the asymmetry of the immediate environment around the Nd^{3+} sites, it demonstrates the reliance of the covalence between Nd^{3+} ions and ligand anions. The ion site becomes less centro symmetrical and forms a stronger ionic chemical interaction with the ligands when the Ω_2 parameter values increase. The current investigation reveals that the values of Ω_2 changes from 5.8593 to 5.7406 and then increases to 6.4213 as the concentration of Nd^{3+} ions increases from 0.1, to 0.5 and to 1.0 mol%, respectively, indicating a rise in the covalent nature of the Nd–O bond. The values obtained in the present study are higher than the values reported in the different composition, presented in the Table 2 [35–38]. Ω_4 is the parameter which defines the bulk characteristics such as rigidity and viscosity of the host glass system comprising of the RE ions. Ω_6 is inversely proportional to the covalency of Nd–O bonds and this can be altered on modification of the host glass structure or composition [40, 41]. J-O intensity parameters for NBS10 glass shows the highest value among the six prepared glasses by varying Nd^{3+} ion content, so we can say that NBS10 possess

Table 2 Judd–Ofelt parameters ($\Omega_\lambda \times 10^{-20}$cm^2) and their spectroscopic quality factor ($\chi = \Omega_4/\Omega_6$) of Nd^{3+} ions in NBS glasses along with various reported hosts

Glass system	Ω_2	Ω_4	Ω_6	χ	Trend	References
NBS01	5.8593	1.6311	7.1709	0.227	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
NBS05	5.7406	1.7295	7.1778	0.240	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
NBS10	6.4213	1.4773	8.7806	0.168	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
NBS15	4.0929	1.2659	5.9441	0.213	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
NBS20	2.9608	1.0277	4.8332	0.212	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
NBS25	2.2733	0.9155	3.8884	0.235	$\Omega_6 > \Omega_2 > \Omega_4$	Present work
SBNACZ	4.71	4.54	5.05	0.899	$\Omega_6 > \Omega_2 > \Omega_4$	[35]
BPN70	3.523	2.984	5.480	0.544	$\Omega_6 > \Omega_2 > \Omega_4$	[36]
CS-TCP	4.80	3.20	5.15	0.621	$\Omega_6 > \Omega_2 > \Omega_4$	[37]
40CaO-60SiO_2	4.26	4.06	4.71	0.861	$\Omega_6 > \Omega_2 > \Omega_4$	[38]
40BaO-60SiO_2	3.82	3.45	3.93	0.877	$\Omega_6 > \Omega_2 > \Omega_4$	[38]

more asymmetry, high covalency, and maximum rigidity as compared with the other Nd^{3+} doped NBS glasses.

Hypersensitive transition for Nd^{3+} doped NBS glasses is $^4I_{9/2} \rightarrow {}^4G_{5/2}$ situated around 582 nm in Fig. 5. This transition is entirely influenced by the RE ion-ligand surrounding and obeys the selection rule $\Delta S = 0$, $\Delta L \leq 2$ and $\Delta J \leq 2$. Along with this, the $^4I_{9/2} \rightarrow {}^4G_{5/2}$ transition possess maximum values for oscillator strengths and matrix elements $\|U^2\|^2$, showing an association with hypersensitivity and Ω_2 parameter (both are directly proportional). Ω_4 and Ω_6 help in calculating the radiative analysis like stimulated emission cross section, transition probability and branching ratio for the active ions. Ω_4 and Ω_6 are quite impactful in considering the neodymium emission intensity. For the present case, $^4F_{3/2} \rightarrow {}^4I_{9/2}$ transition is dependent upon Ω_4 while Ω_6 is accountable for $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition. Spectroscopic quality factor (χ), a ratio between Ω_4 and Ω_6, thats helps in recognizing the path way for the excited metastable state $^4F_{3/2}$ to decay to ground state. Smaller the value of χ, better will be the probability for lasing in NIR corresponding to $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition. According to relation between χ and NIR transitions ($^4F_{3/2} \rightarrow {}^4I_{9/2}$ and $^4I_{11/2}$), if its values are less than 1 then $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition has higher efficiency and is better for lasing. While if its values are more than one then, $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition efficiency is reduced and $^4F_{3/2} \rightarrow {}^4I_{9/2}$ transition efficiency is increased [42]. For the present case study, NBS10 possess minimum value of χ as 0.168 (however, usual reported values are in the range 0.22–1.5) and is shown in Table 2 [43].

3.4 Emission and radiative characteristics

Photoluminescene excitation spectra for 1 mol% (NBS10) Nd^{3+} doped NBS glass at 1068 nm emission wavelength is shown in Fig. 6. Eleven bands arising from $^4I_{9/2} \rightarrow {}^2I_{11/2}$, $^4D_{3/2}$, $^2P_{1/2}$, $^2K_{15/2}$, ($^4G_{9/2}$, $^2K_{13/2}$), ($^4G_{7/2}$, $^4G_{5/2}$), $^2H_{11/2}$, $^4F_{9/2}$, ($^4S_{3/2}$, $^4F_{7/2}$), $^2H_{9/2}$ and $^4F_{5/2}$ which are centered around 293, 358, 431, 475, 526, 585, 627, 683, 738, 800 and 825 nm in UV–vis–NIR range can be seen. Among all the transitions, 585 nm was the most intense peaks observed. Therefore, the emission spectra of all the prepared glass samples were recorded at this excitation wavelength. Figure 7 shows the emission spectra with three peaks in NIR range, arising from the transitions $^4F_{3/2} \rightarrow {}^4I_{9/2}$, $^4I_{11/2}$ and $^4I_{13/2}$ centered around 910, 1068 and 1339 nm respectively. Among these bands, the most intense one is at 1068 nm and so, is best suited for laser transition. In the inset diagram of Fig. 7 the variation of intensity for $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition with the concentration of Nd^{3+} ions as shown in the form of bar graphs. The observed intensity of $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition increases till 1 mol% of the RE ion beyond which, the intensity of the band decreases. This decrement is ascribed due to increased interaction among Nd^{3+}—Nd^{3+} ions and also among Nd^{3+}-host defects, resulting in the cross-relaxation process in the effective ion centres through concentration

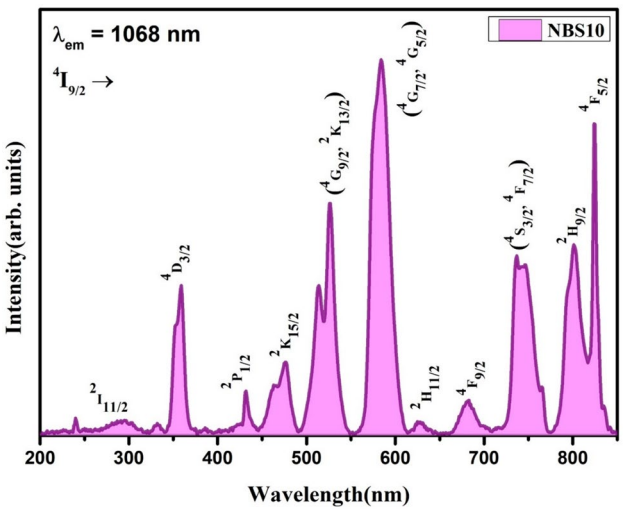


Fig. 6 Excitation spectra of Nd^{3+} ions in NBS10 glass monitored at 1068 nm emission

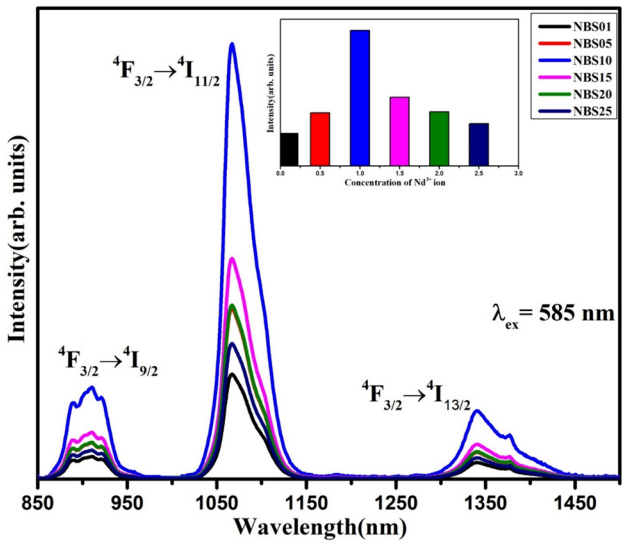


Fig. 7 Emission spectra of Nd^{3+} ions in NBS glasses under 585 nm excitation (*Inset* Variation of $^4F_{3/2}\rightarrow{}^4I_{11/2}$ transition with the concentration of Nd^{3+} ions)

quenching. Figure 8 depicts the diagram for energy level transitions of the Nd^{3+} ions in the NBS glasses emphasizing upon the excitation, emission and cross- relaxation processes. As the Nd^{3+} ions are excited from $^4I_{9/2}$ ground state to $^4F_{5/2}$ excited state, they decay to a metastable state $^4F_{3/2}$ via a non-radiative decay. This $^4F_{3/2}$ level gets strongly populated by Nd^{3+} ions resulting in three radiative transitions: $^4F_{3/2}\rightarrow{}^4I_{9/2}$, $^4I_{11/2}$ and $^4I_{13/2}$. The transition corresponding to $^4I_{11/2}$ is most intense. When the Nd^{3+} ions content is increased in the host glass system, it leads to reduction in the separation between Nd^{3+} ions resulting in increased cross- relaxation mechanism among the Nd^{3+} ions. These processes cumulatively depopulate the Nd^{3+} ions situated at $^4F_{3/2}$ energy level. Cross relaxation channels involved in the process is $^4F_{3/2}+{}^4I_{9/2}\rightarrow{}^4I_{15/2}+{}^4I_{15/2}$ result in loss of energy through non-radiative mechanism. The non radiative energy transfer can occur by either exchange interaction or multipolar interaction. The kind of inter-ionic interaction between the dopant ions can be accessed via Dexter Theory. The interactions can be grouped as dipole–dipole (d–d), dipole–quadrupole (d–q) and quadrupole—quadrupole (q–q). Based upon the relation between emission intensity and concentration of the activator ions, a graph can be plotted using the following relation [16]:

$$\log\frac{I}{c}=\log f-\frac{s\log(c)}{d} \tag{6}$$

Here I, c, f and d are emission intensity, concentration of activator ions, a constant which is independent of activator ions concentration and dimension of the compound (=3) respectively. Interaction between the closest neighbouring ions is predicted by s, where s = 6, 8 and 10 for d-d, d-q and q-q interaction respectively [16]. The plot between log (I/c) and log (c) is presented in Fig. 9 which shows the value of slope (s/d) to be 1.86, hence the calculated data for s = 1.86 *3 = 5.6. Here the obtained data for s is approximately equal to 6 indicating towards dipole–dipole interaction among the activator ions. Hence in the NBS glasses, dipole–dipole interaction is responsible for non-radiative energy transfer among Nd^{3+} ions. Moreover, this non radiative energy transfer leads to the concentration quenching effect in the emission spectra.

The radiative parameters like transition probability (A_R), total transition probability (A_T), radiative lifetime (τ_R) and branching ratio (β_R) corresponding to $^4F_{3/2}\rightarrow{}^4I_{9/2}$, $^4I_{11/2}$ and $^4I_{13/2}$ transitions for Nd^{3+} ions doped NBS glasses with the help of relations provided in literature were also calculated [44] and are represented in Table 3.

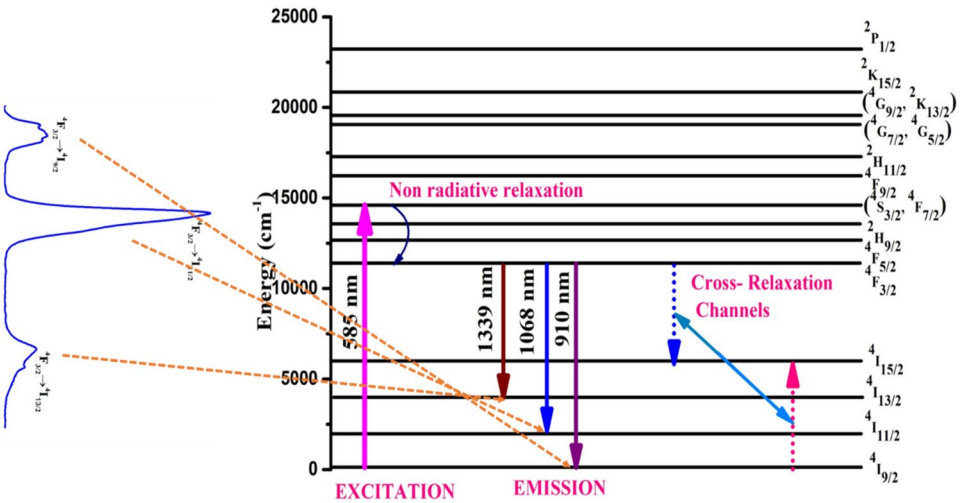


Fig. 8 Partial energy level diagram showing excitation, emission and cross-relaxation mechanism for Nd^{3+} ions in NBS glasses

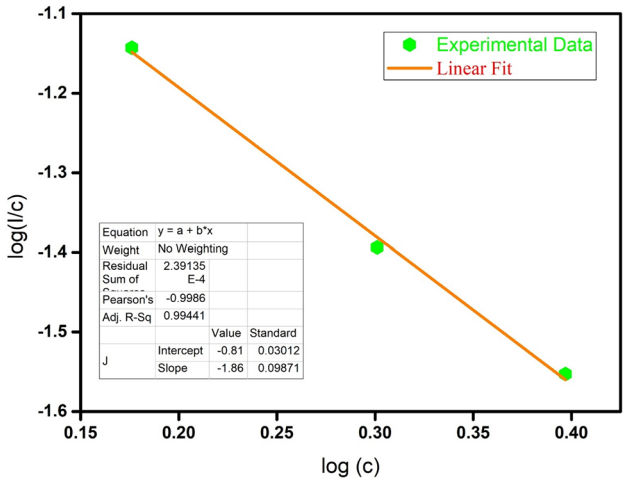


Fig. 9 Plot of log(I/c) v/s log(c) of Nd^{3+} ions in NBS glasses

From the data represented in Table 3, we can observe that the values of branching ratio (β_R) for NBS10 glass are maximum among all prepared Nd^{3+} doped NBS glasses corresponding to $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition. The values for total radiative transition probability increase upto 1 mol% of Nd^{3+} ions and then decrease upon further increment of the RE ions concentration. The total radiative transition probability for $^4F_{3/2} \rightarrow {}^4I_{11/2}$ is highest which is the most intense peaks. The fluorescence branching ratio is an important parameter in identifying the intensity for laser transition ascribed to electric dipole transitions of various excited and lower energy levels. If the value of β_R is ≥ 0.5 then it is very well suited for the lasing job. In the present study the value of β_R is above 0.5 for $^4F_{3/2} \rightarrow {}^4I_{11/2}$ energy level for all Nd^{3+} ions doped NBS making it as possible candidate for the lasing field applications. For further analyzing the utility as an appropriate lasing host, stimulated emission cross-section (σ_{se}) of the glass should be of high value. The value of stimulated emission cross-section suggests the behavior of energy extraction from optical material and is evaluated with the help of relation [1, 11]:

$$\sigma_{se} = \left[\frac{\lambda_p^4}{8\pi c n^2 \Delta\lambda_p}\right] A_R \tag{7}$$

where, the λ_p and $\Delta\lambda_p$ represent the wavelength corresponding to most intense band and linewidth (FWHM). Table 4 shows all the calculated values for stimulated emission cross-sections (σ_{se}), emission peak wavelength (λ_P), line widths ($\Delta\lambda_P$), gain band width ($\sigma_{se} \times \Delta\lambda_P$) and optical gain parameter ($\sigma_{se} \times \tau_R$) parameters for the emission transitions for Nd^{3+} ions in NBS glasses. The value of stimulated emission cross-section for NBS10 glass is maximum, making it superior host for lasing and optical fiber amplifier applications.

3.5 Decay spectral examination

Figure 10 shows the fluorescence decay curves for the Nd^{3+} ions in NBS glasses at 1068 nm emission wavelength for $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition under 585 nm excitation. The measured decay curves for the Nd^{3+} doped NBS glasses show single exponential fit behavior for all concentrations. The experimental decay time (τ_{exp}) for the Nd^{3+} doped NBS

Table 3 Transition probability (A_R) (s^{-1}), measured and experimental branching ratio (β_{exp} *and* β_R), total transition probability (A_T) (s^{-1}) and radiative lifetime (τ_R) (μs) for the observed emission transitions of Nd^{3+} ions in NBS glasses

Sample name	Transition	A_R	A_T	β_R	β_{exp}	τ_R
NBS01	$^4F_{3/2} \rightarrow {}^4I_{9/2}$	3468.18	13,736.8	0.2525	0.1653	72
	$^4F_{3/2} \rightarrow {}^4I_{11/2}$	8203.46		0.5972	0.6704	
	$^4F_{3/2} \rightarrow {}^4I_{13/2}$	1967.83		0.1433	0.1609	
NBS05	$^4F_{3/2} \rightarrow {}^4I_{9/2}$	3862.95	15,042.4	0.2568	0.1640	66
	$^4F_{3/2} \rightarrow {}^4I_{11/2}$	8938.47		0.5942	0.6680	
	$^4F_{3/2} \rightarrow {}^4I_{13/2}$	2135.44		0.142	0.1623	
NBS10	$^4F_{3/2} \rightarrow {}^4I_{9/2}$	4297.01	18,627.7	0.2307	0.1621	53
	$^4F_{3/2} \rightarrow {}^4I_{11/2}$	11,403.2		0.6122	0.6736	
	$^4F_{3/2} \rightarrow {}^4I_{13/2}$	2789.59		0.1498	0.1636	
NBS15	$^4F_{3/2} \rightarrow {}^4I_{9/2}$	3378.27	13,670.5	0.2471	0.1615	73
	$^4F_{3/2} \rightarrow {}^4I_{11/2}$	8214.18		0.6009	0.6743	
	$^4F_{3/2} \rightarrow {}^4I_{13/2}$	1980.17		0.1448	0.1615	
NBS20	$^4F_{3/2} \rightarrow {}^4I_{9/2}$	2967.57	12,024.9	0.2468	0.1617	83
	$^4F_{3/2} \rightarrow {}^4I_{11/2}$	7228.14		0.6011	0.6710	
	$^4F_{3/2} \rightarrow {}^4I_{13/2}$	1743		0.1449	0.1602	
NBS25	$^4F_{3/2} \rightarrow {}^4I_{9/2}$	2632.92	10,308.6	0.2554	0.1654	97
	$^4F_{3/2} \rightarrow {}^4I_{11/2}$	6135.38		0.5952	0.6710	
	$^4F_{3/2} \rightarrow {}^4I_{13/2}$	1467.7		0.1424	0.1559	

Table 4 Emission peak wavelength (λ_P) (nm), linewidths ($\Delta\lambda_P$) (nm), stimulated emission cross-sections (σ_{se}) (cm^2), gain band width ($\sigma_{se} \times \Delta\lambda_P$) ($cm^3$) and optical gain parameter ($\sigma_{se} \times \tau_R$) ($cm^2$ s) parameters for the emission transitions for Nd^{3+} ions in NBS glasses

Spectral parameters	NBS01	NBS05	NBS10	NBS15	NBS20	NBS25
$\lambda_P = 910$ nm ($^4F_{3/2} \rightarrow {}^4I_{9/2}$)						
Δλp	49.217	49.133	49.106	49.019	49.0	49.098
$\sigma_{se}(\times 10^{-22})$	109.491	117.266	126.388	97.204	82.153	71.086
$\sigma_{se} \times \Delta\lambda_P(\times 10^{-28})$	538.884	576.169	620.645	476.487	402.555	349.027
$\sigma_{se} \times \tau_R(\times 10^{-25})$	7.883	7.739	6.698	7.095	6.818	6.895
$\lambda_P = 1068$ nm ($^4F_{3/2} \rightarrow {}^4I_{11/2}$)						
Δλp	37.825	37.803	37.905	37.816	37.787	37.760
$\sigma_{se}(\times 10^{-22})$	639.333	669.081	824.368	581.245	492.292	408.644
$\sigma_{se} \times \Delta\lambda_P(\times 10^{-28})$	2418.302	2529.376	3124.806	2198.067	1860.249	1543.061
$\sigma_{se} \times \tau_R(\times 10^{-25})$	46.032	44.159	43.691	42.430	40.860	39.638
$\lambda_P = 1339$ nm ($^4F_{3/2} \rightarrow {}^4I_{13/2}$)						
Δλp	57.410	57.465	57.643	57.396	57.371	57.088
$\sigma_{se}(\times 10^{-22})$	249.660	259.815	327.658	228.103	193.188	159.760
$\sigma_{se} \times \Delta\lambda_P(\times 10^{-28})$	1433.304	1493.052	1888.749	1309.232	1108.356	912.044
$\sigma_{se} \times \tau_R(\times 10^{-25})$	17.975	17.147	17.365	16.651	16.034	15.496

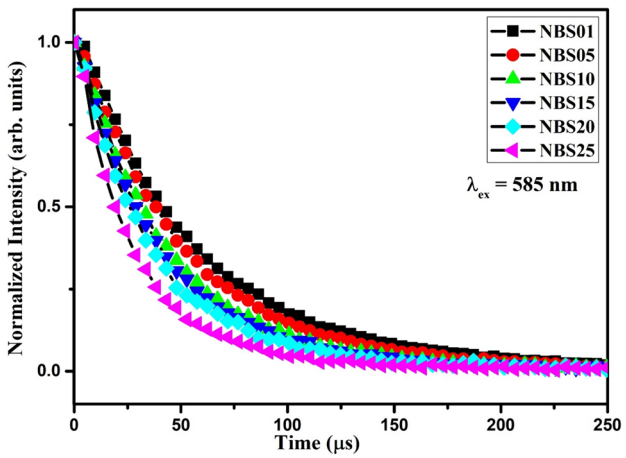


Fig. 10 Decay profiles of Nd^{3+} ions in NBS glasses for $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition under 585 nm excitation

Table 5 Experimental lifetime (τ_{exp}) (μs), radiative lifetime (τ_R) (μs), quantum efficiency (η), and non-radiative decay rates (W_{NR}) (s^{-1}) for Nd^{3+} ions in NBS glasses

Name of sample	τ_{exp}	τ_R	η (%)	W_{NR}
NBS01	56	72	77.7	3968
NBS05	51	66	77.2	4456
NBS10	44	53	83	3859
NBS15	41	73	56.1	10,691
NBS20	36	83	43.3	15,729
NBS25	28	97	28.8	25,405

glasses are calculated initially by applying e-folding time of the intensity for decay curves and with the help of equation provided in literature which are shown in Table 5 [8, 11, 22]. The values of τ_{exp} are found to be decreasing with increment in concentration of Nd^{3+} ions for NBS glasses which may be attributed to concentration quenching effect among the Nd^{3+} ions. The τ_{exp} values are also lower than the corresponding τ_R obtained via J–O theory, which may be ascribed due to specific the non-radiative decay mechanism (W_{NR}). Quantum efficiency (η), which is the ratio between emitted and absorbed light intensity, is an important parameter that describes the utility of the material for lasing applications. Higher the value of quantum efficiency, better will be the lasing properties of studied glass matrix. Following equations are used to determine the values of η and W_{NR} given in literature and are shown in Table 5 [8, 11]:

$$\eta = \frac{\tau_{exp}}{\tau_R} x100\% \tag{8}$$

$$W_{NR} = \left(\frac{1}{\tau_{exp}} - \frac{1}{\tau_R}\right) \tag{9}$$

From the obtained values we can see that the highest quantum efficiency and lowest non radiative decay rate is for 1 mol% of Nd^{3+} ions doped NBS glasses, which further ascertains its suitability for NIR laser emission at 1068 nm [45, 46].

The values of important parameters, such as stimulated emission cross-section, gain band width, optical gain, experiment decay time, quantum efficiency and non-radiative decay rate, of NBS glass are compared to other Nd^{3+}-doped glasses and are shown in Table 6, along with the other reported glasses doped with Nd^{3+} ions. The measured values for optimized NBS10 glasses are maximum from all the other prepared glass as well as other

Table 6 Stimulated emission cross-sections ($\sigma_{se}\times10^{-22}$) ($cm^2$), gain band width ($\sigma_{se}\times\Delta\lambda_P$) ($\times10^{-28}$) ($cm^3$), optical gain parameter ($\sigma_{se}\times\tau_R$) ($\times10^{-25}$) (cm^2 s), experimental lifetime (τ_{exp}) (μs), radiative lifetime (τ_R) (μs), quantum efficiency (η), and non-radiative decay rates (W_{NR}) (s^{-1}) for ($^4F_{3/2}\rightarrow{}^4I_{11/2}$) transition of Nd^{3+} ions in NBS glasses along with other reported values

Name of sample	$\sigma_{se}(\times10^{-22})$	$\sigma_{se}\times\Delta\lambda_P(\times10^{-28})$	$\sigma_{se}\times\tau_R(\times10^{-25})$	τ_{exp}	η (%)	W_{NR}
NBS01	639.333	2418.302	46.032	56	77.7	3968
NBS05	669.081	2529.376	44.159	51	77.2	4456
NBS10	824.368	3124.806	43.691	44	83	3859
NBS15	581.245	2198.067	42.43	41	56.1	10,691
NBS20	492.292	1860.249	40.86	36	43.3	15,729
NBS25	408.644	1543.061	39.638	28	28.8	25,405
LiPbAlBNd [47]	300	1330	4.41	113	77.1	2046
TTNW0 [48]	367	1056.9	-	133	66	-
LTTNd10 [49]	4750	1980	394	68	82	2657

reported values, which marks towards more suitability in various lasing applications.

4 Conclusions

Nd^{3+}doped NBS glasses were prepared via a melt quench technique. The X-ray diffraction pattern shows the amorphous behavior of the undoped glass. The Thermo-gravimetric analysis shows stability of the glass. A thorough spectroscopic analysis of the samples was done via characterization techniques like UV–VIS-NIR absorption, photoluminescence excitation and emission and time decay measurements. Judd–Ofelt analysis was conducted and three intensity parameters Ω_2, Ω_4 and Ω_6 were obtained for all the prepared glass samples from the absorption data. Various important radiative parameters were calculated to help assess the suitability of the glasses in lasing applications. Emission of the Nd^{3+} doped NBS glasses was found in NIR range with three prominent transitions corresponding to $^4F_{3/2}\rightarrow{}^4I_{9/2}$, $^4I_{11/2}$ and $^4I_{13/2}$, wherein $^4F_{3/2}\rightarrow{}^4I_{11/2}$ transition centered at 1068 nm had the maximum intensity. This transition possesses the maximum value of spectroscopic quality factor, branching ratio, stimulated emission cross section and quantum efficiency for the optimized glass (NBS10). Interaction among the activator ions was predicted to be of dipole–dipole nature by utilizing the Dexter theory. The decay curves for NBS glasses show single exponential for all concentrations. The aforementioned results are encouraging enough to ascertain that the 1 mol% of Nd^{3+} ions doped NBS glass is ideal for laser and fiber amplifier applications.

Acknowledgements One of the authors Dr. Yasha Tayal is grateful to the management of ABES Engineering College Ghaziabad Uttar Pradesh for providing a conducive work and research environment. She also acknowledges Prof. A.S. Rao for his impeccable research guidance that helped in achieving the desired results in this work.

Author contribution Y. Tayal: Conceptualization, Formal analysis, writing & editing, Investigation. R.A. Talewar: Formal analysis, review. Sk. Mahamuda: Formal analysis. K. Maheshwari: Formal analysis. S. Kumari: Formal analysis. R. Pilania: Formal analysis. M. Kumar: Formal analysis. A. Prasad: Formal analysis, review. A.S. Rao: Supervision, Writing-review & editing.

Data availability The data that support the findings of this study are available on request from the corresponding author Dr. Yasha Tayal. The data are not publicly available due to restrictions, e.g. their containing information that could compromise the privacy of research participants.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. F. Zaman, G. Rooh, N. Srisittipokakun, S. Ruengsri, H.J. Kim, J. Kaewkhao, J. Non-Cryst, Solids **452**, 307 (2016)
2. C.M. da Silva Jr, L.A. Bueno, A.S. Gouveia-Neto, J. Non- Cryst. Solids **410**, 151 (2015)
3. M.D. Dramićanin, Methods Appl. Fluoresc. **4**, 042001 (2016)
4. C.X. Liu, X.L. Shen, H.T. Guo, W.N. Li, W. Wei, Optik **131**, 132 (2017)
5. N.K. Goel, G. Pickrell, R. Stolen, Opt. Fiber Technol. **20**, 325 (2014)
6. F. Elan, E.L. Falcão-Filho, M.E. Camilo, J.A.M. Garcia, L.R.P. Kassab, C. de Araújo, Opt. Mater. **60**, 313 (2016)
7. S. Mohan, K.S. Thind, Opt. Mater. **57**, 134 (2016)
8. C.R. Kesavulu, H.J. Kim, S.W. Lee, J. Kaewkhao, N. Wantana, E. Kaewnuam, S. Kothan, S. Kaewjaeng, J. Alloys Compd. **695**, 590 (2017)
9. L.R.P. Kassab, D.M. Silva, J.A.M. Garcia, D.S. da Silva, C.B. de Araújo, Opt. Mater. **60**, 25 (2016)
10. H.K. Dan, D.C. Zhou, Z.W. Yang, Z.G. Song, X. Yu, J.B. Qiu, J. Non-Cryst, Solids **414**, 21 (2015)
11. T.F. Xue, L.Y. Zhang, J.J. Hu, M.S. Liao, L.L. Hu, Opt. Mater. **47**, 24 (2015)

12. A. Pandey, S. Som, V. Kumar, V. Kumar, K. Kumar, V. KumarRai, H.C. Swart, Sensors Actuators B **202**, 1305 (2014)
13. D.D. Ramteke, R.E. Kroon, H.C. Swart, J. Non-Cryst, Solids **457**, 157 (2017)
14. L.Y. Zhang, B.H. Yang, L.L. Hu, J. Quant. Spectrosc. Radiat. Transf. **147**, 47 (2014)
15. S.M. Lima, J.A. Sampaio, T. Catunda, A.S.S. de Camargo, L.A.O. Nunes, M.L. Baesso, D.W. Hewak, J. Non-Cryst. Solids **284**, 274 (2001)
16. Y. Tayal, A.S. Rao, Opt. Mat. **107**, 110070 (2020)
17. J.A. Savage, Mater. Sci. Rep. **2**, 99 (1987)
18. H. Yamasaki, K. Minato, D. Nezaki, T. Okamoto, A. Kawamoto, M. Takata, Solid State Ionics **172**, 349 (2004)
19. Y. Tayal, A.S. Rao, Opt. Mat. **117**, 111112 (2021)
20. Y.I. Alivov, D.C. Look, B.M. Ataev, M.V. Chukichev, V.V. Mamedov, V.I. Zineko, Y.A. Agafonov, A.N. Pustovit, Solid State Electron. **48**, 2343 (2004)
21. B.L. Zhu, C.S. Xie, D.W. Zeng, W.L. Song, A.H. Wang, Mater. Chem. Phys. **89**, 148 (2005)
22. Y. Tayal, A.S. Rao, S. Kaur, Solid State Sci. **125**, 106834 (2022)
23. X.F. Wang, Q. Liu, Y.Y. Bu, C.S. Liu, T. Liu, X.H. Yan, RSC Adv. **5**, 86219 (2015)
24. S. Rai, A.L. Fanai, J. Lumin. **170**, 325 (2016)
25. C. Tian, X. Chen, Y. Shuibao, Solid State Sci. **48**, 171 (2015)
26. B. Klimesz, G. Dominiak-Dzik, P. Solarz, M. Zelechower, W. RybaRomanowski, J. Alloys Compd. **403**, 76 (2005)
27. S. Mohan, K.S. Thind, G. Sharma, L. Gerward, Spectro chim. Acta Part A **70**, 1173 (2008)
28. Y. Tayal, R.A. Talewar, S.K. Mahamuda, A. Prasad, K. Maheshwari, M. Kumar, A.S. Rao, Opt. Mater. **142**, 114049 (2023)
29. K. Linganna, R. Narro-García, H. Desirena, E. De la Rosa, Ch. Basavapoornima, V. Venkatramu, C.K. Jayasankar, J. Alloys Compd. **684**, 322 (2016)
30. W.T. Carnall, P.R. Fields, K. Rajnak, J. Chem. Phys. **49**, 4424–4442 (1968)
31. G.S. Ofelt, J. Chem. Phys. **37**, 511 (1962)
32. P. Babu, C.K. Jayasankar, Opt. Mater. **15**, 65 (2000)
33. R. Praveena, V. Venkatramu, P. Babu, C.K. Jayasankar, Phys. B **403**, 3527 (2008)
34. S. Tanabe, T. Ohyagi, N. Soga, T. Hanada, Phys. Rev. B **46**, 3305 (1992)
35. T. Suzuki, H. Nasu, M. Hughes, S. Mizuno, K. Hasegawa, H. Ito, Y. Ohishi, J. Non Cryst. Solids **356**, 2344 (2010)
36. M.B. Saisudha, J. Ramakrishna, Phys. Rev. B **53**, 6186 (1996)
37. J.A. Pardo, J.I. Pena, R.I. Merino, R. Cases, A. Larrea, V.M. Orera, J. Non Cryst. Solids **298**, 23 (2002)
38. H. Takebe, K. Morinaga, T. Izumitami, J. Non-Cryst, Solids **178**, 58 (1994)
39. W.T. Carnall, H. Crosswhite, H.M. Crosswhite, Technical report, Argonne National. Laboratory (1997). https://doi.org/10.2172/6417825
40. D.D. Ramteke, A. Balakrishna, V. Kumar, H.C. Swart, Opt. Mater. **64**, 171 (2017)
41. C.R. Kesavulu, H.J. Kim, S.W. Lee, J. Kaewkhao, N. Wantana, S. Kothan, S. Kaewjaeng, J. Alloys Compd. **683**, 590 (2016)
42. J. Azevedo, J. Coelho, G. Hungerford, N.S. Hussain, Phys. B: Condens. Matter. **405**, 4696 (2010)
43. M. Ajroud, M. Haouari, H.B. Ouada, H. Maaref, A. Brenier, C. Garapon, J. Phys. Condens. Matter **12**, 3181 (2000)
44. R. Cases, M.A. Chamarro, R. Alcala, V.D. Rodriguez, J. Lumin. **48**, 509 (1991)
45. Q.L. Chen, H. Wang, Q.P. Chen, J. Non-Cryst, Solids **391**, 43 (2014)
46. D. Ramachari, L.R. Moorthy, C.K. Jayasankar, Opt. Mater. **36**, 823 (2014)
47. N. Deopa, A.S. Rao, M. Gupta, G.V. Prakash, Opt. Mater. **75**, 127–134 (2018)
48. H. Fares, I. Jlassi, S. Hraiech, H. Elhouichet, M. Ferid, J. Quant. Spectrosc. RA. **147**, 224–232 (2014)
49. M. Venkateswarlu, S.K. Mahamuda, K. Swapna, M.V.V.K.S. Prasad, A.S. Rao, A.M. Babu, S. Shakya, G.V. Prakash, Opt. Mater. **39**, 8–15 (2015)

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

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

Check for updates

Received: 28 February 2024 | Revised: 1 July 2024 | Accepted: 26 July 2024

DOI: 10.1002/jnm.3279

RESEARCH ARTICLE

WILEY

Enhancement of multiplication factor of capacitor using single current follower differential input transconductance amplifier

Ajishek Raj[1] | **Meghana Shrivastava**[2] | **D. R. Bhaskar**[2] | **Pragati Kumar**[2]

[1]Department of Electrical Engineering, National Institute of Technology, Patna, India

[2]Department of Electrical Engineering, Delhi Technological University, Delhi, India

Correspondence
Ajishek Raj, Department of Electrical Engineering, National Institute of Technology, Patna, Bihar 800005, India.
Email: ajishek.ee@nitp.ac.in

Meghana Shrivastava, Department of Electrical Engineering, Delhi Technological University, Delhi, India.
Email: meghanashrivastava_2k21phdee15@dtu.ac.in

Abstract

This paper presents a new circuit approach to realize a capacitance multiplier circuit with positive and negative multiplication factors. Based on this approach, two new implementations of positive and negative grounded capacitance multiplier (GCM) circuits are proposed, which utilize only one current follower differential input transconductance amplifier (CFDITA), in conjunction with only one resistor and one virtually grounded capacitor. The presented GCM circuits can enhance a low capacitance value to a very high value (used in low frequency applications), up to 9202 times its original value. An important aspect of the proposed circuits involves designing a lossy parallel inductor circuit by interchanging the passive elements (RC:CR transformation) with each other. The obtained value of the capacitance and inductance can be controlled independently and electronically through the transconductance of CFDITA. The practical usability of the suggested circuits as first and second order filters is discussed. The functionality of the proposed GCM circuits is validated using CMOS CFDITA implemented with 180 nm TSMC technology parameters. Experimental verification of the proposed circuits and application examples is reinforced through the utilization of CFDITA implemented with readily available ICs AD844 and LM13700. These outcomes emphasize the dependability of the suggested circuits.

KEYWORDS

capacitance multiplier, CFDITA, electronic control, filter, SPICE

1 | INTRODUCTION

Capacitance multipliers (CMs) are circuits used to increase the value of a capacitor, which find various applications in analog systems where it may be used to design low frequency filters, oscillators, and so forth. These circuits are used to alleviate the issue of needing a large area when implementing high valued capacitors in integrated circuit form. Consequently, researchers have suggested and reported a large number of analog CM circuits of lossless and lossy nature employing various active building blocks (ABB), such as op-amps, current feedback op-amps, current conveyors and its variants, and many more. Different types of transconductance amplifiers have also been used in previous research works to design CMs, including operational transconductance amplifiers, voltage differencing transconductance amplifiers, and current follower transconductance amplifiers including CFDITA.

 © 2024 John Wiley & Sons Ltd.

Lossless grounded capacitance multipliers have been previously reported in various research papers, whose aim is to increase the value of a capacitor without introducing any loss or degradation in the circuit's performance. By providing, a brief overview of these previously reported single ABB based lossless GCMs, the proposed work can be better understood and contextualized.

A translinear loop consisting of four transistors has been utilized in Ref. [1] to implement a CM circuit, incorporating impedance and four current sources. The circuit's multiplication factor can be adjusted by varying two currents, and the maximum achievable multiplication factor is 300. Numerous positive and negative CM circuits employing current feedback operational amplifiers (CFOA) and their derivatives have been documented in the literature.[2–8] In reference [2], a CM circuit was reported using a single CFOA in conjunction with two resistors and one capacitor to achieve a positive capacitance multiplication factor (MF). References [3–5] describe CM circuits that utilize two CFOAs along with one capacitor and a minimum of two resistors to attain a positive MF. References [6,7] present CM circuits that yield negative MF using one CFOA, two resistors, and one grounded capacitor. In reference [8], an inverting CFOA is employed to realize a positive CM circuit along with two resistors and one capacitor. Reference [9] discusses two distinct CM circuits. The first circuit employs a single current gain CCII− and one capacitor, while the other circuit utilizes one CCII+, one CCII−, two resistors, and one capacitor to realize a positive CM circuit. Reference [10] reports a CM circuit based on a DXCCII, two MOS resistors, and one capacitor, with a maximum achievable capacitance MF of 50. A grounded capacitance multiplier circuit, based on a single TFTFN and one capacitor, is discussed in Ref. [11] Additionally, the article discusses the conversion of a grounded-type CM circuit into its floating counterparts. In reference [12], both lossy and lossless common-mode circuits have been documented, employing operational transconductance amplifiers (OTA). These circuits allow for a maximum achievable scale factor of 10. Another noteworthy circuit, utilizing OTA and CCII, is presented in Ref. [13] It facilitates both positive and negative impedance multiplier circuits with a maximum MF of up to 400. Various impedance multipliers/capacitance multiplier circuits have been reported in Ref. [14–19]. References [14,15] explore capacitance multiplier circuits based on single voltage differencing transconductance amplifier (VDTA) while[16,17] require two VDTAs to realize CM circuits. In Ref. [14], two circuits have been implemented and both of these circuits use a resistor and one capacitor to realize the CM circuit. The maximum achievable MF for these circuits is up to 20. In contrast, the circuit described in reference [15] does not necessitate a resistor for such implementation, and the maximum MF achievable is up to 115. References [16,17] employ only one capacitor alongside active devices to achieve the function of impedance multiplication/CM. The maximum MF of capacitance obtained in Ref. [16] is 315, while in Ref. [17], this value is 25. Reference [18] introduces a VDTA-based lossy capacitance and inductance simulator employing a single VDTA and one capacitor, without any matching constraints on passive components. In reference [19], an electronically controllable CM circuit is discussed, employing a third-generation current conveyor transconductance amplifier (CCIIITA) along with one capacitor, capable of achieving an MF value of 50. Reference [20] explores the use of a differential buffer and a current conveyor-based CM circuit, discussing both lossy and lossless structures. Reference [21] presents two lossless positive/negative grounded capacitance multipliers, utilizing a current follower transconductance amplifier (CFTA), one resistor, and one capacitor. Using these circuits, an MF value of 10 can be achieved. Reference [22] discusses the existence of two CFTAs, two resistors, and one capacitor, employed in two distinct grounded positive/negative capacitance multiplier circuit configurations. These circuits were verified to attain a maximum MF of 10 in both modes. In Ref. [23], a capacitance multiplier utilizing a differential input buffered transconductance amplifier (DBTA) has been introduced. This circuit incorporates one DBTA, one grounded resistor, and one grounded capacitor. Notably, the MF of this circuit can be electronically adjusted via the transconductance of the DBTA. Furthermore,[24] presents a resistorless, electronically tunable grounded capacitance multiplier circuit employing one voltage differencing inverting buffered amplifier (VDIBA), one capacitor, and one voltage-controlled resistor (VCR). The achieved MF for this circuit is approximately 1.5. Reference [25] discusses a grounded positive capacitance multiplier circuit utilizing a voltage differencing buffered amplifier (VDBA), in conjunction with one resistor and one capacitor. This circuit is capable of enhancing capacitance with a MF of 100. In Ref. [26], two positive capacitance multiplier circuits based on two/three operational amplifiers have been explored. The required passive components for realization are either six or four resistors and one capacitor, and these circuits are capable of providing an MF of 100. Additionally,[27] employs a combination of a translinear stage and OTA stage to design a capacitance multiplier circuit, suitable for low-frequency applications. In Ref. [28], two-capacitance multiplier circuits utilizing a second-generation voltage

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

conveyor (VCII−), two grounded resistors, and one virtually grounded capacitor have been reported. The first circuit yields a positive MF, while the second provides a negative MF. Furthermore,[29] presents an electronically tunable grounded capacitor multiplier circuit with a positive multiplication factor, using a second-generation voltage conveyor (E-VCII+) and a floating capacitor. This same circuit can function as a negative capacitance multiplier by replacing E-VCII+ with E-VCII−. The multiplication factor for both the circuits operates within the range of 1–100. In Ref. [30], a capacitance multiplier circuit has been detailed, utilizing a single operational transresistance amplifier, one unity gain voltage buffer, three resistors, and one capacitor, for both positive and negative applications. Notably, electronic control is facilitated by substituting the resistors in the circuit with their MOS counterparts. Additionally,[31] presents multiple circuits, one of which is designed as a negative capacitance multiplier circuit employing a single current feedback operational amplifier (CFOA), three resistors, and one capacitor. Reference [32] discusses a capacitance multiplier circuit based on a single current differencing transconductance amplifier (CDTA), incorporating one resistor and one capacitor. Moreover, the paper introduces an approach to convert grounded capacitance multiplier circuit into its floating counterpart. Furthermore,[33] highlights two distinct configurations of parallel lossy inductor simulators, each employing a single current differencing buffered amplifier (CDBA), one buffer, three resistors, and one capacitor. These configurations are designed for various applications. To provide a comprehensive overview, Table 1 tabulates the key features of both the previously reported capacitance multiplier circuits and the newly proposed circuit.

From comparison Table 1, it is observed that different configurations of CM multiplier circuits use a single ABB (see the references cited therein:[2,6–11,14,15,19,21,23–26,29,31,32,35,36,38]), providing positive or negative multiplication factors, except in Ref. [14], which provides both positive and negative capacitance multiplier circuits. When comparing the proposed circuit with,[14] it is found that the proposed circuit has a higher multiplication factor using the same number of ABBs and passive elements. In addition to a high multiplication factor, the proposed circuit has a large frequency span of capacitance, no additional requirement for matching constraints, and electronic tunability features. Furthermore, using RC:CR transformation, the proposed CM circuit can be realized as parallel lossy RL simulators.

Recently, the current follower differential input transconductance amplifier has attracted considerable attention in the implementation of a wide range of analog circuits, including immittance simulator circuits, analog filters, instrumentation amplifiers, non-sinusoidal wave generators, bi-stable multivibrator, and more.[35–37,40–44] The CFDITA, characterized as a four-terminal active block, demonstrates current-following behavior from the Z terminal to the F terminal. Subsequently, a differential transconductance amplifier generates an output current proportionate to the voltage differential between the Z and V terminals.[45]

Recently, interesting circuits of grounded capacitance multiplier circuits using CFDITAs have been reported in the literature.[35–37] In both,[35,36] the capacitance multiplier circuits have been implemented using a single CFDITA. The former circuit requires two resistors and one capacitor, while in the latter circuit, three resistors and one capacitor have been employed, resulting in capacitance multiplication factors (MF) of 30 and 2.52, respectively. Conversely, the circuit detailed in Ref. [37] utilized three CFDITAs in conjunction with two capacitors.

From the aforementioned descriptions of the lossless CM circuits, it can be observed that there is no reported circuit using CFDITA as of yet that offers electronically controllable capacitance utilizing a single CFDITA along with canonic number of one resistor and capacitor with a high MF value of up to 9202.

In this communication, new capacitance multiplier and parallel RL simulator circuits capable of both positive and negative multiplication factors are introduced. These circuits utilize a single CFDITA as the ABB in combination with one resistor and one capacitor. The proposed CM circuit can significantly boost the equivalent capacitance, up to 9202 times its original value, by adjusting the resistor. Furthermore, it can be electronically tuned, achieving up to 1800 times enhancement, by varying the transconductance of the CFDITA. To validate the proposed circuits, a CMOS CFDITA implemented with 0.18 μm technology parameters is employed in Cadence Virtuoso tools and conducted experimental tests using CFDITAs realized with commercially available ICs.

The manuscript is organized as follows: In Section 2, the new designs of capacitance multiplier circuits and parallel RL simulators are presented. Section 3 covers the non-ideal analysis, while Sections 4 and 5 provide simulation and experimental results, respectively. Lastly, in Section 6, the conclusions of this paper are discussed.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Comparative table of previously reported CM circuits with proposed circuit.

Ref.	Type of simulated capacitor	ABB used	R	C	Technology used	Multiplication factor	Matching condition required or not?	Frequency range	Electronically tunable?	THD (%)	FoM (kHz/μW)	Number of discrete IC used required
[34]	Positive	CMOS	0	1	0.18 μm	10 to 300	No	100 Hz to 900 kHz	No	NM	NM	NA
[2]	Positive	1 CFOA	2	1	0.13 μm	51	No	100 KHz–6 MHz	No	NM	NM	1 AD844
[3]	Positive	2 CFOAs	2	1	AD844	10	No	500 Hz–10 MHz	No	NM	NM	2 AD844
[4]	Positive	2 CFOAs	2	1	0.13 μm	100	No	10 Hz–10 MHz	No	NM	NM	2 AD844
[5]	Positive	2 CFOAs	3	1	-	200	No	Upto 190 kHz	No	NM	NM	2 AD844
[6] Figure 2&3	Negative	2 CFOAs	2	1	0.35 μm	NM	No	–	No	NM	NM	2 AD844
[6] Figure 4&5	Negative	1 CFOA	2	1	0.35 μm	0.5	No	1 kHz–5 MHz	No	NM	NM	1 AD844
[7] Fig. a	Negative	1 CFOA	2	1	AD844	2001	No	1 kHz–1 MHz	No	NM	NM	1 AD844
[7] Fig. b	Negative	1 CFOA	2	1	AD844	201	No	1 kHz–1 MHz	No	NM	NM	1 AD844
[7] Fig. c	Negative	1 CFOA	2	1	AD844	400	No	1 kHz–1 MHz	No	NM	NM	1 AD844
[8]	Positive	1 ICFOA	2	1	0.18 μm	318	No	1 Hz to 30 kHz	No	0.06	NM	2 AD844
[9] Figure 1	Positive	1 CCII-	0	1	1.2 μm	NM	No	–	No	NM	NM	2 AD844
[9] Figure 2	Positive	1 CCII-, 1 CCII+	2	1	1.2 μm	1000	No	10 Hz–1 MHz	No	NM	NM	3 AD844
[10]	Positive	1DXCCII	0	1	0.35 μm	50	No	1 Hz–10 MHz	Yes	NM	NM	NM
[11] Figure 3	Positive	1 TFTFN	0	1	AD704N, BJT	NM	No	1 Hz–1 MHz	No	NM	NM	–
[12] Figure 3B	Positive	2 DO-OTAs	0	1	Macro model of OTA	10	Yes	–	Yes	NM	NM	4 LM13700
[13]	Positive	1 CFOA, 1 OTA	1	1	AD844, LM13700N	400	No	10 mHz–10 MHz	Yes	NM	NM	1 AD844, 1 LM13700N
[14] Figure 1A	Positive/ Negative	1 VDTA	1	1	0.18 μm	20	No	100 Hz to 100 MHz	Yes	NM	194.7	2 LM13700

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Ref.	Type of simulated capacitor	ABB used	R	C	Technology used	Multiplication factor	Matching condition required or not?	Frequency range	Electronically tunable?	THD (%)	FoM (kHz/μW)	Number of discrete IC used required
[15]	Positive	1 VDTA	0	1	0.18 μm	115	No	–	Yes	3.7	NM	2 LM13700
[16]	Positive	2 VDTAs	0	1	0.18 μm	315	No	–	Yes	NM	NM	4 LM13700
[17] Figure 2A–C	Negative	2 VDTAs	0	1	0.18 μm	25	No	3.235 kHz to 851.138 kHz	Yes	NM	NM	4 LM13700
[19]	Positive	1 CCIIITA	0	1	0.18 μm	50	No	10 Hz to 100 kHz	Yes	NM	NM	6 AD844
[20]	Positive	1 DVB, 1ECCII-	2	1	–	193	No	–	Yes	NM	NM	1 AD8130, 1 EL2082
[21] Figure 6	Positive	1 CFTA	1	1	0.18 μm	10	No	1 kHz–1 MHz	Yes	NM	NM	1 AD844, 1 LM13700
[21] Figure 7	Negative	1 CFTA	1	1	0.18 μm	10	No	1 kHz–1 MHz	Yes	NM	NM	1 AD844, 1 LM13700
[22] Figure 4&5	Positive	2 CFTAs	2	1	0.13 μm	19.7	No	10 kHz–100 MHz	Yes	NM	NM	3 AD844, 4 LM13700
[22] Figure 6 &7	Negative	2 CFTAs	2	1	0.13 μm	20.3	No	10 kHz–100 MHz	Yes	NM	NM	3 AD844, 4 LM13700
[23]	Positive	1 DBTA	1	1	AD844	10	No	–	Yes	NM	NM	3 AD844
[24]	Positive	1 VDIBA	0	1	0.25 μm	3.59	No	1 kHz to 5 MHz	Yes	NM	NM	NM
[25]	Positive	1 VDBA	1	1	BJT	100	No	100 Hz −100 kHz	Yes	NM	NM	NM
[26] Figure 1	Negative	1 Op-Amp	2	1	LF411	1000	No	100mHz–100KHz	No	NM	NM	1 uA741
[26] Figure 3	Positive	3 Op-Amps	6	1	LF411	NM	No	–	No	NM	NM	3 uA741
[26] Figure 4	Positive	2 Op-Amps	4	1	LF411	NM	No	–	No	NM	NM	2 uA741
[27]	Positive	2 OTAs	0	1	0.18 μm	1200	No	1 Hz to 10KHz	Yes	NM	NM	2 LM13700
[28] Figure 3	Positive	1VCII+, 1 VCII-	2	1	0.35 μm	50	No	Upto 10 MHz	No	1.2	NM	3 AD844
[28] Figure 4	Negative	2VCII+	2	1	0.35 μm	50	No	Upto 10 MHz	No	1.2	NM	2 AD844
[29]	Positive	1 E-VCII+	0	1	0.18 μm	100	No	80 Hz–40 kHz	Yes	NM	NM	NM

(Continues)

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Ref.	Type of simulated capacitor	ABB used	R	C	Technology used	Multiplication factor	Matching condition required or not?	Frequency range	Electronically tunable?	THD (%)	FoM (kHz/μW)	Number of discrete IC used required
[29]	Negative	1 E-VCII-	0	1	0.18 μm	NM	No	–	Yes	NM	NM	NM
[30]	Positive	1 OTRA, 1 VF	3	1	0.18 μm	101	Yes	100 Hz–1 MHz	Yes	NM	NM	2 AD844
[30]	Negative	1 OTRA, 1 VF	3	1	0.18 μm	99	Yes	100 Hz–1 MHz	Yes	NM	NM	2 AD844
[31]Figure 8	Negative	1 CFOA	3	1	0.13 μm	NM	Yes	–	No	1.4	NM	1 AD844
[32]Figure 3A	Positive	1 CDTA	1	1	0.7 μm	NM	No	–	Yes	NM	NM	2 AD844, 1 LM13700
[35]	Positive	1 CFDITA	2	1	0.18 μm	30	No	10 kHz to3 MHz	Yes	NM	NM	2 AD844, 1 LM13700
[36]	Positive	1 CFDITA	3	1	0.18 μm	2.52	No	10 kHz to10 MHz	Yes	NM	NM	2 AD844, 1 LM13700
[37] Figure 3B	Positive	3 CFDITAs	0	1	0.18 μm	NM	No	–	Yes	NM	NM	6 AD844, 2 LM13700
[38] Figure 3B	Positive	1 OTA	0	1	0.5 μm	NM	No	–	Yes	NM	2.1	1 LM13700
[39]	Positive	Class-AB current mirror	0	1	**0.18 μm**	11	No	750 Hz–200 kHz/ 45 Hz–5 kHz	Yes	NM	95.6/178.5	NM
This work	**Positive Negative**	**1 CFDITA**	**1**	**1**	**0.18 μm**	**9202**	**No**	**100 mHz–70 MHz**	**Yes**	**4**	**147**	**1 AD844, 1 LM13700**

Abbreviations: C, Capacitor; DXCCII, Dual-x second generation current conveyor; FoM, Figure of Merit; R, Resistor; TFTFN, Tunable-four terminal floating nullors; THD, Total Harmonic Distortion; VF, Voltage follower.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | NEW DESIGN OF GROUNDED LOSSLESS CAPACITANCE MULTIPLIER CIRCUIT

Symbolic representation of CFDITA is illustrated in Figure 1 and the terminal equations of the CFDITA can be given by:

$$\begin{bmatrix} I_V \\ I_Z \\ V_F \\ I_O \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -g_m & g_m & 0 & 0 \end{bmatrix} \begin{bmatrix} V_V \\ V_Z \\ I_F \\ V_O \end{bmatrix} \tag{1}$$

where g_m is the transconductance of CFDITA.

In the literature, various topologies have been reported for implementing positive capacitance multipliers.[7,28,32,46] Most of these circuits utilize voltage feedback mode and current feedback mode approaches, as illustrated in Ref. [46]. To design a negative multiplication factor for capacitance, a synthesis using two different approaches is discussed in Ref. [7]. In this paper, we present a circuit model depicted in Figure 2, which is employed for designing new CM circuits.

Circuit analysis of Figure 2 yields the input impedance expression as:

$$\frac{V_{in}}{I_{in}} = \frac{1}{sC(1+K_1K_2)} = \frac{1}{sC_{eq}} \tag{2}$$

where K_1 and K_2 are the combinations of resistors and/or transconductors.

Figure 3 illustrates the new implementations of grounded positive and negative lossless capacitance multiplier circuits using the approach illustrated in Figure 2 employing a single CFDITA, one resistor, and one virtually grounded capacitor.

The circuit analysis of Figure 3, assuming an ideal CFDITA yields the expressions of the impedance of both the circuits as follows:

$$\left.\frac{V_{in}}{I_{in}}\right|_{Fig.3(a)} = Z_{in}|_{(a)} = \frac{1}{sC(2+g_mR)} = \frac{1}{sC_{eq1}} \tag{3}$$

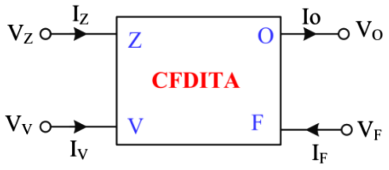


FIGURE 1 Symbolic notation of CFDITA.

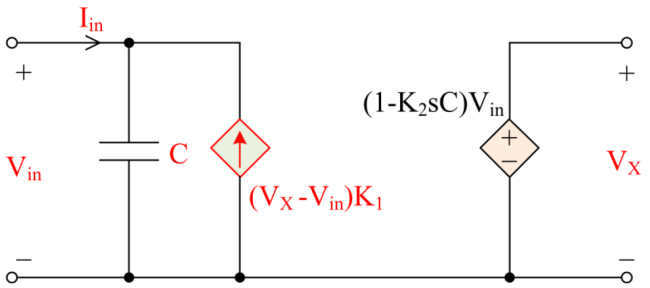


FIGURE 2 Circuit approach to implement new CM circuits.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

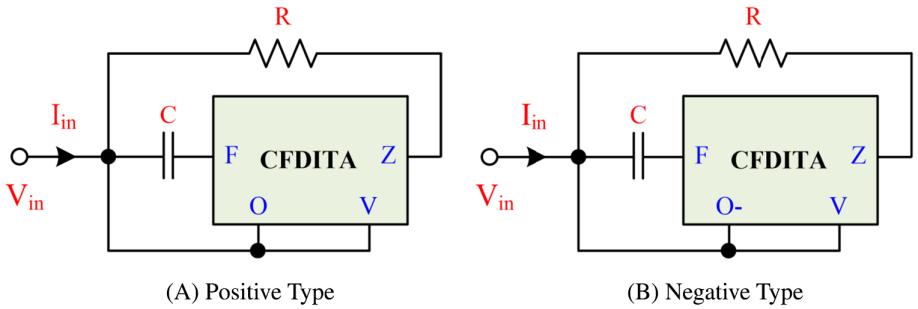


FIGURE 3 Proposed lossless grounded capacitance multiplier circuits.

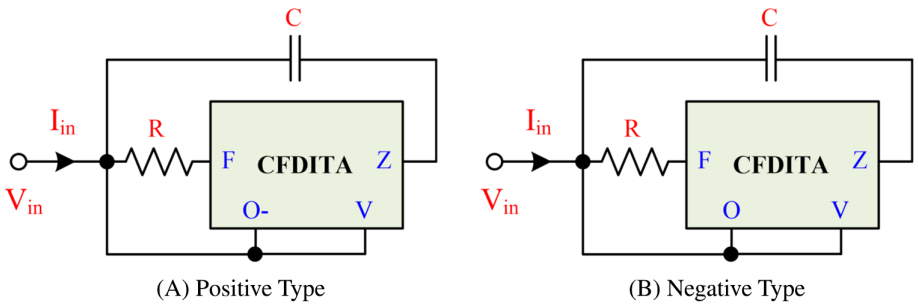


FIGURE 4 Proposed lossy parallel-type RL simulator circuits.

$$\left.\frac{V_{in}}{I_{in}}\right|_{Fig.3(b)} = Z_{in}|_{(b)} = \frac{1}{sC(2-g_mR)} = \frac{-1}{sC_{eq2}} \tag{4}$$

where $C_{eq1} = C(2+g_mR)$ and $C_{eq2} = C(2-g_mR)$.

From equation (3), we can conclude that the equivalent capacitance (C_{eq1}) of the circuit in Figure 3A yields a positive capacitance multiplication factor (MF) of ($2 + g_mR$). Similarly, equation (4) demonstrates that the MF for the circuit in Figure 3B can be either positive or negative. It can provide a positive multiplication factor of up to 2 and a negative MF if the value of g_mR greatly exceeds 2. In both the circuits, the MF can be independently and electronically adjusted using the transconductor of the CFDITA and by varying resistor R, within the constraints of the CFDITA.

A notable feature of the circuits in Figure 3 is its ability to function as a parallel RL simulator circuits when the passive components are interchanged using RC:CR transformation. The modified circuit's implementation is depicted in Figure 4.

Assuming ideal CFDITA, routine analysis of parallel lossy RL circuits of Figure 4 yield the following input admittance:

$$\left.\frac{I_{in}}{V_{in}}\right|_{Fig.4(a)} = Y_{in} = \frac{1}{Z_{in}} = \frac{2}{R} + \frac{g_m}{sCR} = \frac{1}{R_{eq}} + \frac{1}{sL_{eq}} \tag{5}$$

$$\left.\frac{I_{in}}{V_{in}}\right|_{Fig.4(b)} = Y_{in} = \frac{1}{Z_{in}} = \left(\frac{2}{R} - \frac{g_m}{sCR}\right) = \frac{1}{R_{eq}} - \frac{1}{sL_{eq}} \tag{6}$$

The expressions for admittance obtained in equations (5) and (6), it is worth noting that the inductance achieved in the proposed lossy inductance simulator circuits is independently tunable through the transconductance of the CFDITA. This tunability can be further controlled electronically by adjusting the bias current. Additionally, the resistor employed in both circuits is effectively grounded, given that the voltage at terminal F of the CFDITA is zero. This can be further replaced with its MOS counterparts, making it highly suitable for IC integration.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The non-ideal gains of the CFDITA are considered, and the terminal characteristics are modified as provided in equation (7) for the evaluation of the input impedance of the proposed circuits.

$$\begin{bmatrix} I_V \\ I_Z \\ V_F \\ I_O \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 \\ -\beta g_m & \beta g_m & 0 & 0 \end{bmatrix} \begin{bmatrix} V_V \\ V_Z \\ I_F \\ V_O \end{bmatrix} \tag{7}$$

where α and β are non-ideal current and transconductance gains.

The input impedance of the circuits shown in Figures 3 and 4, accounting for non-ideal gains, can be expressed as:

$$\left(\frac{V_{in}}{I_{in}}\right)'\Big|_{Fig.3(a)} = \frac{1}{sC(1+\alpha+\alpha\beta g_m R)} \tag{8}$$

$$\left(\frac{V_{in}}{I_{in}}\right)'\Big|_{Fig.3(b)} = \frac{1}{sC(1+\alpha-\alpha\beta g_m R)} \tag{9}$$

$$\left(\frac{I_{in}}{V_{in}}\right)'\Big|_{Fig.4(a)} = \frac{\alpha+1}{R} + \frac{g_m\alpha\beta}{sCR} \tag{10}$$

$$\left(\frac{I_{in}}{V_{in}}\right)'\Big|_{Fig.4(b)} = \frac{\alpha+1}{R} - \frac{g_m\alpha\beta}{sCR} \tag{11}$$

It is noteworthy from equations (8) to (11) that the non-ideal gain significantly impacts the impedance and admittance of both the GCM and parallel inductance simulator circuits, leading to deviations from unity, especially at higher frequencies. Consequently, the equivalent and inductance values experience notable alterations in the higher frequency range. We quantitatively assessed these changes, deriving approximate values for α and β as 0.96 and 0.98, respectively. Upon substituting these values into equations (8) to (11), we observed that the non-ideal expressions closely approximate the ideal values.

3 | SIMULATION RESULTS

To validate the performances of the proposed capacitance multiplier and lossy parallel RL simulator circuits, CMOS CFDITA shown in Figure 5 has been used. Aspect ratios of the MOSFETs used in Figure 5 have been provided in Table 2.

D.C. power supply voltages for the CFDITA were chosen as $V_{DD} = -V_{SS} = 1.32$ V and the bias voltage V_{Bias} was set to -0.7 V. The bias voltages V_{BIAS} and V_{BB} were set to -0.78 V and 0.7 V respectively to ensure that the transconductance of CFDITA would be 4.6 mS. Simulation results have been provided to analyze frequency and phase responses, time responses of input voltage and current, Monte-Carlo simulation results, and process-corner-voltage-temperature effects. The simulated frequency responses of magnitude and phase of the lossless capacitance circuit along with the theoretical evaluations are demonstrated in Figure 6 for which g_m was set to 4.6 mS and that includes a capacitor and resistor of values 10 pF and 100 kΩ respectively.

The results depicted in Figure 6 clearly demonstrate that the impedance and phase characteristics of the proposed CM closely align with the theoretical values across a broad frequency range spanning from 0.1 Hz to 100 MHz and 1 Hz to 2 MHz respectively.

The transient responses of input voltage and current are demonstrated in Figure 7 for which a sinusoidal input voltage of magnitude 100 mV at 100 kHz is applied.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

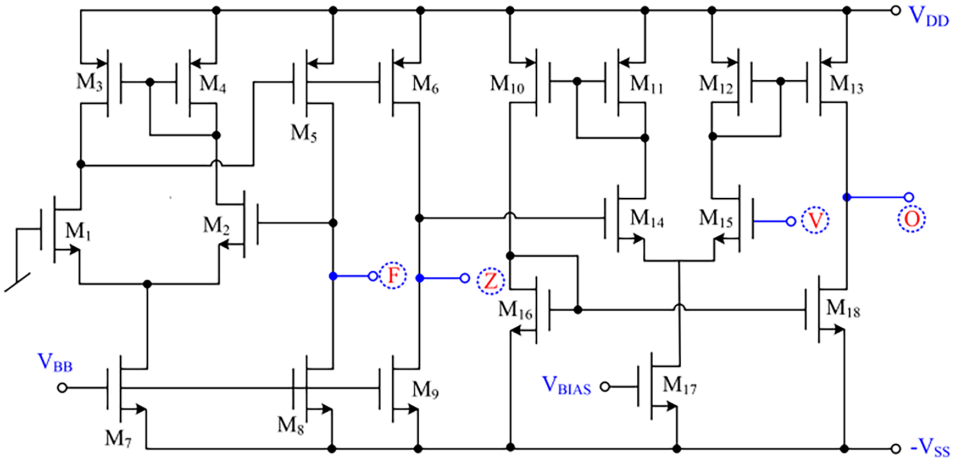


FIGURE 5 Internal architecture of CMOS CFDITA.[44]

TABLE 2 Aspect Ratios of MOSFETs shown in Figure 5.

MOS transistors	W(μm)	L(μm)
M_1, M_2	0.36	0.18
M_3, M_4	1.44	0.18
M_5, M_6	3.6	0.18
M_7–M_9	16.2	0.18
M_{10}, M_{13}	2.88	0.36
M_{11}, M_{12}, M_{16}, M_{18}	1.44	0.36
M14, M15	3.6	0.36
M_{17}	5.44	0.36

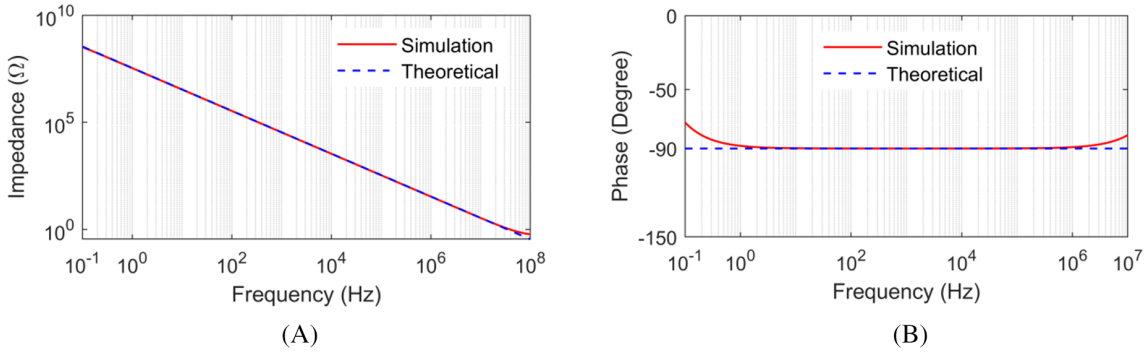


FIGURE 6 Simulated magnitude and phase of the impedance of CM circuit along with theoretical evaluations.

The variation in capacitance of the circuits shown in Figure 3 was also examined by repeatedly measuring transient and frequency responses. The value of resistor R was altered from 100 kΩ to 2MΩ and this adjustment resulted in capacitance values ranging from 4.6 to 92.02 nF while maintaining a constant g_m of 4.6 mS. The frequency responses of capacitance have been shown in Figure 8A. Furthermore, with variation in transconductance (g_m) of the CFDITA results into variation in MF upto 1800 and the corresponding results are illustrated in Figure 8B.

The results shown in Figure 8 indicate that the simulated capacitance obtained through the proposed circuit is in good agreement with the theoretical values.

Monte-Carlo simulations were performed on the proposed capacitance multiplier circuits to test their robustness. The simulations involved applying uniform deviations of 10% in resistance with 200 samples. The frequency responses of magnitude and phase and histogram results of maximum capacitance by varying resistor *R* are displayed in

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

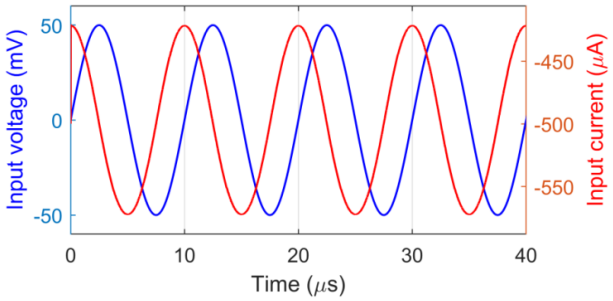


FIGURE 7 Transient input voltage and current of CM circuit at 100 kHz.

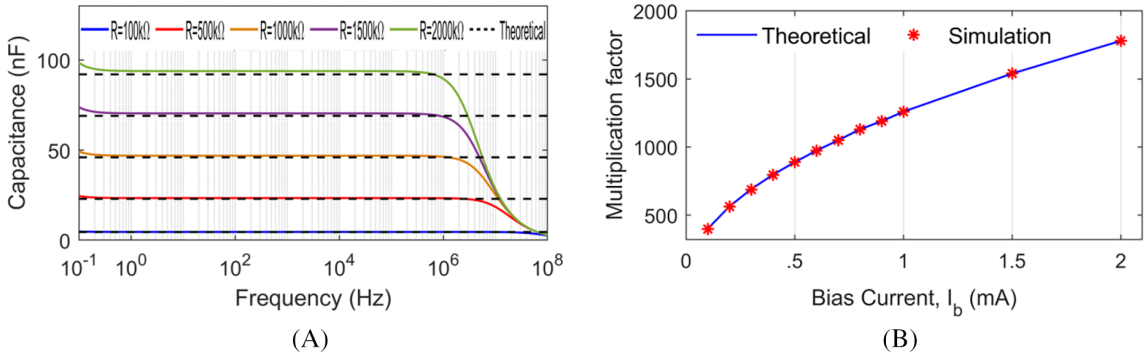


FIGURE 8 (A) Variation in capacitance with Resistor R (B) Variation in MF with I_{Bias}.

Figure 9A–C respectively. Temperature analysis has also been performed on the proposed circuit to assess its performance under different temperature conditions. The simulation results of magnitude and phase under different temperatures are shown in Figure 10.

Based on the results depicted in Figure 10, it is apparent that the magnitude and phase of the impedance have a negligible impact upto 100 °C temperature.

Similar analyses were conducted for the proposed lossy parallel simulator circuit, and the corresponding results are showcased in Figure 11. The transconductance g_m was set to 4.56 mS and the passive elements selected for the results were $R = 10$ kΩ and $C = 100$ pF. For the obtainment of transient responses, a sinusoidal voltage of magnitude 100 mV at a frequency of 100 kHz was applied to the input of the circuit and the simulated input voltage and current are displayed in Figure 11C.

From the results shown in Figure 11A,B, it is observed that the magnitude of the parallel RL circuit closely matches the theoretical predictions within the frequency range of 8 kHz to 100 MHz whereas the phase of the simulator is in good agreement with the theoretical values in the frequency range of 1 kHz–10 MHz.

To evaluate and compare the overall performance of the proposed GCM circuit, the figure of merit (FoM) for the GCMs is expressed as in[39]:

$$FoM_{GCM} = \frac{K.BW}{P_{diss}} = \frac{K(f_2 - f_1)}{P_{diss}} \tag{12}$$

where P_{diss} = Total quiescent power dissipation of the circuit, K = Multiplication factor and BW = Bandwidth over which the quality factor of capacitor has a value $|Q| > 10$.

The FoM for the GCM circuit was evaluated with C = 10 pF, K = 9202, BW approximately equal to 1 MHz, and the value was found to be 147 kHz/μW.

The simulations were also conducted to analyze the total harmonic distortions of the proposed circuit, and the variation of THD with input voltage has been depicted in Figure 12.

From Figure 12, it is observed that the maximum THD is 4% when the input voltage varies between 10 and 100 mV.

The simulations results shown in Figures 6–12 validate the performance of the proposed capacitance multiplier and lossy inductance simulator circuits and demonstrate their robustness under varying conditions.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

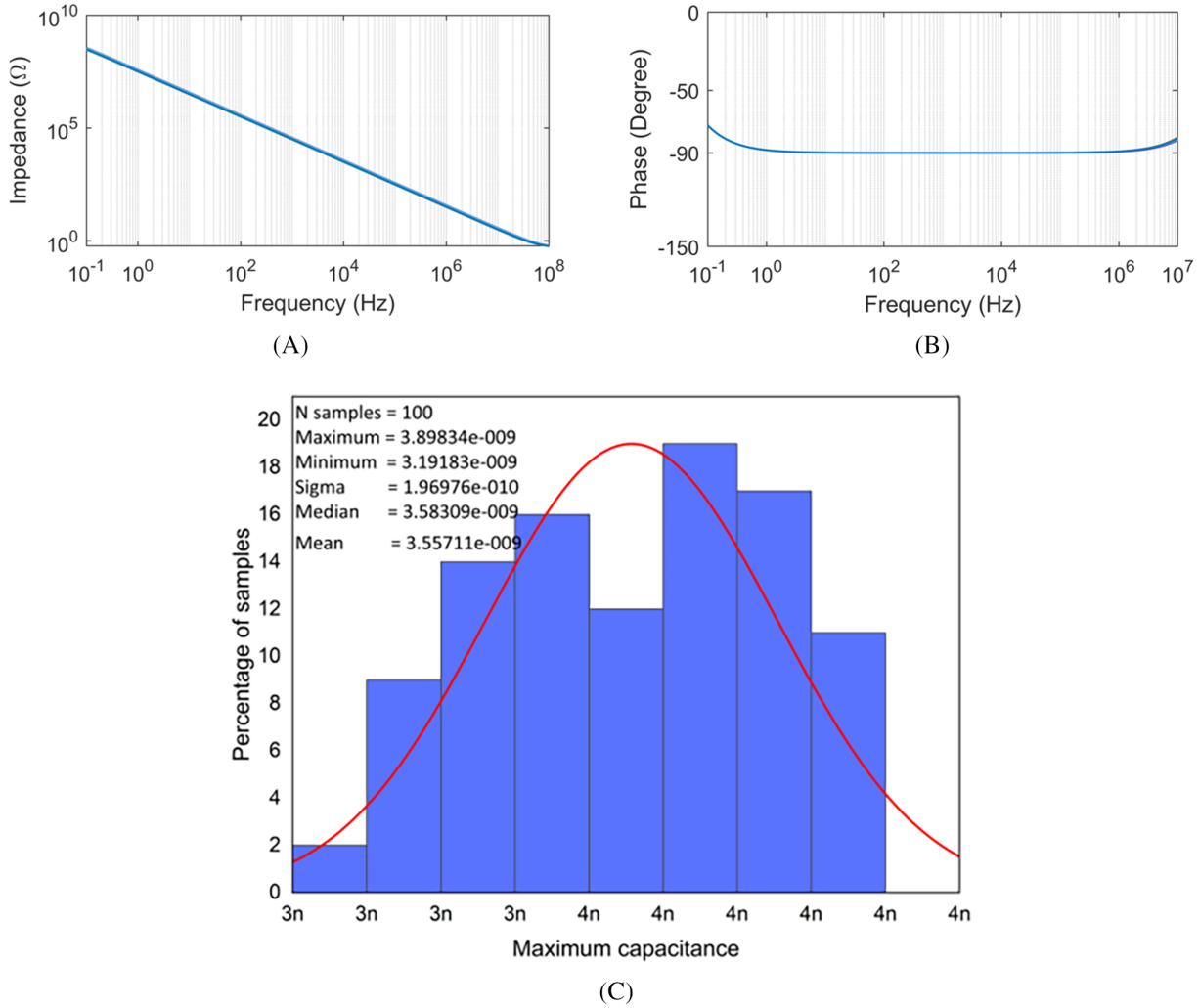


FIGURE 9 Monte-Carlo results (A) Magnitude (B) Phase (C) Histogram.

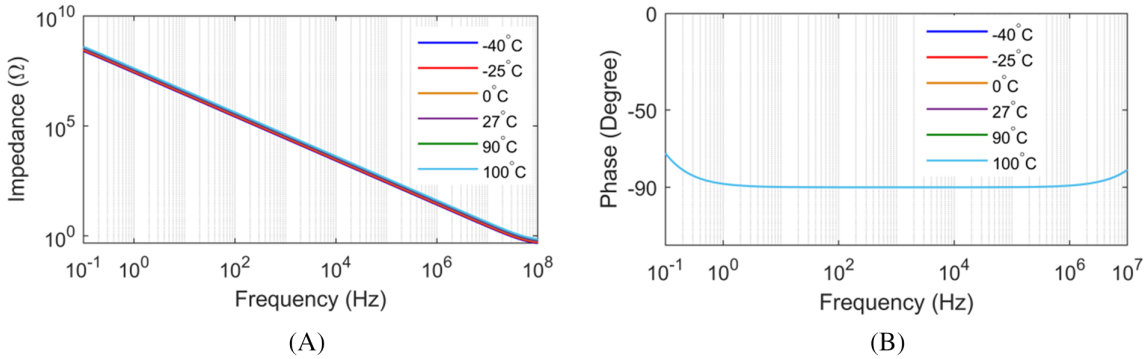


FIGURE 10 Simulated (A) magnitude and (B) phase responses of CM circuit under different temperatures.

4 | EXPERIMENTAL RESULTS

The workability of the presented CM circuits has also been validated experimentally using CFDITA implemented with off-the-shelf ICs AD844 and LM13700. The complete structure of experimental schematic of the circuit of Figure 3 using AD844 and LM13700 is presented in Figure 13.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

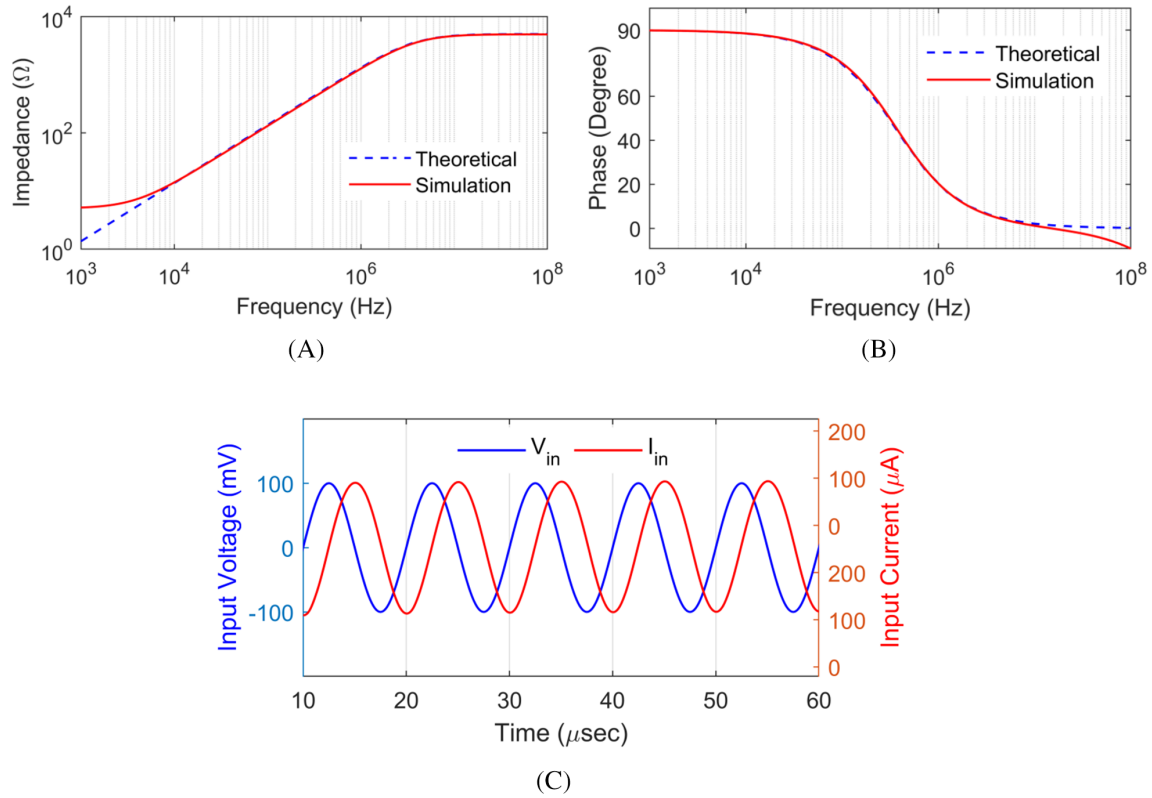


FIGURE 11 Simulated frequency and transient responses of the proposed parallel RL circuit (A) Magnitude responses (B) Phase responses (C) Input voltage and current.

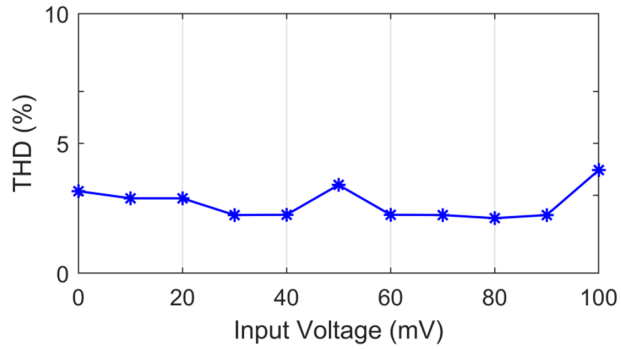


FIGURE 12 Variation in THD with input voltage.

The ICs AD844 and LM13700 used in Figure 13 were biased with power supply voltages of ±12 V. The passive components in the experimental verification circuit were selected as $R = 10$ kΩ, $C = 1$ nF, $R_B = 2.2$ kΩ and the value of transconductance g_m was set to 5.5 mS (for $R_{Bias} = 100$ kΩ) using the formula given in Ref. [47] resulting an equivalent capacitance value of 57 nF. The experimental frequency responses of magnitude and phase of the proposed circuit are shown in Figure 14A,B, while the transient responses to a sinusoidal input voltage of 1 V peak-to-peak at 1.5 kHz are displayed in Figure 14C.

From Figure 14A, it is seen that the experimental magnitude response of the capacitance multiplier aligns with the simulated and theoretical responses within the frequency range of 1–100 kHz. Figure 14C illustrates that the phase difference between input and output voltages is approximately 90 °, characteristic of a capacitor.

The derived parallel RL circuit was tested experimentally using passive component values of $R = 10\ k\Omega$, $R_{Bias} = 100\ k\Omega$, and $C = 1\ nF$. The obtained frequency responses of impedance and phase, as well as the macro model of AD844 and theoretical evaluations, are illustrated in Figure 15A,B. Additionally, Figure 15C showcases time-domain results with a sinusoidal input voltage of magnitude 2 V at a frequency of 100 kHz for further validation.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

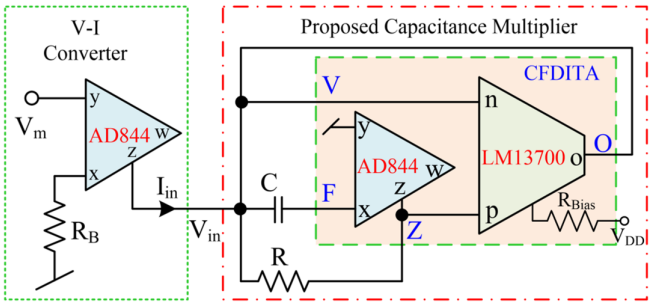


FIGURE 13 Schematic of proposed CM circuit using off-the-shelf ICs.

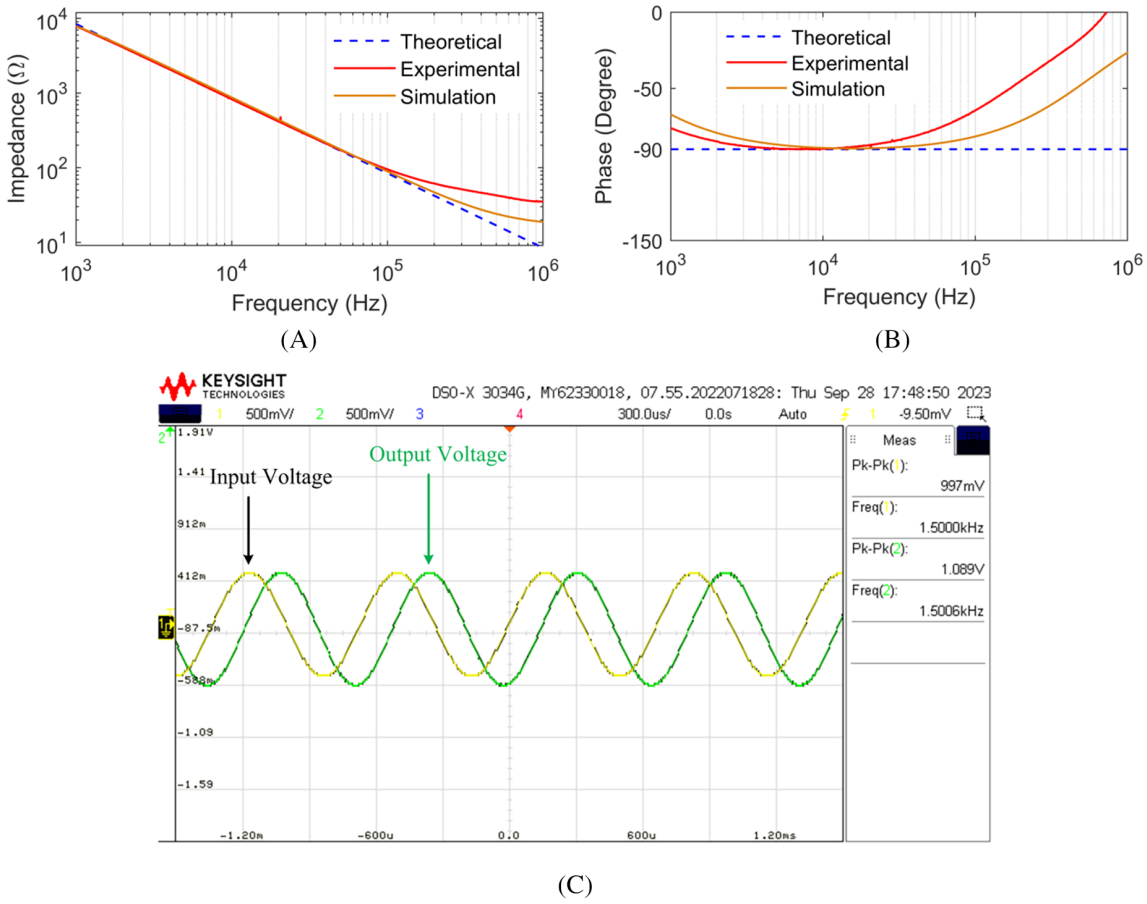


FIGURE 14 Experimentally obtained magnitude, phase and time domain results for CM circuit.

The experimental results depicted in Figure 15A indicate that the experimentally obtained frequency responses of the magnitude of the lossy inductor align with the simulated and theoretical results within the frequency range of 5–300 kHz. From the results illustrated in Figure 15C, it is seen that the phase between input and output voltages is approximately 44.4°.

5 | APPLICATION EXAMPLES

As an application, the proposed positive capacitance multiplier and parallel RL simulator were used to design first-order low-pass filter and second-order high-pass filter respectively whose circuit models and their active realizations using the proposed simulators are illustrated in Figure 16.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

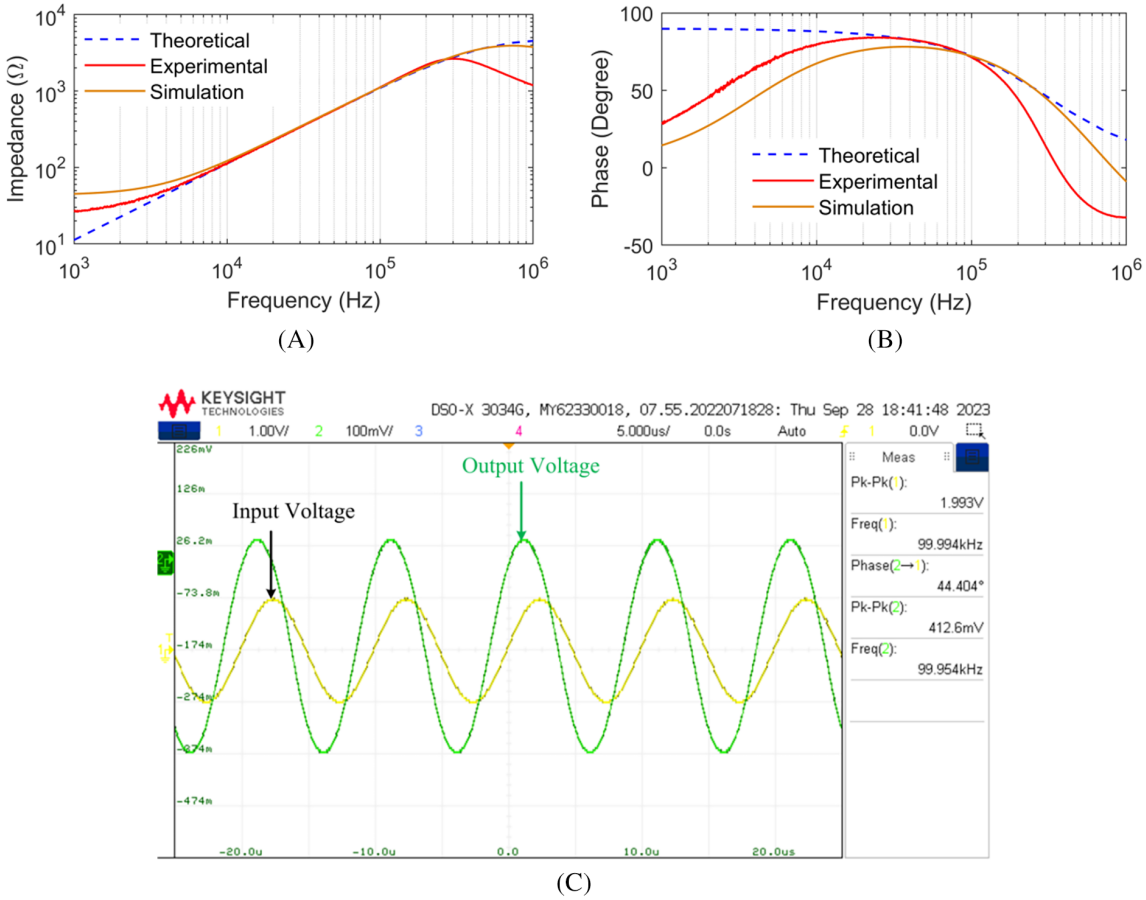


FIGURE 15 Experimentally obtained frequency and transient responses of parallel RL simulator.

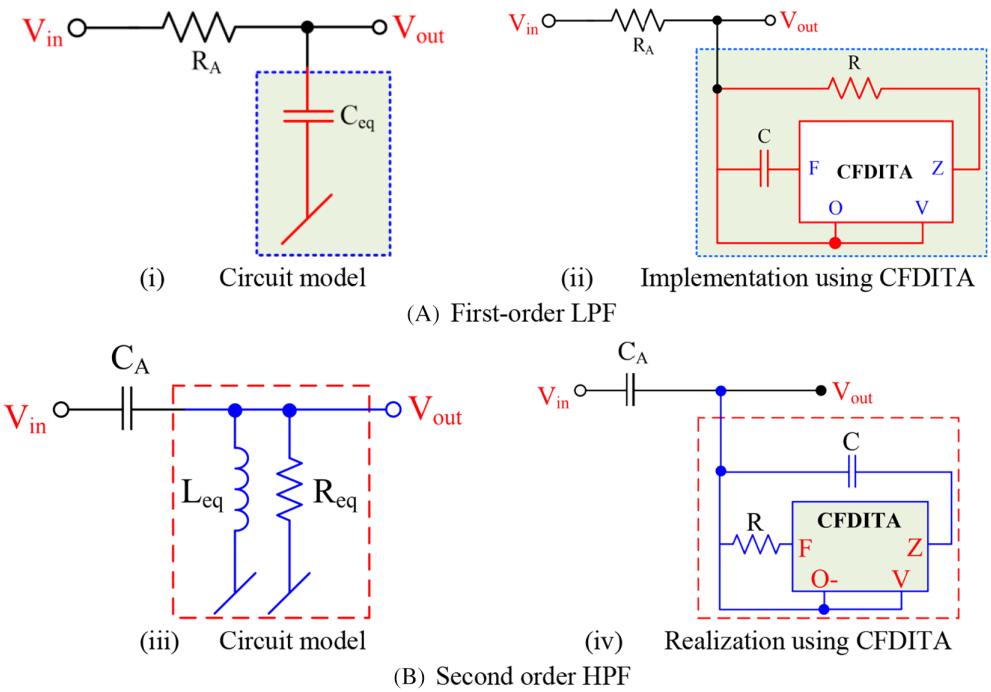
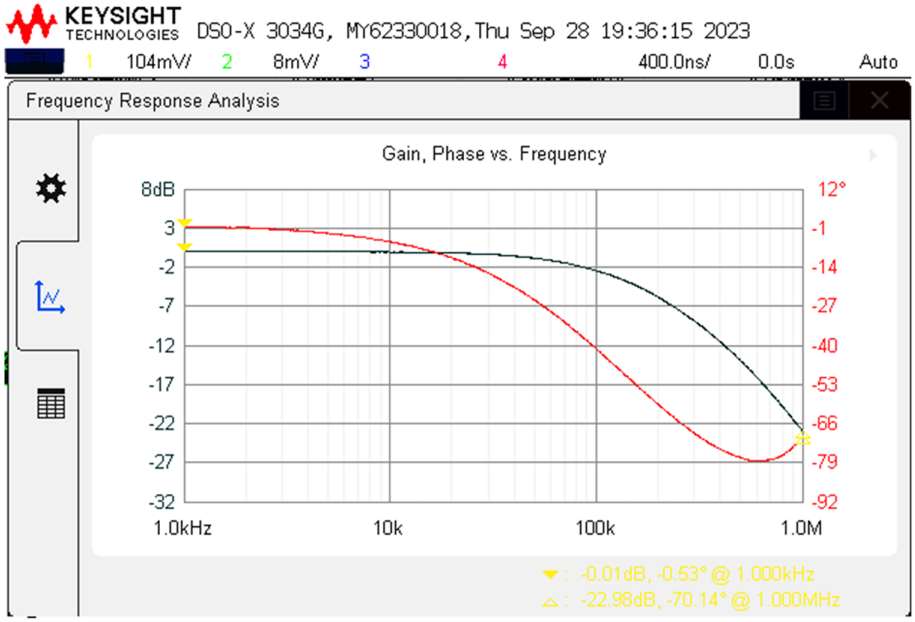
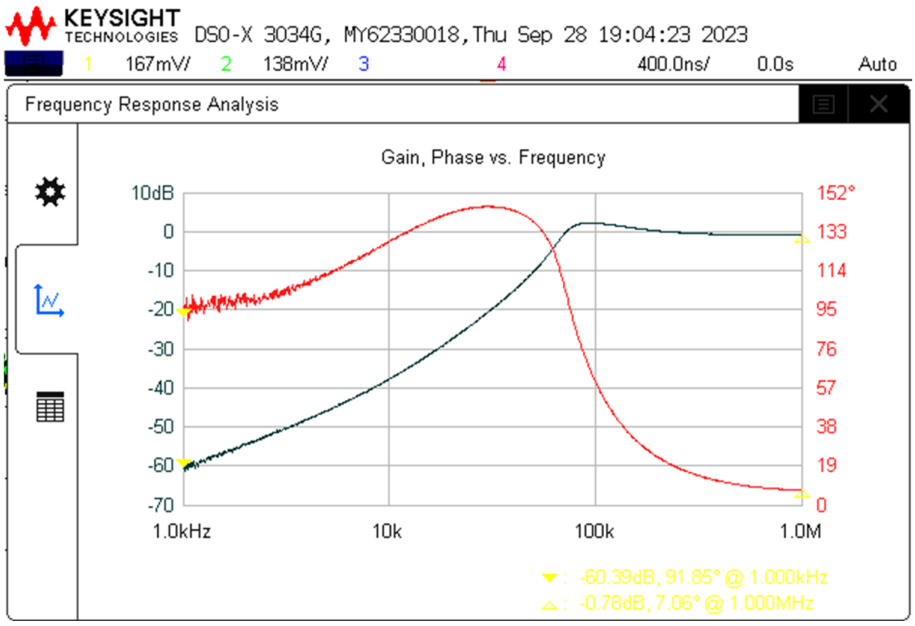


FIGURE 16 Application examples of proposed capacitance multiplier and parallel lossy inductor circuits.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(A) First-order LPF



(B) Second order HPF

FIGURE 17 Experimentally obtained frequency responses of filters shown in Figure 16.

Assuming ideal CFDITA, the transfer functions of the circuits shown in Figure 16 can be expressed as:

$$\text{LPF} \quad \frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{1+sC_{eq}R_A} = \frac{1}{1+sC(2+g_mR)R_A} \tag{13}$$

$$\text{HPF} \quad \frac{V_{out}(s)}{V_{in}(s)} = \frac{s^2L_{eq}R_{eq}C_A}{s^2L_{eq}R_{eq}C_A+sL_{eq}+R_{eq}} = \frac{s^2CC_AR^2}{s^2CC_AR^2+2sCR+g_mR} \tag{14}$$

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The performances of the derived low-pass and high-pass filters were experimentally validated for a nominal frequency of 110 kHz using the following passive component values:

For LPF: R = 2.2 kΩ, C = 100 pF, R_A = 1 kΩ, R_{Bias} = 100 kΩ for g_m = 5.6 mS.

For HPF: R = 1MΩ, C = C_A = 100 pF, R_{Bias} = 100 kΩ for g_m = 5.6 mS.

The experimentally obtained frequency responses of gain and phase of the LPF and HPF are displayed in Figure 17.

The proposed negative capacitance multiplier circuit may find application in capacitance cancelation at the current output port of the active device[19] and the application of grounded parallel positive *R* and negative *L* is open to investigation.

The simulation and experimental results shown in Figures 6–17, thus, validate the effectiveness of the proposed capacitance multiplier and parallel lossy RL simulator circuits.

6 | CONCLUDING REMARKS

In conclusion, this paper introduces an approach for designing capacitance multiplier circuits. Based on this approach, two novel capacitance multiplier circuits capable of providing positive and negative multiplication factors have been presented. Both the circuits utilize a single CFDITA, one resistor, and one virtually grounded capacitor. These circuits enable electronic enhancement of the capacitor's value by up to 1800 times its original magnitude. Additionally, the capacitance multiplication factor can be independently adjusted to reach values of up to 9202 by manipulating the resistor independently. Furthermore, the adaptability of these circuits extends to the implementation of parallel RL simulators, allowing for independent control of inductance in the derived circuits.

The paper also includes practical application examples that highlight the versatility of these proposed circuits. To validate their performance, SPICE simulations and experimental tests were conducted, both of which exhibited good agreement with the theoretical predictions.

The capacitance multiplier circuit employing a single CFDITA with a grounded capacitor and canonic number of passive elements has not been presented in the open literature yet and is open for further investigation.

CONFLICT OF INTEREST STATEMENT

The authors have no competing interest to declare.

DATA AVAILABILITY STATEMENT

We hereby confirm that our manuscript does not have any additional data beyond what is already contained in the manuscript. However, if any reader is interested in having the circuit files to verify any of the results, the same can be obtained from the authors on request.

ORCID

Ajishek Raj https://orcid.org/0009-0000-8659-5129

REFERENCES

1. Al-Absi MA, Al-Suhaibani ES, Abuelma'atti MT. A new compact CMOS C-multiplier. *Analog Integr Circuits Signal Process*. 2017;90(3):653-658.
2. Arslanalp R, Yücehan T. *Capacitance Multiplier Design by Using CFOA*. 23nd Signal Processing and Communications Applications Conference (SIU), IEEE; 2015.
3. Dogan M, Yuce E. A new CFOA based grounded capacitance multiplier. *AEU Int J Electron Commun*. 2020;115:153034.
4. Ozer E, Basak ME, Kacar F. Realizations of lossy and lossless capacitance multiplier using CFOAs. *AEU Int J Electron Commun*. 2020;127:153444.
5. Verma R, Pandey N, Pandey R. Novel CFOA based capacitance multiplier and its application. *AEU Int J Electron Commun*. 2019;107:192-198.
6. Lahiri A, Gupta A. Realizations of grounded negative capacitance using CFOAs. *Circuits Syst Signal Process*. 2011;30(1):143-155.
7. Raj A, Bhaskar DR, Shrivastava M, Kumar P. New negative-grounded capacitance multiplier circuits. *Int J Circuit Theory Appl*. 2023;51(3):1476-1491.
8. Yucehan T, Yuce E. *A New Grounded Capacitance Multiplier Using a Single ICFOA and a Grounded Capacitor*. IEEE Transactions on Circuits and Systems II: Express Briefs; 2021.
9. Di Cataldo G, Ferri G, Pennisi S. Active capacitance multipliers using current conveyors. *IEEE Int Symp Circuits Syst*. 1998;1-4.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10. Myderrizi I, Zeki A. Electronically tunable DXCCII-based grounded capacitance multiplier. *AEU Int J Electron Commun*. 2014;68(9): 899-906.
11. Winai J, Lahiri A, Siripruchyanun M. *Capacitance Multipliers Using Tunable Four Terminal Floating Nullors*. ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. IEEE; 2010.
12. Kamath DV. *Novel DO-OTA Based Current-Mode Grounded Capacitor Multiplier*. 2nd International Conference on Inventive Systems and Control (ICISC). IEEE; 2018.
13. Al-Absi MA, Abuelma'atti MT. A novel tunable grounded positive and negative impedance multiplier. *IEEE Trans Circuits Syst II Express Briefs*. 2018;66(6):924-927.
14. Petrović PB. Single VDTA-based lossless and Lossy electronically tunable positive and negative grounded capacitance multipliers. *Circuits Syst Signal Process*. 2022;41(12):6581-6614.
15. Özenl D, Alaybeyoğlu E, Kuntman H. A grounded capacitance multiplier circuit employing VDTA. *13th International Conference on Electrical and Electronics Engineering (ELECO)*. IEEE; 2021.
16. Aggarwal B, Mittal A. Design of temperature and input voltage insensitive VDTA and impedance multiplier. *Microelectron J*. 2019;85: 34-51.
17. Paul TK, Roy S, Pal RR. Lossy & Lossless Capacitance Multipliers: a series of realization using VDTAs & single grounded capacitor. *Mapana J Sci*. 2022;21(3):1-34.
18. Moonmuang P, Faseehuddin M, Pukkalanun T, Herencsar N, Tangsrirat W. VDTA-based floating/grounded series/parallel RL and RC immittance simulators with a single grounded capacitor. *AEU Int J Electron Commun*. 2023;160:154502.
19. Singh S, Jatin PN, Pandey R. Electronically tunable grounded capacitance multiplier. *IETE J Res*. 2020;68:1-12.
20. Sotner R, Jerabek J, Polak L, Petrzela J. Capacitance multiplier using small values of multiplication factors for adjustability extension and parasitic resistance cancellation technique. *IEEE Access*. 2020;8:144382-144392.
21. Başak ME, Özer E. Electronically tunable grounded inductance simulators and capacitor multipliers realization by using single current follower Transconductance amplifier (CFTA). *Analog Integr Circuits Signal Process*. 2022;1-15:401-415.
22. Ozer E. Electronically tunable CFTA based positive and negative grounded capacitance multipliers. *AEU Int J Electron Commun*. 2021; 134:153685.
23. Vavra J. *A Capacitance Multiplier Based on DBTA*. IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC). IEEE; 2017.
24. Tangsrirat W. Resistorless tunable capacitance multiplier using single voltage differencing inverting buffered amplifier. *Rev Roum des Sci Techn-Électrotechn et Énerg*. 2017;62(1):72-75.
25. Unhavanich S, Onjan O, Tangsrirat W. Tunable capacitance multiplier with a single voltage differencing buffered amplifier. *Proc IMECS*. 2016;1-4.
26. De MA, Ferri G, Stornelli V. *NIC-Based Capacitance Multipliers For Low-Frequency Integrated Active Filter Applications*. Ph D Research in Microelectronics and Electronics Conference IEEE; 2007:225-228.
27. Al-Absi MA, Al-Khulaifi AA. *A New High Multiplication Factor Tunable Grounded Positive and Negative Capacitance Simulator*. Asia Pacific Conference on Circuits and Systems (PCCAS); 2018:564-567.
28. Stornelli V, Safari L, Barile G, Ferri G. A new VCII based grounded positive/negative capacitance multiplier. *AEU Int J Electron Commun*. 2021;137:153793.
29. Stornelli V, Safari L, Barile G, Ferri G. A new extremely low power temperature insensitive electronically tunable VCII-based grounded capacitance multiplier. *IEEE Trans Circuits Syst II Express Briefs*. 2020;68(1):72-76.
30. Bhaskar DR, Mann G, Kumar P. OTRA-based positive/negative grounded capacitance multiplier. *Analog Integr Circuits Signal Process*. 2022;111:469-481.
31. Dogan M, Yuce E. Supplementary single active device based grounded immittance function simulators. *AEU Int J Electron Commun*. 2018;94:311-321.
32. Biolek D, Vavra J, Keskin AU. CDTA-based capacitance multipliers. *Circuits Syst Signal Process*. 2019;38(4):1466-1481.
33. Bhagat R, Bhaskar DR, Raj A, Kumar P. Single CDBA-based grounded parallel lossy inductor simulator circuits. *AEU Int J Electron Commun*. 2023;168:154728.
34. Aparicio R, Hajimiri A. Capacity limits and matching properties of integrated capacitors. *IEEE J Solid State Circuits*. 2002;37(3):384-393.
35. Kumar A, Singh D, Nand D. A Novel CFDITA-based design of grounded capacitance multiplier and its transpose structure. *Circuits Syst Signal Process*. 2022;41:c1-c21.
36. Singh D, Nand D, Kumar A. Newly realized grounded capacitance multiplier using single CFDITA. *In 2021 7th International Conference on Signal Processing and Communication (ICSC)*. IEEE; 2021:362-365.
37. Kumar A, Chaturvedi B, Jagga S. New CMOS compatible realizations of grounded/floating L, C multiplier and FDNC simulators. *Circuits Syst Signal Process*. 2023;42(4):1911-1939.
38. Padilla-Cantoya I, Furth PM. Enhanced grounded capacitor multiplier and its floating implementation for analog filters. *IEEE Trans Circuits Syst II Express Briefs*. 2015;62(10):962-966.
39. Pourashraf S, Ramírez-Angulo J, Montero JMH, González-Carvajal R, Lopez-Martin AJ. ± 0.25-V class-AB CMOS capacitance multiplier and precision rectifiers. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2019;27(4):830-842.
40. Chaturvedi B, Kumar A. CMOS MO-CFDITA based fully electronically controlled square/triangular wave generator with adjustable duty cycle. *IET Circuits Devices Syst*. 2018;12(6):817-826.

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

41. Kumar A, Chaturvedi B, Novel CMOS. CFDITA and its application as electronically tunable bistable multivibrator. *International Conference on Signal Processing and Communication*. ICSC; 2016:374-379.
42. Kumar A, Chaturvedi B. Fully electronically controllable Schmitt trigger circuit with dual hysteresis. *Electron Lett*. 2017;53(7):459-461.
43. Chaturvedi B, Kumar A. Electronically tunable current-mode instrumentation amplifier with high CMRR and wide bandwidth. *AEU-Int J Electron Commun*. 2018;92:116-123.
44. Shrivastava M, Kumar P, Raj A, Bhaskar DR. Single current follower differential input transconductance amplifier based grounded lossy capacitance multiplier with large multiplication factor. *Int J Numer Model: Electron Networks, Devices Fields*. 2024;e3139.
45. Biolek D, Senani R, Biolkova V, Kolka Z. Active elements for analog signal processing: classification, review, and new proposals. *Radioengineering*. 2008;17(4):15-32.
46. Rincon-Mora GA. Active capacitor multiplier in miller-compensated circuits. *IEEE J Solid-State Circuits*. 2000;35(1):26-32.
47. Greeneich EW. *Analog Integrated Circuits*. International Thomson Publishing; 1997.

How to cite this article: Raj A, Shrivastava M, Bhaskar DR, Kumar P. Enhancement of multiplication factor of capacitor using single current follower differential input transconductance amplifier. *Int J Numer Model*. 2024; 37(4):e3279. doi:10.1002/jnm.3279

10991204, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jnm.3279 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Journal of Optics
https://doi.org/10.1007/s12596-024-02152-x

RESEARCH ARTICLE

Ethanol-infiltrated circular photonic crystal fiber for low peak power supercontinuum generation from near-infrared to mid-infrared region

Drishti Singh Tomer[1] · **Ajeet Kumar**[1]

Received: 27 June 2024 / Accepted: 5 August 2024
© The Author(s), under exclusive licence to The Optical Society of India 2024

Abstract
We report computational modeling of a silica based Photonic Crystal Fiber (PCF) infiltrated with ethanol for NIR to MIR supercontinuum generation. Full vectorial finite element method (FEM) has been used to analyse the impact of infiltration. A flat-top dispersion profile has been obtained with a low peak power of 0.55 kW at a pump wavelength of 1.55 µm. A dispersion value of +30.231ps/nm/km is obtained in the anomalous region. For the proposed structure nonlinear coefficient is as high as 281.4 $W^{-1}km^{-1}$ with effective mode area of 4.16 μm^2 at 1.55 µm. A broadband supercontinuum spanning, from 1.25 µm to 3.0 µm is achieved by using a fiber length of 20 cm with 60 fs secant laser pulse source. The suggested PCF design assure several critical applications such as biomedical diagnostics, explosives detection, enhanced gas sensing capabilities and improvised food quality monitoring systems.

Keywords Nonlinear optics · Supercontinuum generation · Photonic crystal fibers · Photonic integrated circuits

Introduction

In recent times, supercontinuum generation (SCG) has become a focal point of extensive research interest [1–4]. In 1970, the concept of SCG was introduced for the first time [5]. Stimulated Raman scattering (SRS), self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM), assimilation of these nonlinear phenomena results in the process of SCG. When a high-intensity pump beam propagates through an optical fiber, these effects collectively interact with incident beam, causing substantial spectral broadening beyond the initial spectrum of the pump beam. This phenomenon arises from the interaction between the Raman effect, the Kerr effect, and the fiber's dispersion characteristics. Within the category of fibers, there exists a unique class known as holey fibers (HFs), photonic crystal fibers (PCFs) or microstructured fibers (MOFs). Their incredible linear and nonlinear properties of these fibers, make them a popular choice among researchers [6–10].

All normal dispersion (ANDi) profile of PCFs offer several advantages, including the preservation of ultrashort pulses, characterized by a uniform power spectral density and temporal profile. The absence of modulation instability (MI) due to soliton dynamics minimizes noise effects, facilitating the generation of coherent flat spectra. This characteristic is advantageous for applications in spectroscopy, metrology, and optical tomography systems including optical coherence tomography (OCT) [11, 12]. But the disadvantage with them they require high input power which can lead to photo thermal effect which can damage the fiber.

However, if we use anomalous dispersion regime with lesser input power broad spectra can be achieved, but coherency will be compromised. In order to balance out this soft glasses like chalcogenide was used [13–15]. However, they exhibit steep normal dispersion and a long zero-dispersion wavelength (ZDW) in the near-infrared (NIR), which restricts the flatness of the dispersion curve and the spectral bandwidth of supercontinuum generation.

To overcome these shortcomings recently liquid infiltrated PCFs are designed which have high nonlinear refractive index like CS_2 where $n_2 = 2.2 \times 10^{-18}$ m^2 W^{-1} at 1032 nm [16] which is quite large in comparision to silica. The advantages of using liquids in the air holes enables

✉ Ajeet Kumar
ajeetdph@dtu.ac.in

Drishti Singh Tomer
drishti.tomer@gmail.com

[1] Department of Applied Physics, Advanced Photonics Simulation Research Lab, Delhi Technological University, Delhi 110042, India

Published online: 30 August 2024

high nonlinear and tunable properties. The enhancement of nonlinear properties results in a broader SC spectrum with low pump power and low attenuation. They also provide extra degree of freedom. Some of the recent published work shows where Chemnitz et al., who reported on experimental evidence for dynamics of hybrid solitons, a new type of solitary wave, by using a CS_2-filled liquid core optical fiber [17]. Kedenburg et al., demonstrated that the PCF filled with CS_2 is capable of generating a spectral broadening ranging from 1.2 to 2.4 μm [18]. Similary, Pniewski et al., numerically demonstrated that the PCF filled with toluene can generate a broadband spectrum over 1000 nm [19]. Dai et al., measured a stimulated Raman scattering response around 0.6 μm in a commercially available hollow silica fiber filled with a toluene/chloroform mixture [20] and Van et al., numerically investigated the broadband spectrum of 700 nm in IR wavelength ranges for all-normal dispersion in hollow core PCFs with toluene [21]. But CS_2 and toluene both are highly toxic. Other organic solvents like nitrobenzene ($C_6H_5NO_2$). tetrachloroethylene (C_2Cl_4), chloroform ($CHCl_3$), benzene (C_6H_6), carbon tetrachloride (CCl_4), and ethanol (C_2H_5OH) etc. are also used which are comparatively less toxic. Vieweg et al., who experimentally demonstrated solitonic SC generation over 600 nm bandwidth with PCFs filled with carbon tetrachloride (CCl_4) [22]; In terms of fabrication techniques multiple options for infiltration have been used depending upon the requirement like photopolymerization [22], Thermal collapse [23], splicer process or UV adhesive process [24], Micro channel milling [25].

Structures with multiple hexagonal ring arrangement around the core has been a popular choice [26, 27].But the spectral bands are not very flat even though the dispersion curve is close to zero. Various designs like square, circular, octagonal, pentagonal and spiral structures were used to overcome this limitation [28, 29]. Circular PCFs demonstrate enhanced suitability for broadband SCG owing to their unique attributes, including (i) the redshift of the ZDW, which facilitates SCG in the MIR, (ii) the high design symmetry leads to the highest concentration of electromagnetic field energy in the core, (iii) a reduced dispersion slope within the anomalous dispersion region results in reduction of noise in the SCG spectrum. Also, group velocity dispersion is low in the normal dispersion regime, leading to longer dispersion lengths and a higher degree of soliton interaction [30].

In our reported design, we are analyzing the impact of infiltration of ethanol on supercontinuum generation. A near-infrared (NIR) to mid-infrared (MIR) SC spectrum is observed at a low input pump power extending form 1.25 –3.0 μm at 1.55 μm pump wavelength using a fiber of length 20 cm injected with 60 femtoseconds laser pulses of 550 W. Low input power is crucial from practical point of view. Especially, with liquid infiltration in PCFs, they cannot stand high pump power as they can be evaporated if high power is used. A higher broadening with low input power is reported in contrast to recently reported result mentioned in Table 1. Furthermore, the challenge of balancing spectral broadening with maintaining coherence has been adeptly managed. The generated SC source is versatile and can be employed in a wide array of applications such as chemical analysis, explosive detection, frequency comb generation, food quality inspection, medical diagnostics and advanced optical imaging techniques including OCT. This paper is organized into five main sections. The first section offers a concise introduction and an overview of the study. The second section delves into the design of PCF. The third section provides a detailed explanation of the numerical methodology employed in the research. The fourth section presents the results and discusses their implications. Finally, the fifth section concludes the paper which is followed by a list of references.

Table 1 Comparison with earlier documented findings of supercontinuum spectra in infiltrated PCFs

References	# PCF	Pump Wavelength (μm)	Peak power (kW)	Fiber length	SCG bandwidth (μm)
[33]	$C_6H_5NO_2$,#F_2	1.56	5.55	5 cm	0.8–2.1
[33]	$C_6H_5NO_2$, #F_3	1.56	0.66	5 cm	1.3–2.3
[34]	C_7H_8	1.03	25	10 cm	0.95–1.1
[35]	C_6H_6, #F_1	1.56	55	1 cm	0.7-2.0
[36]	C_2Cl_4 ,#F_1	1.56	16.67	5 cm	0.8-2.0
[36]	C_2Cl_4 ,#F_2	1.56	16.67	10 cm	1.0–2.0
[36]	C_2Cl_4 ,#F_3	1.03	20.83	10 cm	0.7–2.4
[37]	$C_2H_4Br_2$, #F_1	1.03	12.5	10 cm	0.64–1.7
[37]	$C_2H_4Br_2$, #F_2	1.03	0.75	15 cm	0.7–2.4
[38]	C_7H_8, # I_3_0.3	1.55	7.14	10 cm	1.1–1.75
[38]	C_7H_8, # I_3_0.35	1.55	6.67	10 cm	1.0-1.75
Proposed work	C_2H_6O	1.55	0.55	20 cm	1.25-3.0

Design principle of PCF

Dispersion properties are not the sole determinant of the effectiveness of a nonlinear medium for SCG. It is also crucial to achieve low loss and small effective mode area. Also, the material used should have high nonlinear refractive index (n_2).These factors are influenced by the fiber structure. Figure 1 shows the cross-sectional view of proposed PCF. The PCF structure comprises of silica with micro structured air holes in the form of four circular rings. Each ring is having constant diameter of air holes as $d = 0.7$ μm. The first ring has eight air holes and other rings have larger number of air holes constituting symmetry in the structure given by

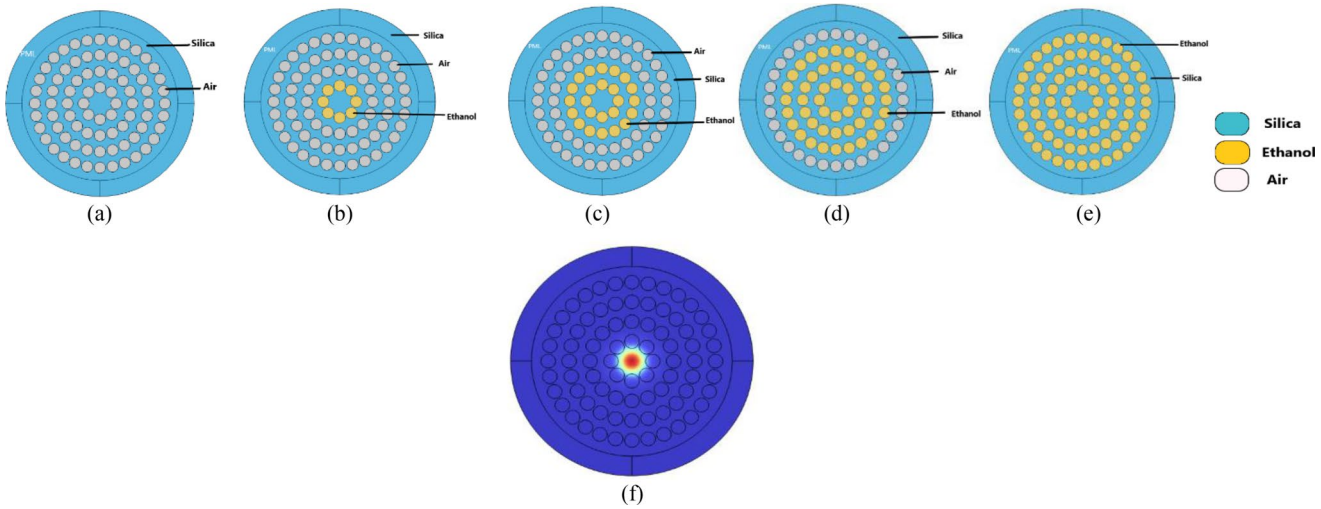


Fig. 1 (*a*), (**b**), (**c**), (**d**), (**e**) depicts the cross-sectional view of the suggested PCF; (**f**) at 1.55 µm the dispersal of electric field for the principal mode

$x = r\ cos\ (\theta)$ *and* $y = r\ sin\ (\theta)$, where r is the distance from the center. The pitch (Λ) is the distance between the two consecutive air holes. The pitch (Λ) is 1 µm. The cylindrical Perfectly matched layer (PML) is generally 10–20% of the diameter of the cladding region, which is implemented at the edges of the cladding for eradicating reflections at the boundary. The circular rings are infiltrated with ethanol gradually (one ring → two rings → three rings → four rings) and its impact is studied.

Numerical methodology

Linear attributes

The PCF design is created and the effective mode index (n_{eff}) of the principle mode is determined using COMSOL multiphysics software which uses FEM technique. The wavelength dependent refractive index of ethanol and silica glass material was calculated by using the Sellmeier equation and the constants are $B_1 = 0.83189$, $B_2 = -0.15582$, $b_1^2 = 0.0093\ \mu m^2$, $b_2^2 = -49.452\ \mu m^2$ for ethanol and for silica glass are $B_1 = 0.69616630$, $B_2 = 0.40794260$, $B_3 = 0.89747940$, $b_1^2 = 0.00467914 \mu m^2$, $b_2^2 = 0.01351206 \mu m^2$, $b_3^2 = 97.93400253 \mu m^2$ with λ as the operating wavelength [31]

$$n^2(\lambda) = 1 + \sum_{k=1}^{N} \frac{B_k \lambda^2}{\lambda^2 - b_k^2} \tag{1}$$

The comprehensive dispersion profile of the intended PCF framework is assessed by employing the following relationship

$$D(\lambda) = -\frac{\lambda}{c} \frac{\partial^2 Re(n_{eff})}{\partial \lambda^2} \tag{2}$$

Re (n_{eff}) denotes the real component of the effective mode index and c is the speed of light in free space. Using the following equation, the effective mode area $(A_{eff})$of the propagating mode has been conducted [32]

$$A_{eff} = \frac{\left(\iint_{-\infty}^{\infty} |E|^2 dx\, dy\right)^2}{\left(\iint_{-\infty}^{\infty} |E|^4 dx\, dy\right)} \tag{3}$$

Where, the amplitude of the electric field of the light is represented by E.

Nonlinear attributes

For the simulations, we utilize an initial input laser pulse modeled by a hyperbolic secant function, as shown in the equation below.

$$A(0, T) = \sqrt{P_0}\, \mathrm{sech} \frac{T}{T_0} \tag{4}$$

The equation presents several key components that define the characteristics of the laser pulse. First, the pulse covering A(0, T) determines the shape and amplitude of the pulse over time. P_0 represents the peak power of the initial input pulse, indicating its maximum intensity. T_0 is a parameter tailored specifically for hyperbolic secant pulses to describe their temporal width, it is calculated as $T_0 = T_{FWHM}/1.7627$, where T_{FWHM} is the full-width-half-maxima. Finally, T denotes the time frame aligned with the group velocity (GV) of the pulse covering, capturing how the pulse propagates through the medium over time.

The following generalized nonlinear Schrödinger equation is solved using Split-Step Fourier method to obtain the widening of the spectral range observed in the generated SC within the PCF structure [33]

$$\frac{\partial \widetilde{A}'}{\partial z} = i\bar{\gamma}(\omega)\exp\left(-\hat{L}(\omega)z\right)\mathcal{F}\left\{\overline{A}(z,T)\int_{-\infty}^{\infty} R(T')\left|\overline{A}(z,T-T')\right|^2 dT'\right\} \quad (5)$$

Where, $\widetilde{A}'$ is envelope or the covering of the output pulse which represents its modulation amplitude or shape, $\bar{A}(z,T)$, the envelope at distance z at time t. $\bar{A}(z,\omega)$, the Fourier transform of $\bar{A}(z,T)$. The frequency-dependent nonlinear coefficient $\bar{\gamma}(\omega)$ is obtained by taking into account its dependency on the frequency

$$\bar{\gamma}(\omega) = \frac{n_2 n_0 \omega}{c n_{eff}(\omega) A_{eff}^{1/4}(\omega)} \quad (6)$$

n_2 corresponds to the nonlinear refractive index, which is specifically 2.89×10^{-20} m^2/W at 1550 nm for silica glass [34]. The refractive index of glass at the specified wavelength is represented by symbol n_0, which is required to calculate n_2 and c is equivalent to 3.0×10^8 m/s. The expressions for the modified variables are given as follows

$$\widetilde{A}'(z,\omega) = \widetilde{A}(z,\omega)\exp(-\widehat{L}(\omega)z) \quad (7)$$

The linear operator $\widehat{L}(\omega)$ is expressed by the following equation as

$$\widehat{L}(w) = i(\beta(\omega) - \beta(\omega_0) - \beta_1(\omega_0)[\omega - \omega_0]) - \frac{\alpha(\omega)}{2} \quad (8)$$

Within the PCF design β, β_1, α and ω_0, represents the propagation constant, the GV inverse of pulse covering, the total loss and the reference frequency respectively. The Raman response function $R(t)$ is defined as

$$R(t) = (1-f_R)\delta(T) + f_R\frac{\tau_1^2+\tau_2^2}{\tau_1\tau_2^2}\exp\left(-\frac{T}{\tau_2}\right)\sin\left(\frac{T}{\tau_1}\right) \quad (9)$$

For the silica glass, Raman parameters f_R,τ_1,τ_2 are 0.245,12.25 fs and 32 fs respectively [31].

Results and its discussion

In a PCF structure the broadening in the supercontinuum spectrum primarily depend upon the nonlinearity of material and the dispersion characteristics [34]. Figure 2 depicts the values of dispersion for different rings. The plot with all air has one Zero dispersion wavelength (ZDW) around 1.3 μm and a dispersion compensating nature. It has been observed that the behavior of plot has largely been changed after infiltration, it has resulted into a more flat and smoother curve. Also the value of dispersion has greatly been reduced. It can be seen for one ring with $d=0.7$ μm and the pitch $(\Lambda)=1.0$ μm it has two ZDWs one at 1.8 μm and another at 2.5 μm. The profile mostly being in the normal and near zero anomalous dispersion region leading to enhancement of the coherent characteristics. Now, again the infiltration was performed for more circular rings two rings, three rings and four rings gradually. The nature of the plot is similar but they all have only one ZDW around 1.8 μm with large value of dispersion as compared to one ring infiltration. Also, they have both normal and anomalous regime. One ring with ethanol is having better dispersion control and hence has been considered for further study. Also, from the fabrication perspective one ring infiltration is simpler. In Fig. 3, Λ

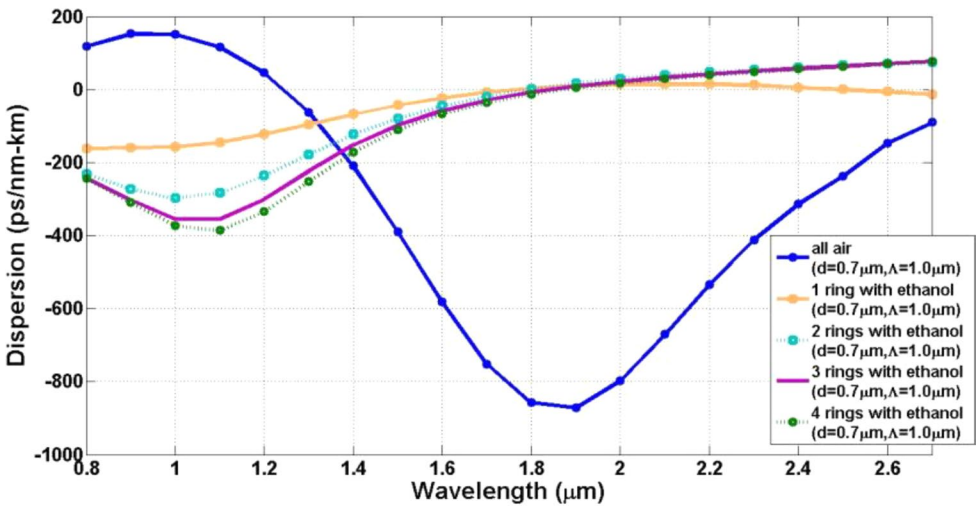


Fig. 2 Effect of subsequent ring infiltrations with Λ=1 μm on dispersion characteristic of the PCF with $d=0.7$ μm

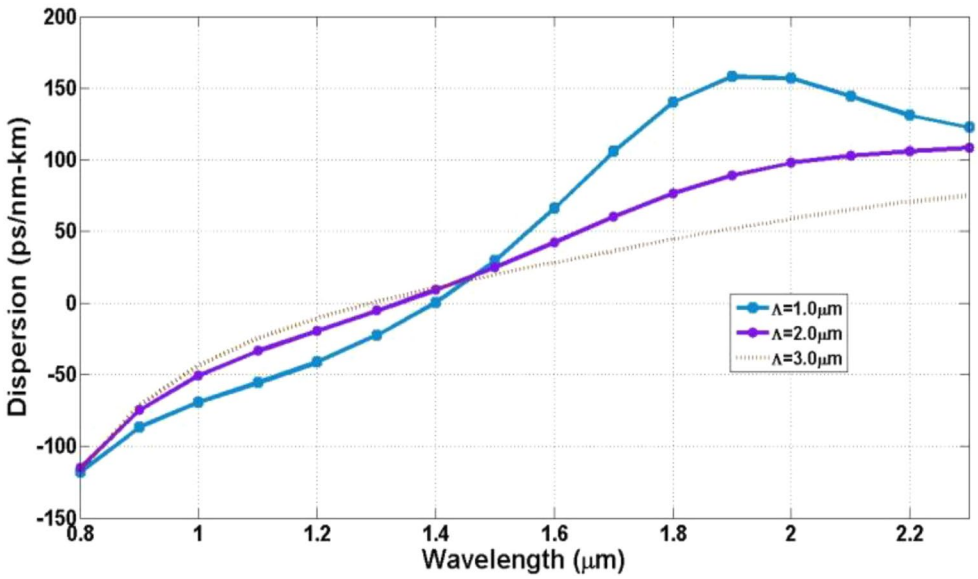


Fig. 3 Effect of Pitch variation on one ring infiltration with Λ=1 μm, 2 μm and 3 μm and $d=0.7$ μm

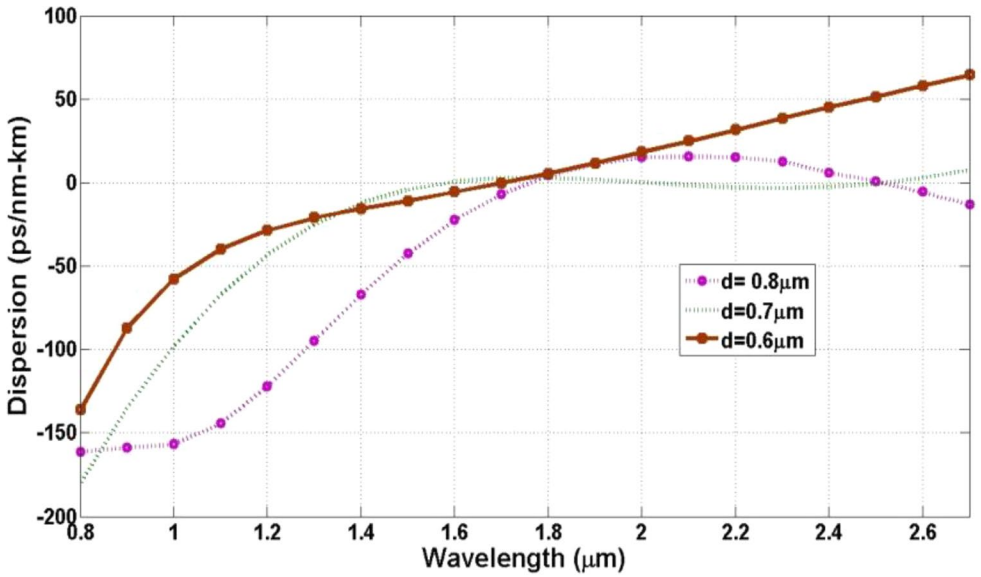


Fig. 4 Effect of air hole diameter variation on the dispersion characteristic of the PCF with one ring infiltration for $d=0.6$ μm, 0.7 μm & 0.8 μm with Λ=1 μm

is optimized for 1.0 μm, 2.0 μm and 3.0 μm. It is observed that the profile with Λ =1.0 μm has lesser value of dispersion as compared to Λ =2.0 μm and Λ =3.0 μm. So the former was considered. In Fig. 4 while doing the optimization of d (=0.6 μm, 0.7 μm and 0.8 μm), $d=0.7$ μm has shown comparatively more flat profile with less dispersion. A flat, normal and lesser dispersion valued profile gives a broader as well as coherent SC spectra. So, as a structural parameter we have considered $d=0.7$ μm as air hole diameters and pitch as 1 μm with one ring infiltration of ethanol. The value of dispersion is +30.231ps/nm/km for $d=0.7$ μm and Λ =1.0 μm. Effective mode area (A_{eff}) & nonlinearity (γ) possesses inverse relation with each other, the trend can be seen in Fig. 5 when both plotted with wavelength. The value of A_{eff} and nonlinearity at 1.55 μm is 4.16 μm^2 and 281.4 $W^{-1}Km^{-1}$ respectively. Keeping in mind the availability of a standard commercial femtosecond laser system, 1.55 μm Er-Fiber laser with 100 MHz repetition rate is used.

The value of confinement loss is 3.19×10^{-2} dB/km, which is less in comparison to the material loss of the silica glass [35]. For multiple values of fiber length, pulse width and pump power, SCG has been assessed at a wavelength of 1.55 μm with 100 MHz repetition rate at −40 dB power level. SCG has been observed for various fiber length 5 cm, 10 cm and 20 cm in Figs. 6 and 7. The spectrum is getting broader with increasing length. The phenomenon responsible for this broadening is SPM. The coherence is also evaluated for 20 cm fiber length, 550 W pump power and 60 fs pulse width in Fig. 8. In Fig. 9 the peak power is changed from 450 W to 650 W, the spectrum gets shifted towards the

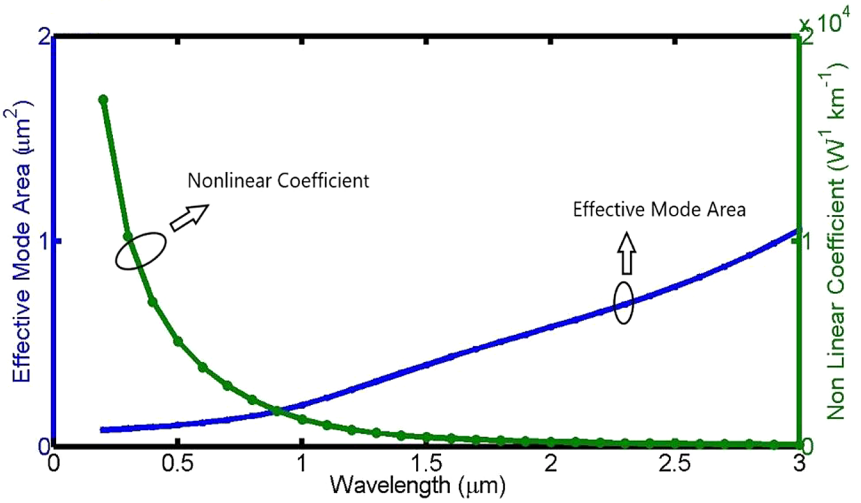


Fig. 5 The variation of the effective mode area and nonlinear coefficient are illustrated across different wavelengths for the PCF

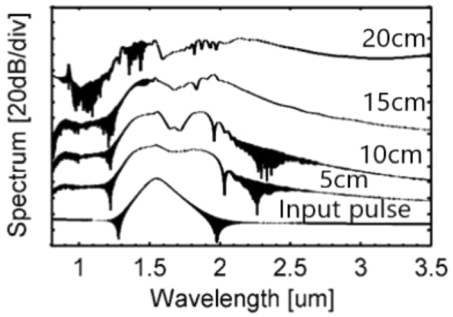


Fig. 6 The progression of spectral supercontinuum (SC) broadening at 1.55 μm due to varying fiber lengths of 5 cm, 10 cm, 15 cm, and 20 cm under conditions of 60 fs and 550 W

longer wavelength region. The symmetrical broadening of SC spectrum is mainly occurred due to the nonlinear SPM effect along the shorter and longer wavelength regions. Simultaneously, at higher peak power, the SC started to deteriorate and the coherency has also started to degrade as noise is increasing. This happened due to the absorption at the coating [36]. Then, the impact of increased pulse widths on SC has been analyzed. In Fig. 10, it is evident that as the T_{FWHM} value increases from 60 femtoseconds to 80 femtoseconds, the resulting spectra narrow down. It is noteworthy that shorter pulses typically yield broader supercontinuum spectra. Therefore, the 60 femtosecond pulses exhibit a stable spectrum. The maximum broadening, ranging from 1.25 to 3.0 μm, is observed with the length of fiber as 20 cm, 550 W peak power, and a pulse width of 60 femtoseconds.

For 60 fs laser pulse at 1.55 μm, β_2 is 1.21×10^{-26}, ($T_0 = T_{FWHM}/1.763$), the nonlinear length ($L_{NL} = (\gamma P_0)^{-1}$) and dispersion length ($L_D = (\beta_2/T_0^2)^{-1}$) for proposed PCF are 6.49×10^{-3} m and 2.975×10^{-1} m respectively. The order of soliton $N \approx L_D/L_{NL} \approx 45$ and the fission length of soliton calculated as $L_{fiss} = L_D/N \approx 6.6$ mm are determined using these relationships. To achieve a wider SCG, it is advantageous to have $L_{NL} \leq L_D$. This ensures that nonlinear effects significantly influence the spectrum of the input pulse, while preventing excessive broadening of the pulse by dispersion before nonlinear effects can occur. Initially, during pulse propagation, spectral broadening primarily occurs due to SPM, where the intensity alters the pulse phase. As the pulse progresses to longer wavelengths, broadening is increasingly influenced by optical wave breaking, a phenomenon where the intensity of the pulse or shape changes due to nonlinear effects in the medium.

Additionally, to assess the coherence level of the obtained spectra, we have performed a coherence calculation for 20 cm fiber length in Fig. 8.The spectrum is coherent for the entire SC broadening range 1.25–3.0 μm. The first-order degree of coherence is given by equation-(10) mentioned below. The electric fields of individual pairs

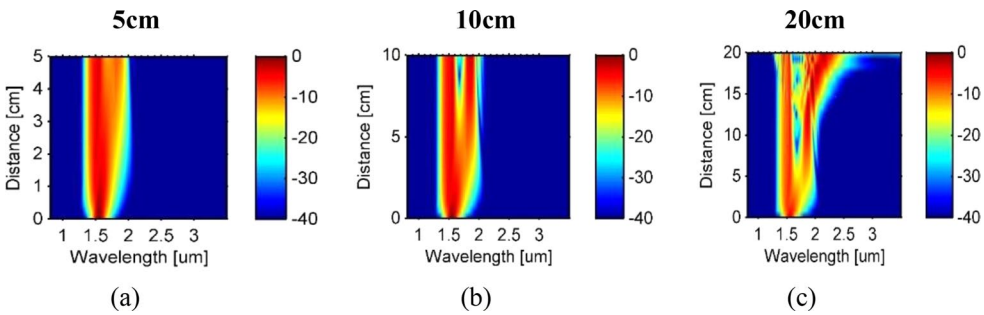


Fig. 7 The spectral evolution of SCG at 1.55 μm induced by (**a**), (**b**), (**c**) length variation of 5 cm,10 cm,20 cm at 60fs & 550 W

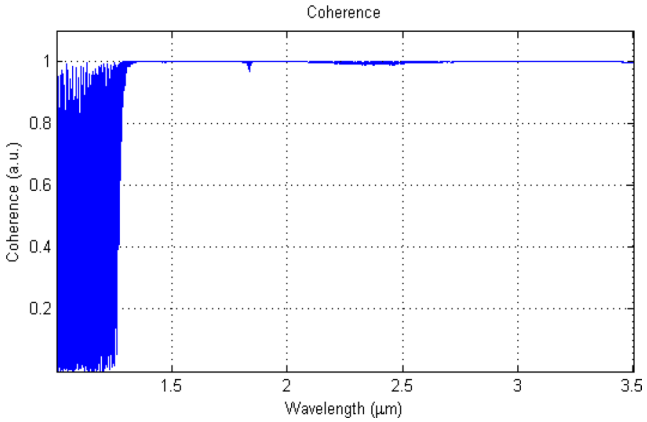


Fig. 8 The coherence evolution at 1.55 micron, using 550 W, 60 fs pulse width along the fiber length of 20 cm

of supercontinuum generation (SCG) are denoted as $E_1(\lambda, t_1)$ and $E_2(\lambda, t_2)$. These fields correspond to immediately subsequent SCG spectra, with angular brackets indicating an average over the entire pulse ensemble. To focus on the reliance of coherence on wavelength, we adopted $\Delta t = t_1 - t_2 = 0$.

$$|g_{12}(\lambda\ (t_1 - t_2)| = \frac{|< E_1^*(\lambda, t_1) E_2(\lambda, t_2) >|}{(< |E_1(\lambda, t_1)^2| >< |E_2(\lambda, t_2)^2| >)^{1/2}} \quad (10)$$

The Degree of coherence function is bounded as $0 \leq |_{g_{12}}(\lambda, t_1 - t_2)| \leq 1$. If the value is 1 it means perfectly coherent spectrum and if it is 0 than no coherency is obtained. A value in between 0 and 1 shows moderate coherency. We have computed the Degree of coherence especially to examine the wavelength dependence of the coherence [36].

In conclusion, our proposed structure demonstrates the capability to achieve a broader and comparable bandwidth for supercontinuum generation (SCG) using lower pump power compared to previously documented results, as outlined in Table 1. As can be seen [37] has comparable peak power but have lesser broadening range, [38–42] the peak power used is very high and the broadening range is also less as compared to the proposed design.

This study focuses on achieving substantial spectral broadening with low peak power, providing multiple benefits for supercontinuum generation (SCG). Operating at lower peak power significantly reduces noise, thereby enhancing the quality of measurements and imaging. Additionally, lower peak power typically results in longer pulse durations, which can be advantageous for applications that require prolonged interaction with the sample, including material processing and biomedical imaging [43]. In summary, employing a lower value of peak power in SCG can produce more stable, high-quality spectra and offers increased safety for sensitive samples.

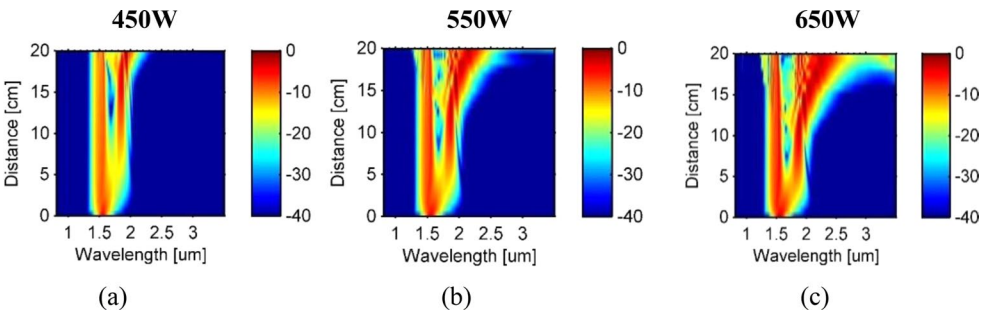


Fig. 9 The spectral evolution of SCG at 1.55 μm induced by (**a**), (**b**), (**c**) peak power variation of 450 W,550 W,650 W at 60fs & 20 cm

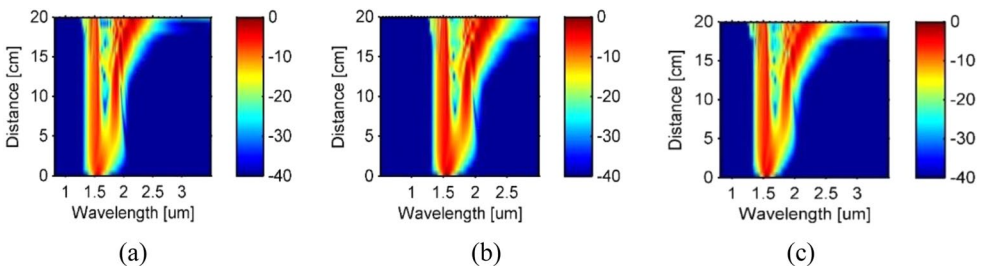


Fig. 10 The spectral evolution of SCG at 1.55 μm induced by (**a**), (**b**), (**c**) pulse width variation of 60 fs,70 fs ,80 fs at 20 cm fiber length and 550 W pump power

Conclusion

In conclusion, our proposed structure demonstrates the capability to achieve a broader and a comparable bandwidth for supercontinuum generation (SCG) using lower pump power of 0.55 kW. Furthermore, the challenge of balancing spectral broadening with maintaining coherence has been adeptly managed. The core focus of this work centres on generation of near to mid-IR supercontinuum with a coherent SC spectrum extending from 1.25 to 3.0 μm in a PCF of 20 cm length, when subjected to 550 W and 60 femtosecond laser pulse at 1.55 μm. From an application perspective, leveraging a PCF-based broadband coherent supercontinuum (SC) light source spanning from the near-infrared (NIR) to the mid-infrared (MIR) spectrum is profoundly impactful across several vital fields. These applications encompass early cancer detection, food quality monitoring, security applications, and sensing technologies. Additionally, the technology facilitates the creation of a stable optical frequency comb essential for precise OCT and spectroscopy.

Acknowledgements Drishti Singh Tomer deeply acknowledges Delhi Technological University, Delhi for providing necessary support to conduct this work.

Data availability The research described in the article uses no data.

References

1. Y. Qu, W. Sun, J. Cao, H. Chen, H. Jia, Coherent ultrabroad orbital angular momentum supercontinuum generation in an $AsSe_2$ -As_2S_5 microstructured fiber with all-normal dispersion. Opt. Commun. **497**, 127191 (2021). https://doi.org/10.1016/j.optcom.2021.127191
2. D.S. Tomer, Design and numerical modeling of chalcogenide parabolic - core waveguide for on - chip supercontinuum generation extending from near - IR region to mid - IR region, (2024) 1–7. https://doi.org/10.1002/mop.34105
3. S. Rohan Sharma, P. Kaur, A. K. Chauhan, Computational design & analysis of GeSe2-As2Se3-PbSe based rib waveguide for mid-infrared supercontinuum generation. Optik. **220**, 0030–4026 (2020). https://doi.org/10.1016/j.ijleo.2020.165032
4. T.S. Saini, A. Kumar, R.K. Sinha, Design and modelling of dispersion-engineered rib waveguide for ultra broadband mid-infrared supercontinuum generation. J. Mod. Opt. **64**(2), 143–149 (2016). https://doi.org/10.1080/09500340.2016.1216190
5. R.R. Alfano, S.L. Shapiro, Emission in the region 4000 to 7000 Å Via Four-Photon Coupling in Glass. Phys. Rev. Lett. **24**, 584–587 (1970). https://doi.org/10.1103/PhysRevLett.24.584
6. P. Chauhan, A. Kumar, Y. Kalra, Optik A dispersion engineered silica-based photonic crystal fi ber for supercontinuum generation in near-infrared wavelength region. Opt. - Int. J. Light Electron. Opt. **187**, 230–237 (2019). https://doi.org/10.1016/j.ijleo.2019.03.107
7. T.S. Saini, T.H. Tuan, T. Suzuki et al., Coherent Mid-IR Supercontinuum Generation using Tapered Chalcogenide Step-Index Optical Fiber: experiment and modelling. Sci. Rep. **10**, 2236 (2020). https://doi.org/10.1038/s41598-020-59288-6
8. T.S. Saini, T.H. Tuan, L. Xing, N.P.T. Hoa, T. Suzuki and Y Ohishi, Appl. Phys. Express. **11 102501** (2018). https://doi.org/10.7567/APEX.11.102501
9. M. Sharma, F. Amirkhan, S.K. Mishra, D. Sengupta, Y. Messaddeq, F. Blanchard, B. Ung, Transmission of Orbital Angular Momentum and Cylindrical Vector beams in a large-bandwidth annular core Photonic Crystal Fiber. Fibers. **8**(4), 22 (2020). https://doi.org/10.3390/fib8040022
10. M. Sharma, F. Amirkhan, S.K. Mishra, D. Sengupta, Y. Messaddeq, F. Blanchard, B. Ung, Annular core photonic crystal fiber for propagation of optical vortices, in Conference on Lasers and Electro-Optics, OSA Technical DigestOptica Publishing Group, (2020), paper STu3R.2
11. T. Sylvestre, E. Genier, A.N. Ghosh, P. Bowen, G. Genty, J. Troles, A. Mussot, A.C. Peacock, M. Klimczak, A.M. Heidt, Recent advances in supercontinuum generation in specialty optical fibers. J. Opt. Soc. Am. B **38**, F90 (2021). https://doi.org/10.1364/JOSAB.439330
12. J.M. Dudley, G. Helsinki, S. Coen, Supercontinuum generation in photonic crystal fiber Rev. Mod. Phys. **78**, 1135 (2006). https://doi.org/10.1103/RevModPhys.78.1135
13. M. Diouf, M.L.M. Mandeng, C. Tchawoua, M. Zghal, Numerical Investigation of Supercontinuum Generation through AsSe2/As2S5 Chalcogenide Photonic Crystal Fibres and Rib structures. J. Lightwave Technol. **37**(22), 5692–5698 (2019). https://doi.org/10.1109/JLT.2019.2934034
14. C. Heuteu, S.K. Boukar, L.M. Mandeng, C. Tchawoua, Supercontinuum generation of truncated airy pulses in a cubic-quintic AsSe2/As2S5 optical waveguide with rib-like structure. J. Opt. **23**(09), 095503 (2021). https://doi.org/10.1088/2040-8986/ac0ff4
15. C.V. Lanh, L.T.B. Tran, Low peak power broadband supercontinuum spectra generated in a square lattice toluene core photonic crystal fiber with different air hole diameters, Laser Phys. 33 095102, (2023).10.1088/1555–6611/ace24d
16. S. Kedenburg, A. Steinmann, R. Hegenbarth, T. Steinle, H. Giessen, ,nonlinear refractive indices of nonlinear liquids: wavelength dependence and influence of retarded response. Appl. Phys. B **117**, 803–816 (2014). https://doi.org/10.1007/s00340-014-5833-y
17. M. Chemnitz, M. Gebhardt, C. Gaida, F. Stutzki, J. Kobelke, J. Limpert, A. Tünnermann, Schmidt,Hybrid Soliton dynamics in liquid-core fibres. Nat. Commun. **8**, 42 (2017). https://doi.org/10.1038/s41467-017-00033-5
18. S. Kedenburg, T. Gissibl, T. Steinle, A. Steinmann, H. Giessen, Towards integration of a liquid-filled fiber capillary for supercontinuum generation in the 1.2–2.4 μm range,opt. Express. **23**, 8281–8289 (2015). https://doi.org/10.1364/OE.23.008281
19. J. Pniewski, T. Stefaniuk, H.L. Van, V.C. Long, L.C. Van, R. Kasz telanic, G. Stepniewski, A. Ramaniuk, M. Trippenbach, R Buczynski,dispersion engineering in nonlinear soft glass photonic crystal fibers infiltrated with liquids. Appl. Opt. **55**, 5033–5040 (2016). https://doi.org/10.1364/AO.55.005033
20. F. Dai, Y. Xu, X. Chen, Enhanced and broadened SRS spectra of toluene mixed with chloroform in liquid-core fiber, opt. Express. **17**, 19882–19886 (2009). https://doi.org/10.1364/OE.17.019882
21. L.C. Van, A. Anuszkiewicz, A. Ramaniuk, R Kasztelanic, K D Xuan, ,V C Long, M Trippenbach and R Buczynski, Supercontinuum generation in photonic crystal fibers with core filled with toluene. J. Opt. **19**, 125604 (2017). https://doi.org/10.1007/s11082-021-02820-3
22. M. Vieweg, T. Gissibl, S. Pricking, B.T. Kuhlmey, D.C. Wu, B.J. Egg leton, H. Giessen, Ultrafast nonlinear optofluid ics in selectively liquid-filled photonic crystal fibers. Opt. Express. **18**, 25232–25240 (2010). https://doi.org/10.1364/OE.18.025232

23. S.A. Adnan, A.W. Abdulwahhab, S.N. Ismail, ‚Fusion splicing: the penalty of increasing the collapse length of the air-holes in ESM-12B photonic crystal fibers opt. Appl. XLVI. **265** (2016). https://doi.org/10.5277/oa160211
24. X. Lin, H. Yan, Y. Ma, Z. Zhou, A construction method of the quasi-monolithic compact interferometer based on UV adhesives bonding. Rev. Sci. Instrum. **94**, 074501 (2023). https://doi.org/10.1063/5.0155637
25. P Fürjes, Controlled Focused Ion Beam Milling of Composite Solid State Nanopore arrays for Molecule Sensing micromachines **10** (2019) 774. https://doi.org/10.3390/mi10110774
26. V.T.M. Ngoc, C.V. Lanh, N.T.H. Phuong, H.V. Thuy, N.T. Hue, L.V. Hieu, Supercontinuum generation in a square-lattice photonic crystal fiber using carbon disulfide infiltration. Optik. **286**, 171049 (2023). https://doi.org/10.1016/j.ijleo.2023.171049
27. N.T. Thuy, H.T. Duc, C.V. Lanh, Comparison of supercontinuum spectral widths in CCl_4-core PCF with square and circular lattices in the claddings. Laser Phys. **33**, 055102 (2023). https://doi.org/10.1088/1555-6611/acc240
28. A.G.N. Chaitanya, T.S. Saini, A. Kumar, Ravindra Kumar Sinha, Ultra broadband mid-IR supercontinuum generation in $Ge_{11.5}As_{24}Se_{64.5}$ based chalcogenide graded-index photonic crystal fiber: design and analysis. Appl. Opt. **55**, 10138–10145 (2016). https://doi.org/10.1364/ao.55.010138
29. T.S. Saini, A. Baili, A. Kumar, R. Cherif, M. Zghal, Sinha,Design and analysis of equiangular spiral photonic crystal fiber for mid-infrared supercontinuum generation. J. Mod. Opt. **62**(19), 1570–1576 (2015). https://doi.org/10.1080/09500340.2015.1051600
30. A. Rjeb, H. Fathallah, S. Chebaane, M. Machhout, Design of novel circular lattice photonic crystal fiber suitable for transporting 48 OAM modes. Optoelectron. Lett. **17** (2021). https://doi.org/10.1007/s11801-021-0158-7
31. H.V. Le, V.L. Cao, H.T. Nguyen, A.M. Nguyen, R. Buczyński, R. Kasztelanic, Application of ethanol infiltration for ultra-flattened normal dispersion in fused silica photonic crystal fibers,Laser Phys. 28(11),(2018) art. 115106. https://doi.org/10.1088/1555-6611/aad93a
32. G.P. Agrawal, Nonlinear Fiber Optics, in: *Nonlinear Sci. Daw. 21st Century*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 195–211. https://doi.org/10.1007/3-540-46629-0_9
33. J.M. Dudley, J.R. Taylor (eds.), *Supercontinuum Generation in Optical Fibers* (Cambridge University Press, 2010). https://doi.org/10.1017/CBO9780511750465
34. J.M. Dudley, G. Genty, S. Coen, Supercontinuum generation in photonic crystal fiber. Rev. Mod. Phys. **78**, 1135–1184 (2006). https://doi.org/10.1103/RevModPhys.78.1135
35. G.P. Agrawal, *Applications of nonlinear fiber optics*, 3rd edn. (Academic Press, 2021), Chap. 1 - Fiber gratings, P. 1–55, ISBN 9780128170403. https://doi.org/10.1016/B978-0-12-817040-3.00007-9
36. P. Chauhan, A. Kumar, Y. Kalra, Numerical exploration of coherent supercontinuum generation in multicomponent $GeSe_2$ -As_2Se_3 -PbSe chalcogenide based photonic crystal fi ber. Opt. Fiber Technol. **54**, 102100 (2020). https://doi.org/10.1016/j.yofte.2019.102100
37. C.V. Lanh, H.V. Thuy, K. Borzycki, D.X. Khoa, T.Q. Vu, M. Trippenbach, R. Buczyn ´ski, and J Pniewski, Broadband supercontinuum generation with low peak power in controllable C7H8-core photonic crystal fibers of characteristic quantities,laser phys. **30** 035105 (2020). https://doi.org/10.1007/s12648-023-02830-9
38. H.V. Thuy et al., All-normal dispersion supercontinuum generation in photonic crystal fibers with large hollow cores infiltrated with toluene. Opt. Mater. Express. **8**(11), 3568 (2018). https://doi.org/10.1364/OME.8.003568
39. L.C. Van, H.V. Thuy, V.C. Long, K. Borzycki, K.D. Xuan, V.T. Quoc, M. Trippenbach, R Buczyn´ski and, J. Pniewski, Broadband supercontinuum generation with low peak power in controllable C_7H_8-core photonic crystal fibers of characteristic quantities. Opt. Eng. **60**, 116109 (2021). https://doi.org/10.1007/s12648-023-02830-9
40. L.V. Hieu, H.V. Thuy, N.T. Hue, C.L. Van, R. Buczyn ´ski, R. Kasztelanic, Supercontinuum generation in photonic crystal fibers infiltrated with tetrachloroethylene. Opt. Quant. Electron. **53**, 187 (2021). https://doi.org/10.1007/s11082-021-02820-3
41. Le H. Van, V.T. Hoang, Le T. Canh, Q.H. Dinh, H.T. Nguyen, N.V. Thi Minh, M. Klimczak, R. Buczynski, R. Kasztelanic, Silica-based photonic crystal fiber infiltrated with 1,2-dibromoethane for supercontinuum generation, Appl Opt. 2021 Aug 20, 60(24) 7268–7278. PMID: 34613015. https://doi.org/10.1364/AO.430843
42. L.C. Van, A. Anuszkiewicz, A. Ramaniuk, R. Kasztelanic, K.D. Xuan, V.C. Long, M. Trippenbach, R. Buczynski, Supercontinuum generation in photonic crystal fibres with core filled with toluene. J. Opt. **9**, 125604 (2017). https://doi.org/10.1088/2040-8986/aa96bc
43. D.S. Tomer, A. Kumar, A Graded Index Hybrid Photonic Crystal Fiber for Supercontinuum Generation using $AsSe_2$- As_2S_5, J Opt (2024). https://doi.org/10.1007/s12596-024-01954-3

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

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

AIP Conference Proceedings

RESEARCH ARTICLE | AUGUST 03 2024

Existence of topological phase in $RbCd_4As_3$: A first principles study

Nitika Sachdeva; Ramesh Kumar; Mukhtiyar Singh

AIP Conf. Proc. 3044, 050006 (2024)
https://doi.org/10.1063/5.0208631

View Online
Export Citation

13 August 2024 07:13:01

Nanotechnology & Materials Science
Optics & Photonics
Impedance Analysis
Scanning Probe Microscopy
Sensors
Failure Analysis & Semiconductors



Unlock the Full Spectrum.
From DC to 8.5 GHz.
Your Application. Measured.
Find out more
Zurich Instruments

AIP Publishing

Existence of Topological Phase in $RbCd_4As_3$: A First Principles Study

Nitika Sachdeva, Ramesh Kumar and Mukhtiyar Singh[a)]
Delhi Technological University, New Delhi, Delhi-110042, India
[a)]Corresponding author: mukhtiyarsingh@dtu.ac.in

Abstract. In this work, we report topological phase in the layered arsenide $RbCd_4As_3$ using *first-principles* calculations. We find that $RbCd_4As_3$, have topological band inversion at Γ point, which verifies the topological nature. Also, a small spin-orbit band gap of 0.1559 eV is observed at Γ point. Further verification of the topological nature is carried out by calculating the Z_2 topological invariants. The parity product of valance bands shows non-zero value of Z_2 topological invariant. Surface density of state (SDOS) study shows existence of Dirac cone near the Fermi level. Conducting surface states near the Fermi level, as observed in SDOS, confirm the topological non-trivial nature of $RbCd_4As_3$.

INTRODUCTION

Topological Materials (TMs) have a broad application in Spintronics[1], Quantum Computing[2], Chemical Catalysis[3], and Thermoelectric Materials[4]. The phenomena exhibited by these materials such as magnetoresistance[5], chiral anomaly [6], and Weyl Fermion quantum transport [7] are increasing their popularity amongst the research community. TMs can be divided into several categories such as Topological Semi-metals (TSM), Topological insulators (TI), Topological Crystalline Insulator (TCI) and many more. TSMs show gapless electronic states exhibiting crossing of valence and conduction bands near the Fermi level (E_F)[8]. They can be categorised into several categories such as Dirac Semi-metals, Weyl semimetals, nodal line semimetals, and more[9–12] based on the degeneracy, co-dimension, dispersion, and origin of the band crossing.

TIs have an insulating bulk and gapless boundary states (2D) and gapless surface states (3D)[13–15]. Non–trivial TIs have odd number of surface Dirac cones whereas even number of cones make them trivial. The non-triviality preserve Time Reversal Symmetry (TRS) and leads to some exotic properties. Another class of TIs in which surface states are protected by crystal symmetries instead of TRS named as topological crystalline insulators (TCIs). There are different types of TCIs which exhibit mirror symmetry as in SnTe family[16,17] and Rotational Symmetry[18] as observed in α-Bi_4Br_4[19]. ARPES experimental study and Density Functional Theory (DFT) analysis of $NaCd_4As_3$ show that it has TCI characteristics at room temperature which converts to TI upon lowering the temperature[20]. This happens due to the broken mirror symmetry leading to a topological phase transition from TCI to TI[20]. Besides, $NaCd_4As_3$ have been reported as a part of two Zintl Phases series namely, ACd_4Pn_3 and AZn_4Pn_3 (A = Na, K, Rb, Cs; Pn = As, P), in which 12 complex compounds have been synthesized from respective elements including layered arsenide like $NaCd_4As_3$, $NaZn_4As_3$, $NaCd_4As_3$, KCd_4As_3, and $RbCd_4As_3$[21,22]. There are structural similarities found in these layered arsenide as established by X-ray Diffraction (XRD) [21,22]. Moreover, the stable crystal structure of $RbCd_4As_3$ have been experimentally synthesized and its electronic structure calculations using the linear muffin-tin orbital (LMTO) method have been reported without considering the effect of spin orbit coupling (SOC)[21]. However, including spin-orbit interaction in band structure calculation can have drastic effects on the properties (topological) of a material as observed in Bi_2Se_3[23]. Even the strength of the spin-orbit interaction can be enhanced by various methods like chemical doping[24], applying magnetic field[25], or increasing hydro-static pressure[26].

Thus, the experimental observation of topological phase transition in $NaCd_4As_3$ and the similarity of structure with $RbCd_4As_3$, makes it interesting to study the electronic and topological properties of $RbCd_4As_3$[20–22] under the effect of SOC. Hence, we have probed the topological structure of the material $RbCd_4As_3$ using the *first-principle* calculations and found that $RbCd_4As_3$ shows topological phase with SOC. Further the topological nature is also verified by surface states visualised in (001) plane.

COMPUTATIONAL DETAILS

To evaluate the structural and electronic properties of $RbCd_4As_3$, we used the Density Functional Theory (DFT) calculations[27,28] as implemented in Vienna ab-initio simulation package (VASP)[29] with projector augmented wave formalism (PAW)[30]. To include the exchange-correlation function as in PAW formalism, we used the Generalized

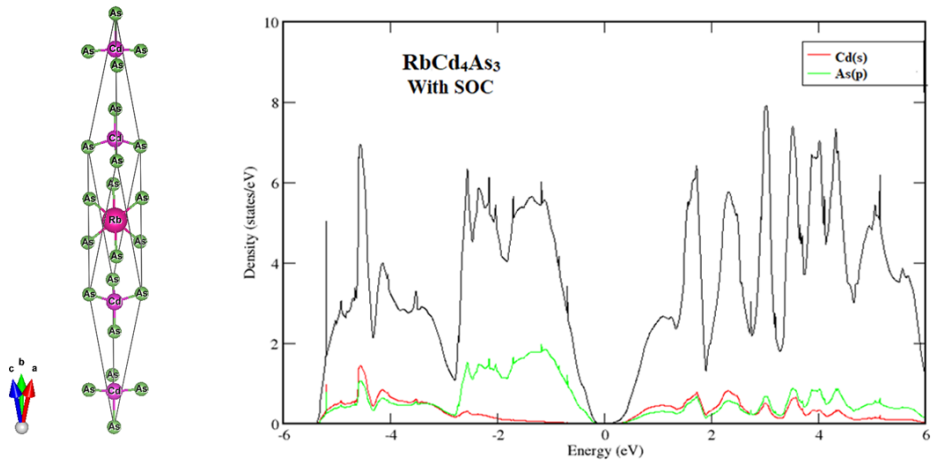
Smart Grid & Electric Vehicle
AIP Conf. Proc. 3044, 050006-1–050006-6; https://doi.org/10.1063/5.0208631
Published under an exclusive license by AIP Publishing. 978-0-7354-4972-5/$30.00

13 August 2024 07:13:01

gradient approximation (GGA) of Perdew-Burke Ernzehrof (PBE)[31,32] with spin-orbit coupling. The structural relaxation of $RbCd_4As_3$ is performed using the strict energy convergence criteria of 1×10^{-6} eV . To calculate structural properties and the electronic band structure, the kinetic energy cut-off of 520 eV is used. For a sampling of Brillion Zone (BZ) in rhombohedral structure, a 7x7x3 size Gamma centered k-mesh is used. Fermi level is broadened with a width of 0.001 eV using Gaussian smearing. All atoms in the unit cell are relaxed with the force of 0.001 eV/A per atom. Using maximally localized Wannier Functions[33], for tight-binding models to produce the band structure under effect of spin orbit coupling (SOC) constructed by Wannier90 code[34]. The resultant tight binding Hamiltonian is used to calculate the surfaces states of $RbCd_4As_3$ using Wannier tools[35].

RESULTS AND DISCUSSION

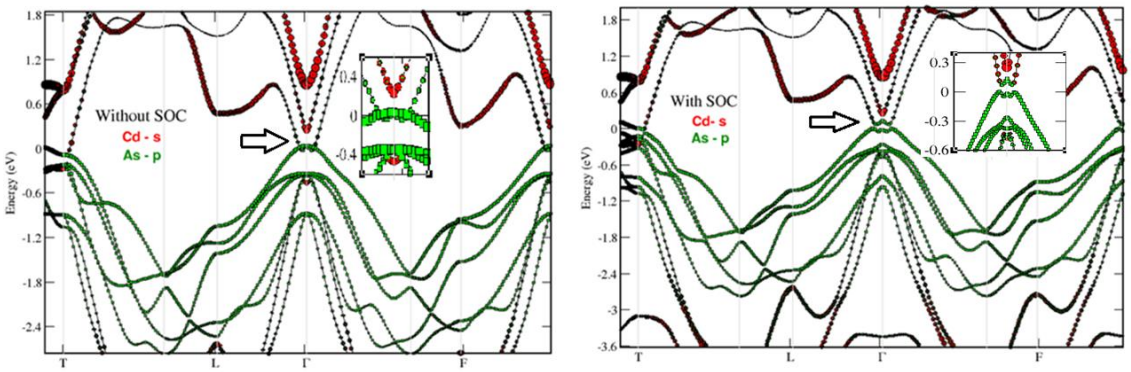
We investigate the electronic structure of the alkali metal layered arsenide $RbCd_4As_3$ for topological nature. $RbCd_4As_3$ is experimentally synthesized in space group $R\bar{3}m$ (166) in the rhombohedral phase. We took the lattice parameters a (4.4752 Å) and c (36.946 Å) from the reported experimental study[21]. For this study, we used the primitive structure to avoid band-folding with atomic positions as shown in Fig. 1(a). To understand our system, we have plotted the projected density of states (PDOS) or atom-resolved density of states for $RbCd_4As_3$ in Fig. 1(b). We observe that in the valence band the DOS majorly consists of *p-orbitals* of As near the fermi level whereas the *s-orbital* of Cd vanishes. On the contrary, in the conduction band both orbitals have similar contribution near the fermi level.
The band structure calculation with and without SOC is shown in Fig. 2. The band structure calculation is performed along high symmetry path Γ, T, H_2, H_0 L, S_0, S_2, and F out of which Γ, T, L, and F are the Time reversal Invariant Momenta (TRIM) points. Under the effect of SOC, $RbCd_4As_3$ shows topological band inversion at Γ high symmetry point. A small spin orbit band gap of 0.1559 eV is also observed at Γ point. The orbital contribution of *Cd(s-orbital)* and *As(p-orbital)* is observed mainly near E_F (Fig. 2) and can also be verified with the help of projected density of states (PDOS) as shown in Fig. 1(b). The spin orbit band gap can also be observed in PDOS of $RbCd_4As_3$ as shown in Fig. 1(b). From Fig. 2(b) it can be verified that at Γ point inverted contribution of *s-orbital* of Cd and *p-orbital* of As (Fig. 2(b) inset) show topological inversion and hence non-trivial nature can be seen. No other inversion is observed on high symmetry points which confirm the existence of single Dirac cone in $RbCd_4As_3$ and hence its topological nature.



(a) (b)

FIGURE 1. (a) Primitive Structure of $RbCd_4As_3$; (b) Projected Density of States (PDOS) of $RbCd_4As_3$ with consideration of contribution from *s-orbital* of Cd and *p-orbital* of As. The Fermi level is set to 0 eV.

13 August 2024 07:13:01



(a) (b)

FIGURE 2. Band structures of $RbCd_4As_3$ (a) without the SOC effect and (b) with SOC. Red spheres show the contribution of *s-orbital* of Cd atom and green spheres show the contribution of *p-orbital* of As atom to the bands near the Fermi level. The Fermi level is set to 0 eV.

To further investigate the topological character of $RbCd_4As_3$, we calculate the parity of the system at TRIM points. For 3D systems there are four Z_2-invariants namely (ν_0: ν_1, ν_2, ν_3) whose values are governed by the following equations as suggested by Kane and Mele[14,36,37]:

$$(-1)^{\nu_0} = \prod_i \delta_i \tag{1}$$

$$(-1)^{\nu_{i=1,2,3}} = \prod_{m_{j\neq i}\&m_{i=1}} \delta_{m_1 m_2 m_3} \tag{2}$$

Where, $\prod_i \delta_i$ denotes the parity products for filled valence bands at all TRIM points and the index value gives the class. These equations are applicable only for systems with inversion symmetry and Time Reversal Symmetry (TRS). There are eight TRIM points which can be calculated using the equation (3)[14,36,37].

$$G_{i=m_1 m_2 m_3} = \frac{m_1 a_1 + m_2 a_2 + m_3 a_3}{2} \tag{3}$$

Here, m_j= 0, 1 and a_1, a_2, a_3 are the reciprocal primitive lattice vectors. The nature or strength of the topological insulator can be determined by the value of ν_0 calculated using equation (1). For ν_0=1, the system is a strong topological insulator and for ν_0 = 0, it is trivial or weak in nature. The values of ν_1, ν_2, ν_3, given by equation (2) are used to distinguish between a weak TI and trivial insulator. The parity for valence bands at all TRIM points for band structure calculations with spin orbit coupling is shown in table 1. It can be observed that the product of all TRIM point is (-). The value of the Z_2 index ν_0 as calculated from equation (1) comes out to be 1. This verifies that $RbCd_4As_3$ have strong topological nature.

13 August 2024 07:13:01

TABLE 1. Parities of all the occupied bands at TRIM points of first BZ of $RbCd_4As_3$ in primitive structure with spin-orbit coupling.

TRIM Points	Parity
Γ	-
3L	-
3F	-
T	+
Total	-

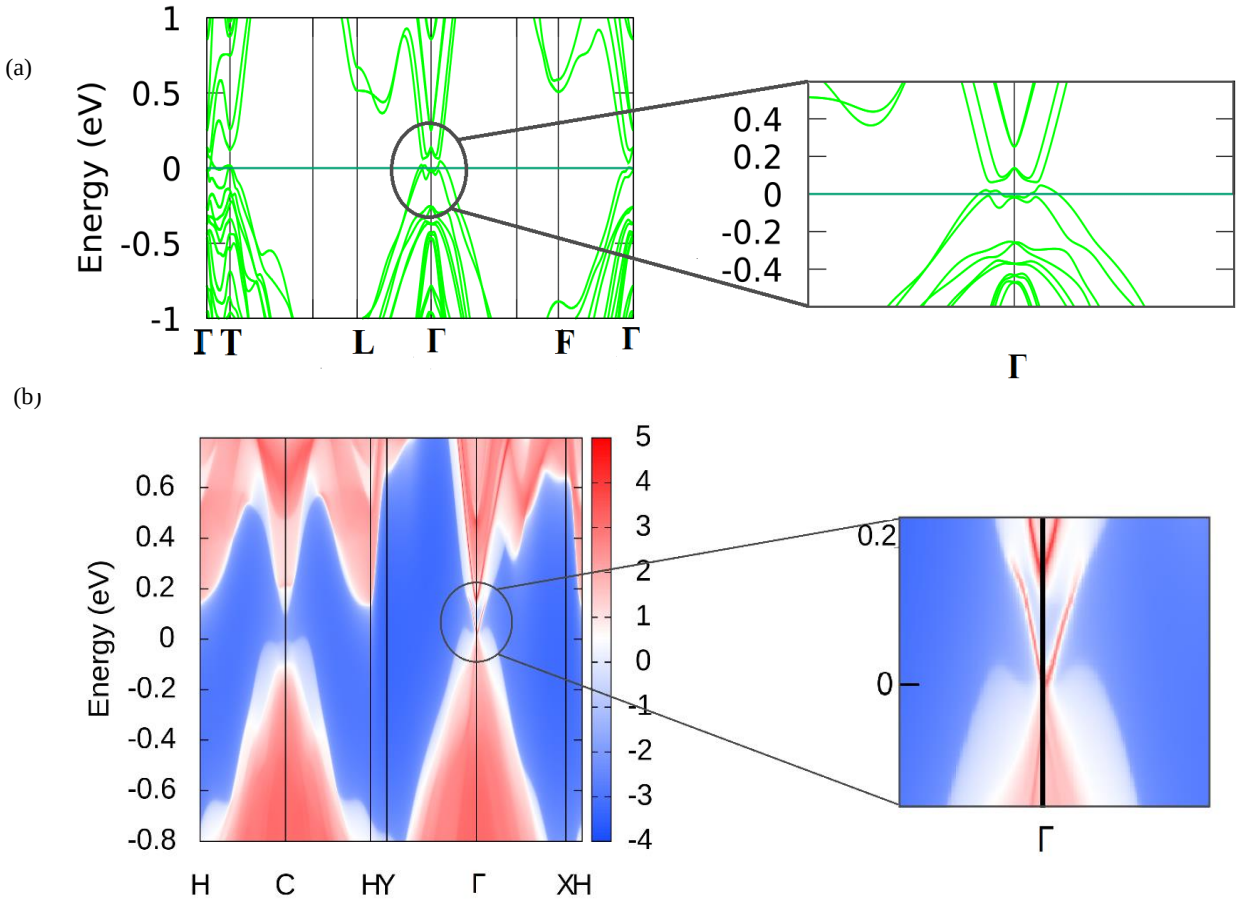


FIGURE 3. (a) Band Structure of $RbCd_4As_3$ in wannier space with band gap induced at Γ due to spin-orbit coupling. Magnified image of the band gap induced at Γ point is shown. (b) Surface Density of States for (001) surface of $RbCd_4As_3$ calculated with spin-orbit coupling. Formation of a single Dirac cone can be observed at Γ point.

Additionally, we calculate maximally localized wannier function (MLWF) from the Bloch energy bands of $RbCd_4As_3$ shown in Fig. 2(b). Since, the MLWFs are independent of the basis set used to obtain the bloch states[34], we can use them to calculate band structure in wannier space and verify the band structure obtained from Bloch states. The calculated band structure in wannier space can be observed in Fig. 3(a). The wannier band structure of $RbCd_4As_3$ (Fig.3(a)) exhibits a spin orbit band gap at *Γ-point* and perfectly coincides with the Bloch band structure, thus verifying it (Fig.2(b)). Further, we used MLWFs to perform surface state calculations for (001) surface of $RbCd_4As_3$ using the green's function approach. From our *first-principle* calculations, we derive a tight-binding model Hamiltonian to calculate the surface states. The resulting Surface Density of States (SDOS) forming a single Dirac cone at *Γ-point* is

13 August 2024 07:13:01

shown in Fig.3(b). This is in agreement with the spin orbit band gap observed at Γ point in Fig. 2(b) and Fig. 3(a). The presence of odd number of Dirac cones on the (001) surface of $RbCd_4As_3$ confirms the presence of conducting surface state on (001) plane of $RbCd_4As_3$.The existence of topological band inversion and conducting surface state along with a surface Dirac cone near Fermi level shows that $RbCd_4As_3$ has non-trivial topological nature. A topological insulator can generally be good thermoelectric material. $RbCd_4As_3$ is a Zintl compound and on preliminary transport calculations Zintl Compounds are suggested to be good thermoelectric materials[21]. Besides, $RbZn_4P_3$ is also a member of this family and is proven to be a thermoelectric material[4]. Another member of this family is $NaCd_4As_3$, which is shown to have a favorable combination of electrical conductivity and Seebeck coefficient[21] along with existence of a topological phase[20], indicating a strong possibility of possessing thermoelectric properties. Hence, it would be worthwhile to investigate $RbCd_4As_3$ and other materials of this family for having thermoelectric properties for which our study can be used.

CONCLUSION

The topological properties of the alkali metal layered arsenide $RbCd_4As_3$ have been studied based on the *first-principles* density functional theory calculations. The stable rhombohedral structure with space group 166 is considered to study the band structure of $RbCd_4As_3$. We have observed a single band inversion at *Γ-point* under the consideration of spin-orbit coupling effect and a small spin-orbit band gap of 0.1559 eV. The orbital contribution during the inversion can be attributed to *s-orbital* of Cd and *p-orbital* of As. We have also calculated the Z_2 topological invariants to verify the topological phase. The product of parities of all occupied bands have shown that the value of Z_2 index $\nu_0 = 1$. This have suggested that $RbCd_4As_3$ is topologically non-trivial. The Surface density of state have also verified the existence of single Dirac cone and conducting surface state in $RbCd_4As_3$. The band inversion and surface state study of $RbCd_4As_3$ have confirmed its topologically non-trivial nature. We hope that the future experimental studies will verify its topological nature as the surface state can be observed with the help of Angle Resolved Photoemission Spectroscopy (ARPES).

13 August 2024 07:13:01

ACKNOWLEDGEMENT

We acknowledge National Supercomputing Mission (NSM) for providing computing resources of 'PARAM SMRITI' at NABI, Mohali, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India.

REFERENCES

1. S.A. Yang, SPIN **6**, 1640003 (2016).
2. C. Nayak, S.H. Simon, A. Stern, M. Freedman, and S. das Sarma, Rev Mod Phys **80**, 1083-1159 (2008).
3. C.R. Rajamathi, U. Gupta, N. Kumar, H. Yang, Y. Sun, V. Süß, C. Shekhar, M. Schmidt, H. Blumtritt, P. Werner, B. Yan, S. Parkin, C. Felser, and C.N.R. Rao, Advanced Materials **29**, 1606202 (2017).
4. Sangeeta, R. Kumar, and M. Singh, J Mater Sci **57**, 10691–10701 (2022).
5. F.F. Tafti, Q.D. Gibson, S.K. Kushwaha, N. Haldolaarachchige, and R.J. Cava, Nat Phys **12**, 272–277 (2016).
6. X. Huang, L. Zhao, Y. Long, P. Wang, D. Chen, Z. Yang, H. Liang, M. Xue, H. Weng, Z. Fang, X. Dai, and G. Chen, Phys Rev X **5**, 031023 (2015).
7. A.A. Zyuzin and A.A. Burkov, Phys Rev B Condens Matter Mater Phys **86**, 115133 (2012).
8. J. Li, Z. Zhang, C. Wang, H. Huang, B.L. Gu, and W. Duan, J Appl Phys **128**, 191101 (2020).
9. H. Gao, J.W.F. Venderbos, Y. Kim, and A.M. Rappe, Annu Rev Mater Res **49**, 153-183 (2019).
10. S.M. Young, S. Zaheer, J.C.Y. Teo, C.L. Kane, E.J. Mele, and A.M. Rappe, Phys Rev Lett **108**, 140405 (2012).
11. H. Weng, C. Fang, Z. Fang, B. Andrei Bernevig, and X. Dai, Phys Rev X **5**, 011029 (2015).
12. A.A. Burkov, M.D. Hook, and L. Balents, Phys Rev B Condens Matter Mater Phys **84**, 235126 (2011).
13. B.A. Bernevig, T.L. Hughes, and S.C. Zhang, Science (1979) **314**, 1757-1761 (2006).
14. C.L. Kane and E.J. Mele, Phys Rev Lett **95**, 146802 (2005).
15. M.Z. Hasan and C.L. Kane, Rev Mod Phys **82**, 3045-3067 (2010).

16. T.H. Hsieh, H. Lin, J. Liu, W. Duan, A. Bansil, and L. Fu, Nat Commun **3**, 982 (2012).
17. Y. Tanaka, Z. Ren, T. Sato, K. Nakayama, S. Souma, T. Takahashi, K. Segawa, and Y. Ando, Nat Phys **8**, 800–803 (2012).
18. C. Fang and L. Fu, Sci Adv **5**, eaat2374 (2019).
19. C.-H. Hsu, X. Zhou, Q. Ma, N. Gedik, A. Bansil, V.M. Pereira, H. Lin, L. Fu, S.-Y. Xu, and T.-R. Chang, 2D Mater. **6**, 031004 (2019).
20. Y.Y. Wang, C. Zhong, M. Li, W.L. Zhu, W.J. Hou, W.H. Song, Q.X. Dong, Y.F. Huang, S. Zhang, Z.A. Ren, S. Wang, and G.F. Chen, Phys Rev B **102**, 115122 (2020).
21. H. He, C. Tyson, and S. Bobev, ChemInform **42**, 8375–8383 (2011).
22. C. Grotz, M. Baumgartner, K.M. Freitag, F. Baumer, and T. Nilges, Inorg Chem **55**, 7764–7776 (2016).
23. D. West, Y.Y. Sun, H. Wang, J. Bang, and S.B. Zhang, Phys Rev B **86**, 121201(R) (2012).
24. A.B. Maghirang, Z.Q. Huang, R.A.B. Villaos, C.H. Hsu, L.Y. Feng, E. Florido, H. Lin, A. Bansil, and F.C. Chuang, NPJ 2D Mater Appl **3**, 35 (2019).
25. J. Macy, D. Ratkovski, P.P. Balakrishnan, M. Strungaru, Y.C. Chiu, A. Flessa Savvidou, A. Moon, W. Zheng, A. Weiland, G.T. McCandless, J.Y. Chan, G.S. Kumar, M. Shatruk, A.J. Grutter, J.A. Borchers, W.D. Ratcliff, E.S. Choi, E.J.G. Santos, and L. Balicas, Appl Phys Rev **8**, 041401 (2021).
26. M. Singh, R. Kumar, and R.K. Bibiyan, Eur Phys J Plus **137**, 633 (2022).
27. A.K. Rajagopal and J. Callaway, Phys Rev B, **7**, 1912-1919 (1973).
28. W. Kohn and L.J. Sham, Physical Review **140**, A1133-A1138 (1965).
29. G. Kresse and J. Furthmüller, Phys Rev B Condens Matter Mater Phys **54**, 11169-11186 (1996).
30. G. Kresse and D. Joubert, Phys Rev B Condens Matter Mater Phys **59**, 1758-1775 (1999).
31. J.P. Perdew, K. Burke, and M. Ernzerhof, Phys Rev Lett **77**, 3865-3868 (1996).
32. J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, and C. Fiolhais, Phys Rev B **46**, 6671-6687 (1992).
33. N. Marzari and D. Vanderbilt, Phys Rev B Condens Matter Mater Phys **56**, 12847-12865 (1997).
34. A.A. Mostofi, J.R. Yates, G. Pizzi, Y.S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput Phys Commun **185**, 2309-2310 (2014).
35. Q.S. Wu, S.N. Zhang, H.F. Song, M. Troyer, and A.A. Soluyanov, Comput Phys Commun **224**, 405-416 (2018).
36. L. Fu and C.L. Kane, Phys Rev B Condens Matter Mater Phys **76**, 045302 (2007).
37. L. Fu, C.L. Kane, and E.J. Mele, Phys Rev Lett **98**, 106803 (2007).

13 August 2024 07:13:01

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/382728537

Experimental Research on Green Concrete

Article · July 2024

DOI: 10.37082/IJIRMPS.v12.i4.230874

CITATIONS
0

READS
40

2 authors, including:



Neda Rasoly
Delhi Technological University
1 PUBLICATION **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Neda Rasoly on 31 July 2024.

The user has requested enhancement of the downloaded file.

Experimental Research on Green Concrete

Neda Rasoly[1], Prof. Kongan Aryan[2]

[1]Student, Civil Engineering Department, Delhi Technological University, Delhi.
[2]Lecturer, Civil Engineering Department, Delhi Technological University, Delhi.

Abstract:
Concrete is a predominant material in global construction due to its versatility and essential role in various infrastructure projects. However, its production is responsible for approximately 8 to 10 % of global CO2 emissions, primarily attributed to the manufacturing of cement. Consequently, there is a critical need to reduce concrete usage and seek eco-friendly alternatives to conventional cement. Green concrete refers to concrete produced by incorporating waste materials such as fly ash, slag, by-products from power plants, recycled concrete aggregate, and waste glass, this approach aims to reduce carbon dioxide emissions associated with traditional concrete production.
This study presents experimental work conducted on both conventional and green concrete samples of M40 mix design. The green concrete samples incorporated fly ash as a partial replacement for cement at 20%, 25%, and 30% levels, and used recycled concrete aggregate (RCA) to replace 50% of both fine and coarse aggregates. The primary objectives were to design a concrete mix that maintains properties comparable to conventional concrete, reduces environmental impact by substituting waste materials for cement, and achieves cost efficiency.

Keywords: Green concrete, Fly ash, Recycle Concrete Aggregate.

INTRODUCTION

Concrete is the second most widely used material in the world after water, contributing approximately 8 to 10 percent of global CO2 emissions [1]. This significant environmental impact underscores the necessity of reducing cement consumption and exploring more sustainable alternatives. Green concrete, introduced in Denmark in 1998, represents a pivotal development in concrete technology. Despite its name, green concrete is not related to colour rather, it is designed to address environmental concerns by minimizing pollution and reducing the carbon footprint of concrete production.

Green concrete is a highly advantageous material for construction due to its superior thermal and fire resistance properties. It also enhances the damping resistance of structural elements. When compared to conventional concrete, green concrete can achieve higher compressive strength with an appropriate water-to-cement ratio while maintaining similar flexural strength. Furthermore, green concrete generally demands less maintenance and repair over its lifespan. In this research, we explored the use of fly ash as a partial replacement for cement at levels of 0%, 20%, 25%, and 30%, and replaced 50% of both fine and coarse aggregates with recycled concrete aggregate, through various parameters such as slump, compressive strength, flexural strength, and cost-effectiveness.

LITERATURE REVIEW

Concrete is increasingly utilized in the construction of buildings and other infrastructure worldwide. However, cement, a critical component of concrete, poses significant environmental challenges due to its production processes. To address this issue and mitigate environmental impact, extensive research has focused on improving concrete's constituent materials without compromising its performance. One promising solution is the development of green concrete, which incorporates eco-friendly materials by partially substituting traditional cement with waste-derived resources. This approach offers a viable strategy for enhancing the sustainability of concrete while maintaining its essential properties. Gauri Aglave et al. (2020) investigated the cost-effectiveness and performance characteristics of conventional versus green concrete. Their study involved partially replacing cement with fly ash and substituting fine aggregate with a blend of 50% glass powder and 50% crushed debris. The findings indicated that the properties of green concrete were comparable

to those of conventional concrete. Notably, green concrete demonstrated a lower economic cost, highlighting its potential as a more sustainable and cost-effective alternative in construction applications. Building on this, Neeraj Agarwal et al. (2018) examined the compressive strength and economic implications of partially replacing cement with glass powder and fly ash. Their research demonstrated that cement can be effectively substituted with these materials up to a replacement level of 45% without compromising concrete quality. However, they noted that exceeding this threshold could negatively impact the performance of the concrete, thus establishing a practical limit for such substitutions. Further expanding on alternative materials, Ronak Malpani et al. (2014) explored the effects of replacing fine aggregate with marble sludge powder and quarry dust. Their study found that using substantial amounts of marble powder and quarry dust resulted in lower slump values, which in turn decreased the workability of the concrete. The authors suggested that the use of admixtures could mitigate this issue. Despite the reduced workability, the study highlighted the cost-saving benefits of incorporating quarry rock dust and marble sludge powder as fine aggregates in concrete production, reinforcing the viability of these materials as sustainable alternatives.

METHODOLOGY

The concrete mix design **Fig. 1** for both conventional and green concrete was performed in accordance with the specifications outlined in IS: 10262-2009. The design focused on achieving an M40 grade of concrete. The mix proportions for both types of concrete are detailed in **Table_1.** For the experimental investigation, concrete samples were prepared as follows:

Conventional Concrete: Six cubes of dimensions 150 mm× 150 mm × 150 mm and six beams of dimensions 100 mm × 100 mm × 500 mm were cast for conventional concrete.

Green Concrete: Six cubes of dimensions 150 mm × 150 mm × 150 mm were prepared with cement replaced by fly ash at 20%, 25%, and 30% levels. Additionally, 50% of both fine and coarse aggregates were replaced with recycled concrete aggregate, and six beams of dimensions 100 mm × 100 mm × 500 mm were cast with 20% of the cement replaced by fly ash and 50% of both fine and coarse aggregates replaced with recycled concrete aggregate.



Fig 1: Process of maxing and casting of specimens

Table_1: Mix design portion

Concrete Type	Material (kg)						
	Cement	Fly ash	Water	Coarse (Fresh)	Fine (Fresh)	Coarse (RCA)	Fine (RCA)
0%	400	0	152	1158	718	0	0
20%	320	80	122	596	346	596	346
25%	300	100	124	600	349	600	349
30%	280	120	106	604	351	604	351

RESULT

The workability of the concrete mixtures was evaluated prior to hardening. The results of these assessments, including slump measurements for all concrete samples, are presented in **Fig 2**. This table provides a comparative analysis of workability between conventional and green concrete mixes. Following the curing

period, compressive and flexural strength tests were conducted to evaluate the structural performance of the concrete samples. The results of these tests provide insight into the mechanical properties and durability of both conventional and green concrete formulations.

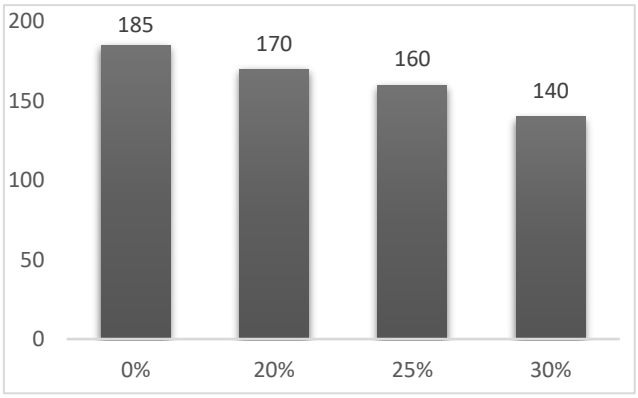


Fig 2: Slump result

Table_2: Compressive strength

Sample	Compressive Strength in 7 Days (N/mm^2)			
	0%	**20%**	**25%**	**30%**
N1	38.49	39.30	38.49	33.40
N2	41.29	41.40	41.29	35.20
N3	39.00	41.10	39.00	35.50
Average	**39.60**	**40.69**	**39.60**	**34.70**

Sample	Compressive Strength in 28 Days (N/mm^2)			
	0%	**20%**	**25%**	**30%**
N4	54.80	58.00	54.80	48.89
N5	56.30	50.10	56.30	50.18
N6	59.60	55.80	59.60	48.49
Average	**56.90**	**54.70**	**56.90**	**49.19**

Table_3: Flexure strength of concrete beams.

Samples	Conventional Concrete		Green Concrete	
	7 Days (N/mm$^{2)}$	**28 Days (N/mm$^{2)}$**	**7 Days (N/mm$^{2)}$**	**28 Days (N/mm$^{2)}$**
N1	3.40	-	3.40	-
N2	4.10	-	4.10	-
N3	3.80	-	3.80	-
N4	-	5.30	-	5.30
N5	-	4.80	-	4.80
N6	-	5.10	-	5.10
Average	**3.80**	**5.10**	**3.80**	**5.10**

Table_4: Cost compression of conventional and green concrete.

Material	Conventional Concrete		Green Concrete	
	Amount (Kg)	Price (INR)	Amount (Kg)	Price (INR)
Cement	400	2600	320	2080
Coarse Aggregate (Fresh)	1158	46321	596	23840

Fine Aggregate (Fresh)	718	32310	346	15570
RCA Aggregate	0	0	942	376
Fly ash	0	0	80	104
Total	**2432**	**81230**	**2284**	**41970**

CONCLUSION

The experimental results indicate that green concrete, which incorporates environmentally friendly materials, presents a viable option for reducing the ecological impact of construction activities.

The compressive strength of green concrete was found to be comparable to that of conventional concrete, with no significant differences observed. Additionally, as illustrated in **Table_3** green concrete demonstrated superior flexural strength compared to its conventional counterpart. Notably, green concrete also offers considerable economic advantages as shown in **Table_4**, with a cost reduction of approximately 52% per cubic meter compared to conventional concrete.

Moreover, green concrete utilizes waste materials such as glass powder, fly ash, and recycled concrete aggregate, which are environmentally benign and do not require additional industrial processes for their production. The use of these materials contributes to reducing carbon dioxide emissions, as each kilogram of cement produced generates approximately 900 kilograms of CO2, representing 8 to 10% of global pollution. By substituting traditional cement with green alternatives, we can significantly reduce CO2 emissions, with an estimated savings of 1 kilogram of CO2 per kilogram of cement replaced.

In conclusion, green concrete not only addresses environmental concerns by utilizing waste materials but also offers economic and performance benefits, making it a promising alternative to conventional concrete.

REFERENCES:

1. D. S. Vijayan, Dineshkumar, S. Arvindan, Thattil Shreelakshmi Janarthanan, Evaluation of ferrock: A greener substitute to cement, Materials Today: Proceedings 22 (2020) 781–787.
2. Sudhakar, Pandey Sangeeta, Green Concrete and its Scope in India, Civil Engineering Department, Darbhanga College of Engineering, Mabbi, Darbhanga, India, 2021.
3. Sarah Khan, Naveen Maheshwari, Gauri Aglave, Rishabh Arora, Experimental design of green concrete and assessing its suitability as a sustainable building material, Materials Today: Proceedings 26 (2020) 1126-1130.
4. Neeraj Agarwal, Nikhil Garg, A Research on Green Concrete, Civil Engineering Department, Krishna Institute of Engineering and Technology, Ghaziabad, India, 2018.
5. Hasmukh D. Makavana, Prof. Kishor B. Vaghela, Dr. J. R. Pitroda, Experimental Study Green Concrete with Bacterial Inclusion, International Research Journal of Engineering and Technology (IRJET), 2020.
6. Faraz Ahmad, Muhsmmad Usama Memon, Mahboob Oad, Bashir Ahmad Memon, Sheeraz Ahmaed Chandio, Effect of Fly Ash on the Compressive Strength of Green Concrete, Engineering Technology and Applied Science Research.
7. S.Panneer Selvam, V.Anupreethi, C.Anusiya, S.Babyshalini, S.K.Keerthika, Experimental Investigations on concrete by using replacement of granite sludge powder, Volume 6, Issue 4 (2019) 1-4 ISSN 2347 - 3258 International Journal of Advance Research and Innovation.
8. S.Paneerselvam, S.A.Iswariya, V.Karthiga, S.Keerthana, R.Eniya, Experimental investigation on fiber reinforced concrete using waste bottle cap fiber, Volume 6, Issue 4 (2019) 1-4 ISSN 2347 - 3258 International Journal of Advance Research and Innovation.
9. K.M. Liew, A.O. Sojobi, L.W. Zhang, Green concrete: Prospects and challenges, Construction and Building Materials 156 (2017) 1063–1095. [10] Swarna Swetha Kolaventi , TP Tezeswi, MVN Siva Kumar, An assessment of construction waste management in India: A statistical approach, Waste Management & Research journal, 2019.
10. Chirag Garg, Aakash Jain, Green Concrete: Efficient & EcoFriendly Construction Materials, Department Of Civil Engineering, Bits-Pilani, Hyderabad Campus, Andhra Pradesh, India, 2014.
11. T. Horiguchi, T. Kikuchi, Y. Nakagawa, K. Shimura, Physical properties of CLSM using high volumes of incineration ash from sewage sludge, Special Publication 242 (2007).

View publication stats

Molecular Biotechnology
https://doi.org/10.1007/s12033-024-01258-8

ORIGINAL PAPER

Exploring Natural Compounds as Potential CDK4 Inhibitors for Therapeutic Intervention in Neurodegenerative Diseases through Computational Analysis

Neetu Rani[1] · **Pravir Kumar**[1,2]

Received: 27 March 2024 / Accepted: 3 July 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

CDK4 is a member of the serine-threonine kinase family, which has been found to be overexpressed in a plethora of studies related to neurodegenerative diseases. CDK4 is one of the most validated therapeutic targets for neurodegenerative diseases. Hence, the discovery of potent inhibitors of CDK4 is a promising candidate in the drug discovery field. Firstly, the reference drug Palbociclib was identified from the available literature as a potential candidate against target CDK4. In the present study, the Collection of Open Natural Products (COCONUT) database was accessed for determining potential CDK4 inhibitors using computational approaches based on the Tanimoto algorithm for similarity with the target drug, i.e., Palbociclib. The potential candidates were analyzed using SWISSADME, and the best candidates were filtered based on Lipinski's Rule of 5, Brenk, blood–brain barrier permeability, and Pains parameter. Further, the molecular docking protocol was accessed for the filtered compounds to anticipate the CDK4-ligand binding score, which was validated by the fastDRH web-based server. Based on the best docking score so obtained, the best four natural compounds were chosen for further molecular dynamic simulation to assess their stability with CDK4. In this study, two natural products, with COCONUT Database compound ID—CNP0396493 and CNP0070947, have been identified as the most suitable candidates for neuroprotection.

Graphical Abstract

Overexpression of CDK4 in NDDs has been considered as prominent target for therapeutics. Possible CDK4 inhibitors have been explored using computational techniques and the COCONUT database, which keeps Palbociclib as the reference drug. Using several criteria, such as molecular docking and Lipinski's Rule of 5, 110 natural compounds have been obtained. In addition to molecular docking, this study utilized Molecular Mechanics Generalised Born Surface Area (MM-GBSA) analysis to further filter and assess the binding affinity and stability of the top 4 natural compounds. Further evaluation was carried

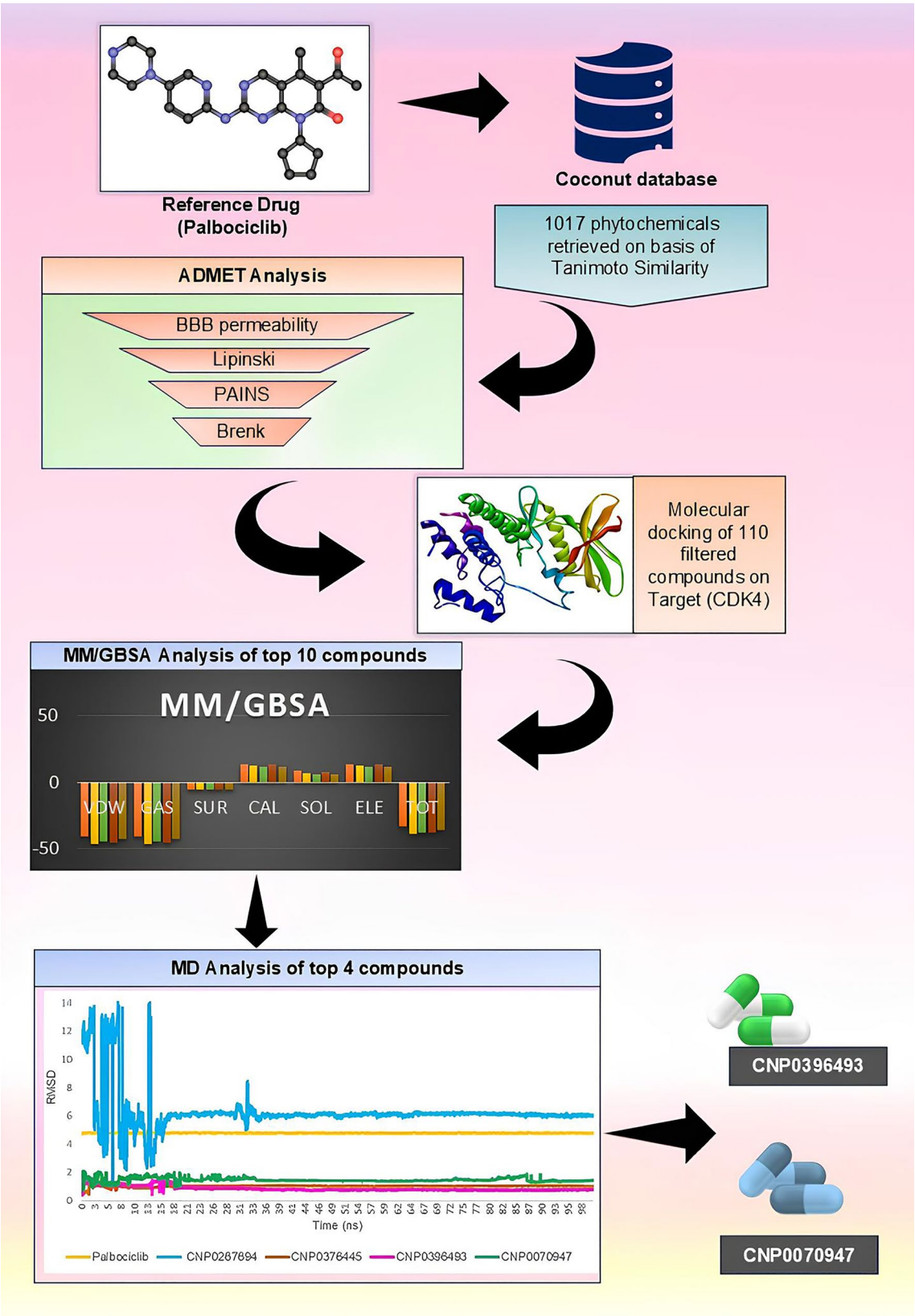
✉ Pravir Kumar
pravirkumar@dtu.ac.in; kpravir@gmail.com

[1] Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University (Formerly DCE), Shahbad Daulatpur, Bawana Road, Delhi 110042, India

[2] Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University, Delhi 110042, India

Published online: 29 August 2024

out by molecular dynamic simulation, demonstrating CNP0396493 and CNP0070947 as the most promising candidates for neuroprotection.



Keywords CDK4 · Virtual screening · Coconut database · MM/GBSA · Binding affinity

Introduction

Neurodegenerative diseases (NDDs) are deteriorating conditions resulting in impairment of cognitive or motor abilities, affecting a substantial percentage of the global population [1]. Alzheimer's disease is among the top 6 fatal diseases in the world, and Parkinson's disease is the second most common NDD in the world, affecting up to 2% of individuals above age 65 [2]. NDDs, namely Alzheimer's disease (AD) [3], Parkinson's disease (PD) [4], Huntington's disease (HD), and spinocerebellar ataxia [3], are manifested by progressive neuronal loss in the central nervous system (CNS) or peripheral nervous system (PNS) [2, 4]. Many studies suggest that oxidative stress, neuroinflammation, ubiquitin–proteasome system dysfunction, aberrant cell cycle, and apoptosis are hallmark manifestations of NDDs [5–8]. In literature studies, elevated levels of cyclin-dependent kinases (CDK4) have been identified which have a deep connection with NDDs and mechanistic correlation between cell cycle regulators and NDDs required to be explored to understand the complex nature of these diseases to avail information about putative therapeutic targets. [9, 10].

The cell cycle is a tightly regulated process responsible for the growth of the cell, controlled by a group of regulators, namely cyclins, CDKs, and CDK inhibitors (CKIs), among which CDK4, an important cell cycle regulator that involves accurate coordination for maintaining the integrity and functionality of cell, chosen as the target for this study [11, 12]. CDK4 is a member of the conserved family of serine-threonine kinase directed by proline, which is known for its partnership with D-type cyclins, thereby forming complexes governing the G1 phase of the cell cycle [4, 13]. Structurally, CDK4 comprises an N-terminal regulatory region and a C-terminal catalytic domain featuring an ATP binding site, the latter being particularly pertinent to its enzymatic activity [14]. CDK4 has the typical bilobed structure as seen in other kinases, comprising of a 5-stranded β-sheet in the N-terminal domain (with 1–96 residues), in which helix-αC sequence PISTVRE is present for cyclin binding and activating T-loop [15]. The active site of CDK4 within its catalytic domain plays a fundamental role in phosphorylating key substrates involved in cell cycle progression [15]. In CDK4 catalytic activity regulation, conformational changes, specifically phosphorylation, occur in the T-loop [17], which is a crucial step in the progression of the cell cycle through the G1 phase and activation of CDK4, mediated by Cyclin D [16]. CDKs and CKIs, along with several cyclin proteins, regulate the progression of the cell cycle [17, 18]. Cyclin D and CDK4 coordinate the transition from the G1 to the S phase. Under normal physiological environment, in the Rb-E2F pathway, the transcription factor E2F is released due to phosphorylation of retinoblastoma (Rb) protein by CDK4. The synthesis of proteins required for DNA replication during the S phase is then triggered by the translocation of E2F into the nucleus [19].

In the present study, the reference drug Palbociclib, a well-known CDK4 inhibitor, has been selected, involving its interaction with the kinase domain at the ATP binding site of CDK4, reducing its kinase activity [20–23]. Residues, including Gln103, Gly101, Asn142, Glu144, and Phe98, are present in this pocket and are essential for regular kinase activity of CDK4 [16]. These residues play a role in the phosphorylation process by binding ATP. Palbociclib potentially interacts with these residues when it binds to the ATP binding pocket, surpassing ATP for binding [24–26]. Palbociclib binds to this pocket to prevent ATP from binding, interfering with CDK4 regular kinase activity [20, 27, 28].

Our study leverages natural compound's huge potential to go beyond conventional medicinal product development. The COCONUT Database is a valuable tool with abundant chemical data from many natural sources. With the assistance of Tanimoto's likeness, we could filter the natural molecules that have structural similarities with Palbociclib, thereby expanding the list of possible candidates with therapeutic implications. Phytochemicals (natural compounds) act as bioactive molecules with potential therapeutic usage, making them valuable in drug development [29, 30]. These plant-derived naturally occurring substances have proven to be an invaluable resource, playing an important role in creating cutting-edge medicinal treatments [31]. The varying chemical structures of phytochemicals facilitate innumerable biological functions, such as anti-inflammatory, anticancer, anti-microbial, and neuroprotective characteristics [32, 33]. A stringent screening process was used in the drug development to identify the putative candidate with alluring pharmacological properties. Several computational and bioinformatics approaches are widely implicated in studying the capitative approach of natural compounds to modify the target associated with NDDs; the complex mechanism involved in their interactions has indicated a positive ray of hope in the field of drug development.

This study investigates the potential of natural substances to inhibit CDK4 using computational approaches. The Tanimoto approach was utilized to identify possible inhibitors comparable to Palbociclib from the COCONUT database [34]. The potential candidates underwent

evaluation using SWISSADME to assess their drug-likeness, blood–brain barrier (BBB) permeability, and compliance with Lipinski's rule [35]. The CDK4 binding scores were predicted using molecular docking, which the fastDRH web-based tool confirmed, resulting in four top possibilities. In this study, two natural products were finally identified as the most suitable candidates for neuroprotection. The current findings may direct interest in natural compounds and may be used to further optimize them into successful dual-targeted drugs to control and halt neuronal damage. Such a comprehensive exploration of established pharmacological agents with nature-inspired compounds holds the promise of reshaping therapeutic strategies for NDDs.

Methodology

Structure Retrieval

In this study, the structure of CDK4 was retrieved from the protein data bank (PDB) database, maintained by RCSB http://www.rcsb.org [15, 36, 37]. The X-ray crystal structure of CDK4 (PDB ID: 2W96) with 2.30 Å resolution was visualized and prepared by web-based server discovery studio https://discover.3ds.com/discovery-studio-visualizer-download [38]. The steps followed for protein preparation included the removal of water molecules and heteroatoms, whereas hydrogen atoms were added, and total charges were generated. The protein was then optimized and energy-minimized using the CHARMM force field to achieve the structural integrity of the final proteins. After the structure was fully prepared, saved and then converted to the pdbqt format required for AutoDock Vina docking studies. This conversion was performed using the Webina web-based tool.

Ligand Screening for Palbociclib Analogue

Palbociclib is a well-established drug for suppressing CDK4 and has an inhibitory role in cell proliferation. The COCONUT database is a freely available web-based server https://coconut.naturalproducts.net/ for the storage, search, and analysis of natural products (NP), gathers information of NP from 50 different open sources [39]. The COCONUT database repository was used to retrieve natural products as ligands for the target protein CDK4 (PDB ID-2W96) based on a 75 percent similarity-based approach considering palbociclib as the reference drug [40].

Ligand Preparation

The COCONUT database is a repository of 4,07,270 natural compounds. Based on the Tanimoto similarity algorithm, we extracted 1017 investigational natural products for their prospective role as ligands against CDK4 [39, 41]. The ligands (.sdf) files were extracted from the COCONUT database and prepared using the Discovery Studio tool for further analysis where hydrogen atoms were added, any solvent molecules, counter ions, or other extraneous entities were removed from the ligand structures to ensure that only the relevant ligand atoms were retained [42]. Hydrogen atoms were added to the ligand structures to ensure correct valence states. This step is crucial for accurate charge assignment and interaction analysis. Appropriate charges were assigned to the ligand atoms using the CHARMM force field available in Discovery Studio. This step ensures that the electronic properties of the ligands are accurately represented. The geometry of the ligands was optimized to relieve any steric clashes and to place the ligands in their lowest energy conformations. This step was performed using the minimization tools available in Discovery Studio. Further, the ligands were fully prepared, saved, and then converted to the pdbqt format required for AutoDock Vina docking studies. This conversion was performed using the Webina web-based tool.

Pharmacokinetics Profiling Using SWISSADME

SWISSADME is a web-based computational tool that was used for pharmacokinetics profiling (ADME and medicinal properties) of natural compounds used as ligands [43]. Drug likeness of ligands was predicted by the SWISSADME tools, which follow Lipinski's rule of 5 [44–47]. Rule of five was proposed by Lipinski and co-workers, which is a well-established filter for the drug-likeliness of a ligand [48]. The criteria used for filtering to determine whether a ligand can be absorbed orally or not, are that the molecular weight should be less than or equal to 500, the water partition coefficient (AlogP) should be less than or equal to 5, a number of hydrogen bond acceptors (HBAs) $\leq$10.6 and number of hydrogen bond donors (HBDs) $\leq$5. Several other chemical characteristics such as TPSA (topological polar surface area), % ABS (absorption percentage) and rotatable bond count are also considered in this analysis. According to Lipinski's rule of five, for a compound to be orally active, the TPSA value should be less than 140 $Å^2$ [49–52]. A few other factors, such as the blood-brain barrier and skin permeability (log Kp), were also considered.

Molecular Docking

Molecular docking is a computational technique used in the field of structural biology and drug discovery to predict the preferred orientation of the ligand when bound to the receptor to form a stable complex. In this study, we used Webina https://durrantlab.pitt.edu/webina/, a web-based

docking tool available for molecular docking, and fastDRH http://cadd.zju.edu.cn/fastdrh/ to perform docking and analyze the result [53, 54]. The input files of receptor and ligands were prepared in the pdbqt format [55], which were uploaded on Webina, followed by the preparation of a docking box using the grid center x (0) y (−2) z (75) and grid size center x (59), y (42) z (50), while grid space was 0.375[56].

MM/GBSA Analysis

In this study for Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) analysis, we have accessed the fastDRH tool [56, 57]. This server predicted the best ligand binding mode and hot spot residue for ligand binding with high speed and accuracy using the Autodock vina docking engine followed by MM/GBSA calculations on an AMBER backend, [58]. The molecular docking study was performed by uploading the CDK4 structure in pdb format and ligand structures in .sdf format. Further, we opted for the Autodock vina docking engine and uploaded the receptor pose file in pdb format for molecular docking, with up to 10 poses selected for further analysis. For rescoring, the forcefield selected for receptor ff99SB with (TIP3P) water model and ligand were GAFF2, respectively [59, 60]. The hotspot prediction was performed using the MM/GBSA method, along with the other parameters applied for rescoring, and the available default parameter was also chosen for the truncation radius. The results were extracted and used for final analysis.

MD Simulation

Molecular dynamic (MD) simulation was accessed to reveal the dynamic behavioral changes of phytochemical and CDK4 binding. The protein-ligand complex structure obtained from docking analysis was retrieved in Gromacs 2023.3 [61]. Charmm27 forcefield was selected for MD simulation for 100 ns. The entire system (phytochemicals and CDK4) was first solvated with a three-point transferable intermolecular potential (TIP3P) orthorhombic box, and their charges were neutralized by adding counter ions, Na and Cl [60]. The Velocity Verlet integrator, known for its stability and energy conservation, was used in MD simulation. Bond lengths are restricted by the LINCS algorithm, thereby setting the time step at 0.002 ps. Energy minimization was carried out using the steepest descent approach prior to the MD simulations to rule out high-energy configurations and ensure a stable starting point. Positional constraints with a force constant of 1000 kJ/mol/nm were placed on the solute's heavy atoms during the first equilibration phase, enabling the solvent and ions to equilibrate around a fixed solute structure. The solvated protein-ligand complex systems were then subjected to 50,000 nsteps of energy minimization followed by 300 K for 100 ps. An equilibration step for pressure was executed for a further 100 ps. Finally, the MD runs were performed for 100 ns, and post-MD simulations were performed for Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of gyration (Rg), hydrogen bond and minimum distance calculation.

Results

Developing effective therapies for NDDs remains a formidable challenge. Recognizing the potential of natural products, traditionally acknowledged for their therapeutic benefits and safety, researchers are increasingly turning to computational studies to screen and predict the potential of these compounds as drug candidates. These computational approaches, employing methods like molecular docking and molecular dynamic simulation, offer a cost-effective means to generate preliminary data, guiding the selection of promising natural products for further in vitro and in vivo studies, facilitating the identification of lead compounds, and expediting the drug discovery process in the quest for safer and more efficacious treatments for NDDs. Our quest is not only to unravel the mysteries of CDK4 but also to pioneer innovative therapeutic modalities that may redefine the landscape of NDDs research and treatment.

Structural Analysis for Predicting Key Residues and Binding Sites of CDK4

Understanding the three-dimensional structure of CDK4 is crucial for unraveling its functional dynamics and identifying potential points of interaction with Palbociclib-like natural products. CDK4 consists of distinct structural domains that confer its specific functions. In this study, we retrieved the CDK4 (2W96) structure from the RCBS database, a complex of chain A, i.e., cyclin, and chain B belonging to CDK4. We selected chain B for our study with a sequence length of 303 amino acids in which the PISTVRE-like cyclin binding domain is present at the 50th–56th position, the protein kinase domain is situated from the 3rd–267th position, and the ATP binding site is at the 12th–35th, as per the data available from Interpro database [62]. According to literature analysis, Lys35 is an important residue in the catalytic domain, and Phe93 acts as a gatekeeper residue and their active site is located at 133rd–145th position as per Interpro result [62]. When we accessed the PDBSUM database for motif analysis of 3rd–267th position, we found that 2W96 consists of N-lobe and C-lobe with 2β sheets A and B, antiparallel in nature with 5β strand in sheet A and 2β strand in sheet 2, 4β hairpins are situated at strand 1 and 2. Locations of β

hairpins are as follows, 1 at (Glu7-Gly13)-(Val20-Arg24), 2 at (Val20-Arg24)-(Phe31-Pro40), 3 at (Leu74-Arg82)-(Glu86-Glu94) and 4 at (Ile146-Val148)-(Val154-Leu156). The location of β strands is denoted as 1st (Glu7-Gly13), 2nd (Val20-Arg24), 3rd (Phe31-Pro40), 4th (Leu74 -Arg82), 5th (Glu86-Glu94), 6th (Ile146-Val148), 7th (Val154-Leu156) [63]. Studies on CDK4 have also shown that certain specific residues are more frequently engaged in binding naturally in the ATP cleft e.g., Ile12, Val20, Ala33, Val77, Phe93, Glu94, His95, Val96, Gln98, Asp99, Thr102, Glu144, Leu147, Ala157, Asp158, whereas Glu144 interacts with the terminal phosphates of ATP as depicted in Figure 1 where it can be seen that CDK4-Palbociclib binding at the active site in which hydrogen bond interaction at GLY15, LYS35 [64].

Database Mining

The current study selected Palbociclib, a CDK 4/6 inhibitor molecule, as the reference drug. Palbociclib can split CDK4/6-cyclin-D complexes, inhibiting pRb phosphorylation and E2F1 release, which results in G1 phase arrest and cell growth reduction, even though CDK4/6 can bind cyclin D and cause pRb hyperphosphorylation [17]. Palbociclib has an elimination half-life of more than 29 hours. The liver is the main metabolic area, where SULT2A1 and CYP3A are involved in metabolism; the kidneys and feces account for 17.5% and 74.1% of excretion, respectively. In the present study, we used the Tanimoto similarity method, which is a widely used metric in cheminformatics for comparing the similarity between two molecular structures [65]. This method calculates

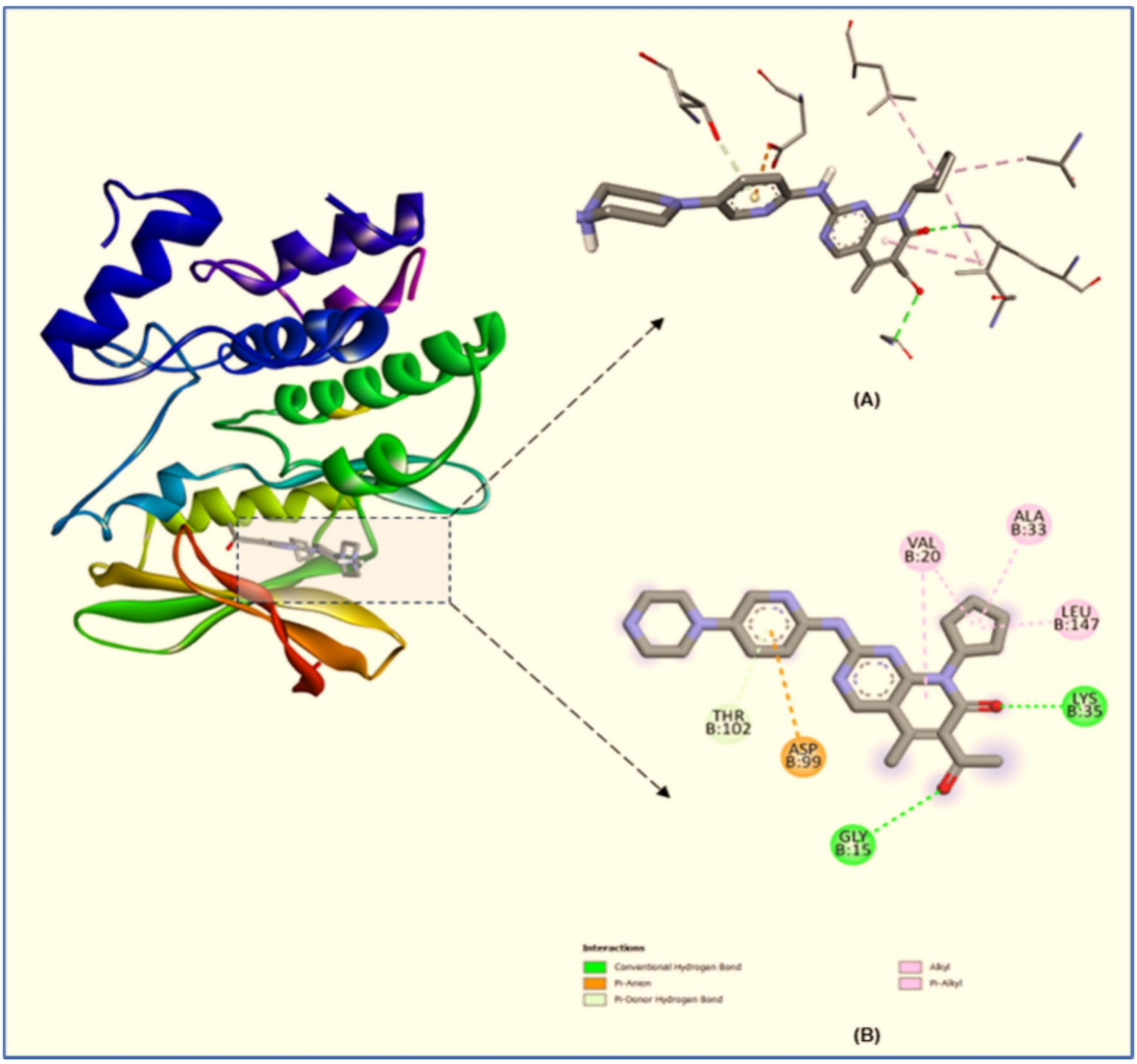


Fig. 1 CDK4-Palbociclib binding at the active site depicting hydrogen bond interaction at GLY15, LYS35, and hydrophobic interaction at VAL20, ALA33, ASP99, THR102, LEU147 (**A**) 3D representation of the native structure, (**B**) 2D Molecular interaction of the predicted docking

the similarity between two sets of binary data, typically representing the presence or absence of certain features in the molecules (such as chemical sub-structures or fingerprints) [34, 40]. In filtering the COCONUT database, the Tanimoto method measures the similarity between a reference drug, palbociclib, and natural compounds. We can filter out compounds that are dissimilar to the palbociclib by setting a threshold Tanimoto coefficient, retaining those with a high degree of similarity. This approach is essential for narrowing down large datasets to a manageable number of relevant compounds for further analysis. We uploaded the smile of Palbociclib in the COCONUT database at structure search and selected a 75% threshold of Tanimoto similarity, which is a measure of similarity between two sets, which often helps in data mining, informatics retrieval, and bioinformatics [65]. A threshold of 75% has been found to provide a good compromise between sensitivity and specificity and minimize the false positive and false negative rate. This threshold is often chosen because it is high enough to ensure meaningful similarity. As a result, we extracted 1017 natural compounds and retrieved .sdf files for further analysis.

ADME Prediction

In this study, physiochemical and pharmacokinetics evaluation of the selected 1017 natural compounds retrieved from COCONUT database were done for further analysis. Drug-likeness was accessed to determine whether the natural products have favorable ADME properties [66]. Drug likeness of natural products depends on Lipinski's rule of five, which was an established filter to determine whether a compound can be absorbed orally, the efficiency of drug absorption from the bloodstream, and factors influencing absorption such as molecular size, lipophilicity, and the presence of efflux transporters. The BBB permeability determines whether a compound can cross the blood-brain barrier and affect the CNS. The P-glycoprotein (Pgp), an efflux transporter belonging to the ATP binding cassette (ABC) transporters, recognizes several different xenobiotics, and many of them are CYP3A4 substrates. The pan-Assay Interference Compounds (PAINS) filter was accessed to investigate and filter out compounds that may interpret false positive results in various bioassays. These compounds are often found to interfere with a wide range of biological assays and can lead to misleading conclusions during drug discovery. The Brenk filter was used to identify potentially problematic chemical structures. These filters are designed to recognize sub-structures that may lead to false positives. Subsequently, we filtered 110 natural compounds from 1017 previously retrieved natural compounds. The filtered compounds were further analyzed by molecular docking study and ten compounds were selected, as represented in Table 1. In this research, we assessed compounds denoted by their CNP identifiers to evaluate their suitability as potential pharmaceutical agents. Among these compounds, CNP0070947 stood out as a highly promising candidate owing to its favorable molecular attributes and pharmacological properties. With a molecular weight (MW) of 322.36, CNP0070947 demonstrated exceptional solubility and notable lipophilicity, as confirmed by its logP values of 2.54 (iLOGP) and 1.14 (XLOGP3), respectively. Additionally, CNP0070947 exhibited favorable ADME characteristics, including permeability across the BBB and its status as a substrate for P-glycoprotein (Pgp), suggesting its potential for efficient delivery and distribution throughout the organism. Despite CNP0070947 presenting the most promising profile, other compounds also displayed significant attributes worthy of consideration. For example, CNP0346604 exhibited moderate to high solubility and favorable lipophilicity

Table 1 Drug likeliness of selected compounds

Compound ID	GI Absorption	BBB	Lipinski violations	PAINS alerts	Brenk alerts	Log $P_{o/w}$	Log S	Log Kp	Synthetic bioavailability
CNP0376445	High	Yes	0	0	2	2.46	−3.83	−6.83	2.16
CNP0352050	High	Yes	0	0	0	3.10	−4.17	−5.90	2.78
CNP0267694	High	Yes	0	0	0	2.27	−3.63	−6.80	2.61
CNP0346604	High	Yes	0	0	0	2.39	−3.26	−7.22	2.85
CNP0396493	High	Yes	0	0	0	2.84	−3.53	−6.74	3.12
CNP0070947	High	Yes	0	0	0	1.80	−2.76	−7.46	3.34
CNP0319774	High	Yes	0	0	1	3.18	−4.29	−5.82	2.18
CNP0396034	High	Yes	0	0	0	2.55	−3.21	−6.94	2.58
CNP0162466	High	Yes	0	0	0	2.50	−3.25	−6.81	2.48
CNP0438514	High	Yes	0	0	0	3.30	−4.43	−5.92	2.63

Table 2 2D structure and binding free energy in Kcal/mol of selected natural compounds

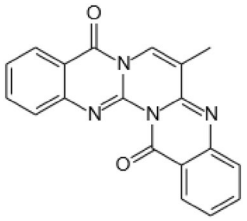
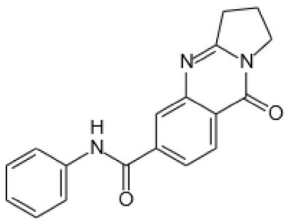
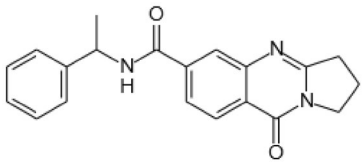
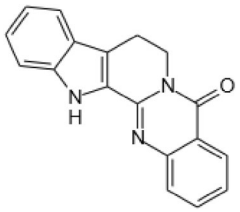
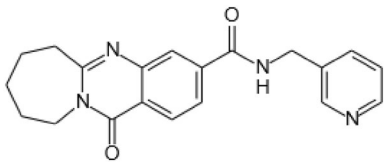
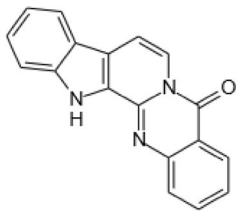
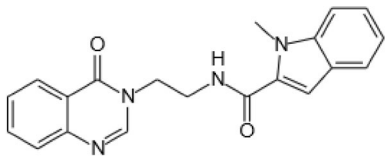
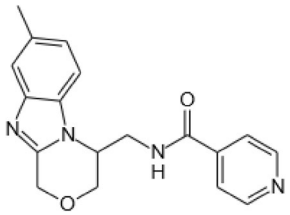
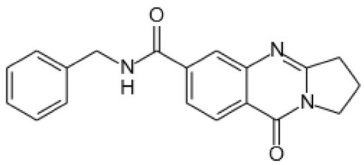
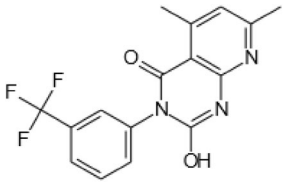
Compound ID	Structure	Affinity score (Kcal/mol)	fastDRH score (Kcal/mol)
CNP0376445		−9.99	−9.90
CNP0162466		−9.83	−8.69
CNP0396493		−9.79	−8.84
CNP0352050		−9.74	−9.46
CNP0346604		−9.72	−8.88
CNP0319774		−9.67	−8.77
CNP0267694		−9.64	−9.09

Table 2 (continued)

Compound ID	Structure	Affinity score (Kcal/mol)	fastDRH score (Kcal/mol)
CNP0070947		−9.55	−8.78
CNP0396034		−9.32	−8.74
CNP0438514		−9.00	−8.44

(with logP values of 2.74 (iLOGP) and 1.7 (XLOGP3), respectively), in addition to interactions with cytochrome P450 enzymes (CYPs). In the same way, CNP0396493 has displayed high solubility and lipophilicity, making it another compound worthy of further examination. Overall, this study showcases valuable insights into the drug-likeliness of these natural compounds forming a base for future research endeavours to utilize their therapeutic potential.

Molecular Docking Analysis

The molecular docking of 110 filtered natural compounds with CDK4 (2W96), was performed for studying their molecular similarity with the reference drug palbociclib. In the present study, Webina tool was used, which involves computational algorithms to interpret the natural compounds that bind specifically to CDK4 (2W96). The interpretation of this study gave the binding energy, which is crucial for understanding the interaction between CDK4 and natural compounds. The strength of binding is often quantified by the binding energy; a high negative binding energy score corresponds to a strong interaction. The filtered 110 drugs were screened for their ability to bind to the CDK4 (2W96), and the binding score was determined. The top 10 natural compounds were determined based on the highest negative binding energy scores, as depicted in Table 1. The molecular docking studies revealed that palbociclib bound with CDK4 with binding energy (or the affinity score) of −8.988 kcal/mol. Based on affinity scores, 37 compounds with affinity score more than palbociclib were observed and we selected top 10 compounds with COCONUT id CNP0376445, CNP0352050, CNP0267694, CNP0346604, CNP0438514, CNP0396493, CNP0070947, CNP0319774, CNP0396034, and CNP0162466 for further analysis and the interpretation is represented in Table 2. In this study, we also used FastDRH for docking study validation. Finally, we filtered the top 4 natural compounds CNP0376445, CNP0267694, CNP0396493, and CNP0070947 as hit compounds for further evaluation.

Interpretation of Interaction of Hit Compounds with CDK4

The top hit compounds retrieved through molecular docking studies have a comparatively higher negative binding affinity as well as key residue interaction with CDK4. In this study, we opted for compounds that have a negative binding score of more than −8.988 kcal/mol from Webina. Further, fastDRH was also utilized as a criterion for the selection of hit compounds on the basis of a docking score of more than −8.57 kcal/mol and a GB1 score of more than −32.5 kcal/mol. Accordingly, taking into consideration the score obtained from Webina and fastDRH, the four natural compounds selected were CNP0070947, CNP0376445, CNP0267694, and CNP0396493 as depicted in Figure 2.

The interactive active site residues in CDK4-Palbociclib complex are ILE12, GLY13, GLY15, GLY18, VAL20,

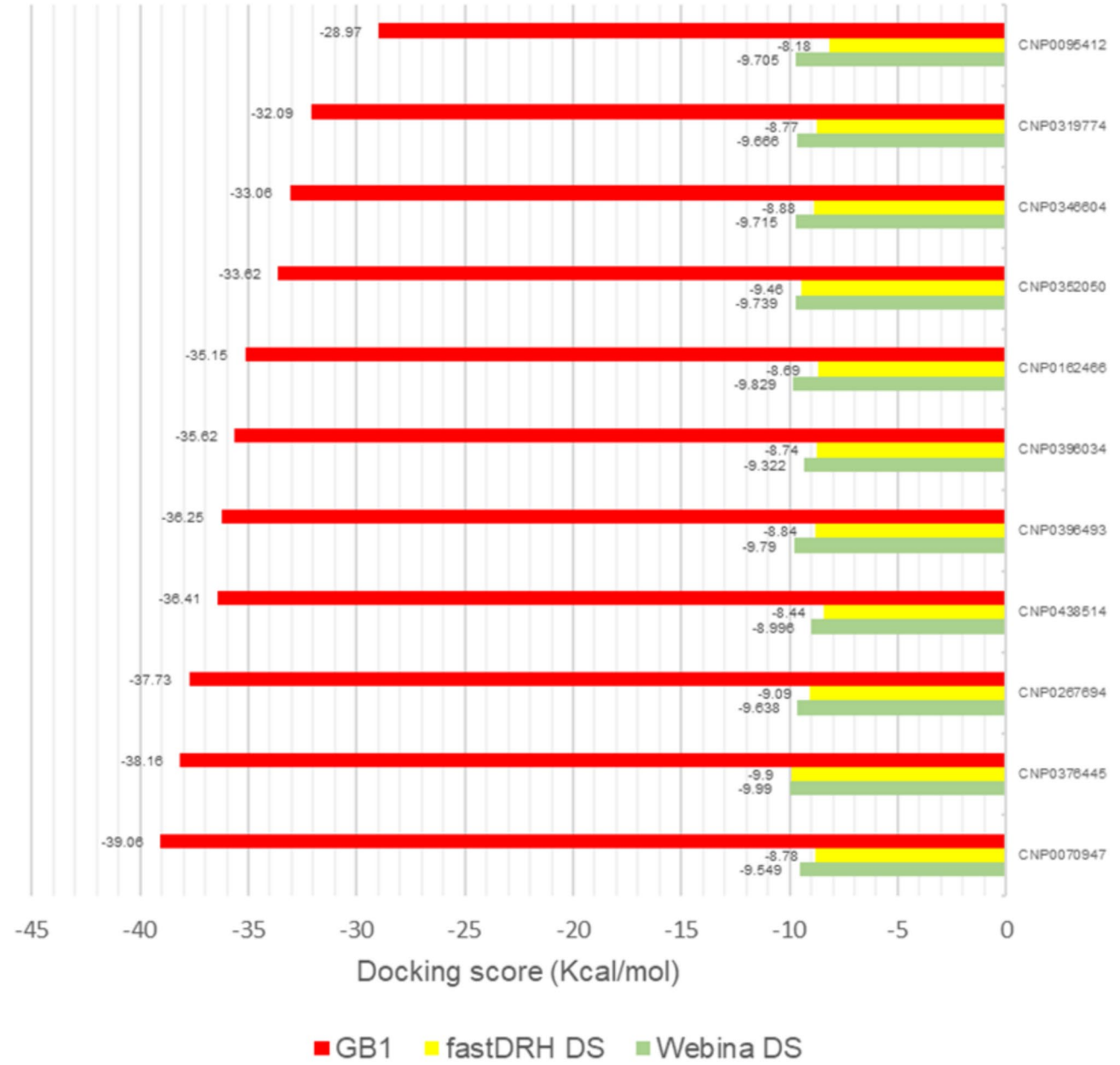


Fig. 2 Docking score interpretation illustrates the selected compounds' binding affinity after filtration through the swissADME tool. Top ten compounds (CNP0376445, CNP0352050, CNP0267694, CNP0346604, CNP0438514, CNP0396493, CNP0070947, CNP0319774, CNP0396034, and CNP0162466) with the highest negative binding energy after molecular docking by different tools such as fastDRH, webina and arranged on the basis of GB1 total binding energy

ALA33, LYS35, HIS95, VAL96, GLY98, ASP99, ARG101, THR102, LEU147, ASP158. These residues are primarily involved in the ATP binding and the kinase activity of CDK4. This complex was stabilized by a Conventional Hydrogen Bond at GLY15, LYS35 position, and several hydrophobic interactions at positions VAL20, ALA33, and LEU147. This palbociclib complex interferes with substrate accessibility and this interferes with its kinase activity. The amino acid residues among hit compounds analysis were with different types of interactions, namely hydrogen bonds, hydrophobic contacts, π-stacks, π alkyl, and alkyl contact. CNP0376445 forms two hydrogen bonds with LYS35, ARG101 while CNP0267694 forms four hydrogen bonds with ILE12, ALA16, ARG101, ASP158. CNP0396493 forms Hydrogen Bond at LYS35, GLU94, HIS95, VAL96, ARG101 and CNP0070947 forms Hydrogen Bond at GLY15, TYR17, LYS35, GLU144. LYS35 is the common residue involved in the Hydrogen bond formation in all hit

Table 3 Ligands binding interactive residue site of CDK4 and hit selected compounds

Compound ID	Ligand binding sites	Residues involved in hydrogen bond formation (Distance in Å)	Number of Hydrogen bonds formed	Residues involved in hydrophobic Interactions	Binding Affinity (Kcal/mol)
CNP0267694	ARG101, ASP158, ILE12, ALA16, LEU147, ALA33	ARG101 (3.30) ASP158 (3.63) ILE12 (3.67) ALA16 (3.72)	04	LEU147, ALA33	−9.63
CNP0376445	LYS35, ARG101, LEU147, LYS142, ILE12, VAL20, ALA33	LYS35 (3.02) ARG101 (3.25)	02	LEU147, LYS142 ILE12, VAL20, ALA33	−9.99
CNP0396493	LYS35, HIS95, ARG101, GLU94, ASP158LEU147, ALA33, VAL72, ALA157,ALA16, PHE93,VAL96	LYS35 (3.42) HIS95 (3.36) VAL96 (3.82) ARG101 (3.07) GLU94 (3.32)	05	LEU147, ALA33, VAL72, ALA157, ALA16, PHE93, ASP158	−9.79
CNP0070947	GLY15, TYR17 GLU144, LYS35 ASP158, EU147, ALA33, ALA157, ALA16, PHE93	GLY15 (3.29) TYR17 (3.19) GLU144 (3.21) LYS35 (4.00)	04	ASP158, LEU147 ALA33, ALA157 ALA16, PHE93	−9.55

compounds except CNP0267694. Hydrophobic interaction at LEU147 is common in all the hit compounds, whereas ALA33 is also involved in all hit compounds in hydrophobic contacts except CNP0267694. The ligand binding sites and hydrogen bonds formed by these hit compounds are shown in Table 3 and Figure 3.

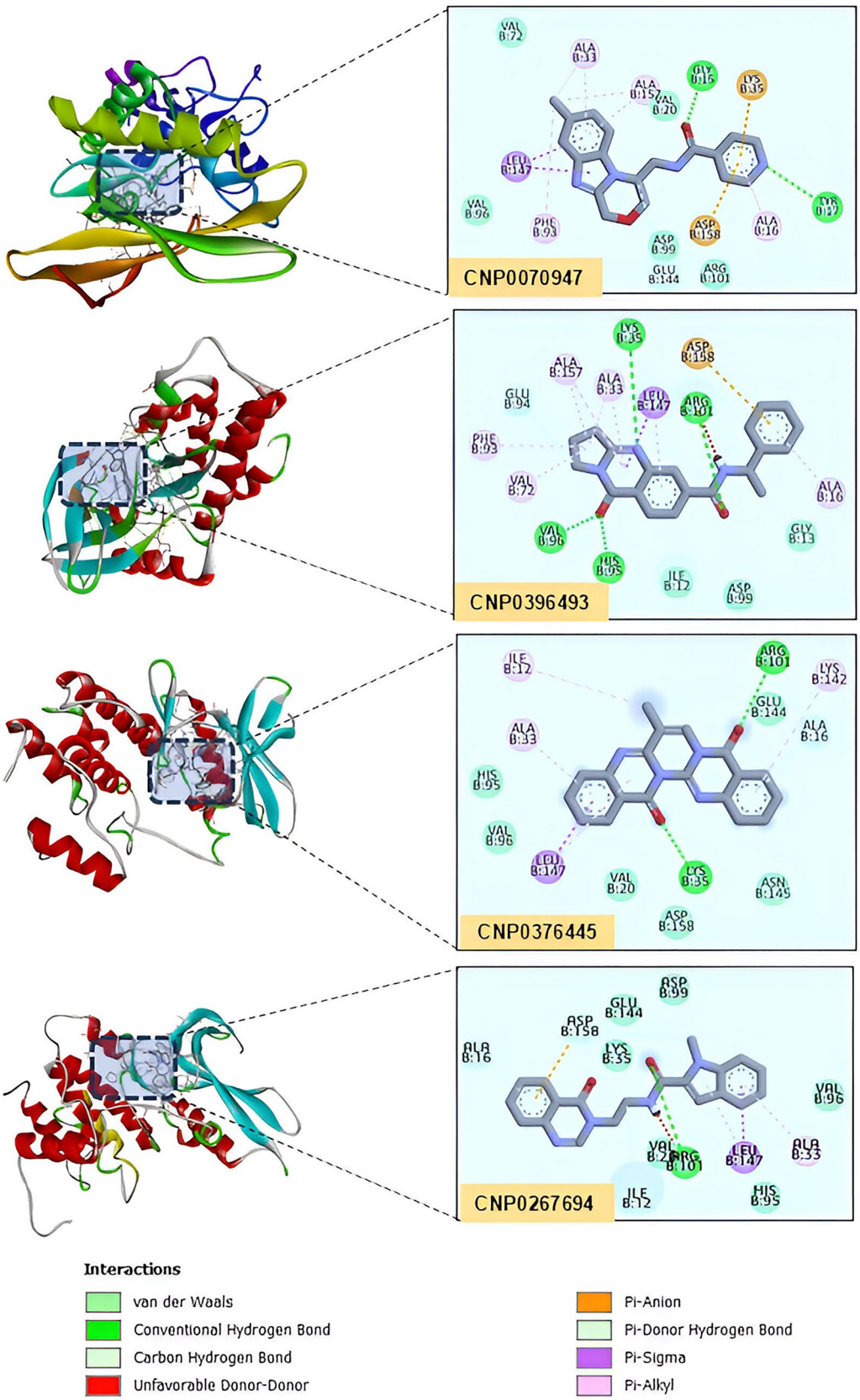
To further validate the ligand molecule complex, MM/GBSA and MM/PBSA were used for the calculation of free energy, which was performed by the fastDRH tool. This tool determines the electrostatic energy (ELE), van der Waals (VDW), and internal energy (INT) as calculated by the MM forcefield; total gas phase energy (GAS) sum of ELE, VDW, and INT, PBSUR/GBSUR, based on empirical model to calculate solvation free energy contributed by non-polar solvation; PBCAL/GB solvation free energy contributed by electrostatic; PBSOL/GB/SOL, solvation based on the sum of polar and non-polar; PBTOT/GBTOT, and the final binding energy (Kcal/mol). In the present study, we accessed fastDRH for determining the decomposition energy of Palbociclib, CNP0070947, CNP0376445, CNP0267694, and CNP0396493 with CDK4 and the result is depicted in Table 4. Based on the data provided by fastDRH, ΔEVDW was considered as important parameter for molecular complex formation, and it was observed that CDK4 molecular complex with palbociclib has ΔEVDW value of −41.21 Kcal/mol and with the natural compounds CNP0070947, CNP0376445, CNP0267694, and CNP0396493 this value of ΔEVDW (Kcal/mol) have been found to be −46.09, −44.21, −45.29, and −42.54, respectively. The result interpretated that total binding energy (TOT) (Kcal/mol) for Palbociclib, CNP0070947, CNP0376445, CNP0267694, CNP0396493 with CDK4 was −32.72, −39.16, −38.16, −37.73, and −36.25 respectively. Decomposition Energy Figure 4.

Molecular Dynamics Simulation

The use of molecular dynamics simulation enables the computation and prediction of compound stability. As a fundamental computational tool, molecular dynamics is highly effective in capturing molecular and atomistic changes. It plays a vital role in understanding the stability of protein-ligand complexes by analyzing deviations and fluctuations, as well as examining intermolecular interactions. Using conventional methods like molecular docking and virtual screening, the design of drugs based on structure provided a list of selected compounds. MD simulation is crucial for gaining insights into the dynamic properties of the ligand and its interaction with the protein. The CDK4 and hit compounds like CNP0070947, CNP0376445, CNP0267694, and CNP0396493 complex were analyzed after MDS for 100 ns, and post MDS analysis result has interpreted based on RMSD, RMSF, and Rg, minimum distance and hydrogen bond.

RMSD Estimation

To determine the difference between the backbone and final position of a protein, one can use the RMSD by analyzing the deviations generated during that simulation, one can determine whether the protein remains stable in


CNP0070947
CNP0396493
CNP0376445
CNP0267694
Interactions
van der Waals
Conventional Hydrogen Bond
Carbon Hydrogen Bond
Unfavorable Donor-Donor
Pi-Anion
Pi-Donor Hydrogen Bond
Pi-Sigma
Pi-Alkyl

◂**Fig. 3** Two-dimensional interaction between CDK4 and selected hit compounds depicting different types of interactions such as van der Waals, conventional hydrogen bond, carbon-hydrogen bond, pi anion, pi -donor hydrogen bond, pi sigma, pi alkyl bond. In CNP0070947, active residues involved in an interaction

terms of its shape. The less deviation there are, the more stable the protein structure. To check the stability of both systems, the RMSD value of the C-α backbone was determined for 100 ns. The protein C-α backbone RMSD of control, along with reference and other hit compounds, were analyzed as displayed in Figure 5. The RMSD of the control protein was approximately 0.27 nm, and Palbociclib represented stability with an average RMSD value of 0.52 nm and maintained stability in the entire simulation. It was observed that control, CNP0267694, and CNP0070947 compounds showed stability throughout the entire simulation process at an average RMSD value of 0.37 and 0.32 nm, while CNP0396493 showed RMSD value of 0.4 nm till 15 ns after that, it stabilized at the average RMSD fluctuation till 20 ns after that, it remained stable till the end of the simulation and showed an average RMSD value of 0.85 nm. CNP0376445 showed consistent stability from the starting point till 100 ns with an average RMSD of 0.52 nm.

The ligand RMSD for Palbociclib, CNP0070947, CNP0376445, CNP0267694, and CNP0396493 represented; these simulated systems displayed steady-state stability after 10 ns of MD run [67]. Palbociclib represented stability with an average RMSD value of 4.79 nm and remained constant throughout the entire simulation. CNP0396493 represented an average RMSD value of 0.85 nm till 100 ns simulation whereas CNP0376445 and CNP0070947 indicated an average RMSD value of 1.03 and 1.47 nm, respectively till 100 ns, whereas, a slight unstability was observed for CNP0267694 till around 32 ns, afterwards it attained stability till 100 ns MD run. However, after 32 ns, the RMSD values of all the complexes remained unchanged until the final MD run. Broadly, the entire simulated complex displayed an RMSD value of more than 0.74 nm. The results interpret that all hit compounds are tightly bound and not affected by the protein topology. The stability of

Table 4 MM/GBSA score (Kcal/mol) CDK4-selected compounds complex along with reference drug

Complex formed with CDK4 (2W96)	VDW	GAS	PBSUR/GBSUR	PBCAL/GBCAL	PBSOL/GBSOL	PBELE/GBELE	PBTOT/GBTOT
Palbociclib	−41.21	−41.21	−5.32	13.81	8.49	13.81	−32.72
CNP0070947	−46.09	−46.09	−5.6	12.64	7.04	12.64	−39.06
CNP0376445	−44.21	−44.21	−5.19	11.25	6.05	11.25	−38.16
CNP0267694	−45.29	−45.29	−5.67	13.23	7.56	13.23	−37.73
CNP0396493	−42.54	−42.54	−5.7	11.98	6.28	11.98	−36.25

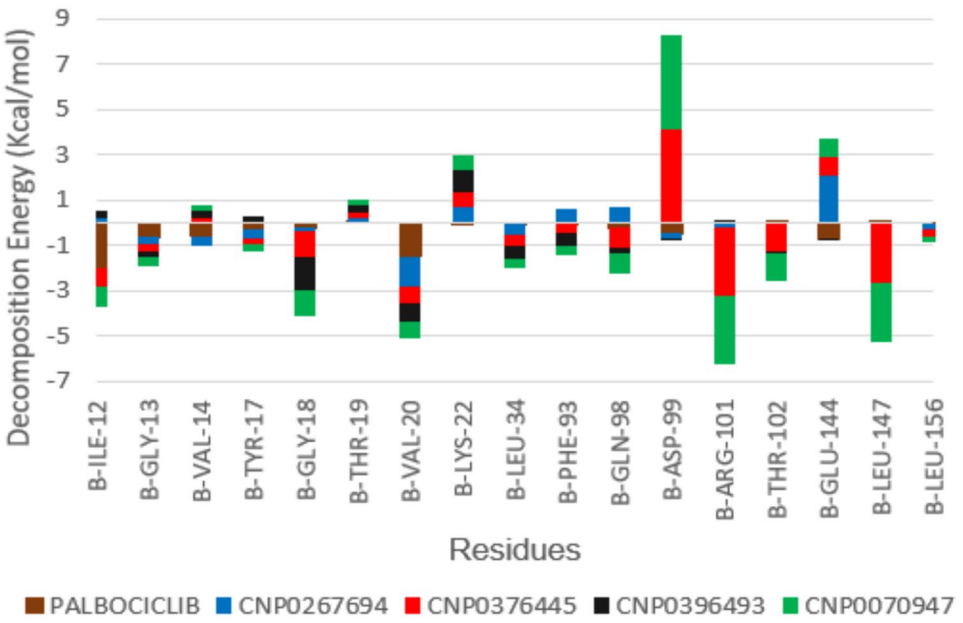


Fig. 4 Decomposition energy per residue of CDK4 with reference and selected phytochemicals. Val 20 and ARG 101 are the best hotspot residues involved in active site binding in each complex

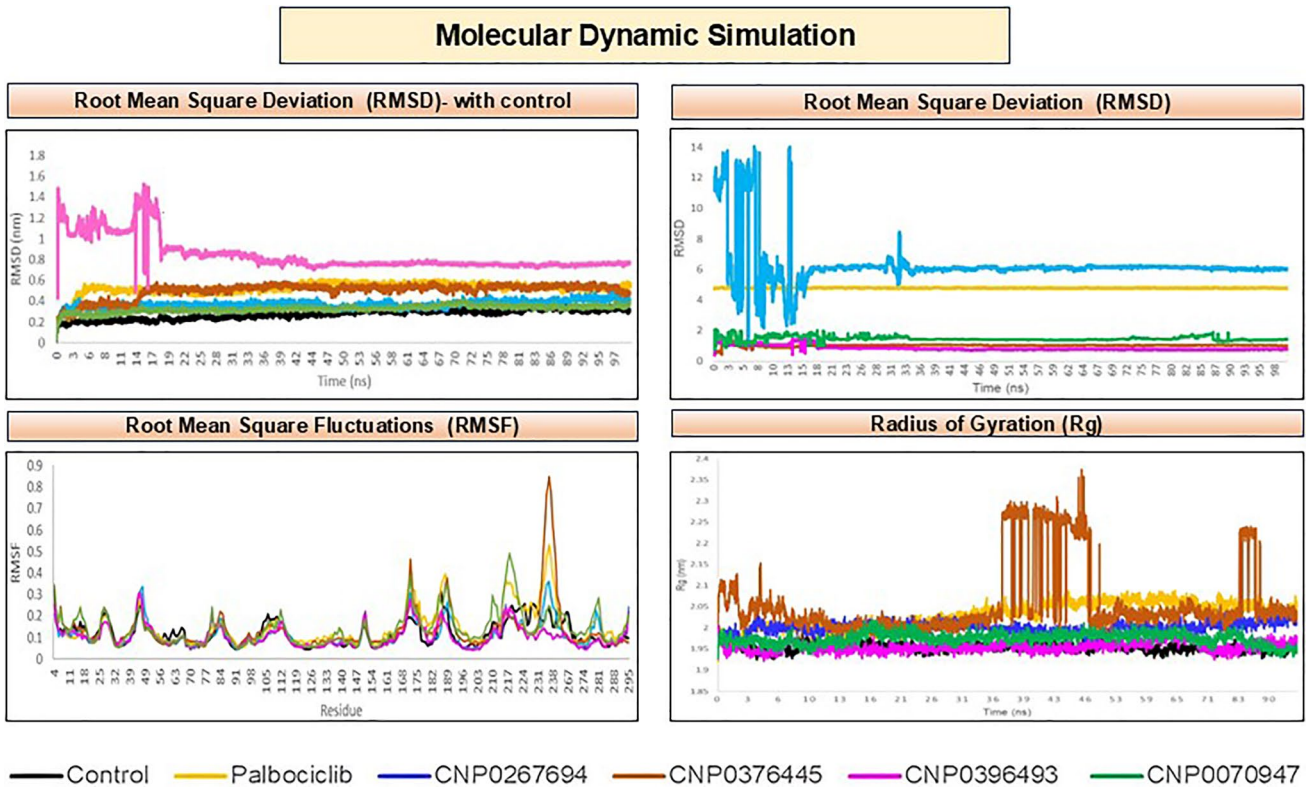


Fig. 5 The CDK4 and the selected hit natural compounds CNP0070947, CNP0376445, CNP0267694, and CNP0396493 complex were analyzed after MDS for 100 ns, and post MDS analysis result was interpreted based on RMSD protein backbone, RMSD, RMSF, Rg. The complex of CDK4 with reference drugs Palbociclib in brown, CNP0376445 in red, CNP0267694 in blue, CNP0396493 in black, and CNP0070947 in green. RMSD graph illustrated Palbociclib stability with an average RMSD value of 4.79 nm for the entire MD run while CNP0396493 represented an average RMSD value of 0.85 nm, CNP0376445 and CNP0070947 indicated an average RMSD value of 1.03 and 1.47 nm, respectively, till 100 ns whereas CNP0267694 remained unstable till 32 ns and stabilized afterward with average RMSD value 0.74 nm. The average RMSF of control is 0.11 nm while CDK4 with reference drug and other hit compounds is 0.15 nm, 0.12, 0.14, and 0.11 nm, respectively. Rg values of the control is 1.95 nm, and the complexes are in the same order: 2.03, 2.06, 2.00, 1.98, and 1.95 nm, respectively, indicating the compactness of the structure

CNP0396493 and CNP0070947 is significantly higher than that of CNP0376445 and CNP0267694, which had very minor deviations and were somewhat like Palbociclib.

RMSF Analysis

RMSF is one of the essential parameters for the characterization of rigid and flexible areas of protein and is used for assessing the flexibility of the protein structure. The Control protein showed an average RMSF value of 0.11 nm. The RMSF was calculated for the four-hit compounds and reference drug palbociclib. The average RMSF value of the reference drug was 0.15 nm. CNP0267694, CNP0376445, CNP0070947, and CNP0396493 represented the average RMSF value of 0.12, 0.14, 0.14, and 0.11 nm, respectively, as illustrated in Figure 5.

Radius of Gyration (Rg)

In MD simulation studies, Rg depicts the compactness and flexibility of a protein inside the biological environment. The control protein showed an average Rg value of 1.95 nm. The lower the value of Rg, the more firmly packed and stable the structure is during the simulation. The control protein showed an average Rg value of 1.95 nm, while the reference palbociclib average Rg value was observed at 2.03 nm, whereas CNP0376445, CNP0267694, CNP0070947, and CNP0396493 had an average Rg value recorded as 2.06, 2.00, 1.98, and 1.95 nm, respectively, as represented in Figure 5. CNP0376445 results indicated considerable bumps

during the simulation. However, the graph of CNP0267694, CNP0070947 and CNP0396493 was considerably stabilized.

Hydrogen Bond and Minimum Distance Analysis

Furthermore, the hit compounds formed the hydrogen bonds with the CDK4, and the survival time of these H-bonds and the number of H-bonds reflected the degree of binding between the ligand and CDK4. As indicated in Figure 6A, palbociclib had four H- bonds for a longer period in the MD simulation, while the compound CNP0267694 made two H-bonds for a short period, which are lesser in comparison to the reference, CNP0376445 represented 3 H-bond. On further analysis, as evident from Figure 6A, compounds CNP0396493 and CNP0070947 formed 3 and 7 H-bonds in the simulation with a longer survival rate. Further, the minimum distance of ligand from the CDK4 binding site was also analyzed as an important post MDS analysis interpretation, and it was seen that except CNP0267694, all other hit compounds show distance below the range of 0.35 nm throughout the simulation after 20 ns as illustrated in Figure 6B. CNP0396493, CNP0070947, and CNP0376445 indicated minimum distance in the range of 0.23 and 0.24 nm.

Discussion

CDK4 inhibitors have been increasingly used in clinical studies for various NDDs in recent years [68]. Nevertheless, these inhibitors exhibit several drawbacks, including gastrointestinal irritation and drug resistance, which ultimately decrease their effectiveness [43]. Currently, there is a growing focus on using natural products for therapeutic research and development. The scientific community has shown particular interest in natural products due to their distinct benefits [30]. Hence, our work seeks to evaluate the effect of natural products filtered from the COCONUT database, on CDK4 molecules.

In this study, we have comprehensively explored potential CDK4 inhibitors by employing various computational tools and methodologies. On exploring the COCONUT database for natural compounds exhibiting molecular similarity with the target drug – Palbociclib, having a considerable therapeutic role in the management of NDDs by targeting CDK4, around 1017 natural compounds were retrieved, which were further filtered to 110 natural compounds using the swissADME tool. We further restricted our study to the ten most suitable natural compounds based on molecular docking studies to assess their binding affinity with CDK4. Notably, we employed stringent criteria, including negative binding scores exceeding −8.988 kcal/mol from Webina and −8.57 kcal/mol from FastDRH, along with the GB1 score calculated by FastDRH, to select the top four-hit compounds: CNP0070947, CNP0376445, CNP0267694, and CNP0396493.

Our findings revealed promising total binding energies for each compound, indicating their strong binding affinity with CDK4. Further analysis of the decomposition energy profiles has shed light on the contribution of individual energy components, providing insights into the stability and interaction dynamics of the ligand-protein complexes. CNP0070947 and CNP0396493 have the potential to be dominant CDK4 inhibitors due to their chemical interactions with key amino acid residues of CDK4, including LYS35, GLU94, HIS95, VAL96, and ARG101. These hydrogen bonding and hydrophobic interactions are crucial for binding ligands and suppressing CDK4 activity, showing that both these compounds may have therapeutic qualities.

Molecular dynamics and MM/GBSA play a crucial role in verifying the stability of the interaction between hit compounds and CDK4. Following the 100 ns simulation, we obtained the RMSD and RMSF data from the trajectory. This data was analyzed to verify the stability and flexibility of the interaction between the hit compounds and CDK4. The interaction and stability dynamics of control, CDK4 protein-ligand complexes involving Palbociclib and the hit compounds CNP0376445, CNP0070947, CNP0396493, and CNP0267694 have been demonstrated by the molecular dynamic simulations involving them. RMSD protein backbone analysis showed that control, palbociclib, and all the compounds are stabilized in the entire simulation after 10–20 ns. In the RMSD analysis, all these protein-ligand complexes demonstrated steady-state stability after an initial equilibration phase. During the entire simulation, the reference drug Palbociclib kept a stable RMSD, whereas the hit compounds also demonstrated such stability, baring few occasional variances, with CNP0070947 and CNP0396493 having the highest stability levels, whereas CNP0267694 and CNP0376445 having comparatively higher deviations, but maintaining the stability levels following an initial adjustment phase. Secondly, the RMSF analysis highlighted regions of flexibility within the protein structure. While all compounds exhibited comparable average RMSF values, CNP0376445 showed the highest fluctuation at the 230th residue with 0.84 nm and the 166th residue with 0.45 nm. Whereas, CNP0267694 showed the fluctuation at the 44th residue with 0.33 nm. These residues showed significant fluctuations, suggesting regions of flexibility that could be targeted for further optimization. Further, the Rg analysis indicated the compactness and stability of the protein-ligand complexes. Lower Rg values were observed for CNP0070947, CNP0396493, and CNP0267694, indicating a tighter packing and greater stability than Palbociclib and CNP0376445. Additionally, the analysis of hydrogen bonds, revealing varying degrees of interaction between the ligands

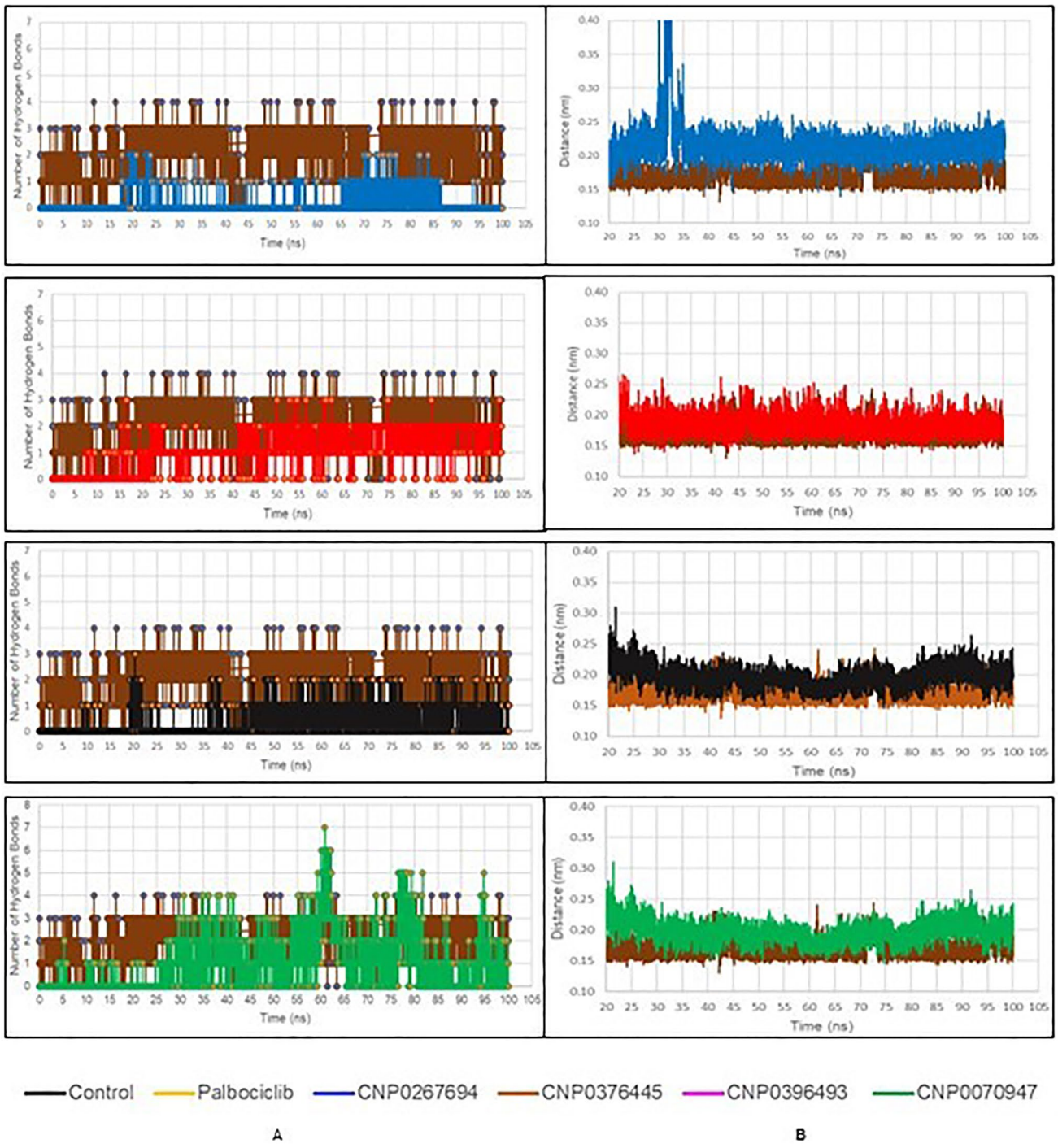


Fig. 6 The CDK4 and hit compounds like CNP0267694, CNP0376445, CNP0396493, CNP0070947 complex were analyzed after MDS for 100 ns (**A**) represented the hydrogen bond number between ligands (CNP0267694, CNP0376445, CNP0396493 and CNP0070947) 2, 3,3,7 and CDK4 in comparison of reference that form 4. (**B**) indicates the minimum distance between the CDK4 protein and the ligand

and CDK4 and minimum distances between protein binding sites with the ligand, further elucidated the binding interactions between the hit compounds and CDK4. Palbociclib formed four hydrogen bonds consistently throughout the simulation, indicating stable binding interactions. A lower binding affinity is suggested by the fact that CNP0267694 only created two hydrogen bonds for a brief period; this transient nature of hydrogen bonds raises the question about

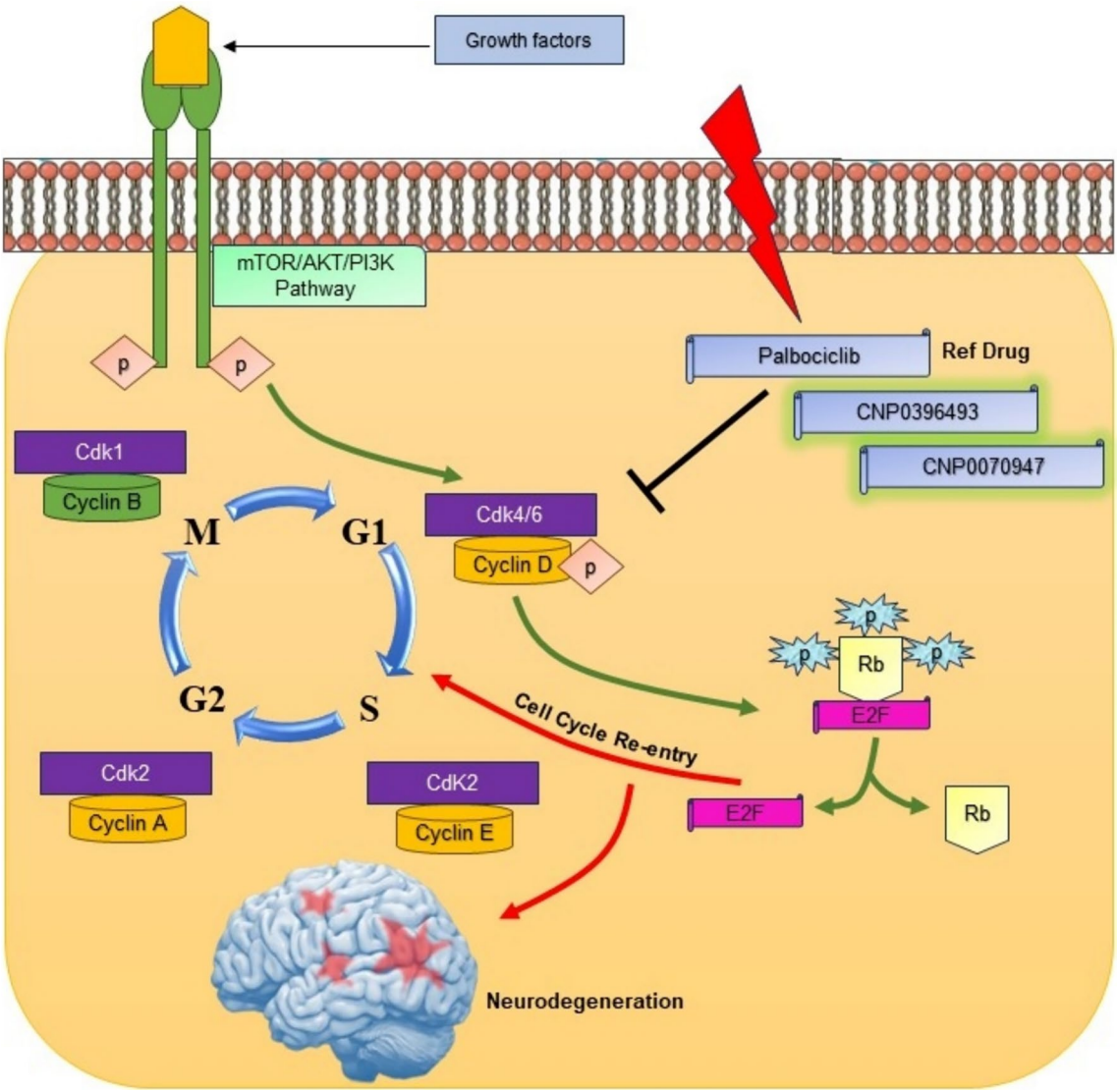


Fig. 7 Representation of Inhibitory mechanism of proposed inhibitors targeting CDK4 in NDDs. The cell cycle is tightly regulated by various regulators, which can act positively and negatively. Various growth factors initiate different kinds of signaling, such as mTOR/PI3K/AKT, which induces the cyclin-D /CDK4/6 complex formation, an important factor involved in the phosphorylation of Rb protein and the release of E2F transcription factor. E2F allows cell proliferation, which means cell cycle re-entry, ultimately moving toward neurodegeneration. Palbociclib, the reference drug, inhibits the cyclin-D/CDK4/6 complex. It is hypothesized that palbociclib analog natural compounds CNP0396493 and CNP0070947 are promising candidates as CDK4 inhibitors and have been validated through MM/GBSA and molecular dynamic simulation

its binding efficiency with CDK4, while CNP0396493 and CNP0070947 had the greatest number of hydrogen bonds—three and seven, respectively—CNP0376445 established three hydrogen bonds, indicating robust and long-lasting interactions with CDK4. After the first equilibration phase, all hit compounds—aside from CNP0267694—maintained minimum distances below 0.35 nm for the whole duration of the simulation. CNP0396493 and CNP0070947 demonstrated minimum distances between 0.23 and 0.24 nm, which is a sign of proximity to the active site and robust binding interactions. Additionally, CNP0376445 demonstrated a minimum distance of 0.24 nm, indicating a sustained binding to CDK4. Based on their greater number of hydrogen bonds, minimum distance, and significant binding affinity to CDK4, our results support the stability and viability of CNP0396493 and CNP0070947. Overall, the results suggest that the hit compounds, particularly CNP0396493 and CNP0070947, exhibit promising binding affinities and stability within the CDK4 protein active site.

The study's methodology and results provide a foundation for future research and optimization in developing efficient CDK4 inhibitors, eventually enhancing our knowledge and

treatment of NDDs associated with CDK4 dysregulation, as demonstrated in Fig 7, However, further experimental validation such as kinase assay and structural studies may be required to corroborate these computational findings and guide the development of novel CDK4 inhibitors with improved efficacy and therapeutic potential.

Conclusion

Our study highlights the understanding of the potential of bioinformatics coupled with database screening in identifying novel therapeutic candidates for NDDs by targeting CDK4. We have demonstrated that CNP0396493 and CNP0070947 exhibit promising binding affinities and stability within the CDK4. These findings provide a valuable perspective of the molecular mechanisms underlying the antiproliferative effects of CNP0396493 and CNP0070947 as CDK4 inhibitors and lay the groundwork for developing and utilizing hit compounds as neuroprotective drugs. The result of our study further highlights the significance of interdisciplinary approaches in drug discovery and emphasizes the potential of natural products in addressing complex NDDs. Future experimental validation is warranted to elucidate these compound's therapeutic efficacy and safety profiles, paving the way for the development of novel neuroprotective agents.

Acknowledgements We would like to thank senior management of Delhi Technological University (DTU) and University Grants Commission (UGC) for their constant support and encouragement.

Author Contribution PK: conceptualized, designed, and supervised this work. NR: drafted the manuscript, analyzed, and interpreted the data, and did the artwork. All authors have read and given consent for submission.

Funding Not applicable.

Data availability The authors confirm that the data supporting the findings of this study are available within the article.

Declarations

Competing interest All authors have read the manuscript and declared no conflicts or competing interests.

References

1. Lamptey, R. N. L., Chaulagain, B., Trivedi, R., Gothwal, A., Layek, B., & Singh, J. (2022). A review of the common neurodegenerative disorders: Current therapeutic approaches and the potential role of nanotherapeutics. *International Journal of Molecular Sciences, 23*(3), 1851. https://doi.org/10.3390/ijms23031851
2. Wilson, D. M., Cookson, M. R., Van Den Bosch, L., Zetterberg, H., Holtzman, D. M., & Dewachter, I. (2023). Hallmarks of neurodegenerative diseases. *Cell, 186*(4), 693–714. https://doi.org/10.1016/J.CELL.2022.12.032
3. Paulson, H. L. (2009). The Spinocerebellar ataxias. *Journal of Neuro-Ophthalmology, 29*(3), 227. https://doi.org/10.1097/WNO0B013E3181B416DE
4. Diaz, M., & Sirin, S. (2023). Lactic acid bacteria-derived exopolysaccharides mitigate the oxidative response via the NRF2-KEAP1 pathway in PC12 cells. *Current Issues in Molecular Biology, 45*(10), 8071–8090. https://doi.org/10.3390/CIMB45100510
5. Guzman-Martinez, L., Maccioni, R. B., Andrade, V., Navarrete, L. P., Pastor, M. G., & Ramos-Escobar, N. (2019). Neuroinflammation as a common feature of neurodegenerative disorders. *Frontiers in Pharmacology*. https://doi.org/10.3389/FPHAR.2019.01008
6. Michalska, P., & León, R. (2020). When it comes to an end: oxidative stress crosstalk with protein aggregation and neuroinflammation induce neurodegeneration. *Antioxidants, 9*(8), 1–34. https://doi.org/10.3390/ANTIOX9080740
7. Fu, H., Hardy, J., & Duff, K. E. (2018). Selective vulnerability in neurodegenerative diseases. *Nature Neuroscience, 21*(10), 1350–1358. https://doi.org/10.1038/S41593-018-0221-2
8. Aziz, T., Nadeem, A. A., Sarwar, A., Perveen, I., Hussain, N., Khan, A. A., Daudzai, Z., Cui, H., & Lin, L. (2023). Particle nanoarchitectonics for nanomedicine and nanotherapeutic drugs with special emphasis on nasal drugs and aging. *Biomedicines, 11*(2), 354. https://doi.org/10.3390/BIOMEDICINES11020354
9. Ippati, S., Deng, Y., van der Hoven, J., Heu, C., VanHummel, A., Chua, S. W., Paric, E., Chan, G., Feiten, A., Fath, T., Ke, Y. D., HaassIttner, N. K., & Ittner, L. M. (2021). Rapid initiation of cell cycle reentry processes protects neurons from amyloid-β toxicity. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/PNAS.2011876118
10. Tsujioka, Y., Takahashi, M., Tsuboi, Y., Yamamoto, T., & Yamada, T. (1999). Localization and expression of cdc2 and cdk4 in Alzheimer brain tissue. *Dementia and Geriatric Cognitive Disorders, 10*(3), 192–198. https://doi.org/10.1159/000017119
11. Frade, J. M., & Ovejero-Benito, M. C. (2015). Neuronal cell cycle: The neuron itself and its circumstances. *Cell Cycle, 14*(5), 712–720. https://doi.org/10.1080/15384101.2015.1004937
12. Dong, A., Moreno-Aliaga, J., Ramirez, M. J., García-Osta, A., Dong, J., Jesús Moreno-Aliaga, M., & Ramirez, M. J. (2022). p27, the cell cycle and Alzheimer´s disease. *International Journal of Molecular Sciences, 23*(3), 1211. https://doi.org/10.3390/IJMS23031211
13. Baker, S. J., Poulikakos, P. I., Irie, H. Y., Parekh, S., & Reddy, E. P. (2022). CDK4: a master regulator of the cell cycle and its role in cancer. *Genes & Cancer, 13*, 21. https://doi.org/10.18632/GENESANDCANCER.221
14. Takaki, T., Echalier, A., Brown, N. R., Hunt, T., Endicott, J. A., & Noble, M. E. M. (2009). The structure of CDK4/cyclin D3 has implications for models of CDK activation. *Proceedings of the National Academy of Sciences of the United States of America, 106*(11), 4171. https://doi.org/10.1073/PNAS.0809674106
15. Day, P. J., Cleasby, A., Tickle, I. J., O'Reilly, M., Coyle, J. E., Holding, F. P., McMenamin, R. L., Yon, J., Chopra, R., Lengauer, C., & Jhoti, H. (2009). Crystal structure of human CDK4 in complex with a D-type cyclin. *Proceedings of the National Academy of Sciences of the United States of America, 106*(11), 4166. https://doi.org/10.1073/PNAS.0809645106
16. Zhang, W., Liu, Y., Jang, H., & Nussinov, R. (2023). Cell cycle progression mechanisms: slower cyclin-D/CDK4 activation and faster cyclin-E/CDK2. *bioRxiv*. https://doi.org/10.1101/2023.08.16.553605
17. Rani, N., Sahu, M., Ambasta, R. K., & Kumar, P. (2024). Triaging between post-translational modification of cell cycle regulators

and their therapeutics in neurodegenerative diseases. *Ageing Research Reviews, 94*, 102174. https://doi.org/10.1016/J.ARR.2023.102174
18. García-Osta, A., Dong, J., Moreno-Aliaga, M. J., & Ramirez, M. J. (2022). p27, The cell cycle and Alzheimer´s disease. *International Journal of Molecular Sciences, 23*(3), 1211. https://doi.org/10.3390/IJMS23031211
19. Koch, J., Schober, S. J., Hindupur, S. V., Schöning, C., Klein, F. G., Mantwill, K., Ehrenfeld, M., Schillinger, U., Hohnecker, T., Qi, P., Steiger, K., Aichler, M., Gschwend, J. E., Nawroth, R., & Holm, P. S. (2022). Targeting the retinoblastoma/E2F repressive complex by CDK4/6 inhibitors amplifies oncolytic potency of an oncolytic adenovirus. *Nature Communications*. https://doi.org/10.1038/S41467-022-32087-5
20. Braal, C. L., Jongbloed, E. M., Wilting, S. M., Mathijssen, R. H. J., Koolen, S. L. W., & Jager, A. (2021). Inhibiting CDK4/6 in breast cancer with palbociclib, ribociclib, and abemaciclib: similarities and differences. *Drugs, 81*(3), 317. https://doi.org/10.1007/S40265-020-01461-2
21. Advani, D., Gupta, R., Tripathi, R., Sharma, S., Ambasta, R. K., & Kumar, P. (2020). Protective role of anticancer drugs in neurodegenerative disorders: A drug repurposing approach. *Neurochemistry International, 140*, 104841. https://doi.org/10.1016/J.NEUINT.2020.104841
22. Yuan, K., Wang, X., Dong, H., Min, W., Hao, H., & Yang, P. (2021). Selective inhibition of CDK4/6: A safe and effective strategy for developing anticancer drugs. *Acta Pharmaceutica Sinica B, 11*(1), 30–54. https://doi.org/10.1016/J.APSB.2020.05.001
23. Toogood, P. L., Harvey, P. J., Repine, J. T., Sheehan, D. J., VanderWel, S. N., Zhou, H., Keller, P. R., McNamara, D. J., Sherry, D., Zhu, T., Brodfuehrer, J., Choi, C., Barvian, M. R., & Fry, D. W. (2005). Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6. *Journal of Medicinal Chemistry, 48*(7), 2388–2406. https://doi.org/10.1021/jm049354h
24. Roskoski, R. (2016). Cyclin-dependent protein kinase inhibitors including palbociclib as anticancer drugs. *Pharmacological Research, 107*, 249–275. https://doi.org/10.1016/J.PHRS.2016.03.012
25. Ammazzalorso, A., Agamennone, M., De Filippis, B., & Fantacuzzi, M. (2021). Development of CDK4/6 inhibitors: A five years update. *Molecules*. https://doi.org/10.3390/MOLECULES26051488
26. Shinde, R. N., & Sobhia, M. E. (2013). Binding and discerning interactions of PTP1B allosteric inhibitors: Novel insights from molecular dynamics simulations. *Journal of Molecular Graphics and Modelling, 45*, 98–110. https://doi.org/10.1016/J.JMGM.2013.08.001
27. Ding, L., Cao, J., Lin, W., Chen, H., Xiong, X., Ao, H., Yu, M., Lin, J., & Cui, Q. (2020). The roles of cyclin-dependent kinases in cell-cycle progression and therapeutic strategies in human breast cancer. *International Journal of Molecular Sciences*. https://doi.org/10.3390/IJMS21061960
28. Guiley, K. Z., Stevenson, J. W., Lou, K., Barkovich, K. J., Kumarasamy, V., Wijeratne, T. U., Bunch, K. L., Tripathi, S., Knudsen, E. S., Witkiewicz, A. K., Shokat, K. M., & Rubin, S. M. (2019). P27 allosterically activates cyclin-dependent kinase 4 and antagonizes palbociclib inhibition. *Science*. https://doi.org/10.1126/SCIENCE.AAW2106
29. Asiamah, I., Obiri, S. A., Tamekloe, W., Armah, F. A., & Borquaye, L. S. (2023). Applications of molecular docking in natural products-based drug discovery. *Scientific African, 20*, e01593. https://doi.org/10.1016/J.SCIAF.2023.E01593
30. Naveed, M., Ali, N., Aziz, T., Hanif, N., Fatima, M., Ali, I., Alharbi, M., Alasmari, A. F., & Albekairi, T. H. (2024). The natural breakthrough: phytochemicals as potent therapeutic agents against spinocerebellar ataxia type 3. *Scientific Reports, 14*(1), 1–13. https://doi.org/10.1038/s41598-024-51954-3
31. Aziz, T., Qadir, R., Anwar, F., Naz, S., Nazir, N., Nabi, G., Haiying, C., Lin, L., Alharbi, M., & Alasmari, A. F. (2024). Optimal enzyme-assisted extraction of phenolics from leaves of pongamia pinnata via response surface methodology and artificial neural networking. *Applied Biochemistry and Biotechnology*. https://doi.org/10.1007/S12010-024-04875-w
32. Kim, J., Lee, H. J., & Lee, K. W. (2010). Naturally occurring phytochemicals for the prevention of Alzheimer's disease. *Journal of Neurochemistry, 112*(6), 1415–1430. https://doi.org/10.1111/J.1471-4159.2009.06562.X
33. Riaz, M., Nawaz, M., Qadir, R., Hussain, S., Roheen, T., Afzal, M., Perviaz, M., Ali, A., Aziz, T., Alharbi, M., Albekairi, T. H., & Alasmari, A. F. (2023). Characterization and antioxidant potential of white mustard (Brassica hirta) leaf extract and stabilization of sunflower oil. *Open Chemistry*. https://doi.org/10.1515/CHEM-2023-0175
34. Maggiora, G., Vogt, M., Stumpfe, D., & Bajorath, J. (2014). Molecular similarity in medicinal chemistry. *Journal of Medicinal Chemistry, 57*(8), 3186–3204. https://doi.org/10.1021/jm401411z
35. Hameed, A. R., Ali, S. F., Alsallameh, S. M. S., Muhseen, Z. T., Almansour, N. M., Alsuhaymi, N., & Allemailem, K. S. (2023). Structural dynamics of P-Rex1 complexed with natural leads establishes the protein as an attractive target for therapeutics to suppress cancer metastasis. *BioMed Research International*. https://doi.org/10.1155/2023/3882081
36. Abola, E. E., Manning, N. O., Prilusky, J., Stampf, D. R., & Sussman, J. L. (1996). The Protein Data Bank: Current status and future challenges. *Journal of Research of the National Institute of Standards and Technology, 101*(3), 231. https://doi.org/10.6028/jres.101.025
37. Yousif, F. A., Alzain, A. A., Alraih, A. M., & Ibraheem, W. (2023). Repurposing of approved drugs for targeting CDK4/6 and aromatase protein using molecular docking and molecular dynamics studies. *PLoS ONE, 18*(9), e0291256. https://doi.org/10.1371/JOURNAL.PONE.0291256
38. Anifowose, L. O., Paimo, O. K., Ogunyemi, O. M., Akanos, R. O., Hammad, S. F., & Ghazy, M. A. (2023). Molecular docking appraisal of dysphania ambrosioides phytochemicals as potential inhibitor of a key triple-negative breast cancer driver gene. *In Silico Pharmacology*. https://doi.org/10.21203/RS.3.RS-2854790/V1
39. Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: collection of open natural products database. *Journal of Cheminformatics, 13*(1), 1–13. https://doi.org/10.1186/S13321-020-00478-9/FIGURES/4
40. Kuwahara, H., & Gao, X. (2021). Analysis of the effects of related fingerprints on molecular similarity using an eigenvalue entropy approach. *Journal of Cheminformatics, 13*(1), 27. https://doi.org/10.1186/s13321-021-00506-2
41. Bajusz, D., Rácz, A., & Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics, 7*(1), 1–13. https://doi.org/10.1186/S13321-015-0069-3/FIGURES/7
42. Khan, S., Hussain, R., Khan, Y., Iqbal, T., Aziz, T., & Alharbi, M. (2024). Correlation between in vitro anti-urease activity and in silico molecular modeling approach of novel imidazopyridine–oxadiazole hybrids derivatives. *Open Chemistry*. https://doi.org/10.1515/CHEM-2023-0210
43. Naveed, M., Abid, A., Aziz, T., Saleem, A., Hanif, N., Ali, I., & Alasmari, A. F. (2024). Comparative toxicity assessment of fisetin-aided artificial intelligence-assisted drug design targeting epibulbar dermoid through phytochemicals. *Open Chemistry*. https://doi.org/10.1515/CHEM-2023-0197

44. Wu, F., Zhou, Y., Li, L., Shen, X., Chen, G., Wang, X., Liang, X., Tan, M., & Huang, Z. (2020). Computational approaches in preclinical studies on drug discovery and development. *Frontiers in Chemistry, 8*, 546712. https://doi.org/10.3389/FCHEM.2020.00726/BIBTEX
45. Luo, X., Zhao, Y., Tang, P., Du, X., Li, F., Wang, Q., Li, R., & He, J. (2021). Discovery of new small-molecule cyclin-dependent kinase 6 inhibitors through computational approaches. *Molecular Diversity, 25*(1), 367–382. https://doi.org/10.1007/S11030-020-10120-3/FIGURES/7
46. Sarma, S., Dowerah, D., Basumatary, M., Phonglo, A., & Deka, R. C. (2023). Inhibitory potential of furanocoumarins against cyclin dependent kinase 4 using integrated docking, molecular dynamics and ONIOM methods. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2023.2300755
47. Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*. https://doi.org/10.1038/srep42717
48. Naveed, M., ul Ain, N., Aziz, T., Saleem, A., Aqib Shabbir, M., Ali Khan, A., & Albekairi, T. H. (2024). Integrated track of nano-informatics coupling with the enrichment concept in developing a novel nanoparticle targeting ERK protein in Naegleria fowleri. *Open Chemistry*. https://doi.org/10.1515/CHEM-2023-0198
49. Blay, V., Tolani, B., Ho, S. P., & Arkin, M. R. (2020). High-throughput screening: Today's biochemical and cell-based approaches. *Drug Discovery Today, 25*(10), 1807–1821. https://doi.org/10.1016/J.DRUDIS.2020.07.024
50. Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., Connolly, D. T., & Shoichet, B. K. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *Journal of Medicinal Chemistry, 45*(11), 2213–2221. https://doi.org/10.1021/JM010548W
51. Eynde, V., Barretta, R., Kralj, S., Jukič, M., & Bren, U. (2023). Molecular filters in medicinal chemistry. *Encyclopedia, 3*(2), 501–511. https://doi.org/10.3390/ENCYCLOPEDIA3020035
52. Chauhan, A., Singh, J., Sangwan, N., Singh, H., Prakash, A., Medhi, B., & Avti, P. K. (2023). Designing the 5HT2BR structure and its modulation as a therapeutic target for repurposing approach in drug-resistant epilepsy. *Epilepsy Research, 194*, 107168. https://doi.org/10.1016/J.EPLEPSYRES.2023.107168
53. Kochnev, Y., Hellemann, E., Cassidy, K. C., & Durrant, J. D. (2020). Webina: An open-source library and web app that runs AutoDock Vina entirely in the web browser. *Bioinformatics, 36*(16), 4513–4515. https://doi.org/10.1093/BIOINFORMATICS/BTAA579
54. Sharma, S., & Kumar, P. (2023). Dissecting the functional significance of HSP90AB1 and other heat shock proteins in countering glioblastomas and ependymomas using omics analysis and drug prediction using virtual screening. *Neuropeptides, 102*, 102383. https://doi.org/10.1016/J.NPEP.2023.102383
55. Jiang, C., & Jin, X. (2017). Quick way to port existing C/C++ Chemoinformatics toolkits to the web using emscripten. *Journal of Chemical Information and Modeling, 57*(10), 2407–2412. https://doi.org/10.1021/ACS.JCIM.7B00434
56. Wang, Z., Pan, H., Sun, H., Kang, Y., Liu, H., Cao, D., & Hou, T. (2022). fastDRH: A webserver to predict and analyze protein–ligand complexes based on molecular docking and MM/PB(GB)SA computation. *Briefings in Bioinformatics, 23*(5), 1–10. https://doi.org/10.1093/BIB/BBAC201
57. Fatima, I., Alshabrmi, F. M., Aziz, T., Alamri, A. S., Alhomrani, M., Alghamdi, S., Alghuraybi, R. A., Babalghith, A. O., Bamagous, G. A., Alhindi, Z., Dablool, A. S., Alhhazmi, A. A., & Alruways, M. W. (2024). Revolutionizing and identifying novel drug targets in Citrobacter koseri via subtractive proteomics and development of a multi-epitope vaccine using reverse vaccinology and immuno-informatics. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2024.2316762
58. Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling, 61*(8), 3891–3898. https://doi.org/10.1021/ACS.JCIM.1C00203/SUPPL_FILE/CI1C00203_SI_002.ZIP
59. Dutta, K., Shityakov, S., & Maruyama, F. (2023). DSF inactivator RpfB homologous FadD upregulated in Bradyrhizobium japonicum under iron limiting conditions. *Scientific Reports*. https://doi.org/10.1038/s41598-023-35487-9
60. Bharatham, N., Chi, S.-W., Sup Yoon, H., & Uversky, V. N. (2011). Molecular basis of Bcl-XL-p53 interaction: Insights from molecular dynamics simulations. *PLoS ONE, 6*(10), 26014. https://doi.org/10.1371/journal.pone.0026014
61. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry, 26*(16), 1701–1718. https://doi.org/10.1002/JCC.20291
62. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research, 51*(D1), D418–D427. https://doi.org/10.1093/NAR/GKAC993
63. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science, 27*(1), 129–134. https://doi.org/10.1002/PRO.3289
64. Islam, R., Rahaman, M., Hoque, H., Hasan, N., Prodhan, S. H., Ruhama, A., & Jewel, N. A. (2021). Computational and structural based approach to identify malignant nonsynonymous single nucleotide polymorphisms associated with CDK4 gene. *PLoS ONE*. https://doi.org/10.1371/JOURNAL.PONE.0259691
65. Zahoránszky-Kőhalmi, G., Bologa, C. G., & Oprea, T. I. (2016). Impact of similarity threshold on the topology of molecular similarity networks and clustering outcomes. *Journal of Cheminformatics, 8*(1), 16. https://doi.org/10.1186/s13321-016-0127-5
66. Naveed, M., Ishfaq, H., Rehman, S. U., Javed, A., Waseem, M., Makhdoom, S. I., Aziz, T., Alharbi, M., Alshammari, A., & Alasmari, A. F. (2023). GC–MS profiling of Bacillus spp. metabolites with an in vitro biological activity assessment and computational analysis of their impact on epithelial glioblastoma cancer genes. *Frontiers in Chemistry, 11*, 1287599. https://doi.org/10.3389/FCHEM.2023.1287599/BIBTEX
67. Kumar, V., Parate, S., Thakur, G., Lee, G., Ro, H. S., Kim, Y., Kim, H. J., Kim, M. O., & Lee, K. W. (2021). Identification of cdk7 inhibitors from natural sources using pharmacoinformatics and molecular dynamics simulations. *Biomedicines, 9*(9), 1197. https://doi.org/10.3390/BIOMEDICINES9091197/S1
68. Alov, P., Stoimenov, H., Lessigiarska, I., Pencheva, T., Tzvetkov, N. T., Pajeva, I., & Tsakovska, I. (2022). In silico identification of multi-target ligands as promising hit compounds for neurodegenerative diseases drug development. *International Journal of Molecular Sciences, 23*(21), 13650. https://doi.org/10.3390/ijms232113650

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

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

Check for updates

View Article Online
View Journal

RSC Sustainability

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: K. K. Tin, W. Taweepreda, A. Singh, N. K. Wagri and A. Kumar, *RSC Sustain.*, 2024, DOI: 10.1039/D4SU00356J.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

Sustainabilty Spotlight Statement

View Article Online
DOI: 10.1039/D4SU00356J

Using fallen leaves for hydrogen production supports SDG 7 by providing reliable, affordable, and modern energy. Despite improvements, the Energy Progress Report 2024 highlights the shortfall in meeting SDG 7 targets by 2030. Expanding energy access, scaling renewables, enhancing efficiency, and increasing financial flows to developing countries are vital. Hydrogen from waste organic matter is crucial for decarbonization and achieving net-zero emissions by 2050, generating clean energy and aiding waste management. Green hydrogen promotes environmental sustainability, economic growth, and social progress, particularly in developing countries, aligning with the UN's Sustainable Development Goals.

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.



RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Fallen Leaves to Sustainable Energy Solution: Review on Hydrogen Production

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Kyu Kyu Tin[1]

[1]Environmental Management, Faculty of Environmental Management, Prince of Songkla University, Hat-Yai, Songkhla 90110, Thailand

ORCID ID: 0000-0002-8620-7197

Email: 6410930099@email.psu.ac.th

Wirach Taweepreda[2]

[2]Polymer Science Program, Division of Physical Science, Faculty of Science, Prince of Songkla University, Hat-Yai, Songkhla 90110, Thailand

ORCID ID: 0000-0002-0838-6353

Email: wirach.t@psu.ac.th

Akanksha Singh[3]

[3]Department of Zoology, Swami Shraddhanand College, University of Delhi, Delhi-110036, India

ORCID ID: 0009-0008-6704-5218

akanksha@ss.du.ac.in

Naresh Kumar Wagri[4]

[4]Department of Material Science and Engineering, KTH Royal Institute of Technology, Stockholm, Sweden SE-10044,

ORCID ID: 0000-0002-3342-6257

Email: wagri@kth.se

Anil Kumar[5,6]*

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

[6]Division - Clean Energy: Nodal Centre of Excellence in Energy Transition (NCEET), Delhi Technological University, Delhi-110042, India

ORCID ID: 0000-0002-2846-4347

Email: anilkumar76@dtu.ac.in

***Corresponding Author:** Anil Kumar

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Abstract

Fallen leaves represent a significant feedstock for hydrogen production due to their high cellulose content, abundance, and minimal sulfur content. These characteristics make them suitable for various hydrogen production technologies, including biohydrogen and biomass-derived liquid reforming processes, contributing to sustainable energy production. This comprehensive literature review explores fallen leaves as a low-cost biomass feedstock for hydrogen production in the pursuit of zero-carbon and sustainable energy solutions. Steam methane reforming, while cost-effective and having high production capacity, results in substantial carbon emissions. Electrolysis, leveraging renewable resources, is attractive but requires significant energy input. Biomass gasification and thermochemical processes show promise for sustainable hydrogen production, though further technological advancements are necessary. Additionally, anaerobic fermentation by microorganisms can directly produce hydrogen from biomass (including fallen leaves), offering an energy-efficient method that utilizes organic waste. This review evaluates hydrogen production concerning energy efficiency, economics, and environmental impact. The findings contribute to the global transition from fossil fuels to renewable energy sources, aligning with climate commitments and the goal of carbon neutrality.

Keywords: Fallen leaves; Hydrogen production; Economics; Environment; Sustainability

1. Introduction

To survive the winter months, trees shed their leaves each fall. At the base of the leaf stem, an abscission layer forms, creating a dense layer of cells that cuts off the veins supplying nutrients and water to the leaves. This causes the stem to weaken, the leaf to die, and eventually fall from the tree. The colors of fallen leaves—orange, red, yellow, and brown—depend on leaf pigments, precipitation, temperature, and daylight length.

Once on the ground, fallen leaves start to decompose with the help of animals, fungi, and bacteria, transforming into simple water and carbon dioxide, fundamental components of the carbon cycle. [1]However, when dead leaves are converted into fertilizer, they release greenhouse gases such as methane, carbon dioxide, and nitrous oxide. If fallen leaves enter waterways, they can degrade water quality, harm aquatic life, and obstruct sunlight. Leaves in water demand 200% more oxygen compared to those on land.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Traditional disposal methods for dead leaves include landfilling, composting, and incineration. Landfilling emits refractory leachate and methane, composting is time-consuming, and incineration produces vast amounts of noxious particulates, gases, and carbon dioxide[2] (Fig. 1). Achieving sustainable development requires meeting human needs without harming the environment. Considering the drawbacks of traditional techniques, developing effective and ecological approaches for managing fallen leaves is highly attractive.

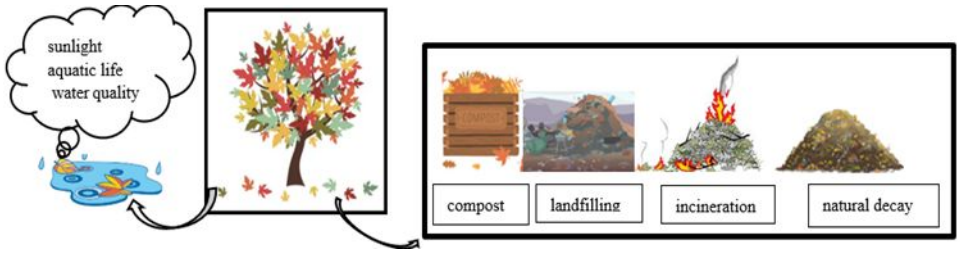


Figure 1. Waste dead leaves and its consequences

The scope of this literature review is to examine hydrogen production through the utilization of biomass, particularly fallen leaves. Often regarded as waste, fallen leaves are renewable and abundant, and can be converted into hydrogen through various processes, supporting sustainable and eco-friendly hydrogen generation technologies. This review will explore these processes, their efficiencies, and their environmental impacts, providing a comprehensive overview of this promising sector. The objectives of this literature review (Fig. 2) include:

- Understanding the role of fallen leaves or biomass in hydrogen production and the fundamental principles involved.
- Evaluating different methods of hydrogen production, analyzing their efficiencies and environmental effects.
- Identifying the limitations and challenges associated with these methods, particularly when using fallen leaves.
- Highlighting recent developments and advancements in hydrogen production.
- Discussing prospects and future research directions in this sector.

Using fallen leaves to produce hydrogen can contribute significantly to the transition to a more sustainable energy system.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

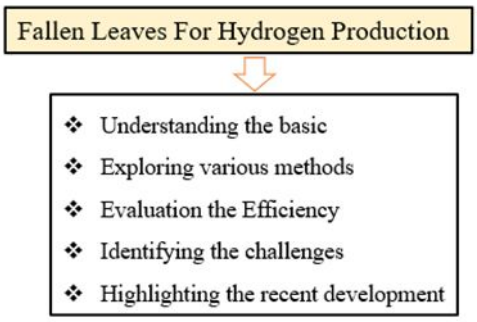


Figure 2. Objectives of review

The highlights of this study emphasize the potential of converting dead leaves, often considered waste, into valuable energy through hydrogen production. This underscores the importance of fallen leaves as a renewable and sustainable feedstock for biohydrogen production, exploring various methods to produce renewable energy with a focus on sustainability. Utilizing fallen leaves in energy processes not only reduces waste but also promotes sustainability, although the installation of such systems faces economic, technical, and environmental challenges. The novelty of this research lies in the largely unexplored potential of dead leaves as a feedstock for hydrogen production, particularly through steam gasification and the integration of advanced methods for utilizing the generated hydrogen. This approach emphasizes environmental friendliness and sustainability, contributing to the global transition towards renewable energy sources. To enhance the introduction, a clearer statement of the novelty and specific contributions compared to existing literature should be included. The utilization of dead leaves as a feedstock for hydrogen production represents a novel approach that has not been extensively explored in the existing literature. This study highlights the potential of steam gasification combined with advanced methods to efficiently convert fallen leaves into hydrogen. The literature review emphasizes the environmental benefits and sustainability of this approach, contributing to the global transition towards renewable energy sources. By addressing the economic, technical, and environmental challenges associated with this process, the study provides valuable insights and practical solutions that can advance the field of renewable energy. This work not only integrates existing knowledge but also introduces new perspectives on the use of biomass for sustainable energy production, making a significant contribution to the literature on renewable energy and waste management.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

2. Innovative Energy Solutions from Fallen Leaves

Instead of discarding fallen leaves, numerous methods exist to reprocess them. Currently, dead leaves are used for synthesizing graphene, [3] carbon materials for adsorbing heavy metals and dyes, [4] batteries,[5]supercapacitors,[6]feedstock for biogas production,[7,8] pellets,[9]shrimp preservation,[10]packaging,[11] textiles (fabric, natural dyes, printing on fabric, leather),[12,13] cellulose ethanol,[14] carbon nanotubes,[15] preparation of bio-based reaction mediums, active multifunctional materials such as photocatalysts, bioplastics, and evaporators.[16]

These examples highlight environmentally friendly and innovative ways to utilize dead leaves for energy generation, agricultural purposes, and material creation, contributing to a more circular and sustainable economy. By adopting these methods, we can achieve greater environmental sustainability through the economical treatment of large volumes of fallen leaves (Fig. 3).

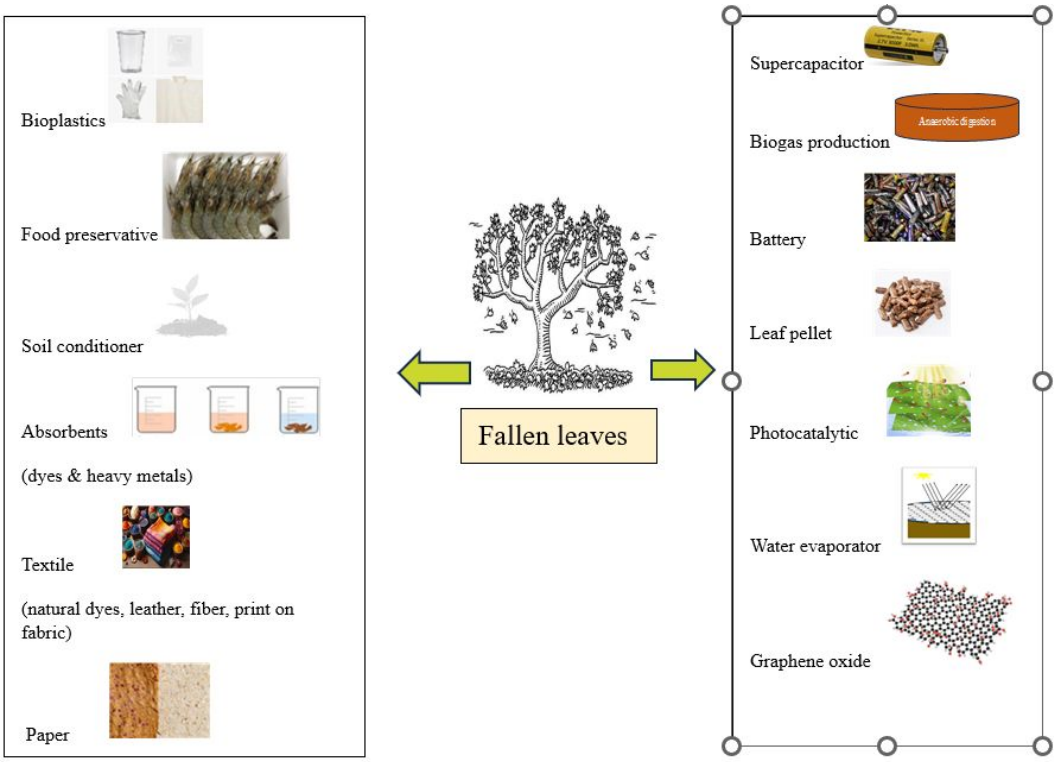


Figure 3. Application of fallen leaves

The mechanism of converting fallen leaves into energy encompasses various methods, including turning them into organic capacitors, utilizing them as feedstock for biogas generation, and

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

transforming them into other energy storage devices such as supercapacitors and batteries. One promising alternative to traditional capacitors is organic capacitors, achieved by introducing a composite material made from graphene oxide and leaf pellets. Graphene oxide provides excellent electrical conductivity, while the leaf pellets offer a high surface area. The combination results in a supercapacitor with both high power density and energy density. This method has been successfully demonstrated by Chinese scientists who innovated supercapacitors from fallen autumn leaves. [17,18]

Moreover, using fallen leaves as feedstock for biogas generation involves breaking down the leaves in an anaerobic digester to generate biogas—a mixture of carbon dioxide and methane—that can be used as fuel. This process helps reduce air pollution by preventing the burning of leaves and contributes to a more sustainable energy solution. Additionally, fallen leaves can act as a natural photocatalyst under sunlight to split water into oxygen and hydrogen, providing a sustainable method to produce hydrogen fuel using solar energy. Furthermore, fallen leaves can be converted into pellets, offering a transportable and compact form of biomass-derived energy, thereby contributing to the circular economy.

Converting fallen leaves into energy can have both economic and environmental advantages (Fig. 4). Environmentally, it supports more efficient waste management by reducing leaf burning and decay. For instance, creating organic capacitors from leaves emits significantly less carbon dioxide, while using dead leaves as feedstock for biogas generation decreases the amount of organic waste in landfills. Economically, in Amiens, France, a mechanization factory converts fallen leaves into biogas to power homes in the region. The more energy and biogas generated at the plant, the lower the community's treatment costs, with the potential for buses to run entirely on biogas in the future, reducing reliance on electricity imports and fossil fuels.[19,20]

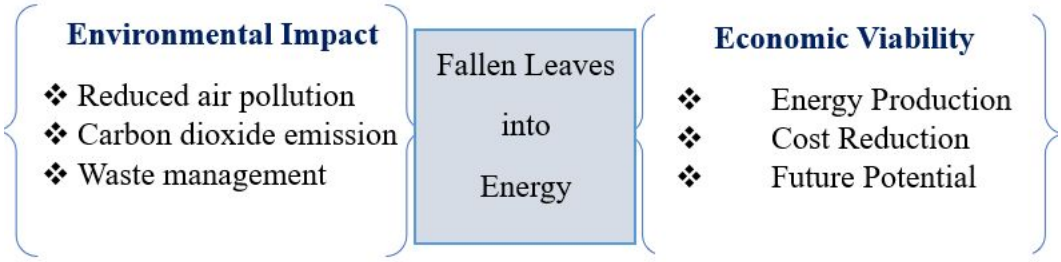


Figure 4. Perspectives of environment and economic turning fallen leaves into energy

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

The quest for efficient, clean, and sustainable energy production is one of the most pressing challenges, with hydrogen often regarded as the "fuel of the future." When utilized in fuel cells, hydrogen generates water as the only byproduct, making it a truly clean energy source. However, hydrogen production is a complex process and often relies on fossil fuels, which can mitigate its environmental benefits. Various methods exist, including conventional ones like electrolysis and steam methane reforming (SMR), as well as renewable methods such as thermolysis, biological methods, and photolysis.[21,22] Each method has its strengths and weaknesses. SMR is cost-effective but requires carbon capture for emission reduction, while electrolysis produces renewable hydrogen but is energy-intensive. Renewable technologies like biological and photolysis methods show promise for sustainable hydrogen generation. These diverse strategies contribute to the evolving landscape of hydrogen production, paving the way for a cleaner, greener future.[23-25]

2. Advancement in Hydrogen Production from Fallen Leaves

The utilization of biomass, particularly fallen leaves, often recognized as waste, is renewable and abundant and can be transformed into hydrogen gas through various processes, promoting eco-friendly and sustainable methods for hydrogen production. The most interesting methods for turning dead leaves into hydrogen involve biohydrogen production and multifunctional development (Table 1). Fallen leaves, especially from species like Mango, Mahogany, and Teak, have a high potential for biogas and biohydrogen production due to their favorable chemical composition. Co-fermentation of fallen leaves with sewage sludge provides a feasible technique for sustainable hydrogen production using waste biomass.

The water extracts from teak leaf ash (WET) introduce a green material for utilizing fallen leaves in chemical processes. Scientists in Germany have demonstrated the use of autumn tree leaves as feedstock for hydrogen production.[26] The optimal mixing ratio of fallen leaves to sewage sludge was observed to be 20:80 (on a volatile solids basis), yielding 37.8 mL/g-VS_{added} of biohydrogen, which is significantly higher compared to the mono-fermentation of fallen leaves (30.5 mL/g-VS_{added}) or the mono-fermentation of sewage sludge (10.3 mL/g-VS_{added}). [27,28]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 1. Hydrogen production methods from fallen leaves

Methods	Advantages	Disadvantages	References
Multifunctional Material Development	Diversifies the utilization of dead leaves beyond conventional waste	Based on the specific processing methods and applications, yield and efficiency may vary; Need specialized equipment	[12]
Biohydrogen Production	Co-fermentation of fallen leaves with sewage sludge offers a feasible method for hydrogen generation. Contributes to sustainable energy production by utilizing waste biomass for hydrogen production.	The process may require careful optimization to achieve high hydrogen yields. Challenges related to scalability and cost-effectiveness may need to be addressed for large-scale implementation.	[29]

Under visible light and sunlight irradiation, the active multifunctional material derived from fallen leaves exhibits high performance in photocatalytic hydrogen production. Hydrogen production rate over the active material reached 12.4 μmol h^{-1} cm^{-2} under simulated sunlight (1kWm^{-2}), which is 4.2 times higher than that of the original fallen leaves. Even under visible light ($\lambda > 400$ nm, 0.85 kW m^{-2}), the hydrogen generation rate reached 8.4 μmol h^{-1} cm^{-2}, demonstrating the material's visible-light-responsive photocatalytic activity.

The conversion of fallen leaves into hydrogen as biobased reaction materials primarily involves thermochemical and biological processes (Table 2). Enzymatic hydrolysis, utilizing ligninases and cellulases, efficiently breaks down fallen leaves into sugars for hydrogen production, preserving sugar integrity and avoiding inhibitory byproducts. However, challenges such as enzyme stability and cost remain. Advances in enzyme technology and biomass pre-treatment methods aim to increase sugar yields and hydrogen production potential.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

In anaerobic digestion, various factors influence the rate of hydrogen production, including microbial community composition, biomass composition, and digester conditions. Strategies to enhance hydrogen concentration include inhibiting hydrogen-consuming microorganisms and continuously removing hydrogen to maximize production efficiency.

Photo-fermentation shows promise for sustainable hydrogen production but faces challenges in economic viability and scalability. Optimization of biomass pre-treatment and fermentation conditions—such as substrate concentration, pH, and fermentation modes—is ongoing to improve competitiveness for large-scale hydrogen production.

Dark fermentation offers a flexible biohydrogen production technique that operates without light and can utilize various biomass sources. However, separating hydrogen from other gases generated adds complexity and cost. Research focuses on enhancing yield and efficiency through process optimization and genetic engineering.[30-33]

Supercritical water gasification (SCWG), pyrolysis, and gasification are advanced techniques for hydrogen production from biomass. Gasification is efficient for large-scale applications but requires sophisticated equipment and high temperatures. Pyrolysis offers flexibility and produces multiple products but requires further purification for hydrogen extraction. SCWG shows promise for wet biomass and generates hydrogen-rich syngas but requires ongoing development to address complexity, cost, and sustainability challenges.

Each method represents a pathway toward sustainable hydrogen production, with ongoing research aimed at enhancing efficiency, reducing costs, and advancing environmental sustainability[34-37].

Collecting fallen leaves presents several challenges, including the effort and time required for raking and collection, the environmental impact of using power tools or disposing of leaves in landfills, and potential negative effects on wildlife habitats and soil health if leaves are removed from gardens. However, there are also advantages to collecting fallen leaves, such as reducing landfill waste, improving soil health by adding organic matter, and preventing the smothering of small plants in garden beds.

Promoting sustainable leaf disposal and collection involves implementing mechanical leaf collection in areas with heavy leaf accumulation, organizing drop-off events for bagged leaves, and implementing composting programs to divert organic waste from landfills. These practices

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

aim to enhance waste management practices, support sustainable energy production, and preserve wildlife habitats and local ecosystems.

Table 2. Techniques for the conversion of fallen leaves into hydrogen

Method	Advantages	Disadvantages	References
Enzymatic hydrolysis	High specificity for lignin and cellulose breakdown; Operated under mild circumstances; Produces fermentable sugars for hydrogen generation	Expensive enzymes; Enzymes challenges under operational conditions; Require pretreatment biomass to enhance efficiency	[30]
Anaerobic digestion	Low energy and low-cost requirements; Apply organic waste as feedstock; Generate biogas with high hydrogen composition	Slow rate of hydrogen generation compared to other techniques; Require pre-treatment of biomass; Process optimization needed for higher yields	[31]
Photo-fermentation	Uses renewable solar energy; Clean process with no toxic emission; Can be combined with waste management	Require consistent and strong light resources; Basically, lower hydrogen yields than other methods; Needs artificial lighting systems	[32]
Dark-fermentation	No need for light, faster than anaerobic digestion, can handle different conditions; Can process a wide range of biomass types	Separating hydrogen from other gases generated can be complicated; Lower hydrogen yields than other methods; By-products may need further treatment	[33]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Supercritical water gasification	Process wet biomass without the need for drying; Generates hydrogen-rich syngas; High gasification efficiency and hydrogen selectivity	Still in the experimental stage Needs high temperature and pressure; Needs further improvement to become commercially viable	[34][41][42]
Pyrolysis	Can use a wide range of biomass types; Produces multiple valuable products (bio-oil, char, hydrogen); Flexible and can be combined into existing waste management systems	The quality of hydrogen can be different, requiring purification; High temperatures needed that may enhance energy consumption; The equipment and process can be costly and complex	[35]
Gasification	Efficient conversion of large amounts of biomass materials; Produces a versatile syngas reasonable for different applications; Can combine with carbon capture for cleaner energy	Requires high temperatures, enhancing energy input; Requires sophisticated and expensive; May generate greenhouse gases if not managed properly	[36][37]

To provide a comprehensive understanding of hydrogen production, it is essential to compare various methods in terms of operating conditions, performance metrics, and challenges. Steam Methane Reforming (SMR) operates at high temperatures (700-1,000°C) and pressures, yielding high amounts of hydrogen but also significant CO_2 emissions, as it requires fossil fuels. Electrolysis, which splits water into hydrogen and oxygen using electricity, produces high-purity hydrogen, with efficiency depending on the electricity source; however, it faces challenges of high energy consumption and cost variability based on electricity prices. Biomass gasification operates at high temperatures (800-1,000°C) with a gasifying agent, offering a renewable hydrogen source with moderate efficiency, but it is challenged by feedstock variability, tar formation, and gas cleanup requirements. Photocatalytic water splitting uses sunlight and photocatalysts to split water,

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

currently exhibiting low efficiency and still in experimental stages, with the main challenge being the development of efficient and stable photocatalysts.[32] Hydrogen production from fallen leaves involves processes like pyrolysis or fermentation, utilizing waste biomass and offering potential for carbon-neutral production, but it faces challenges in technology development, feedstock collection, and processing. By comparing these methods, the manuscript can highlight the advantages and limitations of each approach, providing a clearer picture of the feasibility and sustainability of hydrogen production from fallen leaves.[35]

Promoting sustainable leaf collection and disposal involves implementing mechanical leaf collection in areas with heavy leaf accumulation, organizing drop-off events for bagged leaves, and implementing composting programs to reduce landfill waste. These initiatives can support sustainable energy production and improve waste management practices, while also benefiting local ecosystems and wildlife habitats. [38-40]

The advantages of using fallen leaves for hydrogen production, compared to other renewable biomass sources, include availability, reduced waste, lower water usage, enhanced hydrogen generation efficiency, versatility, lower energy input, and reduced greenhouse gas emissions. Fallen leaves are widely available, making them readily accessible for biomass-based hydrogen production. Their utilization can decrease waste and enhance rural livelihoods, contributing to overall energy security.

Dry anaerobic fermentation of fallen leaves requires less water compared to wet anaerobic fermentation, making it a more sustainable and efficient process. Additionally, synergistic dry fermentation of fallen leaves with other waste materials, such as garden or food waste, can further enhance hydrogen generation efficiency. This approach offers flexibility in hydrogen production methods, either using fallen leaves as a sole substrate or co-fermenting them with other waste materials.

Moreover, fallen leaves require lower energy input compared to other biomass sources, making them a cost-effective and sustainable option for hydrogen production. Utilizing sustainable and renewable biomass sources also contributes to reducing greenhouse gas emissions.[10]

The age of fallen leaves can affect their suitability for hydrogen production. Older leaves can more easily convert into simple sugars, but as leaves age, their lignin content increases, making them more resistant to pretreatment. Components like xylan, cellulose, and arabinan (a type of

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

hemicellulose in plant cell walls) are crucial for hydrogen production and can be transformed into simple sugars through pretreatment processes.[11,12]

The following documents collectively provide a comprehensive overview of various hydrogen production methods, highlighting their advantages, limitations, and potential for integration with sustainable practices (Table 3).

Table 3. Author's contributions to various hydrogen production methods

Document	Contribution	Review Focus	References
A concise review of recent biohydrogen production technologies	This review provides an overview of the latest advancements in biohydrogen production technologies, highlighting the efficiency, scalability, and sustainability of different methods.	The paper discusses various biohydrogen production methods, including dark fermentation, photo fermentation, and microbial electrolysis cells, emphasizing their potential for sustainable energy production.	[41]
Hydrogen Roadmap Europe: A Sustainable Pathway for the European Energy Transition	This roadmap outlines a strategic plan for the large-scale deployment of hydrogen and fuel cells across Europe, aiming to achieve a sustainable energy transition.	The document focuses on the socio-economic impacts of hydrogen adoption, the necessary infrastructure developments, and policy recommendations to support the hydrogen economy.	[42]
A comprehensive review on hydrogen production	This comprehensive review analyzes the current	The review covers various hydrogen	[43]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

and utilization in North America: Prospects and challenges	state and future prospects of hydrogen production and utilization in North America, addressing technological, economic, and policy challenges.	production methods, including steam methane reforming, electrolysis, and biomass gasification, and discusses their feasibility, efficiency, and environmental impacts.	
Feasibility Analysis of Green Hydrogen Production from Oceanic Energy	This study evaluates the technical, economic, and environmental feasibility of producing green hydrogen using oceanic energy sources, such as offshore wind and wave energy.	The analysis includes a detailed assessment of the potential for integrating oceanic energy with hydrogen production technologies, highlighting the benefits and challenges of this approach.	[44]
Comparison of biohydrogen production processes	This paper compares different biohydrogen production processes, providing insights into their efficiency, scalability, and environmental impacts.	The comparison includes dark fermentation, photo fermentation, two-stage processes, and biocatalyzed electrolysis, with a focus on optimizing conditions for maximum hydrogen yield.	[45]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Hydrogen production from biomass through integration of anaerobic digestion and biogas dry reforming	This review explores the integration of anaerobic digestion and biogas dry reforming for hydrogen production, emphasizing the potential for utilizing biomass waste.	The paper discusses the technical and economic aspects of combining these processes, including the challenges of feedstock variability and gas cleanup requirements.	[46]
Hydrogen production from phototrophic microorganisms: reality and perspectives	This review examines the potential of phototrophic microorganisms for hydrogen production, discussing the current state of research and future perspectives.	The focus is on the biological mechanisms of hydrogen production in phototrophic microorganisms, the efficiency of different species, and the challenges of scaling up these processes for industrial applications.	[47]

To evaluate the hydrogen production potential from fallen leaves, a detailed methodology is essential. Initially, fallen leaves are collected from various locations to ensure a diverse sample, sorted to remove non-leaf material, air-dried to reduce moisture content, and ground into a fine powder to increase surface area for better reaction efficiency. The leaf biomass is then characterized through proximate analysis to determine moisture, volatile matter, ash content, and fixed carbon, ultimate analysis to measure elemental composition (carbon, hydrogen, nitrogen, sulfur, and oxygen), and calorific value assessment using a bomb calorimeter. For hydrogen production, two processes are employed: pyrolysis and fermentation. In pyrolysis, a reactor heats the leaf biomass in the absence of oxygen, gradually increasing the temperature to 500-700°C, with gaseous products collected and analyzed for hydrogen content using gas chromatography. In fermentation, anaerobic fermentation is conducted in a bioreactor with specific microbial cultures,

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

where the leaf biomass is mixed with water and inoculated with microbes, maintaining a temperature of 30-37°C, and the biogas produced is collected and analyzed for hydrogen content using gas chromatography. Data analysis involves calculating the hydrogen yield per unit mass of leaf biomass, evaluating the energy efficiency by comparing the energy content of the produced hydrogen with the energy input and assessing the environmental impact by analyzing by-products and emissions. Optimization is achieved by varying parameters such as temperature, pressure, and microbial culture to optimize hydrogen yield, and using statistical tools to analyze the data and determine the optimal conditions for maximum hydrogen production. This comprehensive methodology ensures a clear and detailed evaluation of the hydrogen production potential from fallen leaves.[48]

To determine the hydrogen yield from a known quantity of leaf biomass, follow this detailed process:

Step 1: Weigh the Leaf Biomass

Measure the initial mass of the dried and ground leaf biomass, denoting it as $m_{biomass}$ (in grams)

Step 2: Collect the Produced Hydrogen

During the hydrogen production process (such as pyrolysis or fermentation), collect the hydrogen gas produced. Ensure that the collection system is sealed and calibrated to prevent gas loss.

Step 3: Measure the Volume of Hydrogen

Utilize a gas collection system, like a gas syringe or flow meter, to measure the volume of hydrogen gas produced, denoted as V_{H_2}(in liters).

Step 4: Standardize the Conditions

Make sure the volume of hydrogen gas is measured under standard temperature and pressure (STP) conditions (0°C and 1 atm). If measurements are taken under different conditions, convert the measured volume to STP using the ideal gas law.

Step 5: Calculate the Moles of Hydrogen

Use the ideal gas law, PV = nRT, to calculate the number of moles of hydrogen gas produced[58]. Rearranging gives:

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

$$n = \frac{PV}{RT} \quad (1)$$

At STP, where P=1 atm and T=273.15, this simplifies to:

$$n = \frac{V_{H_2}}{22.414} \quad (2)$$

where 22.414 L is the molar volume of an ideal gas at STP.

Step 6: Calculate the Hydrogen Yield

Hydrogen yield is the amount of hydrogen produced per unit mass of leaf biomass, calculated using the formula:

$$\text{Hydrogen Yield} = \frac{n \times M_{H_2}}{m_{biomass}} \quad (3)$$

where:

- n is the number of moles of hydrogen
- $M_{H_2}$is the molar mass of hydrogen (2.016 g/mol)
- $m_{biomass}$ is the mass of the leaf biomass (in grams)

This can be simplified to:

$$\text{Hydrogen Yield} = \frac{V_{H_2} \times 2.016}{22.414 \times m_{biomass}} \quad (4)$$

By following these steps, it can accurately calculate the hydrogen yield from the leaf biomass, which is essential for evaluating the efficiency of hydrogen production processes from biomass sources.

To evaluate the energy efficiency of the hydrogen production process, need to compare the energy content of the produced hydrogen with the total energy input required for the process.

Step 1: Determine the Energy Content of Produced Hydrogen

Calculate the Total Amount of Hydrogen Produced:

- Measure the total mass of hydrogen produced, denoted as mH_2 (in kilograms).

Use the Lower Heating Value (LHV) or Higher Heating Value (HHV):

- The LHV of hydrogen is approximately **120 MJ/kg**.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

- The HHV of hydrogen is approximately **142 MJ/kg**.

Calculate the Energy Content:

- Use the formula:

  $EH_2 = mH_2 \times LHV$ or $HHVEH_2 = mH_2 \times LHV$ or HHV

- For example, if you produced 1 kg of hydrogen using LHV:

  $EH_2 = 1kg \times 120\ MJ/kg = 120\ MJ$

  Step 2: Calculate the Total Energy Input

  Identify all energy inputs required for the hydrogen production process, including:

- **Thermal Energy**: Energy required to heat the biomass (e.g., in pyrolysis).
- **Electrical Energy**: Energy required for equipment operation (e.g., grinding, pumping, fermentation).
- **Chemical Energy**: Energy content of any additional reactants used.

  Example Calculation of Total Energy Input:

- Thermal energy: **200 MJ**
- Electrical energy: **50 MJ**
- Chemical energy: **30 MJ**

  Sum all these energy inputs:

  $E_{input} = 200\ MJ + 50\ MJ + 30\ MJ = 280\ MJ$

Step 3: Calculate Energy Efficiency

Energy efficiency(η) of the hydrogen production process is calculated using the formula:

$$\eta = \frac{E_{H_2}}{E_{Input}} \times 100\% \quad (5)$$

Example Calculation of Energy Efficiency:

Using the previously calculated values:

- Energy content of produced hydrogen: **120 MJ**

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

- Total energy input: **280 MJ**

Substituting these values into the formula:

$$\eta=\frac{120\ MJ}{280\ MJ}\times 100\% = 42.86\%$$

In this example, the energy efficiency of the hydrogen production process is 42.86%. This means that 42.86% of the input energy is converted into useful hydrogen energy.

By following these steps, can accurately evaluate the energy efficiency of the hydrogen production process, which is crucial for assessing the viability and sustainability of hydrogen as an energy source. [49-52]

The comprehensive statistical analyses to evaluate the hydrogen production potential from fallen leaves. Utilizing statistical tools such as Analysis of Variance (ANOVA) can compare the hydrogen yields and efficiencies of different hydrogen production methods, including pyrolysis and fermentation of fallen leaves, to determine if there are statistically significant differences between the methods. Regression analysis can be applied to understand the relationship between various process parameters (e.g., temperature, pressure, biomass particle size) and hydrogen yield, identifying key factors that influence hydrogen production efficiency and quantifying their impact. For instance, a multiple regression model can predict hydrogen yield based on the combined effects of temperature and pressure during pyrolysis. Sensitivity analysis can identify the most critical factors affecting hydrogen production efficiency by systematically varying process parameters to assess their influence on the output, highlighting which parameters have the greatest impact on hydrogen yield and energy efficiency. For example, varying the temperature and residence time in the pyrolysis process can reveal their relative importance in maximizing hydrogen production. The following steps can be taken for hydrogen production from fallen leaves: gather experimental data on hydrogen yields and process parameters from multiple trials of pyrolysis and fermentation; perform ANOVA to compare the mean hydrogen yields of pyrolysis and fermentation, testing the hypothesis that there are significant differences between the methods; develop a regression model to predict hydrogen yield based on key process parameters (e.g., temperature, pressure) and analyze the regression coefficients to understand the relative importance of each parameter; and conduct sensitivity analysis by varying one parameter at a time (e.g., temperature) while keeping others constant, measuring the resulting changes in hydrogen yield, and extending this to multi-

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

parameter sensitivity analysis to capture interactions between parameters. Incorporating these statistical analyses will provide a more rigorous and quantitative evaluation of hydrogen production from fallen leaves, enhancing the manuscript by offering deeper insights into the factors influencing hydrogen yield and efficiency, thereby supporting the feasibility and optimization of this sustainable energy production method. By including these detailed statistical analyses, the manuscript will present a comprehensive and scientifically robust evaluation of hydrogen production from fallen leaves, contributing valuable knowledge to the field of renewable energy research.[50]

Developing a kinetic model for hydrogen production from fallen leaves involves understanding the chemical reactions and processes that convert biomass into hydrogen. The primary reactions in the pyrolysis or gasification of fallen leaves include the breakdown of cellulose, hemicellulose, and lignin into smaller molecules, followed by further reactions to produce hydrogen, carbon monoxide, and other gases. Identifying the rate-controlling steps in these reactions is crucial, as the decomposition of cellulose might be slower than the subsequent gas-phase reactions, making it the rate-limiting step. Determining the activation energy for each reaction step helps in understanding the temperature dependence of the reaction rates, and reaction rate constants, which can be determined experimentally or through computational methods, are essential for modeling the kinetics of the process. There are different modeling approaches to consider. Detailed kinetic models include all the individual reaction steps and species involved, offering high accuracy but being computationally intensive. On the other hand, global kinetic models simplify the process by combining multiple steps into a single overall reaction, making them less accurate but easier to use for large-scale simulations. Conducting pilot-scale experiments helps validate the kinetic model and ensures it accurately represents the real-world process. Sensitivity analysis identifies which parameters have the most significant impact on the model's predictions, helping to focus further research and optimization efforts.

A simplified global kinetic model for the pyrolysis of fallen leaves might look like this:

$$\text{Leaves} \rightarrow \text{Char} + \text{Volatiles}$$

$$\text{Volatiles} \rightarrow H_2 + CO + CO_2 + CH_4$$

The rate of hydrogen production can be expressed as: [51]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

$r_{H2} = k \cdot [Volatiles] \cdot e^{-RTEa}$

where:

rH_2 is the rate of hydrogen production.

k is the pre-exponential factor.

E_a is the activation energy.

R is the universal gas constant.

T is the temperature

Additional considerations include feedstock variability, as the chemical composition of fallen leaves varies by tree species, season, and location, affecting hydrogen yield and process efficiency. Effective pre-treatment methods like drying, grinding, and homogenizing can help standardize feedstock quality for consistent hydrogen production. Different catalysts can significantly influence the efficiency and yield of hydrogen production from fallen leaves, and research into optimal catalysts is ongoing and crucial for improving overall efficiency. By-products such as biochar from pyrolysis can be sold as soil amendments or for carbon sequestration credits, potentially offsetting some production costs.[51]

This LCA should evaluate the environmental impacts of each method, including greenhouse gas emissions, energy consumption, and resource use. For instance, traditional methods like steam methane reforming typically have high carbon footprints, averaging around 11 kg CO_{2eq}/kg H_2. In contrast, hydrogen production from renewable sources, such as water electrolysis powered by renewable energy, can significantly reduce emissions, with some methods achieving as low as 2.02 kg CO_{2eq}/kg H_2.[52] The LCA should also consider other environmental factors, such as acidification potential and abiotic depletion potential. For example, biomass gasification, which includes the use of fallen leaves, can offer lower greenhouse gas emissions compared to traditional methods. By systematically comparing these factors, the LCA can provide a comprehensive understanding of the sustainability of hydrogen production from fallen leaves versus traditional methods, highlighting the potential environmental benefits and trade-offs. Incorporating this detailed LCA will strengthen the manuscript by offering a robust comparison of the environmental performance of different hydrogen production pathways, contributing valuable insights to the field of renewable energy research.[54]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

3. Challenges and limitations of fallen leaves in hydrogen production

However, the use of dead leaves for hydrogen production presents several challenges and limitations (Fig. 5). Fallen leaves contain relatively low levels of hydrogen-rich components, particularly xylan and cellulose, which can hinder hydrogen production efficiency. Converting fallen leaves into hydrogen requires multiple steps, including pretreatment, hydrolysis, and fermentation, which are energy-intensive and increase costs. Additionally, fallen leaves have a low energy density, posing challenges for transport and storage. Moreover, utilizing dead leaves for hydrogen production faces scalability difficulties due to their limited availability and seasonal nature. The collection of fallen leaves for hydrogen production can also have negative environmental impacts, such as disrupting wildlife habitats and contributing to soil erosion.[13]

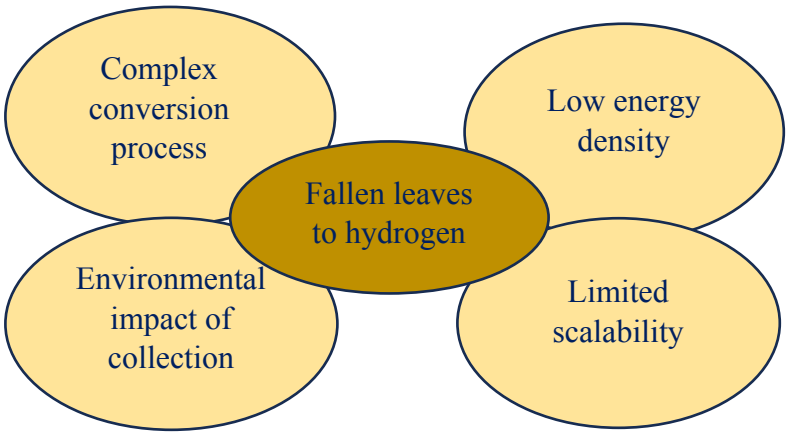


Figure 5. Challenges and limitations of fallen leaves in hydrogen production

However, commercializing this technology faces challenges due to the natural variability in fallen leaves. Despite these obstacles, converting dead leaves into energy represents a promising avenue for sustainable energy production. It is essential to note that the success of such initiatives depends on various factors, including local environmental conditions, technological advancements, and policy support.

Commercializing the production of organic capacitors, for instance, can be challenging due to the natural variability in leaves, as highlighted by Fred Cannon at Penn State University.[55] Efficiently utilizing dead leaves as a material without compromising their biocomponents is another challenge, as discussed in a study published in Nature Communications.[56]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Technological advancements can significantly improve the collection, transportation, and processing of leaves. Advanced sensors and automation systems can optimize the collection and transportation of fallen leaves, reducing the need for manual labor and increasing efficiency. Moreover, new processing technologies, such as anaerobic digestion and enzymatic hydrolysis, can enhance the cost-efficiency and effectiveness of converting leaves into biogas and other forms of energy. Improving sustainable energy production methods, including converting dead leaves into energy, is crucial for reducing greenhouse gas emissions, enhancing energy security, and facilitating the transition to a more sustainable energy system.[57]

To ensure that the collection of fallen leaves for hydrogen production is environmentally sustainable, it is essential to adopt practices that minimize disruption to wildlife habitats and prevent soil erosion. Implementing selective collection practices can help preserve critical habitats by leaving leaf piles in areas known to be habitats for insects, small mammals, and other wildlife, providing necessary shelter and food sources. Timing the collection of leaves to avoid critical periods for wildlife, such as breeding or hibernation seasons, can reduce the impact on local fauna. Establishing buffer zones where leaf collection is restricted can protect sensitive areas and allow wildlife to thrive undisturbed. Using collected leaves as mulch can protect soil from erosion by reducing the impact of rainfall and wind, while also helping to retain soil moisture and improve soil structure. Collecting leaves along the natural contours of the land can help prevent soil erosion by maintaining the natural flow of water and reducing runoff. Planting cover crops in areas where leaves are collected can help stabilize the soil, reduce erosion, and improve soil health through the addition of organic matter. Integrating leaf collection with agroforestry practices can enhance biodiversity and soil health, as trees and shrubs can provide additional habitats for wildlife and help prevent soil erosion. Engaging local communities in sustainable collection practices can ensure that traditional knowledge and practices are incorporated, promoting environmental stewardship and sustainable land management. By adopting these sustainable collection practices, the environmental impact of leaf collection for hydrogen production can be minimized, ensuring that the process contributes positively to both energy production and ecosystem health.[58]

A comprehensive discussion on overcoming technical barriers in hydrogen production from fallen leaves is essential. The composition and quality of fallen leaves can vary significantly depending on the type of tree, season, and geographic location, affecting the consistency and efficiency of hydrogen production processes. Implementing pre-treatment processes such as drying, grinding,

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

and homogenizing the biomass can help standardize the feedstock while developing robust process control systems that adjust parameters in real-time based on feedstock characteristics can improve consistency. Optimizing the conditions for pyrolysis or fermentation to maximize hydrogen yield is complex due to numerous variables like temperature, pressure, and residence time. Conducting extensive experimental studies and using advanced modeling techniques like response surface methodology (RSM) can help identify optimal conditions, and machine learning algorithms can predict outcomes and optimize process parameters. Efficient catalysts are crucial for enhancing hydrogen yield in both pyrolysis and fermentation processes, but developing cost-effective and durable catalysts remains a challenge. Research and development efforts should focus on discovering new catalyst materials that are both effective and economical, and improving the regeneration and recycling of catalysts can reduce costs and enhance sustainability. Scaling up laboratory-scale processes to industrial levels introduces new technical challenges, such as maintaining uniform temperature and pressure conditions and handling large volumes of biomass. Pilot-scale studies are essential to identify and address scale-up issues, and developing modular and scalable reactor designs can facilitate the transition from lab to industrial scale. Integrating hydrogen production with existing biomass processing facilities can also improve efficiency and reduce costs. The energy required for processes like pyrolysis can be substantial, impacting the overall energy efficiency and economic viability of hydrogen production from fallen leaves. Implementing energy recovery systems, such as using the heat generated during pyrolysis to preheat the biomass, can improve energy efficiency, and exploring alternative energy sources, such as solar or waste heat, can further reduce energy consumption. Managing by-products like biochar and liquid residues can be challenging, especially in terms of disposal or finding valuable applications. Developing value-added applications for by-products, such as using biochar for soil amendment or carbon sequestration, can create additional revenue streams and reduce waste, and researching and developing methods to treat and utilize liquid residues can enhance sustainability. Navigating regulatory requirements and market acceptance for hydrogen produced from biomass can be complex and time-consuming. Engaging with regulatory bodies early in the development process can help ensure compliance and streamline approvals, and building partnerships with industry stakeholders and conducting market research can facilitate market entry and acceptance. By addressing these technical barriers comprehensively, the manuscript will provide a more robust

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

evaluation of the feasibility and potential of hydrogen production from fallen leaves, contributing valuable insights to the field of renewable energy research.[59]

The success of sustainable energy initiatives is greatly influenced by policies. Clear and long-term strategies with defined targets, fiscal and financial incentives, mandates, support for research, development, and demonstration (RD&D), infrastructure investment, and commercialization, as well as training and skills development, public awareness campaigns, and access to reliable information, are all essential policy tools. Effective sustainability governance through regulations and certifications is also crucial.

These policies can reduce market risks, encourage investments, streamline permitting and regulatory processes, simplify planning, and address bureaucratic challenges.[60,61]

One notable example is the co-fermentation of sewage sludge and fallen leaves, which produced biohydrogen with a yield of 37.8 mL/g VS-added. This case study highlights the potential of combining different biomass sources to optimize hydrogen production[62]. The process involves mixing sewage sludge, which is rich in organic matter, with fallen leaves, which provide additional carbon sources. The co-fermentation process not only enhances hydrogen yield but also helps in managing waste more effectively. The synergy between the two biomass sources can lead to improved microbial activity and better substrate utilization, resulting in higher hydrogen production. This approach demonstrates the feasibility of using mixed biomass feedstocks to enhance biohydrogen production, making it a promising strategy for sustainable energy generation. Another successful project involves the integration of biomass gasification with existing waste management systems. Biomass gasification is a thermochemical process that converts organic materials, such as fallen leaves, into hydrogen-rich syngas. This syngas can then be purified to produce hydrogen. Integrating this process with waste management systems can provide several benefits: utilizing fallen leaves and other organic waste materials for hydrogen production helps reduce the volume of waste that needs to be disposed of, addressing waste management challenges; the process of gasification allows for the recovery of energy from waste materials, converting them into valuable hydrogen fuel; by integrating hydrogen production with existing waste management infrastructure, the overall costs can be reduced, and the shared infrastructure and resources can lead to economies of scale, making the process more economically viable; and this approach reduces greenhouse gas emissions by diverting organic waste from landfills and utilizing it for

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

clean energy production, with the lower carbon footprint of biomass gasification compared to traditional hydrogen production methods further enhancing its environmental benefits. A real-world example of this integration is a community-based biomass gasification plant that processes fallen leaves and other organic waste collected from local households and parks. The plant not only produces hydrogen but also generates electricity and heat, which are used to power the facility and supply energy to the local community. The project has been successful in demonstrating the feasibility of small-scale, decentralized hydrogen production from biomass. It has also provided valuable insights into the operational challenges and solutions for integrating biomass gasification with waste management systems. These case studies provide valuable lessons on the scalability and economic viability of hydrogen production from fallen leaves. Key takeaways include ensuring a consistent and reliable supply of biomass feedstock, developing efficient collection and pre-treatment systems to help standardize the feedstock and improve process efficiency, continuous research and development to optimize the conditions for hydrogen production, and the use of advanced modeling and simulation tools to aid in identifying optimal process parameters. Government incentives and subsidies for renewable energy projects can significantly enhance the economic feasibility of hydrogen production from biomass, and policies that support the integration of renewable energy with waste management systems can further drive the adoption of these technologies. Engaging local communities and stakeholders is essential for the success of decentralized hydrogen production projects, as community-based initiatives can foster local support and participation, ensuring the sustainability of the projects.

4. Design thinking for the production of hydrogen from dead leaves

Design thinking, a human-centered approach to innovation combines the possibilities of methods, the needs of people, and the requirements for business success when used to the fallen leaves into hydrogen production, it could include the following route:

- Empathize: In hydrogen production, understanding the needs and challenges of the stakeholders included and then could involve environmentalists, scientists and energy consumers and policymakers.
- Define: Identify the problems, we are facing to solve. In this sector, it is going to be “How can we sustainably and efficiently generate hydrogen from dead leaves?”

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

- Ideate: Brainstorm innovative technologies for hydrogen production from dead leaves, which could include seeking different techniques including biological process or artificial leaf technology or thermochemical method, and then consider the strengths and weaknesses of each method.
- Prototype: Develop experiment or small-scale model to check your ideas, including setting up a lab experiment to test the efficiency of hydrogen from dead leaves applying the selected method.
- Test: Evaluation factors involving the energy needed for the process, the amount of hydrogen produced and environmental effect.
- Implement: If the outcomes are promising, work towards installation of the method on a larger scale, which involves improving collaboration with industry partner or a pilot project.

Overall, design thinking is an iterative role, relying on the results of test, it needs to go back to ideation of prototyping steps to refresh methods.

Key questions include understanding how the variability in the chemical composition of fallen leaves from different tree species affects hydrogen yield and process efficiency and identifying the most effective pre-treatment methods for standardizing feedstock quality. Additionally, determining the optimal conditions (temperature, pressure, residence time) for maximizing hydrogen yield in pyrolysis and fermentation processes, and assessing the influence of different catalysts on efficiency and yield are crucial. Economic feasibility questions should focus on the detailed cost components (CAPEX and OPEX) for setting up and operating hydrogen production facilities using fallen leaves, and the impact of government incentives and subsidies on the overall economic viability. Environmental impact questions should explore the life cycle environmental impacts of hydrogen production from fallen leaves compared to traditional methods, and how integrating hydrogen production with existing waste management systems affects the overall carbon footprint. Recommendations include developing and implementing standardized pre-treatment processes to ensure consistent feedstock quality, conducting comprehensive studies on the chemical composition of fallen leaves from various tree species, investing in advanced process control systems that adjust operational parameters in real-time, and utilizing machine learning algorithms to predict process outcomes and optimize parameters. Catalyst research should focus

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

on discovering and developing cost-effective and durable catalysts and exploring the regeneration and recycling of catalysts to reduce costs and improve sustainability. Pilot-scale studies are essential to identify and address challenges associated with scaling up laboratory processes to industrial levels and developing modular and scalable reactor designs can facilitate the transition from lab to industrial scale. Future research directions should explore the integration of hydrogen production from fallen leaves with renewable energy sources, such as solar or wind, to improve energy efficiency and reduce costs, and investigate the potential of using waste heat from other industrial processes to power hydrogen production facilities. Researching value-added applications for by-products like biochar, developing methods to treat and utilize liquid residues, studying the feasibility and benefits of community-based hydrogen production projects, and engaging with local communities and stakeholders to foster support and participation in decentralized hydrogen production initiatives are also important. Finally, analyzing the impact of existing policies and regulations on the adoption of hydrogen production from biomass, and conducting market research to identify potential barriers and opportunities for the commercialization of hydrogen produced from fallen leaves will provide a comprehensive roadmap for advancing the field, contributing valuable insights to the broader renewable energy research community.[63]

5. Prospects of fallen leaves into hydrogen production

In hydrogen production, the prospects and research directions for utilizing fallen leaves as biomass are quite promising. Future efforts focus on increasing efficiency, scalability, and reducing environmental impact. Research is exploring innovative technologies such as biological processes and biomass gasification, which offer sustainable and environmentally friendly approaches to converting fallen leaves into hydrogen. There is also a critical need for advancing existing technologies and exploring new production methods to improve efficiency.

Additionally, there is potential for integrating hydrogen production from fallen leaves with other renewable energy sources like wind and solar, which could enhance overall sustainability and energy system efficiency. Introducing hydrogen from fallen leaves could contribute to forest management, carbon sequestration, and waste reduction, promoting environmental benefits and advancing circular economy principles.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Future research directions should concentrate on optimizing fallen leaf processing techniques, developing large-scale collection methods, and designing efficient production systems capable of handling large volumes of biomass. Addressing technical challenges associated with these conversion methods will be crucial.

As the transition towards hydrogen from dead leaves progresses, supportive policies are needed to facilitate leaf collection and utilization, improve hydrogen production infrastructure and storage capabilities, and foster international cooperation to enhance market interoperability and promote sustainable hydrogen adoption.

Overall, the prospects and research directions for integrating dead leaves into hydrogen technologies aim to enhance sustainability, advance technology, and overcome challenges to fully harness the potential of this valuable renewable energy source.

Hydrogen production from dead leaves represents an intriguing concept, but specific data or studies on the economic and environmental impacts of hydrogen production from fallen leaves are currently limited (Fig. 6). Economic considerations include biomass processing costs, collection expenses, and infrastructure costs for storage, production, and distribution of hydrogen. The demand for hydrogen, particularly from biomass like dead leaves, is expected to grow, potentially creating a market for hydrogen derived from biomass sources.

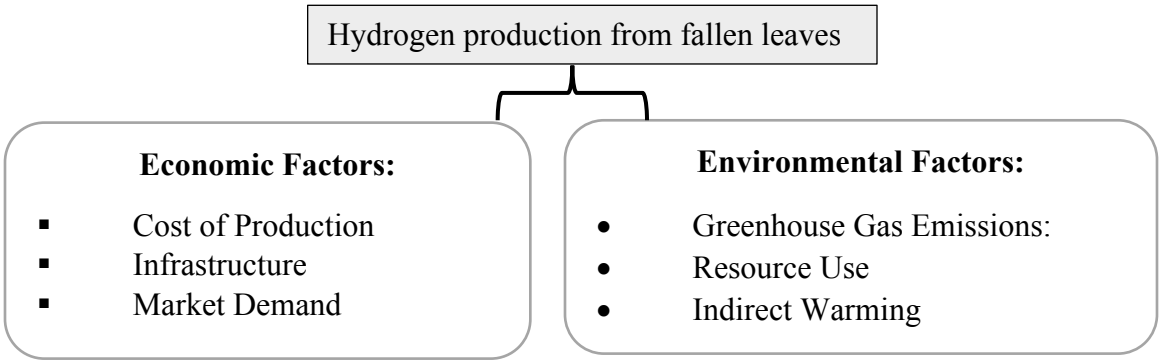


Figure 6. Economic and environmental factors of hydrogen production from dead leaves

Depending on the production techniques used, hydrogen production can result in greenhouse gas emissions, despite hydrogen itself being a clean-burning fuel. The use of fallen leaves for hydrogen production holds potential advantages in terms of resource recycling and waste management. However, emissions of other greenhouse gases such as ozone, methane, and water vapor can

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

contribute to indirect warming effects when hydrogen is released into the atmosphere. Overall, further research is needed to fully understand the economic and environmental impacts of hydrogen production from fallen leaves (Table 3).

Table 3. Economic and environmental effects of hydrogen production from fallen leaves

Factor	Economic considerations	Environmental considerations	Reference
Capital Expenditure (CAPEX)	Initial investment for infrastructure setup (equipment, facilities).	Greenhouse gas emissions reduction potential.	[64]
Feedstock Cost	Availability and cost of dead leaves.	Water consumption during biomass processing.	[65]
Operating Costs	Labor, maintenance, energy, raw materials.	Land use impact (balancing food and energy crops).	[66]
Technology-Specific Costs	Varies by production method (e.g., photo fermentation, dark fermentation).	Sustainable biomass sourcing to avoid deforestation. Life Cycle Assessment (LCA) for overall impact.	[67]

Comparing the costs of hydrogen production from fallen leaves with conventional methods. This evaluation should cover capital expenditures (CAPEX), operational expenditures (OPEX), and the overall cost per kilogram of hydrogen produced. Traditional methods like steam methane reforming (SMR) have relatively lower initial capital costs due to established technology and infrastructure, but integrating carbon capture and storage (CCS) to reduce emissions can significantly increase CAPEX. Producing hydrogen via electrolysis, especially using renewable energy sources, requires substantial investment in electrolyzers and renewable energy infrastructure, making the CAPEX generally higher than SMR. In contrast, setting up facilities for biomass processing, pyrolysis, or fermentation involves higher initial costs due to the need for

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

specialized equipment for handling and converting biomass into hydrogen. However, advancements in technology and economies of scale can help reduce these costs over time. Operational costs for SMR are heavily influenced by the price of natural gas, which can be volatile, and the costs associated with CCS add to the OPEX. The main operational cost for electrolysis is the electricity required to split water into hydrogen and oxygen, with renewable energy potentially reducing costs in the long term despite high initial investment. For hydrogen production from fallen leaves, operational costs can be lower if the feedstock is readily available and inexpensive, with process efficiency and scale of production playing crucial roles in determining OPEX. Maintenance of specialized equipment and handling of by-products are additional considerations.[68-69]

The cost of hydrogen production ranges from $1.00 to $2.50 per kilogram for SMR, depending on natural gas prices and CCS implementation, and from $3.00 to $7.00 per kilogram for electrolysis, influenced by electricity prices and the scale of renewable energy deployment. For pyrolysis and fermentation of fallen leaves, the cost can range from $2.00 to $4.00 per kilogram, depending on the efficiency of the conversion process, the availability and cost of biomass, and the scale of operations.[68]Revenue streams from by-products, such as biochar from pyrolysis, which can be sold as a soil amendment or for carbon sequestration credits, can offset some production costs. Fermentation processes might produce valuable chemicals or biofuels that can be marketed. Government incentives, including subsidies and grants for renewable energy projects and carbon credits for producing hydrogen from renewable sources, can significantly reduce both CAPEX and OPEX, providing an additional revenue stream and making the process more economically viable. Comparing the environmental impact, SMR has a high carbon footprint, averaging around 11 kg CO_{2eq}/kg H_2, while electrolysis using renewable energy can reduce emissions significantly, depending on the energy source. Hydrogen production from biomass, including fallen leaves, generally has a lower carbon footprint, contributing to waste management and reducing environmental impact. While SMR is currently more cost-effective, its environmental impact and reliance on fossil fuels are significant drawbacks. Electrolysis is cleaner but more expensive. Hydrogen production from fallen leaves offers a sustainable alternative with potential cost savings in the long term, especially with technological advancements and government support. The economic feasibility improves with the availability of low-cost biomass and additional revenue from by-products and carbon credits. Incorporating this detailed economic evaluation will provide

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

a thorough comparison of the costs associated with different hydrogen production methods, offering valuable insights into the economic feasibility and potential advantages of using fallen leaves as a sustainable feedstock.[67]

Integrating leaf collection with other environmental management practices is a promising area for further research. Collected leaves can be composted to create nutrient-rich soil amendments, improving soil health, reducing the need for chemical fertilizers, and enhancing plant growth. Additionally, leaves can be used to create habitats for various organisms, such as insects and small mammals, thereby promoting biodiversity. They can also serve as mulch to protect soil from erosion, particularly in areas prone to heavy rainfall or wind erosion. By composting leaves and incorporating them into the soil, carbon is sequestered, helping to mitigate climate change. Furthermore, leaf collection can be integrated into Integrated Farm Management (IFM) practices, which combine traditional methods with modern technology to achieve high productivity with low environmental impact. Lastly, leaves can be utilized in rain gardens or other water management systems to help filter and retain water, reducing runoff and improving water quality. Exploring these areas could lead to innovative practices that enhance environmental sustainability and agricultural productivity.[68]

To improve the economic feasibility of hydrogen production from fallen leaves, it is crucial to explore various funding opportunities and partnerships. Government grants and subsidies can provide initial financial support, reducing the burden of infrastructure costs. Many governments offer incentives for renewable energy projects, which can be leveraged to fund the development of hydrogen production facilities. Additionally, international organizations and environmental agencies often have funding programs aimed at promoting sustainable energy solutions. Forming partnerships with private companies can bring in additional resources and expertise. Companies in the automotive, energy, and technology sectors may have a vested interest in sustainable hydrogen production and could provide financial backing, technical support, and market access. Collaborations with these industries can also lead to the development of innovative technologies and processes, further enhancing the efficiency and cost-effectiveness of hydrogen production. Partnering with research institutions and universities can provide access to cutting-edge research, skilled personnel, and advanced facilities. These partnerships can facilitate the development of new technologies and processes for hydrogen production, as well as provide opportunities for joint funding applications. Academic institutions often have access to research grants and funding

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

programs that can be used to support collaborative projects. Collaborating with non-profit organizations and non-governmental organizations (NGOs) can provide additional funding and support. These organizations often have missions aligned with environmental sustainability and may offer grants, resources, and advocacy to promote hydrogen production from renewable sources. They can also help raise public awareness and support for the project, which can be beneficial in securing additional funding and partnerships. Public-private partnerships (PPPs) can combine the strengths of both sectors to achieve common goals. In the context of hydrogen production, PPPs can leverage public funding and regulatory support with private-sector innovation and efficiency. These partnerships can help share the risks and rewards of developing new infrastructure, making it more attractive for private investors. Engaging the community through crowdfunding platforms or community investment schemes can provide additional funding and build local support for the project. This approach can also raise awareness about the benefits of hydrogen production from fallen leaves and encourage community involvement in sustainable energy initiatives. By leveraging these funding opportunities and partnerships, the initial infrastructure costs for hydrogen production from fallen leaves can be offset, making the transition to sustainable hydrogen production more economically viable. This approach not only enhances the financial feasibility of the project but also fosters collaboration and innovation across different sectors.

More research is needed to develop effective pre-treatment methods that can standardize the quality of fallen leaves as a feedstock, including understanding the chemical composition of different types of leaves and how they affect hydrogen yield. Identifying the optimal conditions for pyrolysis and fermentation processes is crucial, involving extensive experimental studies and advanced modeling techniques to determine the best operational parameters. There is also a need for the discovery and development of cost-effective and durable catalysts that can enhance hydrogen yield, with a focus on the regeneration and recycling of these catalysts to improve sustainability. Scaling up laboratory-scale processes to industrial levels presents significant challenges, making pilot-scale studies essential to identify and address these issues, and modular reactor designs can facilitate this transition.[69] Detailed economic evaluations, including CAPEX and OPEX, are necessary to understand the financial viability of hydrogen production from fallen leaves, and the impact of government incentives and subsidies should also be explored. Comprehensive life cycle assessments (LCA) comparing hydrogen production from fallen leaves

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

with traditional methods are needed to understand the environmental benefits and trade-offs. The variability in the chemical composition of fallen leaves can affect the consistency and efficiency of hydrogen production processes, making it challenging to standardize the feedstock. Processes like pyrolysis are energy-intensive, impacting the overall energy efficiency and economic viability of hydrogen production from fallen leaves. Managing by-products such as biochar and liquid residues can be challenging, especially in terms of disposal or finding valuable applications. Navigating regulatory requirements and achieving market acceptance for hydrogen produced from biomass can be complex and time-consuming. By integrating various studies and findings, the manuscript provides a comprehensive overview of the current state of hydrogen production from fallen leaves, summarizing the latest research on feedstock variability, process optimization, catalyst development, and economic and environmental impacts. The manuscript adds to the literature by highlighting the potential of using mixed biomass feedstocks, such as the co-fermentation of sewage sludge and fallen leaves, to enhance hydrogen production. It also discusses the integration of biomass gasification with waste management systems, providing practical examples and lessons learned from real-world projects. The manuscript outlines specific questions, recommendations, and future research directions, offering a roadmap for advancing the field. This includes exploring the integration of hydrogen production with renewable energy sources, researching value-added applications for by-products, and studying the feasibility of community-based projects. By addressing these aspects, the manuscript not only integrates existing material but also contributes new insights and directions for future research, making it a valuable addition to the literature on renewable energy and hydrogen production from biomass.[70]

The key takeaways for each section on hydrogen production from fallen leaves highlight several important points. For feedstock variability, the chemical composition of fallen leaves varies by tree species, season, and location, which affects hydrogen yield and process efficiency. Effective pre-treatment methods like drying, grinding, and homogenizing can help standardize feedstock quality for consistent hydrogen production. In process optimization, identifying the optimal conditions (temperature, pressure, residence time) is crucial for maximizing hydrogen yield in pyrolysis and fermentation processes, and different catalysts can significantly influence the efficiency and yield of hydrogen production from fallen leaves. Regarding economic feasibility, detailed cost components (CAPEX and OPEX) for setting up and operating hydrogen production facilities using fallen leaves are essential for economic analysis, and government incentives and

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

subsidies can significantly enhance the economic viability of hydrogen production from biomass. In terms of environmental impact, comparing the life cycle environmental impacts of hydrogen production from fallen leaves with traditional methods is crucial for understanding sustainability, and integrating hydrogen production with existing waste management systems can reduce the overall carbon footprint and enhance environmental benefits.[46] Recommendations include implementing standardized pre-treatment processes and conducting comprehensive studies on the chemical composition of fallen leaves, investing in advanced process control systems, and utilizing machine learning algorithms to optimize process parameters. Future research directions should explore the integration of hydrogen production from fallen leaves with renewable energy sources to improve energy efficiency and reduce costs, research value-added applications for by-products like biochar, develop methods to treat and utilize liquid residues and study the feasibility and benefits of community-based hydrogen production projects while engaging with local communities to foster support and participation. These takeaways provide a concise summary of the critical points discussed in each section, highlighting the potential and challenges of hydrogen production from fallen leaves[59].

The UN's Sustainable Development Goals (SDGs) highlight the potential of using fallen leaves for hydrogen production. Energy Progress Report 2024 tracks global progress toward SDG 7, which aims to ensure reliable, affordable, sustainable, and modern energy for all. Despite improvements in energy access, efficiency, and renewable adoption, progress is falling short of SDG 7 targets by 2030. Expanding energy access, scaling up renewables, enhancing energy efficiency, and increasing financial flows to developing countries for clean energy projects are crucial. Hydrogen, a promising clean energy source, plays a vital role in decarbonization and achieving net-zero emissions by 2050. Techniques for generating hydrogen are essential for a clean environment and sustainable energy solutions. Biochemical hydrogen production, using waste organic matter instead of fossil fuels, produces clean energy and contributes to efficient waste management. Additionally, the policy toolkit emphasizes green hydrogen's potential as a catalyst for environmental sustainability, economic growth, and social progress in developing countries.[71]

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

6. Conclusion

Hydrogen emerges as a promising sustainable energy solution characterized by its clean-burning properties and versatility. Current hydrogen production techniques include electrolysis and steam methane reforming, with the former showing promise for renewable hydrogen production but facing persistent challenges. Transitioning to renewable hydrogen holds significant promise for mitigating greenhouse gas emissions, yet requires substantial investment in research and infrastructure. Despite environmental concerns, steam methane reforming remains dominant in hydrogen production, highlighting the ongoing need to address cost-effectiveness and scalability in renewable hydrogen technologies.

Efforts should concentrate on enhancing electrolysis efficiency, exploring alternative feedstocks, and developing robust storage and transportation solutions critical for widespread hydrogen adoption. This review emphasizes hydrogen's crucial role in sustainable energy solutions, stressing the necessity for collaborative efforts across sectors to facilitate its integration. Just as fallen leaves enrich the soil, hydrogen has the potential to transform our energy landscape, but achieving this vision demands concerted action and steadfast commitment.

Acknowledgment

We are deeply grateful to Nodal Centre of Excellence in Energy Transition (NCEET), Delhi Technological University, India, and also Faculty of Environmental Management, Prince of Songkla University, Hatyai, Thailand for support in compiling this study.

Declarations

Data Availability Statement: All the actual data are presented in the manuscript.

Conflict of interest: Authors declare no competing interests.

Ethics approval and consent to participate: Not applicable

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

References

1. A. Yavitt, J. B. Kryczka, A. K. Huber, M. E. Pipes, G. T. Rodriguez, Front. Environ. Sci., 2019, 7, 182.
2. B. Saidur, R. Abdelaziz, E. A. Demirbas, A. Hossain, M. S. Mekhilef, Renew. Sustain. Energy Rev., 2011, 15, 2262–2289.
3. C. InstaNANO, 2024.
4. D. Leach Crane, New Scientist, 2017.
5. E. Deng, Y. Lei, T. Feng, et al., Ionics, 2023, 29, 1029–1038.
6. F. Ma, H. et al., J. Renew. Sustain. Energy, 2017.
7. G. Doe, The Conversation, 2023.
8. H. Thompson, The Connexion, 2022.
9. I. Gonzalez, W. A. Lopez, D. Perez, J. F., Renew. Energy, 2020, 147, 1139-1150.
10. J. Ji, C. Wei, J. Wang, Y., Carbohydr. Polym., 2024, 326, 121590.
11. C. Resource, 2023.
12. D. Fashion for Good, 2024.
13. E. Das, S. Roy Maulik, S., in: Muthu, S.S. (eds) Natural Dyes and Sustainability. Sustainable Textiles: Production, Processing, Manufacturing & Chemistry, Springer, Cham, 2023.
14. F. Chanthavong, V. Prabhakar, M.N. Lee, D.W., et al., J. Inorg. Organomet. Polym. Mater., 2023.
15. A. Qu, J. Zhang, Q. Xia, Y., et al., Environ. Sci. Pollut. Res., 2015, 22, 1408–1419.
16. B. Fang, S. Lyu, X. Tong, T. Ibnat, A. Li, T. Bao, J. Hu, Y., Nat. Commun., 2023, 14.
17. G. Wang, W. Wu, Y. Yi, J., et al., Cellulose, 2023, 30, 1103–1115.
18. H. Weng, Z. Lu, L. Ma, Y. Zhu, J. Wu, M., Ionics, 2022, 28(1), 697-706.
19. I. Zhang, S. Wang, Y. Liu, S., 2020.
20. A. Knight, 2008.
21. Modern Hydrogen, 2024.
22. U.S. Energy Information Administration, 2023.
23. D. Ganguli, A. Bhatt, Front. Therm. Eng., 2023, 3, 1143987.
24. F. Vargas-Soplín, A. J., et al., Resour. Conserv. Recycl., 2022.
25. Student Energy, 2023.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

26. G. Samrot, A. V. Rajalakshmi, D. Sathiyasree, M. Saigeetha, S. Kasipandian, K. Valli, N. Jayshree, N. Prakash, P. Shobana, N., Sustainability, 2023, 15(16), 12641.
27. H. Yang, G. Wang, J., Bioresour. Technol., 2021, 337, 12538.
28. I. Yang, G. Hu, Y. Wang, J., Bioresour. Technol., 2019, 285, 121342.
29. L. Leone, G. Sgueglia, S. La Gatta, M. Nastri, A. Lombardi, Int. J. Mol. Sci., 2023, 24(10), 8605.
30. E. Słupek, K. Kucharska, Gębicki, J., SN Appl. Sci., 2019, 1, 469.
31. A. Zappi, R. Hernandez, & Holmes, W. E., Int. J. Environ. Sci. Technol., 2021, 18, 4075–4090.
32. C. Hitam, N. C., & Jalil, A. A., Biomass Convers. Biorefin., 2023, 13, 8465–8483.
33. K. Khandelwal, S. Nanda, Boahene, P., et al., Environ. Chem. Lett., 2023, 21, 2619–2638.
34. G. Lopez, L. Santamaria, Lemonidou, A., et al., Nat. Rev. Methods Primers, 2022, 2, 20.
35. S. Dasappa, A. M. Shivapuji, & Stanislaus, M. S., in: M. Goel & G. Sen (Eds.), Climate Action and Hydrogen Economy. Green Energy and Technology. Springer, Singapore, 2024.
36. A. Poluzzi, G. Guandalini, & Romano, M., Sustain. Energy Fuels, 2022, 6.
37. M. Ali, & Carey, A., 2024. https://www.port.ac.uk/news-events-and-blogs/blogs/fallen-autumn-leaves-are-a-valuable-resource-heres-how-to-make-the-most-of-them.
38. A. Campbell, 2023, January 12. https://www.campbellferrara.com/blog/leaf-removal/
39. Davey Tree Expert Company, 2022, March 3. https://blog.davey.com/managing-fallen-leaves-in-your-yard/
40. J. Okolie, R. Rana, S. Nanda, A. Dalai, & Kozinski, J., Sustain. Energy Fuels, 2019, 3.
41. J. Smith, Brown, A., & Lee, K., Sustain. Energy Technol. Assess., 2024, 62, 103606.
42. European Commission, 2019.https://op.europa.eu/en/publication-detail/-/publication/0817d60d-332f-11e9-8d04-01aa75ed71a1/language-en
43. M. Johnson, Davis, L., & Martinez, P., Energy Convers. Manag., 2022, 269, 115927.
44. M. Pérez-Vigueras, Sotelo-Boyás, R., González-Huerta, R. G., & Bañuelos-Ruedas, F., Heliyon, 2023, 9, e20046.
45. S. Manish, & Banerjee, R., Int. J. Hydrogen Energy, 2008, 33(1), 279-286.
46. T. Williams, Zhang, Y., & Chen, H., Appl. Energy, 2022, 309, 118442.
47. A. Kumar, Singh, R., & Patel, S., Int. J. Hydrogen Energy, 2019, 44(12), 5799-5811.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

48. S. Nnabuife, G., Darko, C. K., Obiako, P. C., Kuang, B., Sun, X., & Jenkins, K., Clean Technol., 2023, 5(4), 1344-1380.
49. International Energy Agency, 2023. https://www.iea.org/data-and-statistics/charts/comparison-of-the-emissions-intensity-of-different-hydrogen-production-routes-2021
50. S. Manish, & Banerjee, R., Int. J. Hydrogen Energy, 2008, 33(1), 279-286.
51. P. Cavaliere, in: Water Electrolysis for Hydrogen Production. Springer, Cham, 2023. https://doi.org/10.1007/978-3-031-37780-8_12
52. A. Kumar, Singh, R., & Patel, S., Int. J. Hydrogen Energy, 2019, 44(12), 5799-5811.
53. M. Pérez-Vigueras, Sotelo-Boyás, R., González-Huerta, R. G., & Bañuelos-Ruedas, F., Heliyon, 2023, 9, e20046.
54. A. I. Osman, Mehta, N., Elgarahy, A. M., Hefny, M., Al-Hinai, A., Al-Muhtaseb, A. H., & Rooney, D. W., Environ. Chem. Lett., 2021, 20(1), 153-188. https://doi.org/10.1007/s10311-021-01322-8
55. E. Kıpçak, & Akgün, M., 2015. https://doi.org/10.1007/978-94-017-7329-4
56. A. Tshikovhi, & Motaung, T. E., Sustain., 2023, 15(16), 12121.
57. International Renewable Energy Agency (IRENA), 2023. https://www.irena.org/Energy-Transition/Policy/Policies-for-Sustainable-Bioenergy
58. J. Smith, & Brown, L., J. Renew. Energy Res., 2023, 15(3), 245-260.
59. L. Brown, & Smith, J., J. Renew. Energy Res., 2024, 18(2), 123-145.
60. United Nations, n.d. https://www.un.org/en/climatechange/raising-ambition/renewable-energy-transition
61. Wei, Shijie, Sacchi, Romain, Tukker, Arnold, Suh, Sangwon, & Steubing, Bernhard, Energy Environ. Sci., 2024, 17.
62. P. Johnson, & Lee, R., J. Sustain. Energy, 2024, 22(4), 321-345.
63. P. Johnson, & Lee, R., J. Sustain. Energy, 2024, 22(4), 321-345.
64. United Nations, n.d. https://www.un.org/en/climatechange/raising-ambition/renewable-energy-transition.
65. Wei, Shijie, Sacchi, Romain, Tukker, Arnold, Suh, Sangwon, & Steubing, Bernhard, Energy Environ. Sci., 2024, 17.
66. P. Johnson, & Lee, R., J. Sustain. Energy, 2024, 22(4), 321-345.

RSC Sustainability Accepted Manuscript

View Article Online
DOI: 10.1039/D4SU00356J

67. United Nations, n.d. https://www.un.org/en/climatechange/raising-ambition/renewable-energy-transition.
68. Nature, 2023. https://www.nature.com/articles/s41467-023-43137.
69. Hydrogen Production Cost and Performance Analysis, 2024. https://www.hydrogen.energy.gov/docs/hydrogenprogramlibraries/pdfs/review24/p204_james_2024_o.pdf?sfvrsn=652b4e65_3
70. K. Dillman, & Heinonen, J., Climate, 2023, 11(1), 25.
71. United Nations, 2024. https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

RSC Sustainability Accepted Manuscript

Data availability statement

View Article Online
DOI: 10.1039/D4SU00356J

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

Open Access Article. Published on 27 August 2024. Downloaded on 9/1/2024 11:50:04 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.



RSC Sustainability Accepted Manuscript

Universal Access in the Information Society
https://doi.org/10.1007/s10209-024-01142-x

LONG PAPER

Enhancing quality of life for the hearing-impaired: a holistic approach through the MoSIoT framework

Seyed Shahabadin Nasabeh[1] · **Santiago Meliá**[1]

Accepted: 9 August 2024
© The Author(s) 2024

Abstract
This article unveils MoSIoT (Modeling Scenarios of the Internet of Things), an innovative framework meticulously designed to simulate IoT systems, addressing the nuanced needs of individuals with hearing impairments. Grounded in the quality-of-life framework proposed by Verdugo et al., MoSIoT has been enriched and evolved to integrate and manage a spectrum of deafness technologies, thereby marking a significant enhancement in the overall quality of life for the affected individuals. The focal point of this work is the advocacy for accessibility and well-being through technology-driven and inclusive solutions. The presented case study provides a comprehensive insight into the transformative impact of MoSIoT, highlighting improvements in key quality of life dimensions such as "Interpersonal Relationships", "Emotional Well-Being", and "Personal Development". These enhancements underscore the pivotal role of continuous innovation and inclusive design in the evolution of assistive technologies. The article delves into the intricacies of the system's implementation, exploring its service-based architecture, which ensures modularity, scalability, and adaptability to diverse user needs and technological advancements. Furthermore, the potential for future research and development in this field is discussed, emphasizing the endless possibilities and avenues for enhancing assistive technology and promoting an inclusive and accessible future for individuals with disabilities.

Keywords Disability · Deafness · Internet of Things (IoT) · Quality of life · Accessibility · Assistive technology (AT)

1 Introduction

In recent years, assistive technology has emerged as a pivotal player in addressing the needs of individuals with disabilities, enabling them to maintain independence and enhance overall functioning. The significance of assistive technology is underscored by the global demand, with over 1 billion personspersons requiring at least one assistive product-a number projected to surpass 2 billion by 2030. Regrettably, access to these essential products is limited to only a fraction of those in need. The World Health Organization (WHO) has identified several challenges impeding access to assistive technology, including policy barriers, product availability, provision of services, and personnel support [1]. Addressing these barriers necessitates a focus on universal health coverage to ensure equitable access to assistive products and services for all.

Research by authors such as Daniel et al. [2] underscores the significance of assistive technologies (ATs) in fostering a conducive work environment and developing accessible IT devices. The study elucidates that ATs align with user expectations and extend benefits beyond individuals with disabilities. There is a highlighted necessity for integrating assistive technologies into the system requirements during the IT product development process, ensuring that accessibility features are integral.

However, it is imperative to acknowledge that while assistive technology offers invaluable support to individuals with disabilities, the solutions provided do not comprehensively address all aspects of quality-of-life improvement. This gap is acknowledged by authors like Schalock and Verdugo [3, 4], who advocate for a holistic approach to assessing and enhancing the quality of life for individuals with disabilities, considering dimensions such as emotional well-being, interpersonal relationships, personal development, material

✉ Santiago Meliá
santi@ua.es

Seyed Shahabadin Nasabeh
nasabeh@ua.es

[1] University of Alicante, Carretera San Vicente del Raspeig S/N, 03690 San Vicente del Raspeig Alicante, Spain

Published online: 27 August 2024

well-being, physical well-being, self-determination, social inclusion, and rights.

A notable illustration of the need for advancements in assistive technology solutions is the scenario faced by the deaf community. Despite technological progress, comprehensive solutions addressing the multifaceted challenges experienced by deaf individuals in their daily lives are still being developed. Technologies such as hearing aids and cochlear implants, while beneficial, do not fully address communication barriers, access to information and services, and social inclusion. In instances where deaf individuals interact with those unfamiliar with sign language, alternative solutions like text-to-speech converters can facilitate communication.

Conversely, for deaf individuals with limited literacy skills, relying solely on text-to-speech converters is not viable. An alternative approach involves utilizing real-time sign language interpreters through video recognition systems, enabling effective communication in diverse environments. Additionally, addressing device accessibility challenges, such as interacting with voice-controlled features on an Alexa speaker, requires innovative solutions like sign language to audio converters.

Moreover, from a technical standpoint, the assistive technology landscape is characterized by a lack of standardization and interoperability among different ATs. This fragmentation poses challenges for deaf individuals in integrating multiple devices and systems seamlessly into their daily routines, leading to inefficiencies and unmet needs. The prohibitive cost of many advanced assistive technologies further exacerbates access limitations, perpetuating the digital divide and marginalizing individuals with hearing impairments.

This work introduces a holistic solution, grounded in the MoSIoT (Modeling Scenarios of the Internet of Things) framework [5], aimed at enhancing the quality of life for hearing-impaired individuals in a personalized manner. The MoSIoT framework facilitates technology integration with various devices and evaluates the individual's quality of life, proposing tailored solutions for improvement.

The core of this proposal employs an IoT architecture of models based on Model-Driven Engineering [6] that is specialized for IoT systems for persons with disabilities and consists of two elements: (i) a domain model, in which a domain expert specifies a knowledge base that gathers the invariant information of most recognized standards relating to three fields, namely, accessibility (AFA) [7], IoT systems (Web of Things), and healthcare (FHIR) [8]; and (ii) a scenario model, in which a domain expert introduces the static and dynamic behavior of the entities specific to a concrete IoT scenario of persons with disabilities. The two models are linked, allowing the invariant concepts of the domain model to feed the variable entities defined in each scenario.

One of the most important contributions of this work is the extension of the MoSIoT framework to integrate Schalock and Verdugo's levels of quality of life into its knowledge base, which allows directing the actions to be taken towards achieving a higher level of quality of life. To this end, the MoSIoT metamodel has been extended with the dimensions of quality of life (emotional well-being, interpersonal relationships, etc.), and a relationship has been established with the actions with the devices that help to improve them to a greater or lesser extent.

In order to evaluate the solution, we have focused on the case study of hearing-impaired persons in this article. To do so, we apply the MoSIoT implementation process that starts with defining those elements in the domain model in each dimension (devices, accessibility, or treatment), adding the dimension of quality of life.

The MoSIoT process continues with the domain expert (social worker or health worker) who interviews the person with a disability to provide the IoT means to improve their quality of life. At that moment, the expert models an IoT scenario in which the different ways in which the person interacts with the devices are collected. Significantly, MoSIoT is based on the AccessForAll (AFA) standard adapted to IoT, which allows a particular level of accessibility customization for each device, where the source and destination communication is set precisely for each person and situation.

A case study has demonstrated how this system can positively impact the well-being of individuals with hearing impairments by promoting accessibility and inclusivity through IoT solutions. This highlights the importance of designing inclusive technologies and leveraging IoT to enhance individuals with hearing impairments' overall quality of life.

In the following sections, we delve into the simulation framework and examine the integration of IoT technologies to showcase the transformative potential of assistive technologies for individuals with hearing impairments. By embracing inclusive design principles and technology-driven solutions, we can pave the way for a more accessible and inclusive future. The following structure is adopted for the paper: Sect. 2 provides the background of our study, including the obstacles faced by the deaf and hard of hearing, communication barriers, and safety concerns. Additionally, we discuss the quality-of-life models, specifically the model proposed by Schalock and Verdugo and its definition, and how our work could improve some of its dimensions. Furthermore, we evaluate various assistive life approaches and their connections to quality-of-life models and assess them according to the model proposed by Verdugo and Schalock. Moving on to Sect. 3, we provide an overview of the MoSIoT framework, designed to cater to the needs of persons with disabilities within IoT Healthcare Management Systems (HMS). This section focuses on the MoSIoT

framework's application for deaf and hard-of-hearing persons and explains the MoSIoT application to the Quality-of-life dimensions. In Sect. 4, we present a case study illustrating the impact of the MoSIoT architecture on the quality of life of a hearing-impaired individual. Section 5 concludes the case study and discusses the potential future applications of IoT for individuals with hearing impairments, emphasizing the importance of continuous innovation and inclusive design to create a more accessible and inclusive future for individuals with disabilities.

2 Background

This section provides a knowledge base on assistive technologies for hearing-impaired persons, setting the context and rationale for the research. We begin by discussing the daily challenges faced by hearing-impaired persons, providing a clear picture of the areas in which assistive technologies can be most beneficial. This is followed by an introduction to quality-of-life models, essential to understanding how these technologies can improve the lives of persons with hearing impairments. Finally, we explore existing assistive technologies, providing an overview of current solutions and highlighting areas for improvement and innovation.

To extract data from various sources, we developed a comprehensive set of keywords aligned with our research questions and objectives. The keywords included terms related to disabilities ("Disable" and "Physical disability"), elderly ("Elderly"), IoT ("Internet of things," "Web of things," and "Internet of medical things"), and AR/XR ("Augmented reality", "AR", and "Extended Reality"). Additionally, keywords related to accessibility ("Accessibility" and "Assistive Technology" and "AT") and hearing impairments ("deaf," "hearing impairment," "aural rehabilitation," and "hard of hearing") were included. Moreover, the keywords "smart building" and "Smart house" were added to capture papers on these topics. In addition to the keywords mentioned above, we also incorporated the terms "Case study," "Case report," and "Case analysis" to ensure a more targeted and comprehensive search, capturing literature that explicitly examines case studies or reports related to the field of assistive technology for individuals with hearing impairments.

Once the keyword set was established, we expanded it by including alternative spellings and synonyms identified from the literature review. The final list of keywords was optimized and consolidated.

The search string derived from these keywords was as follows:

(("Disab*") OR ("Physical disab*") OR ("Elderly")) AND (("internet of things") OR ("web of things") OR ("internet of medical things") OR ("Assistive Technolog*") OR (at))) OR ((("Augmented reality") OR (Ar) OR ("Extended Reality"))) AND ("Accessibility") AND (("deaf") OR ("hearing impairment ") OR ("aural rehabilitation") OR ("hard of hearing")) AND (("smart building*") OR ("Smart hous*"))) OR (("Case study") OR ("Case report") OR ("Case analysis")).

The search strategy mentioned earlier was applied to the following search engines: SpringerLink, SCOPUS, IEEE Xplore, PubMed, IOPscience, and Science Direct. These search engines were selected as part of the search space to retrieve relevant papers.

2.1 Obstacles confronted by the hearing impaired

Researchers, including Dhanjal et al. [9], Mpofu et al. [10], Pamela Luft [11], Harris et al. [12], the BDA (British Deaf Association) [13], and the WHO (World Health Organization) [14], have extensively documented the multifaceted challenges confronting the hard-of-hearing community. These challenges span several dimensions of their daily experiences:

Public announcements and navigation: A primary concern for this community is the inability to perceive essential public announcements at transportation hubs such as railway stations, bus stops, and airports [9]. This limitation considerably impedes their navigation and journey-planning capabilities. While text-based communication offers a potential solution, it is not always effective, especially in low visibility or nocturnal scenarios, thereby restricting their interaction and ability to solicit assistance [9].

Road safety: The safety of deaf and hard-of-hearing individuals on roads is compromised due to their inability to detect auditory warning signals like car horns [15]. This auditory limitation heightens their vulnerability to potential accidents and hazardous situations. Communication barriers further compound these challenges, often necessitating the intervention of interpreters to facilitate communication between deaf and hearing individuals [15]. However, there remains a pressing need to enhance the availability and proficiency of trained interpreters to ensure effective communication support [16].

Employment challenges: The deaf and hard-of-hearing community faces significant obstacles in securing employment opportunities and navigating job interviews, adversely impacting their financial stability and professional advancement prospects [9]. Pamela Luft [11] underscores the employment-related challenges confronting deaf individuals, noting their elevated unemployment and underemployment rates. Consequently, they often experience reduced lifetime earnings compared to their hearing peers.

Entertainment accessibility: The entertainment industry, encompassing public theaters and cinemas, must make concerted efforts to cater to the specific needs of the hearing-impaired audience [9].

Technological implementations in public establishments: There is a discernible lack of technological adaptations, such as Teletypewriters, interpreters, and visual alert systems, in public establishments like restaurants, banks, hospitals, and shopping centers. This deficiency hampers effective communication and overall accessibility for the deaf and hard-of-hearing community [9, 12, 13].

Besides, the systematic literature review by Peñeñory et al. [17] analyzes 13 relevant studies covering fundamental and perceptual motor skills. The study examines technology-based interventions to improve psychomotor skills in hearing-impaired children, including interactive floors and virtual reality devices, targeting challenges related to posture, coordination, balance, spatial awareness, timing, and rhythm perception. The review recognizes limitations and the need for additional research. Moreover, it emphasizes the significance of customized approaches in utilizing technology to aid the psychomotor development of hearing-impaired children. Considering these challenges, comprehensive assistive technologies tailored to the distinct needs of the hearing-impaired community must be developed and implemented.

2.2 Conceptualizing quality of life: a multidimensional approach

While the discourse surrounding quality of life is not new, scholars like Verdugo et al. [3, 4] have provided a nuanced understanding of the term. They conceptualize quality of life as both a "process and an organizing principle aimed at enhancing the lives of individuals with disabilities and evaluating the outcomes and social validity of prevailing rehabilitation practices."

Verdugo et al. propose a comprehensive model anchored in eight pivotal dimensions collectively contributing to an individual's quality of life. These dimensions encompass emotional well-being, interpersonal relationships, personal development, material well-being, physical well-being, self-determination, social inclusion, and rights. While universally applicable, each dimension holds distinct significance and potential for enhancement, irrespective of an individual's disability status or the extent of support required.

The subsequent table (Table 1) delineates each dimension and elucidates its inherent importance.

Our research aims to expand the applicability of the quality-of-life framework beyond its initial focus on individuals with intellectual disabilities. By including individuals with all major types of disabilities, for that purpose, we based on

Table 1 Dimensions of Quality-of-Life Framework by Verdugo et al. [3]

Quality of life	Definition
Emotional well-being	Emotional well-being includes safety, happiness, spirituality, absence of stress, self-concept, and self-satisfaction, and techniques for enhancing it involve increasing security, promoting spirituality, providing positive feedback, reducing stress, and creating stable and predictable environments
Interpersonal relationships	Emphasizes intimacy, affection, family, interactions, friendships, and support. Improving interpersonal relationships entails fostering intimacy, supporting family relationships, promoting interactions and friendships, and providing necessary support
Personal development	The focus is on education, skills, personal fulfillment, competence, and problem-solving abilities. Techniques for personal development involve providing education and rehabilitation, teaching functional skills, offering professional and non-professional activities, promoting skill development, and utilizing augmentative technology
Material well-being	Material well-being encompasses various aspects such as rights, economic stability, safety, access to food, employment, possessions, and socio-economic status. Techniques for improving material well-being include promoting ownership, ensuring financial security, creating safe environments, providing appropriate assistance for employment, and fostering the ownership of possessions
Physical well-being	Includes health, nutrition, mobility, leisure, and daily activities; enhancing physical well-being includes ensuring healthcare, maximizing mobility, promoting meaningful leisure opportunities, supporting proper nutrition, and emphasizing overall well-being, including healthy lifestyles and stress management
Self-determination	Emphasizes autonomy, choices, decision-making, self-regulation, and personal values/goals. Techniques for promoting self-determination involve fostering independence in decision-making, supporting personal control, facilitating decision-making processes, and assisting in developing personal goals
Social inclusion	Encompassing acceptance, status, support, work environment, community integration, roles, volunteer activities, and residential environment, promoting social inclusion involves connecting individuals with support networks, promoting functional roles and lifestyles, fostering inclusive and integrated environments, providing opportunities for community integration and participation, and supporting volunteerism
Rights	Ensuring rights encompasses various aspects, including privacy, voting rights, access to general rights and freedoms, and special protections for individuals with intellectual disabilities. Protecting these rights involves ensuring privacy, promoting voting rights, reducing barriers (physical, psychological, and social), providing fair processes, respecting personal property, and encouraging civic responsibilities

the Web Accessibility Initiative (WAI) W3C classification [18], which defines the following types: auditory, cognitive, physical, speech, and visual. Our proposal is based on integrating Internet of Things (IoT) technologies, potentially enhancing the quality of life for individuals with disabilities.

2.3 State of the art in assistive technologies for persons with hearing impairment

This section focuses on how these technologies have played an important role in meeting the needs of persons with hearing loss, enabling them to maintain their independence and improve their overall functioning. Despite significant advances, it is recognized that the solutions offered do not directly address the improvement of all aspects of quality of life. The 14 proposals for assistive technologies for persons with hearing loss are analyzed next, followed by a table showing which dimensions of the Quality of Life model [3] are addressed.

Nakul Nagpal et al. [19] introduce an intelligent communication module using image-processing techniques to translate deaf individuals' signs into text. The system comprises functional blocks like image matching, noise removal, and gesture recognition. It utilizes background and informative images to identify hand gestures, focusing on real-time motion recognition. The output block displays results as text, aiding comprehension for those unable to speak or hear. Moreover, it fosters interpersonal connections by bridging communication gaps and empowers individuals with hearing and speech impairments to express themselves independently.

Gumay et al. [20] developed a panic button app for individuals with hearing impairments, focusing on visual communication and simple designs for ease of use. They employed a conceptual model and wireframes, using the USE Questionnaire method to evaluate the app. The study cited previous research emphasizing the importance of audio, video, animation, and text in such apps. Addressing self-determination and social inclusion dimensions, the work aims to enhance accessibility during emergencies, empower individuals to engage in response efforts and foster community inclusion.

Nugroho et al. [21] developed a mobile application using the Ionic framework to provide an alternative means of digital learning for persons who are deaf or hard of hearing. This application enables teachers and parents to help deaf students recognize words and object shapes, making learning more engaging and interactive. The authors used a prototype model consisting of stages such as Communication, Quick Plan, Modelling Quick Design, Construction of Prototype, and Deployment Delivery and Feedback. By helping deaf students learn word recognition and object shapes, the application directly addresses the personal development dimension of quality of life for persons with hearing disabilities. It enhances their communication skills and educational opportunities, thus contributing to their personal growth.

Mohammdi et al. [22] introduced a system for translating Arabic Sign Language (ArSL) to facilitate communication between persons who are deaf or hard of hearing and the general public. The system enables speech-to-ArSL and ArSL-to-speech translation and outlines its architectural design. It addresses speech recognition and tokenization challenges while presenting sign images or sequences with text values as output. The educational application facilitates ArSL learning in academic settings, aiming to bridge communication gaps between deaf and hearing individuals and enhance interpersonal relationships.

Agrawal et al. [23] developed a sign language recognition system using CNN to interpret hand gestures from a live stream captured by a webcam. They modified and augmented the Sign Language MNIST dataset from Kaggle for improved variety. Two models were trained: one without a base model and the other using transfer learning with Inception V3. Achieving high accuracy, the system recognizes 13 letters of the sign language alphabet in real-time, addressing the dimension of interpersonal relationships in the quality-of-life model by Schalock and Verdugo. This facilitates communication between deaf and mute individuals and others, enhancing their interpersonal relationships.

Samonte et al. [24] present the development of BridgeApp, an offline mobile application that enhances everyday communication for individuals with diverse communication abilities. The app incorporates features like text-to-speech, Speech-to-Text, Text-to-Image, Haptic Feedback, and American Sign Language (ASL) and Filipino Sign Language (FSL) to cater to the needs of persons who are deaf, hard of hearing, mute, and persons without disabilities. The study encompasses user acceptance testing and covers essential FSL signs, validating the app's usability. Additionally, the article highlights the prevalence of disabilities, hearing loss, and assistive technology. The app directly improves the quality of users' interpersonal relationships by facilitating communication and interaction.

Chuckun et al. [25] developed the mobile application HandyApps to improve the lives of individuals who are visual, speech, or deaf and hard of hearing. This application incorporates text and face recognition, sign language, speech-to-text, and text-to-speech capabilities, following a comprehensive analysis of existing mobile applications for the visually and hearing impaired. The app promotes material well-being by fostering independence and efficiency in daily activities, offering functionalities like text, object recognition, and face detection. Additionally, its features, including detecting alarms, doorbells, and sounds, enhance the physical well-being of hearing-impaired individuals by

ensuring safety and facilitating appropriate responses to their environment.

Syahriza et al. [26] discuss the development of a sign language interpreter tool emphasizing the use of Kansei Engineering. They propose using the Data Glove concept for accurate finger movement detection, addressing the limitations of existing interpreters. The hardware design includes an Arduino master, flex sensor units, DFplayer Mini, LCD, and a Buzzer Speaker. This innovative approach aims to accurately detect finger movements for sign language interpretation, promoting personal development and social inclusion. By bridging communication gaps, the tool enhances interaction, reduces barriers, and enriches socialization opportunities for deaf and mute individuals.

Reza Naqvi et al. [27] highlight the challenges deaf users face with audio and video interfaces, stressing affordable, accessible designs. They emphasize user feedback for developer improvement and adherence to sign language rules. Design principles like learnability and satisfaction aim to enhance communication and social interaction skills, aiding deaf individuals in overcoming barriers.

Alireza Darvishy [28] led a special thematic session on IoT applications for persons with disabilities and the elderly, showcasing innovative solutions and successful case studies. The session featured a brain-computer interface (BCI) driven domotic control system, BUZZBAND vibrating wristband for sound alerts, alternative input methods for hands-free interaction in augmented reality, and Ontenna, a tactile feedback device for hearing-impaired individuals. These technologies target improving personal development, independence, and engagement in daily activities for their users.

Jindal et al. [29] propose a method to convert sign language to text, utilizing MATLAB and AlexNet to identify seven movements with 70% accuracy. They also introduce a Python-based Convolutional Neural Network (CNN) model recognizing 20 actions with 98% training accuracy and 94% testing accuracy. Both models use real-time webcam image capture and pre-processing techniques, including thresholding, segmentation, and edge detection. Feature extraction involves contour, angle, color, and histogram methods, with MATLAB employing k-means clustering and Python utilizing skin color filtering and canny edge detection. Morphological operations like dilation, erosion, and Gaussian smoothing are employed to remove noise. This research directly addresses the communicative aspect of quality of life for persons with hearing/speech impairments, aiming to improve the communication capabilities of deaf and dumb persons.

Mubin et al. [30] proposed a framework for enhancing the learning experiences of deaf and mute children using Extended Reality (XR) technology. The framework involves three phases: Analysis, Design Development, and Evaluation. It gathers learners' concerns, utilizes XR components like VR and AR, and evaluates effectiveness. The aim is to optimize capabilities in a classroom by creating an immersive, user-friendly virtual environment. The research directly targets personal development, aiming to enhance learning experiences for the target group through customizable XR technologies.

Cano et al. [31] investigate Internet of Tangible Things (IoTT) integration with Tangible User Interfaces (TUIs) for children with hearing impairments, assessing its impact on social, emotional, cognitive, and visual skill development. The article examines how IoTT can foster the development of crucial social, emotional, cognitive, and visual skills through four case studies involving children with cochlear implants and those using sign language. Each case study delves into specific IoTT applications, from cognitive rehabilitation mini-games to interactive toys and educational tools. The study underscores considering social interaction and emotional feedback in IoTT design, highlighting IoTT's educational benefits, including personalized learning and real-time progress tracking.

Although some works do not directly address specific dimensions of quality of life as the Schalock and Verdugo model proposes, they indirectly address some dimensions. Firstly, in terms of personal development, the researchers seek to create a technology that bridges the language gap, potentially opening avenues for personal growth within the deaf community. Secondly, in terms of empowerment, the proposed technologies aim to reduce the reliance on intermediaries for communication, thus enabling hearing-impaired persons to express their needs and wishes more independently. Finally, improved communication accessibility can improve interpersonal relationships between the deaf community and hearing individuals, fostering better understanding, empathy, and social inclusion.

Regarding the dimensions of quality of life outlined in Verdugo's model, we can analyze the above works as follows. To better understand the specific dimensions targeted and indirectly impacted by each work, we categorized them as directly addressed, indirectly addressed, or not addressed. Table 2 summarizes the study results based on their direct and indirect impact or lack thereof. The table highlights how MoSIoT fills the gaps left by other proposals and offers significant advances and improvements in addressing the quality of life of individuals with hearing loss. While the studies above have made valuable contributions to the field, our work surpasses them in terms of comprehensiveness and coverage of different dimensions of QoL. The MoSIoT framework goes beyond this by integrating specific health monitoring systems that address the physical well-being dimension of QoL. It includes features such as tracking vital signs, monitoring medication adherence, providing alerts for medical emergencies, and empowering persons with hearing loss to take control of their health and well-being. This

Table 2 Comparison of Research Studies Based on Coverage of Quality-of-Life Dimensions

Related works	Quality of life							
	Emotional well-being	Interpersonal relationships	Personal development	Material well-being	Physical well-being	Self-determination	Social inclusion	Rights
Nakul Nagpal	∝	✓	∝	×	×	∝	✓	×
Gumay	∝	∝	×	∝	×	✓	✓	×
Nugroho	∝	∝	✓	×	×	×	∝	×
Mohammadi	∝	✓	×	×	×	×	∝	×
Agrawal	∝	✓	∝	×	×	∝	∝	×
Samonte	∝	✓	∝	×	×	∝	∝	×
Chuckun	∝	∝	∝	✓	✓	∝	∝	×
Syahriza	∝	✓	×	×	×	∝	✓	×
Naqvi	∝	✓	∝	×	×	∝	✓	×
Darvishy	∝	✓	✓	∝	∝	∝	∝	×
Jindal	∝	✓	∝	×	×	∝	✓	×
Mubin	∝	∝	✓	×	×	∝	∝	×
Cano	✓	∝	∝	×	×	∝	✓	×
MoSIoT	✓	✓	✓	∝	✓	∝	∝	×

✓ Direct address, ∝ ndirect address, × No impact

comprehensive approach ensures that persons with hearing loss access comprehensive healthcare support. Sections three and four of the present paper discuss how our system enhances interaction and improves accessibility for persons with hearing loss. By offering a complete solution, we address multiple dimensions of quality of life and provide a more holistic and enriched experience for persons with hearing loss. In addition, our work goes beyond the scope of a single application or system. We present a simulation framework for IoT systems that highlights the potential of IoT technologies to positively impact the lives of persons with hearing loss. This framework integrates and manages different assistive technologies, facilitating personalized and tailored solutions for individuals with specific needs.

In summary, our work goes beyond the existing studies discussed in the previous findings by providing a comprehensive and integrated approach to addressing the quality of life of individuals with hearing loss. Our complete MoSIoT system incorporates specific health monitoring systems, improves interaction, and enhances accessibility, thereby covering multiple dimensions of quality of life in more detail. This is further discussed in section four of our article.

3 A MoSIoT overview

MoSIoT is a framework that focuses on the representation of different scenarios faced by an IoT Health Monitoring System (HMS) for persons with disabilities. It uses a Model Driven Engineering (MDE) approach to generate a data-intensive simulator, allowing it to inspect the complex behavior of the IoT HMS under different hypotheses or scenarios. The framework comprises a Domain Model, where a domain expert specifies a knowledge base that gathers invariant information from recognized standards, and a Scenario Model allows a medical expert to introduce the static and dynamic behavior of specific entities for a given scenario. These interconnected models allow the invariant concepts of the domain model to feed the variable entities defined in each scenario. In addition, MoSIoT incorporates a Trigger-Action Programming paradigm, allowing the creation of rules and recipes that further enhance the accessibility and usability of the HMS IoT for individuals with disabilities.

The MoSIoT framework comprises several packages within the Domain Model: (1) The Patient profile package takes center stage in addressing the unique requirements of individuals with disabilities. It draws inspiration from accessibility standards such as AccessForAll (AFA) [7] to ensure that users' accessibility needs and preferences are adequately represented. This package encompasses access modes and diverse adaptations tailored to different disabilities. The Patient profile also gathers detailed information such as preferred language, region, conditions, and disabilities, which define the person's Access Mode elements within the IoT system. This customization ensures precise communication methods and language preferences. For example, devices can adapt to various cultural norms, gender-specific choices, and age-related usability issues, utilizing features like voice interpreters and smartwatches. During the Scenario Model definition phase, domain experts collaborate directly with individuals with disabilities, such as social workers or healthcare professionals, tailoring IoT scenarios to their

specific needs. This includes considering cultural nuances, gender identities, and age-related factors to customize IoT devices and services accordingly. For instance, a Spanish-born individual who is deaf might use sign language with the Spanish version, while a young American woman with acquired deafness might prefer textual language in American English; (2) The Device package focuses on defining IoT device templates that are essential to the operation of the framework. These templates describe the characteristics of the devices, including behaviors, commands, properties, and telemetry. The idea is to have a standardized representation of devices used within the health monitoring system. This package considers established standards such as the W3C Web of Things to ensure interoperability and adoption of standard practices. In practice, this means that a domain expert can configure and define new devices within the system, ensuring they are tailored to specific patient needs and align with industry best practices and standards; (3) The Healthcare package within the MoSIoT framework focuses on providing comprehensive care for persons with disabilities. This package leverages several health information standards, including ISO13606, OpenEHR, and HL7 CDA, with a particular emphasis on the HL7 Fast Healthcare Interoperability Resources (FHIR) [8]. Using the FHIR-based CarePlanTemplate, healthcare providers can define clinical or administrative activities, set goals, and design communication strategies tailored to the specific needs of patients with disabilities. The idea is to have a standardized and personalized approach to care, ensuring that interventions and treatments are appropriate and effective for each individual. This means that care plans can seamlessly integrate with health information systems, ensuring efficient coordination and patient-centered care.

One of the most important components of MoSIoT is the Scenario Model. This multifaceted tool simulates various situations in IoT Health Monitoring Systems (HMS) and communication environments involving SAAC (Augmentative and Alternative Communication Systems). This simulation capability is essential to anticipate and adapt interventions in various scenarios. For example, it is possible to simulate how a deaf person uses an SAAC to interact and communicate in a building, using augmented reality tools that provide visual and contextual information. In addition, the intuitive design and high level of abstraction of the Scenario Model make it easy for medical experts and professionals, such as social workers and educators working with persons with disabilities, to define and customize these scenarios. By raising the level of abstraction, MoSIoT ensures that these professionals focus on patient's needs and well-being according to their cultural norms, gender dynamics, and age-related preferences. At the same time, the system manages the technical complexities, offering tailored and personalized solutions for each situation.

Generally, MoSIoT stores the expert's knowledge within its domain model. Within this model, a feedback loop and initial expert insights lead to creating various templates. These templates encompass the User profile, which archives information from past patients; the Device Template, housing data on previously defined devices; and the CarePlan Template, which catalogs care plans for individuals with disabilities and other conditions, including their activities. This framework enables experts to utilize previous information when defining new scenarios, streamlining the process by leveraging insights from different experts and past experiences.

The MoSIoT framework transforms IoT connectivity using an IoT Hub. The Hub connects IoT devices with cloud services as a centralized bridge while managing and securing communication. The Hub facilitates interoperability by abstracting protocols, enabling devices to communicate across diverse networks like MQTT, AMQP, and HTTP. Moreover, it ensures compatibility by converting messages into various formats for efficient data processing by cloud applications. Scalable and adaptable, the framework accommodates deployments of any scale, from thousands to millions of devices, while integrating with cloud services like data analytics and AI for comprehensive insights. Robust security measures, including per-device authentication and encrypted data transmission, safeguard against potential threats, ensuring the integrity and confidentiality of IoT data.

3.1 Applying MoSIoT to the quality-of-life dimensions

Based on our research, we establish that MoSIoT can improve four dimensions of Verdugo's quality of life directly:

1. Personal development: We explore how integrating IoT technologies and a health monitoring system can empower individuals with disabilities to enhance their skills and knowledge [32] and overall personal growth. By providing accessible tools and resources, we aim to enable individuals to expand their capabilities and achieve their full potential. In this sense, emerging technologies such as virtual reality (VR) or augmented reality (AR) are tools with enormous potential in the field of education due to their motivational aspect and because they make teaching content possible that would otherwise be unfeasible. To further support their development, we propose using a personalized time-tracking system. This system would encourage consistent learning by tracking reading and writing practice hours; this tracking system would offer insights into their progress. We can better understand their personal development journey by recording and analyzing time spent on these

activities – manually or automatically, for example, by smart pens, smartphones, or smart watches;

2. Physical well-being: indicators include health condition, nutrition, mobility, leisure, activities of daily living, and Health monitoring system. MoSIoT provides scenario modeling, where not only the wearable devices that collect health information are set up but also a care plan where the care activities that the person with a disability must perform are planned, be it nutrition, medication, communication with the doctor, or physical activities. For instance, we use smartwatches to measure sleep duration, enabling us to calculate and enhance the dimension of physical well-being;
3. Interpersonal relationships: By leveraging IoT solutions, we develop communication platforms and devices that facilitate seamless communication between individuals with hearing impairments, friends, family, and the broader community. In addition, we recognize the importance of positive relationships, including friendships, family bonds, and romantic partnerships, in promoting well-being and overall quality of life [33, 34]. At this point, MoSIoT prioritizes the implementation of a SAAC (Systems of Augmentative and Alternative Communication); SAACs [35] refer to the set of strategies, including technological devices, that promote the autonomy of individuals with communication disorders, enhance their social abilities, adaptive behavior, and family relationships. SAAC might include real-time translation capabilities, enabling seamless interactions between individuals with hearing impairments and their hearing friends, family, and the broader community. This fosters closer relationships, reduces social isolation, and promotes overall well-being. The MoSIoT platform is a channel for instant conversations, showing ongoing dialogues and counting message exchanges. With the total messages exchanged, we match this data with a set care plan target to get a proper measure. This measure helps us understand how deeply the person with a hearing disability interacts with persons, supporting our aim to assess interpersonal relationship levels;
4. Emotional well-being is another dimension we address through our research. There are two ways to help emotional well-being: by monitoring stress values and through health monitoring systems, which monitor blood pressure, pulse, and temperature, they can predict whether their stress level is altered. On the other hand, using recreational activities such as reading, games, or medical apps based on mobile phones or augmented reality will improve the emotional level of persons with disabilities [36, 37]. Our platform encompasses a range of care activities, all contributing to enhancing and improving emotional well-being by adding appropriate activities, which we monitor through smartwatches. Stabilizing these indicators signifies a tangible enhancement in the quality of life.

In the following, we indicate how the use of MoSIoT can indirectly help the other dimensions:

5. Social inclusion: By leveraging accessibility for persons with disabilities in IoT systems, we aim to break down barriers and promote equal participation in various social settings. We facilitate individual engagement in community activities through assistive technologies, enabling active involvement and fostering a sense of belonging [38, 39]. For example, it might facilitate participation in group activities by providing real-time sign language or speech-to-text interpretation, ensuring individuals can actively engage in discussions, meetings, and community events. Additionally, it could enable deaf individuals to access public announcements and information through IoT-connected devices, fostering a more inclusive and integrated environment.
6. Material well-being: The IoT-enabled smart home security systems with cameras, motion detectors, door/window sensors, real-time monitoring, and alerts can enhance safety [40]. Additionally, smart communication devices facilitate seamless participation in remote work, virtual meetings, and collaborative tasks, promoting employment opportunities. These systems also enable access to essential information and services through speech-to-text-equipped devices, enhancing accessibility to employment, finances, and rights-related resources. Furthermore, IoT systems facilitate supportive social and employment environments by enabling real-time sign language interpretation and enhancing interactions at work.
7. Self-determination: IoT systems offer a pathway to autonomy, choice-making, and personal control, allowing individuals with hearing impairments to express preferences with clarity through communication-enhancing devices, such as real-time translation-capable smart communication tools. Smart home automation furnishes environmental management, while wearable devices facilitate well-informed decisions. Tasks like reminders and routines become self-managed through IoT devices, and customizable settings in these systems harmonize with personal values. Meanwhile, platforms elevate social interactions and financial independence, and IoT-enabled event information fosters community engagement. This seamless integration emboldens individuals with hearing impairments, empowering them to assert agency, nurture independence, and align with their aspirations for self-determined growth.

4 A case study: improving the quality of life of a hearing-impaired person

Hearing impairment is a widespread disease affecting millions of persons worldwide [41]; it is very important to implement effective interventions to improve the quality of life of those affected. As explored in the study by Holman et al. [42], individuals with hearing impairment encounter various challenges that significantly impact their daily lives. One major issue is pervasive fatigue, which stems from various sources. The heightened mental effort required during conversations leads to cognitive fatigue, where individuals must concentrate intensely and focus on non-verbal cues to understand spoken words. Emotion-driven fatigue adds to the burden, as negative emotions like frustration, stress, and anxiety arise from the struggle to hear and actively participate in social interactions. Even without explicit verbal expressions, the necessity to take breaks and recover in social situations is evidence of the underlying mental fatigue experienced during conversations. Sleep disruption, linked mainly to tinnitus, contributes to ongoing tiredness. The perceived relationship between hearing impairment and fatigue varies among individuals, suggesting a complex interplay. Considering the intricate balance of these challenges, the study underscores the need for tailored strategies and support to enhance the well-being of individuals with hearing impairment.

In an extensive examination undertaken by Maila Trujillo [43], the focus was placed on assessing the quality of life of 50 persons with hearing disabilities. This in-depth study produced results concerning this demographic's perceived quality of life. Interestingly, the findings disclosed that 40% of the participants reported a notably low quality of life. Moreover, 34% of the subjects considered their quality of life less than satisfactory. Furthermore, 24% of them described their quality of life as acceptable, while only 2% reported experiencing a positive, good quality of life (Table 3).

The aim is to detail which aspects of quality of life significantly impact the lives of deaf individuals with low levels of hearing. In the study of Maila Trujillo [43], the different dimensions of quality of life provided by the GENCAT scale [44] were measured. This scale facilitates greater objectivity in measuring users' quality of life. The results of the participants with disabilities are presented below. According to the study by Maila Trujillo [43], it is essential to recognize that the low quality of life experienced by individuals with hearing impairments is closely tied to factors such as inadequate research, the influence of coping strategies, the presence of limited social support, and the imperative need for heightened awareness and targeted interventions to address these complex challenges.

A simple way to visualize a person's quality of life is to use a radar chart (Fig. 1). This graph visually portrays the values extracted from different dimensions of the table, depicting a polygon with vertices aligned to the weights assigned to each dimension. It highlights areas characterized by varying low, medium, and high satisfaction levels across different quality-of-life dimensions.

4.1 Applying MoSIoT for improving the quality of life of persons with hearing impairment

Following the use case proposed in this work, we apply the MoSIoT framework to model and obtain IoT scenarios for persons with hearing impairment. To do so, we first start by indicating the process that is followed from the introduction of the generic templates in the domain model through the definition of the simulated scenario and, finally, the real scenario. Next, we discuss one of the main contributions of this work, which is the introduction of the quality-of-life framework within MoSIoT itself, thus

Table 3 The quality of life of hearing-impaired persons on the GENCAT scale [44]

Dimension of Quality of Life	Low	Medium	High	Average	Number of participants	Description
Physical Well-Being	46 (92%)	4 (8%)	0 (0%)	0,03	50	Low satisfaction level of 92%
Emotional Well-Being	7 (14%)	17 (34%)	26 (52%)	0,65	50	The level is high, with a representation of 52%
Social Inclusion	16 (32%)	18 (36%)	16 (32%)	0,48	50	The population assumes a medium satisfaction level of 36%
Material Well-Being	44 (88%)	6 (12%)	0 (0%)	0,05	50	They expressed low satisfaction in 88% of the total population
Interpersonal Relationships	3 (6%)	23 (46%)	24 (48%)	0,67	50	Most of the population has a high level of satisfaction, with 48%
Personal Development	6 (12%)	21 (42%)	23 (46%)	0,64	50	A high level of satisfaction was evidenced in 46%
Self-Determination	3 (6%)	38 (76%)	9 (18%)	0,53	50	Manifested a medium satisfaction level of 76%
Rights	47 (94%)	3 (6%)	0 (0%)	0,02	50	Low level of satisfaction of 94%

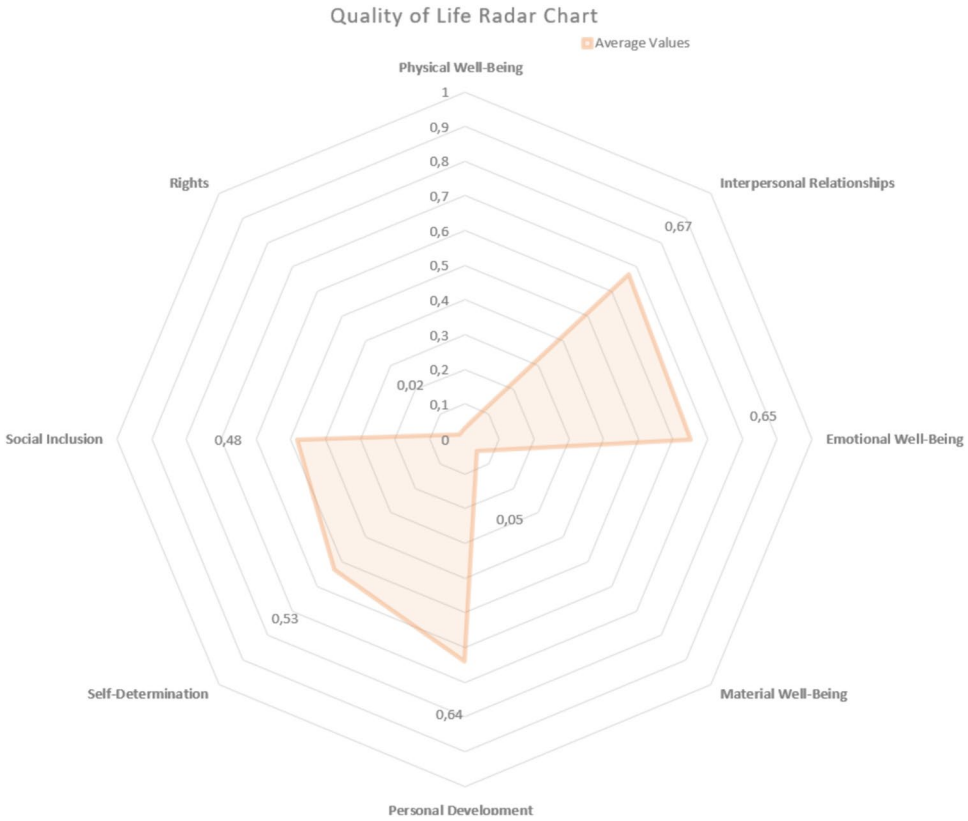


Fig. 1 The Quality-of-Life Radar Chart of Hearing-Impaired Persons with Initiative Values

allowing well-defined objectives to be set for the improvement of different dimensions of quality of life thanks to the proposed IoT system.

We started with the MoSIoT process (see Fig. 2) to define the use case for the hearing impaired. On the left side of Fig. 2, we can observe the configurations previously made by the MoSIoT Admin. At this point, the admin introduces the concepts that can be used in different cases of persons with disabilities based on previous cases and the literature. This is the case of the Alexa voice interpreter device, used to activate or deactivate devices; we start from the fact that it introduces a screen to be able to send text-to-speech messages to it and that it is also in charge of transcribing the voice messages to text on that screen.

Moving to the middle section, attention shifts to the Domain Expert, who has persistently pursued the completion of the system by integrating additional devices and fine-tuning the accessibility, care plan, and care activity configurations.

In addition to configuring the voice interpreter device for accessibility, the Domain Expert has introduced another device: the smartwatch. The access mode for the smartwatch is tailored to the user's preferences, including textual accessibility as an 'AdaptationRequest,' vibration as an 'AdaptationTypeRequired,' and synthesis as an 'AdaptationTypeDetail'.

Two care plans have also been implemented to address stress and fatigue. These plans include specific goals, such as walking for two hours daily and ensuring seven hours of sleep each night.

Regarding care activities, the Domain Expert relies on the appointment with the therapist, which the MoSIoT administrator created, but a different type of care activity has been added, focusing on medication, with a daily teaspoon of magnesium to promote well-being.

One of the main contributions of this work is the ability of MoSIoT to assess the quality of life of a person with or without a disability. To this end, we proceed to introduce in the MoSIoT domain model, specifically in the Healthcare package, the elements derived from Verdugo's framework [3, 4], i.e., the different dimensions (Physical Well-being, etc..); however, we will focus only on those that we have indicated in Sect. 3.1, which are directly improvable through the use of IoT technologies, i.e., Personal Development, Physical Well-being, Interpersonal relationship, and Emotional well-being. The remaining dimensions may benefit indirectly, as indicated above.

To apply the MoSIoT assessment of a person's quality of life, we rely on one of the most important domain model packages of the proposal, namely the Healthcare package, which contains the care plans from the FHIR standard. In them, on the one hand, the Domain expert introduces a care plan based on a series of activities, including physical activities, nutrition, medication, and communication. The care plan is linked to the person's condition, and in this

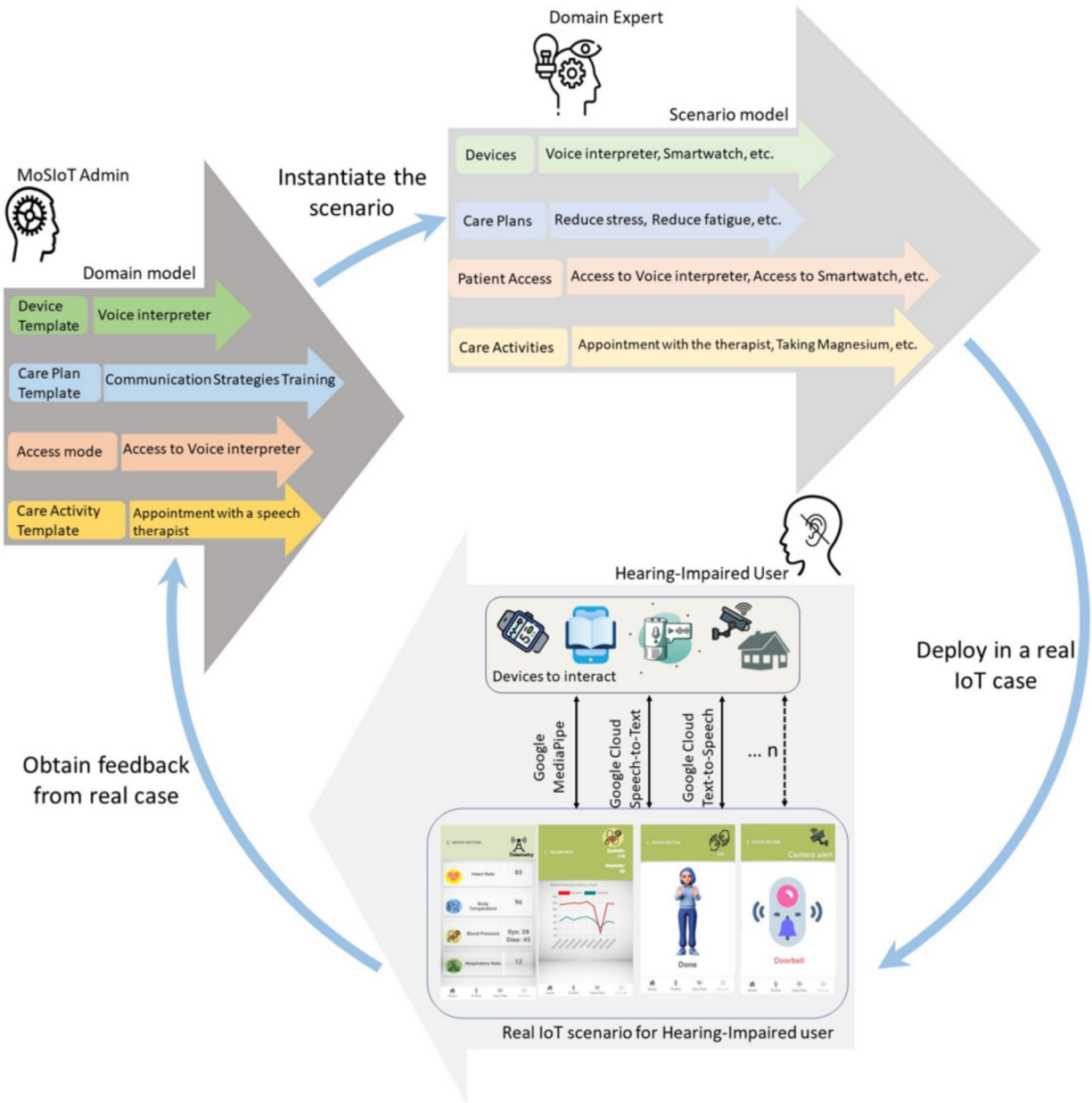


Fig. 2 MoSIoT Process Cycle with Admin, Domain Expert, and User, illustrating the progression from generic template introduction to real scenarios, incorporating personalized IoT solutions

condition, in addition to indicating the person's illness/s and disability/s, the specification of the dimensions of quality of life is introduced as a novelty. However, in order to establish a measurement of the improvement in the quality of life, we need a Goal, or the entity where the objective to be achieved with the care plan is specified, to be directly linked to the dimension or dimensions of quality of life to be achieved. This goal could be expressed as "improve physical activity by 40%", thus linking the goal to the physical well-being dimension of quality of life. On the other hand, the Target, the entity where the objective is specified in a measurable form, would indicate that the measure is to be linked to the telemetry "a number of steps," thus indicating an "increase by 40% the number of steps". Finally, such telemetry is linked to the smartwatch device within the IoT system.

4.2 Domain model for persons with hearing impairment

Before defining the scenarios for hearing-impaired patients, it is necessary to introduce generic concepts or templates in the MoSIoT domain model, their responsibilities, and

relationships that apply to any case of a deaf person and are also independent of any technology used. As stated in [5], the MoSIoT domain model is divided into three distinct packages: (1) the patient profile package, which allows for the definition of the adaptation profiles of patients based on the types of disabilities and conditions; (2) the device package, which defines the device templates or types of devices used in these systems with their characteristics and the types of telemetries; (3) The healthcare package, which proposes different care plan templates with activities, goals, and communications for patients with specific conditions and disability types.

In this way, we will focus on the most significant elements of the hearing-impaired person's case study. The first step is to enter the data corresponding to the patient profile package because it will collect the common characteristics of disability and accessibility that often occur with hearing-impaired persons.

In Fig. 3, the image is divided into three distinct sections. On the left side, the admin has provided patient profile details, including a general situation overview and the patient's preferred language and region. In the central portion of the image, Diabetes and Depression are introduced as diseases afflicting the patient, each accompanied by their respective descriptions. Finally, the right side of the image is dedicated to the classification of disabilities based on the Web Accessibility Initiative (WAI) W3C guidelines [18]. In this context, the admin has selected the "auditory" disability and described its characteristics. Additionally, the severity of this auditory disability has been specified as "severe."

Although no standards to address disability for IoT currently exist, MoSIoT relies on initiatives such as the ISO/IEC 24751 AccessForAll (AFA) specification, which proposes a reference model that represents users' personal accessibility needs and preferences and controls the presentation of the information.

Focusing on the accessibility side, the admin creates different configurations for different hearing levels or degrees of knowledge presented by these persons. For example, two traditional ways to access a device that emits some kind of sound, either a voice interpreter or alarm, are a speech-to-text transcription for deaf persons who have adequate reading skills or a real-time transcription to a sign language for those who only know that form of communication.

According to Caitlin McKeown et al. [45], the typical graduation level of deaf students from high school corresponds to a fourth-grade reading proficiency, with around 20% concluding their education at a second-grade reading level or below.

The main element of the Domain Model that defines how a person interacts with a device is the access mode. Within each AccessMode, provide the framework with general adaptation elements for any hearing-impaired person.

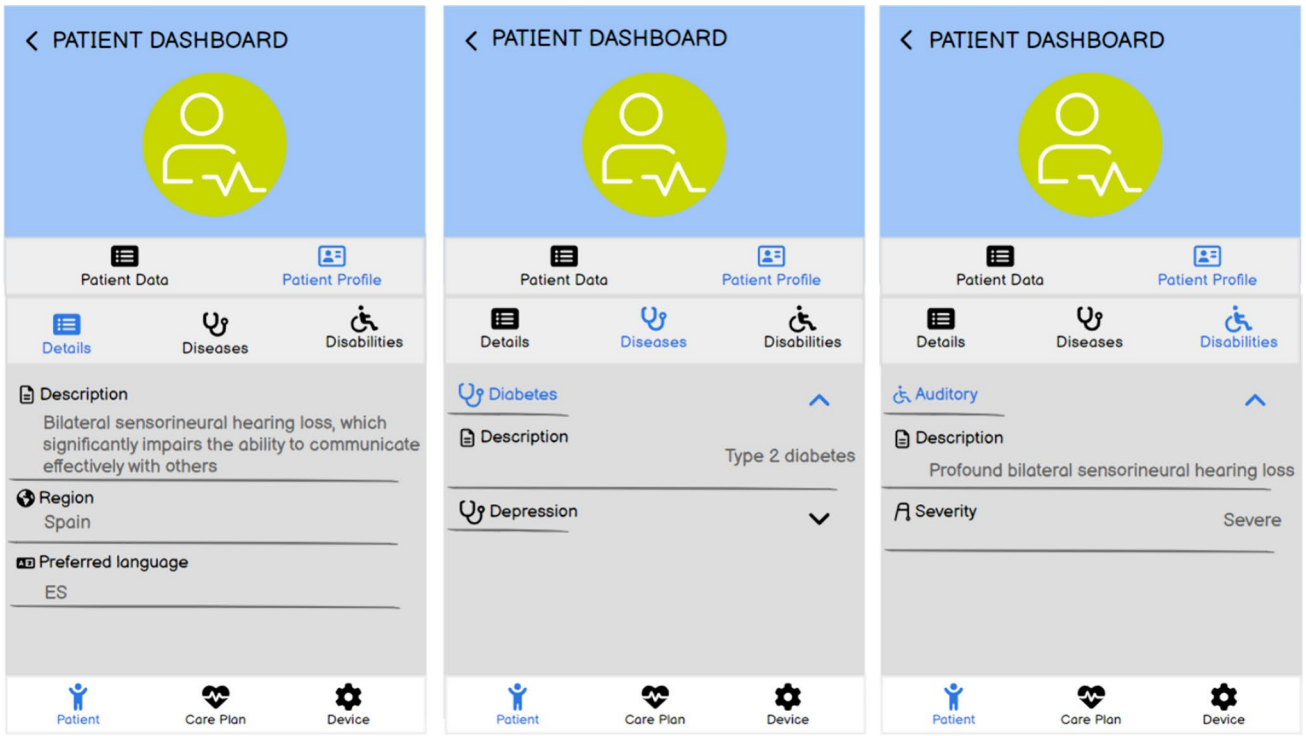


Fig. 3 The domain model for Patient Profile detailing general essential patient information with specific details of diseases and disabilities

Figure 4 displays the three key elements of AccessMode values in the Detail access mode tab.

In the left section of Fig. 4, we can observe the **AdaptationRequest**, configuring the preferred access mode for devices relying on auditory or voice interactions with deaf individuals. The admin introduces the Textual mode in this section, enabling users to receive text-based information instead of relying on auditory cues. This not only eliminates the reliance on auditory cues but also ensures inclusivity.

In the center section of Fig. 4, we find **AdaptationTypeRequired**, based on deaf accessibility. This element configures the required adaptation type. In this scenario, the admin assigns the required adaptation type as Sign language, allowing users to receive information both textually and through advanced sign language interpretation.

Finally, in the right section of Fig. 4, we have **AdaptationTypeDetail**. This element optimizes the settings for real-time interpretation within the Sign language adaptation, ensuring real-time communication for users with hearing disabilities.

By leveraging the AFA-based MoSIoT framework, the admin ensures that devices adequately meet the needs of deaf accessibility, enabling them to interact with the devices using text-based information complemented by real-time sign language interpretation.

The next step is to include the elements in the device package that allow the definition of IoT device templates for persons with hearing disabilities. Figure 5 shows that a version of Amazon Alexa is introduced that features a display on which a deaf person can visualize text information that is also provided by voice. After configuring the new device template and introducing its data (such as Amazon Alexa, etc.), the Simulator admin assigns the access mode to the device, which has previously been created. It allows the person to enter messages to the Alexa display in textual mode to give orders to the devices or receive information from them in real time by sign language.

The last element we are going to create is a care plan template for hearing hearing-impaired disability. The health package's core element is the CarePlanTemplate, which is based on the HL7 FHIR CarePlan [8] to define all the elements (Activities, Communications, and Goals) that will enable the establishment of an IoT-assisted care plan. The CarePlanTemplate allows you to define a generic treatment plan for a specific disability; in this example, we define a care plan for persons with severe hearing impairment (see Fig. 6).

As can be seen in Fig. 6, an audiological assessment is defined as detailing audiograms and specific hearing thresholds or levels of hearing loss.

For instance, Fig. 6 shows a care plan to enhance communication strategies training. The primary focus is integrating communication devices to facilitate interactions across diverse settings. This Goal element (see Fig. 6. b) is linked to the dimension of quality of life and interpersonal Relationships, and the objective is that if we start from 67%

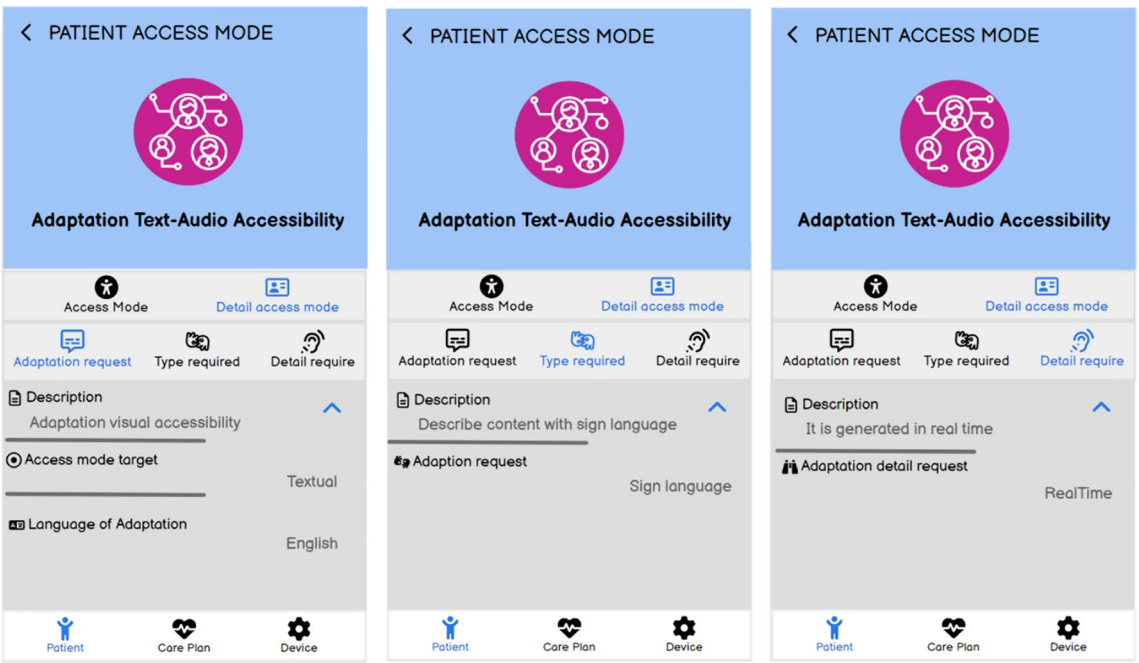


Fig. 4 The domain model Access mode configures preferred access modes for deaf individuals, including textual and sign language options, in real-time

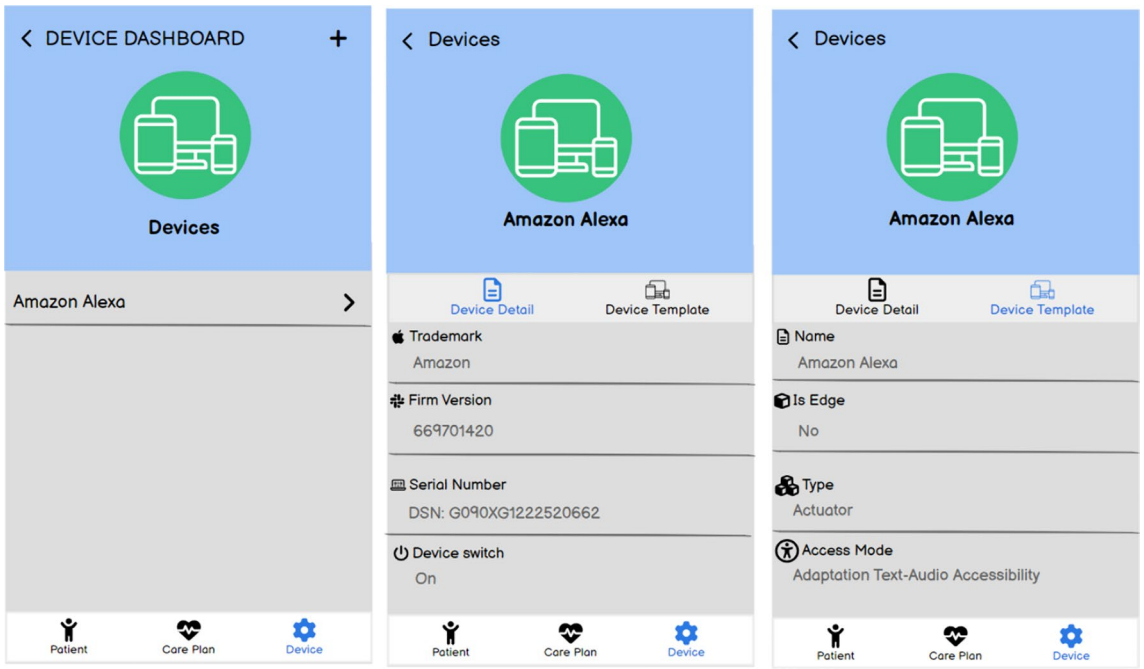


Fig. 5 The domain model device views with Amazon Alexa device with a display for deaf individuals, enabling real-time interaction through text and sign language

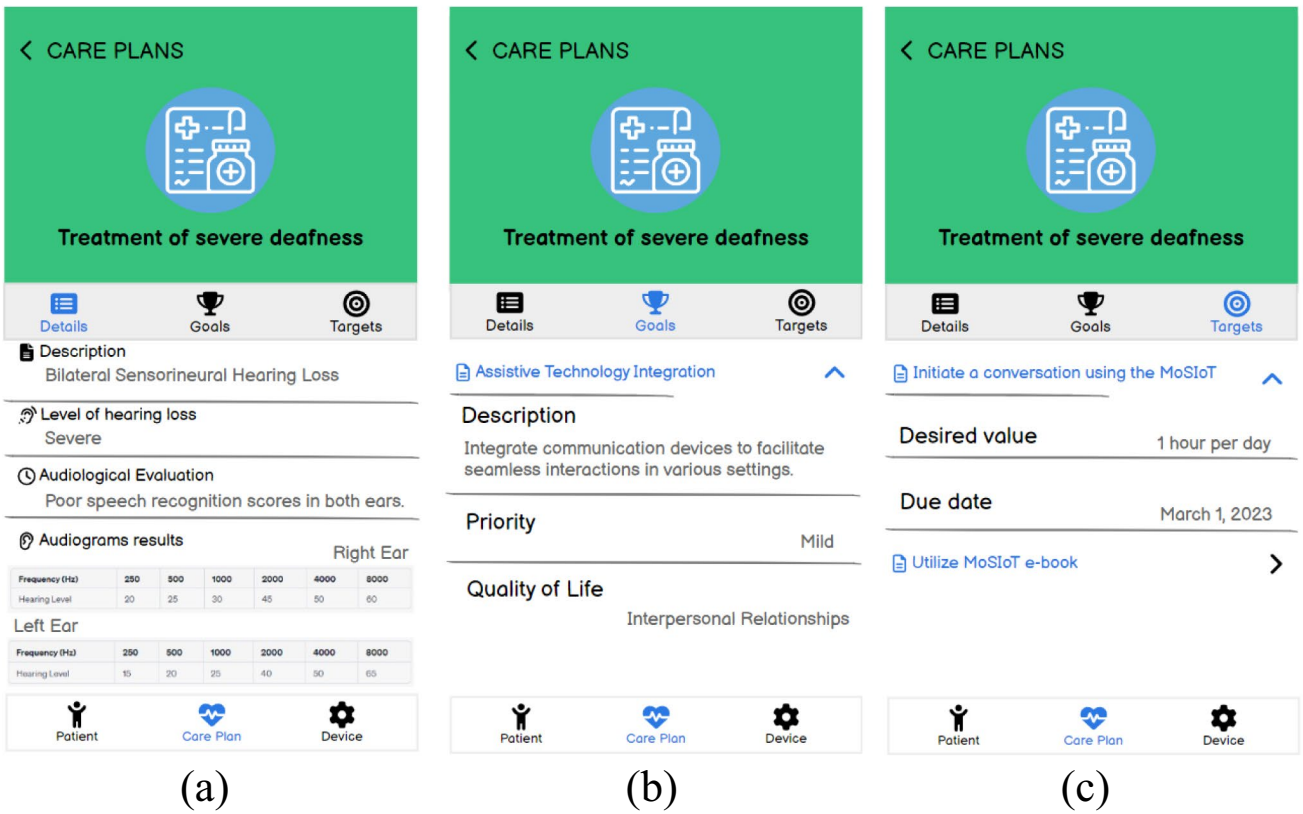


Fig. 6 A Care Plan template for severe hearing impairment detailing the treatment of severe deafness strategies, targets, and goals linked to the quality-of-life dimensions

satisfaction, taking the average value, we want to increase satisfaction by 20%. This objective involves initiating at least one hour of daily conversation through the MoSIoT messenger. The care plan emphasizes utilizing the MoSIoT e-book feature as an additional goal, which is the way the MoSIoT framework can measure the time that persons communicate over the platform. Incorporating this resource encourages individuals to explore and engage with content in an accessible format, fostering greater understanding and knowledge acquisition.

Ultimately, the Care plan template also includes a set of care activities. Specifically, Fig. 7 presents a care activity of type appointment that establishes a set of appointments with a speech therapist to be scheduled twice monthly. It is located in the Hospital of San Vicente in Spain. With support from speech therapists, patients can receive personalized strategies and exercises that enable hearing-impaired persons to navigate diverse communication situations more effectively. In the following section, a scenario model will be defined, which allows the establishment of a specific IoT scenario for a person.

4.3 A scenario model of a hearing-impaired person

Once the domain model contains the generic templates for hearing-impaired persons, the next step is to model the scenario model to define the IoT elements that address a person's specific case. In this case, it is the domain expert who will be a caregiver or medical expert with knowledge of hearing-impaired persons who know a patient with special needs, either through an interview or by observation, will be in charge of defining the scenario based on existing templates or defining them from scratch.

In Fig. 8, we present a data flow diagram that encompasses the IoT elements within the context of the persons with hearing disability scenario. The scenario includes the patient equipped with a smartwatch and a smartphone exhibiting accessible interfaces. The system also incorporates an e-book, equipped with an adaptive user interface, into the user's mobile device. This adaptation captures the user's duration with reading and writing materials, which we collect to generate telemetry. On the other hand, the voice assistant collects ambient audio data and relays it to the user's mobile device. This not only aids the individual in comprehending their surroundings but also enables Domain experts to understand the percentage of social interaction between the patient and their environment (work, relatives, and friends). Various data streams encompass the quartet of devices-the e-book, the smartwatch, the voice assistant, and the Smartphone. These devices collaboratively generate multiple data, including Heart Rate

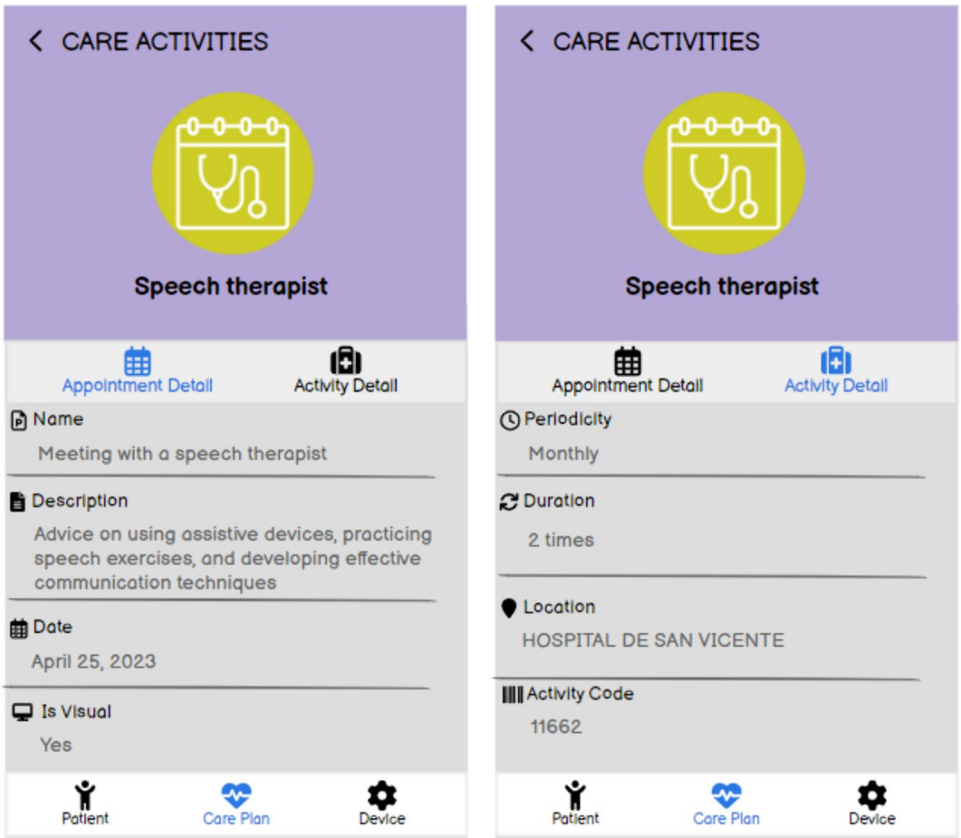


Fig. 7 Scheduler Simulator admin Care Activity view of appointments with speech therapists, who provide personalized strategies for effective communication for hearing-impaired individuals

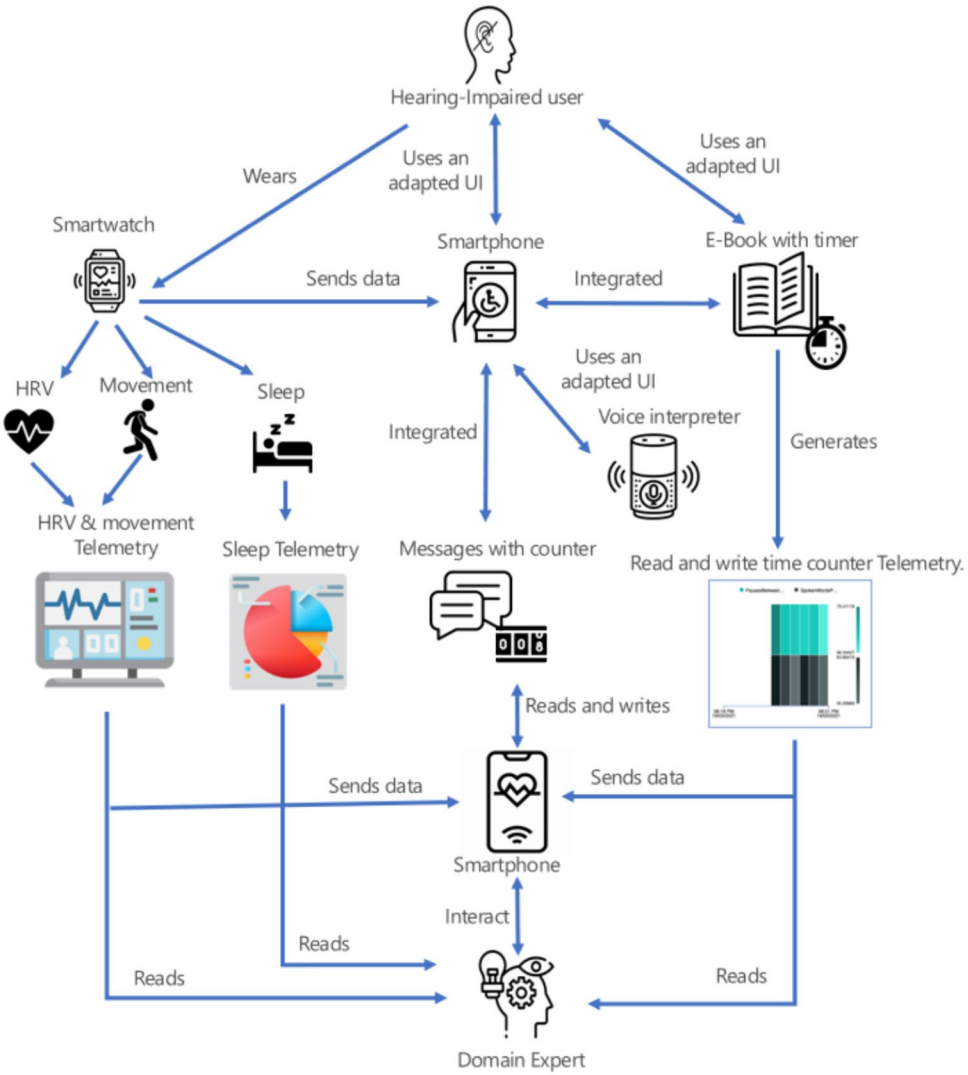


Fig. 8 Data flow diagram illustrating IoT elements in a scenario for persons with hearing disabilities, showcasing device interactions and data generation to enhance user experience and quality of life

Variability (HRV) and movement data, sleep-related metrics, and the count of read-and-write operations.

The Domain Expert plays a crucial role in interpreting the data and providing insights into the patient's condition. This interdependent relationship between data generation and interpretation highlights the clinical significance of this framework. Furthermore, the diagram captures the dynamic interaction between the Domain Expert and the patient. Communication occurs through messages, serving as a bridge for exchanging information.

Within this scenario, we define the trigger-action receipt element dictating how their devices should interact based on specific events or conditions. For instance, they represent a rule: "When I wake up in the morning, activate the coffee machine, display the most important news visually, and deliver a message about the weather forecast through vibration." Similarly, they establish another rule: "When I leave home for work, deactivate all devices, secure the door, and send a smartphone message detailing the traffic situation." Utilizing MoSIoT, they simulate these rules to validate correctness and feasibility before deploying them onto their devices. By incorporating machine learning into MoSIoT, predictive capabilities can be harnessed to anticipate events in an IoT scenario catering to deaf individuals, considering their specific attributes and contexts. For instance, the system could forecast potential risks or emergencies impacting individuals with hearing disabilities and provide visual or vibrating alerts. It might also predict communication requirements and furnish sign language translation or interpretation tools. Moreover, artificial intelligence could predict entertainment or educational preferences and offer tailored content catering to their comprehension and accessibility needs. The seamless integration of these IoT devices into their daily life significantly elevates their overall quality of life. Effective communication, convenient information access, and assured safety become tangible benefits. These devices not only assist them in overcoming the challenges posed by

hearing disabilities but also enrich their everyday experiences, enabling them to lead a more fulfilling life.

4.4 Results analysis of MoSIoT case study for hearing-impaired persons

In addition to instantiating all IoT elements, the scenario model has associated goals that are either derived from the care plan template or predefined by the domain expert. As indicated in Fig. 6, we can measure through the Target elements if, indeed, by the telemetry received, in this case, the hours dedicated to communication, the person is fulfilling a Goal associated with improving a dimension of the quality of life. For this purpose, the application updates the progress quantitatively and represents it in a quality-of-life assessment report. This evaluation report involves assigning weights to different dimensions of quality of life, which are then visualized through a radar chart. This visualization aids the Domain Expert in identifying the dimensions that have experienced positive changes and the areas that still require further enhancement.

In Fig. 9, an initial evaluation report primarily tailored to individuals with hearing disabilities has shown a limited overall impact on the quality of life due to its broad applicability.

Finally, in Fig. 10, we can see the progress of a patient who has improved in several quality-of-life dimensions by following the activities established in the scenario model's care plan.

Table 4 compares the initial values of various quality-of-life dimensions and the improvements observed in a patient's condition following intervention. Each dimension is evaluated based on specific criteria, with higher values indicating better quality of life. The patient's value improvements represent the positive changes achieved in each dimension after the intervention.

Highlighted cases include the dimensions of interpersonal relationships, emotional well-being, and personal development, where notable improvements were observed. Interpersonal relationships are assessed through the message counter of the MoSIoT messenger. Emotional well-being is evaluated using biometric values such as HRV (Heart Rate Variability) and hours of sleep.

Furthermore, personal development is significantly improved, as evidenced by the increased reading hours suggested in the care plan.

4.5 MoSIoT architecture design for deafness case study scenario

Figure 11 illustrates the application of the MoSIoT framework, leveraging Model-Driven Engineering (MDE) principles, within an enterprise architecture that relies on web services. This figure, describes the critical components of this architecture and their roles in the system.

In the center of the image, we have the core Web services, representing the heart of the MoSIoT architecture. These services have object-oriented business logic that controls the domain model's functionalities [46]. They also offer a secure REST façade for external access and ensure database persistence for storing and mapping domain-related data. To facilitate data management in the database, we employ

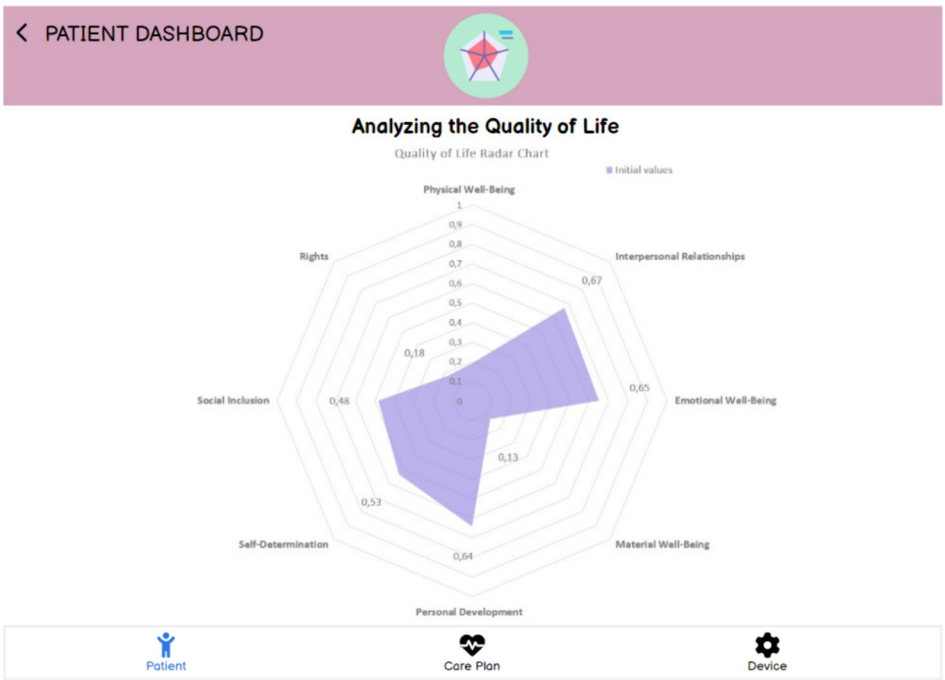


Fig. 9 The initial quality-of-life evaluation report, primarily tailored to individuals with hearing disabilities, shows a limited overall impact on the quality of life

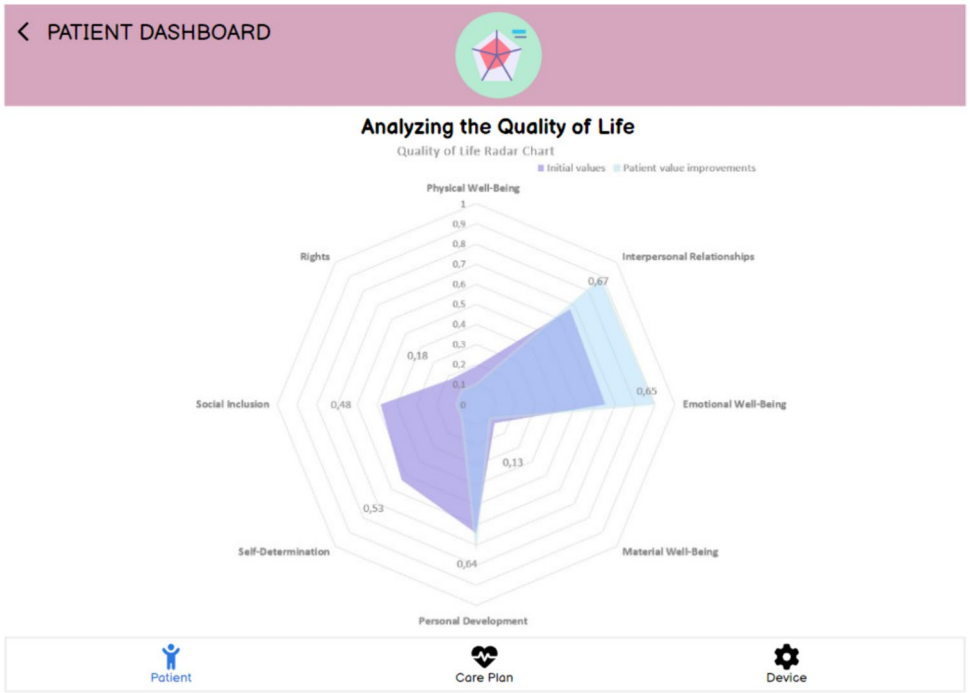


Fig. 10 The progress of a patient who has improved in several quality-of-life dimensions by following the activities established in the care plan of the scenario model

Table 4 Comparison of Initial Quality of Life Values and Patient Value Improvements

Dimension	Initial values	Patient value improvements	Explanation
Physical Well-Being	0.19	0.1	Slight improvement in Physical well-being
Interpersonal Relationships	0.67	0.89	Significant improvement in interpersonal relationships
Emotional Well-Being	0.65	0.9	Emotional well-being greatly improved post-intervention
Material Well-Being	0.13	0.1	Slight improvement in material well-being
Personal Development	0.64	0.72	Noticeable improvement in personal development
Self-Determination	0.53	0.1	Only minor improvement in self-determination
Social Inclusion	0.48	0.1	Limited improvement in social inclusion
Rights	0.18	0.1	Minor improvement in rights

an Object-Relational Mapping Framework (e.g., Hibernate). Each web server scenario encapsulates the system's behavior and includes a prediction module that feeds data to the IoT Hub.

To enhance the system's capabilities, we integrate Azure IoT Central, a potent IoT accelerator that provides a REST API for sending and retrieving IoT-related data. This integration facilitates remote communication and seamless access to the IoT system's information. To manage the backend implementation of the domain model efficiently and promote rapid updates, we apply the OOH4RIA [47] approach.

Moving to the left side of the image, we find the "Simulator or MoSIoT Admin," a web app component designed for managing information within the simulator, implemented according to the specifications outlined in Reference [48]. Developed using Angular, a Progressive Web Application framework, this tool enables the simulator admin to input generic information through provided forms, efficiently constructing the MoSIoT domain model.

We have a mobile app tailored to domain experts, which has been successfully implemented as described in Reference [49] and built using Ionic on the bottom left of the image. The app connects to the central core of the simulator and retrieves default information related to user profiles, connected devices, and care plan templates, specifically for Healthcare Monitoring Systems (HMS). Model-driven web Engineering (MDWE) transformations in adjusting entities within the IoT scenario, resulting in the dynamic conversion of the model into a MoSIoT simulator.

On the right side is the design and implementation of a mobile application intended for users with disabilities, which has been realized based on the details provided in Reference

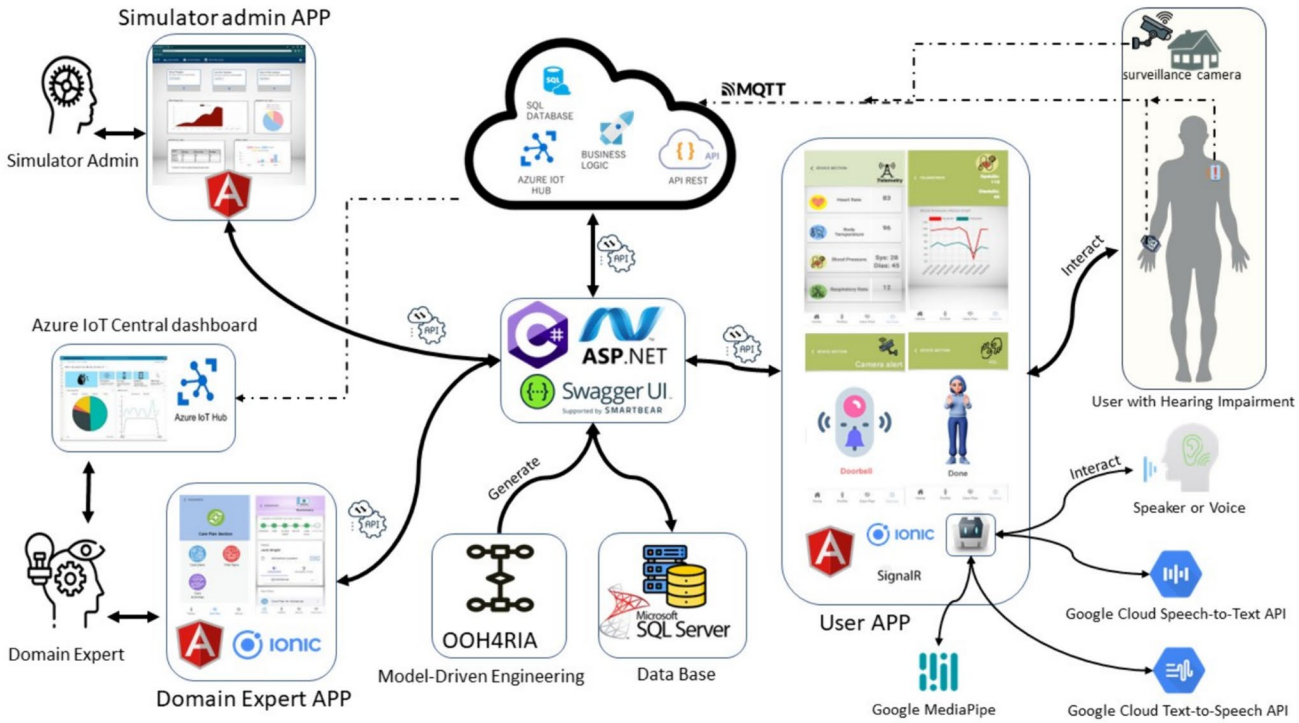


Fig. 11 MoSIoT Architecture Design for Deafness Case Study Scenario

[50], specifically in this case study focused on persons with deafness impairment. The application allows users to interact effectively with the system, taking their disabilities into account. Additionally, for the HMS scenario targeting persons with deafness, users receive information about their latest medical conditions through caregivers and connected devices with variable accessibility solutions.

As part of the simulator's output, all the dynamic elements and data within a specific IoT scenario are injected into an IoT Cloud platform (such as Azure IoT Central) [51] using REST APIs.

To further enrich the system's capabilities, consider that integrating technologies such as Google Cloud Speech-to-Text [52], Google Cloud Text-to-Speech [53], Google MediaPipe library [54], and SignalR [55] can be highly beneficial to develop an IoT system specifically designed for persons with hearing impairments. These technologies open avenues for advanced functionalities.

Google Cloud Speech-to-Text provides real-time speech recognition capabilities, converting spoken language into written text. This allows individuals with hearing impairments to communicate through speech, which can be converted into text for further processing or display. Google Cloud Text-to-Speech can convert any sign language or text into audible speech. This service enables the system to relay messages from a person with hearing impairments to a third person by voice, providing an alternative channel for communication. The Google MediaPipe library offers tools and components for building perception-based applications, including computer vision capabilities. By utilizing MediaPipe, visual-based tasks such as sign language recognition can be implemented, allowing the interpretation and translation of sign language gestures into meaningful actions or textual representations. SignalR, a real-time web application framework, facilitates bi-directional communication between clients and servers. It can deliver real-time notifications, alerts, or updates to various devices or user interfaces within the IoT system. This ensures that users with hearing impairments receive timely information and stay informed about relevant events or changes.

5 Conclusion and future work

This work has elucidated a pioneering approach by simulating Internet of Things (IoT) systems, focusing on enhancing the quality of life for individuals with hearing impairments. The MoSIoT (Modeling Scenarios of the Internet of Things) system, evaluated and extended through the quality-of-life framework proposed by Schalock and Verdugo, has showcased the transformative potential of integrating and managing existing deafness technologies. The presented case study underscores the profound impact of assistive technologies and champions the imperative of inclusive design and

technology-driven solutions in promoting accessibility and holistic well-being.

Significant improvements were observed in key quality-of-life dimensions, namely "Interpersonal Relationships," "Emotional Well-Being," and "Personal Development." These enhancements were attributed to the MoSIoT system's ability to facilitate enriched social interactions, foster emotional stability, and contribute to the personal growth of individuals with hearing impairments. The system's integration of advanced communication technologies and personalized interaction mechanisms played a pivotal role in breaking down barriers and fostering a sense of community and belonging among users. Furthermore, the emphasis on user-centric design and adaptive learning enhanced emotional well-being and personal development, thereby illuminating the transformative potential of IoT systems in enriching lives.

Our future work involves expanding the MoSIoT framework to cater to a broader range of disabilities, including visual, mobility, and cognitive impairments and their combinations, necessitating the integration of technologies tailored to address these specific needs. In the foreseeable future, we plan to integrate auditory augmented reality (AAR) and artificial intelligence (AI) technologies into the MoSIoT framework to enhance individuals' visual impairments' interaction and navigation abilities. We are developing components like Object Detection, Scene Recognition, Object Measurement, and Augmented Reality Geolocation to deliver detailed auditory cues about the surroundings, bolster spatial awareness, and facilitate outdoor navigation. These technologies rely on AI vision capabilities for advanced image processing and analysis, enabling accurate recognition of objects and scenes.

We will also explore how AI technologies can be further utilized to develop innovative solutions for a wide range of disabilities. Additionally, leveraging the synergy between IoT systems and intelligent environments, such as smart homes or buildings, can significantly enhance the accessibility and usability of assistive technologies. By seamlessly integrating with these environments, assistive devices can offer improved communication, environmental control, and safety features.

Evaluations with real-world case studies are essential for understanding assistive technologies' efficacy and tangible benefits. Empirical studies can provide valuable insights into the effectiveness of IoT solutions in improving the quality of life for individuals with disabilities and help identify areas for refinement and improvement. Addressing the lack of standardization and interoperability among different assistive technologies remains a crucial challenge. Therefore, our future efforts should establish common standards, protocols, and interfaces to ensure seamless integration and compatibility between IoT devices and systems.

In conclusion, the advancements made in this work signify a notable progression in addressing the multifaceted needs of individuals with hearing impairments through the innovative application of IoT systems. The ongoing developments in assistive technology harbor the promise of significantly elevating the quality of life and fostering full societal participation for individuals with disabilities.

Author Contributions Conceptualization, S.N. and S.M; methodology, S.N. and S.M.; software, S.N. and S.M.; validation, S.N.; investigation, S.N. and S.M.; resources, S.M; writing—original draft preparation, S.N. and S.M.; writing—review and editing, S.N and S.M.; supervision, S.N. and S.M.; project administration, S.M; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was supported by the Spanish Ministry of Science and Innovation under contract PID2019-111196RB-I00, called "Development of IoT Systems for Persons with Disabilities" (Access@IoT), and also was partially funded by the GVA through the AICO/2020/143 project.

Declarations

Conflicts of interest The authors declare no competing interests.

Ethics approval This paper is a part of MSc thesis No: 980084440533 which approved by research council of Islamic Azad University E-Campus. In the current study, we first obtained the approval of the Department of Psychology, Measurement (Psychometrics), Faculty of Humanities, Islamic Azad University E-Campus. Then obtained informed consent forms from the participants. Also, we Distributed and collected anonymous scales among the participants. All methods were carried out in accordance with relevant guidelines and regulations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. World Health Organization. (2023, May 15). Assistive technology. Retrieved from Accessed on 17 July 2023
2. Sannomia, D., Ferreira, T. B., & Ferreira, M. C. (2019). Evaluation of the use of accessibility tools in the information technology industry: a case study. In 2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 173–180). IEEE Retrieved from Accessed on 19 May 2023
3. Verdugo, M.A., Schalock, R.L., Arias, B., Jordán De Urríes, F.D.B. (2013). Calidad de Vida. In M.A. Verdugo and R.L.

Schalock (Eds.), Discapacidad e Inclusión (1st ed., Chapter 19). Amarú.
4. Modelos de calidad de vida. Modelo de la satisfacción. Retrieved from https://1library.co/article/modelos-calidad-vida-modelo-satisfacci%C3%B3n.zx506vwq Accessed on 26 May 2023
5. Meliá, S., Nasabeh, S., Luján-Mora, S., Cachero, C.: MoSIoT: Modeling and Simulating IoT Healthcare-Monitoring Systems for People with Disabilities. Int J Environm Res Public Health **18**(12), 6357 (2021)
6. Bézivin, J.: In search of a basic principle for model driven engineering. Novatica J. **5**, 21–24 (2004)
7. IMS Global Learning Consortium, Inc. (2012, September 13), IMS Access For All Version 3.0. http://imsglobal.org/accessibility Accessed on May 2023
8. HL7 FHIR (2011): Fast Healthcare Interoperability Resources. http://hl7.org/implement/standards/fhir/ Accessed on May 2023
9. Dhanjal, A. S., & Singh, W. (2019). Tools and Techniques of Assistive Technology for Hearing Impaired PeoplePersons. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th-16th Feb 2019 (pp. 1–6). IEEE. ISBN: 978–1–7281–0211–5/19
10. Mpofu, J., Chimhenga, S.: Challenges faced by hearing impaired pupils in learning: A case study of King George VI Memorial School. IOSR J Res Method Ed (IOSR JRME) **2**(1), 69–74 (2013)
11. Luft, P. (2000). Communication barriers for deaf employees: Needs assessment and problem-solving strategies. Work, 14(1), 51–59. PMID: 12441540. Retrieved from https://pubmed.ncbi.nlm.nih.gov/12441540/ Accessed on 12 July 2023
12. Harris, J., Bamford, C.: The Uphill Struggle: Services for Deaf and hard of hearing people - issues of equality, participation and access. Disability & Society **16**(7), 969–979 (2001). https://doi.org/10.1080/09687590120097854
13. British Deaf Association. (2014). Accessing Public Services: Issues for Deaf People in London Boroughs of Merton, Newham, Southwark, and Waltham Forest. Retrieved from https://bda.org.uk/wp-content/uploads/2017/03/BDA_Accessing_Public_Services-Issues_for_Deaf_People-London_Boroughs_12-2014.pdf Accessed on 12 June, 2022
14. World Health Organization, (2021, March 3). World report on hearing. Retrieved from https://www.who.int/publications/i/item/world-report-on-hearing (p. 144) Accessed on 11 June 2022.
15. Hersh, M., Ohene-Djan, J., Naqvi, S.: Investigating Road Safety Issues and Deaf People in the United Kingdom: An Empirical Study and Recommendations for Good Practice. J. Prev. Interv. Community **38**, 290–305 (2010). https://doi.org/10.1080/10852352.2010.509021
16. Wheatley, M., & Pabsch, A. (2012). Sign Language Legislation in the European Union (pp. 20–22). Brussels, Belgium: European Union of the Deaf. Printed in Brussels, Belgium. ISBN 978–90–816–3390–1
17. Peñeñory, V.M., Manresa-Yee, C., Riquelme, I., Collazos, C.A., Fardoun, H.M.: Scoping review of systems to train psychomotor skills in hearing impaired children. Sensors **18**(8), 2546 (2018). https://doi.org/10.3390/s18082546
18. WAI W3C: Web Accessibility Initiative (2017, May 15). Diverse Abilities and Barriers. Available online: https://www.w3.org/WAI/people-use-web/abilities-barriers/ (accessed on 14 August 2021).
19. Nagpal, N., Mittra, A., Agrawal, P.: Hand Motion-Recognition in an Intelligent Communication Module. Int J Innovative Technol Exploring Eng (IJITEE) **8**(10), 151 (2019)
20. Gumay, M.R.A., Effendy, V., Junaedi, D.: Modeling user interface design for panic button application for deaf people using user-centered design method. IOP Conf Ser: Mater Sci Eng. **830**, 022097 (2020)
21. Nugroho, K., Muljono, M., Marutho, D., & Murdowo, S. (2020). Mobile App for Word Recognition and Visualization of Objects Using Indonesian Language Google Speech To Text for Deaf Students. In Proceedings of the 2020 International Seminar on Application for Technology of Information and Communication (iSemantic) (pp. 12–21). Jakarta, Indonesia: AMIK Jakarta Teknologi Cipta, Universitas Dian Nuswantoro https://ieeexplore.ieee.org/document/9234206/ Accessed on 11 July 2023
22. Mohammdi, H. M., & Elbourhamy, D. M. (2021). An intelligent system to help deaf students learn Arabic Sign Language. Interactive Learning Environments. Advance online publication. https://doi.org/10.1080/10494820.2021.1920431 Retrieved from https://www.tandfonline.com/doi/full/https://doi.org/10.1080/10494820.2021.1920431 Accessed on 13 July 2023
23. Agrawal, M., Ainapure, R., Agrawal, S., Bhosale, S., & Desai, S. (2020). Models for Hand Gesture Recognition using Deep Learning. Paper presented at the 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Galgotias University, Greater Noida, UP, India, October 30–31. Retrieved from https://ieeexplore.ieee.org/document/9250846/ Accessed on 16 July 2023
24. M. J. C. Samonte, R. A. Gazmin, J. D. S. Soriano, and M. N. O. Valencia (2019), "BridgeApp: An Assistive Mobile Communication Application for the Deaf and Mute," in 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 16–18 October 2019. https://doi.org/10.1109/ICTC46691.2019.8939866 Retrieved from Accessed on 16 July 2023
25. Chuckun, V., Coonjan, G., & Nagowah, L. (2019). Enabling the Disabled using mHealth. In Proceedings of the Next Generation Computing Applications (NextComp) (pp. 1–6). IEEE.. Retrieved from Accessed on 16 July 2023
26. Syahriza, S., Zuhri, S., & Adilah, M. (2020). Communication Support Tool With Data Glove Concept For People With Hearing Disability Using Kansai Engineering Method. IOP Conference Series: Materials Science and Engineering, 931, 012003. https://doi.org/10.1088/1757-899X/931/1/012003 Retrieved from https://iopscience.iop.org/article/https://doi.org/10.1088/1757-899X/931/1/012003 Accessed on 16 July 2023
27. Naqvi, M. R., & Malik, M. (2021). Adaptivity is a matter of Concern: Design and Evaluation of a Mobile Interface for Deaf Users. In Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA) (pp. 1–2). Tarbes, France: INP-ENIT, University of Toulouse. Retrieved from https://ieeexplore.ieee.org/document/9682368/ Accessed on 17 July 2023
28. Darvishy, A. (2022). Internet of Things – Services and Applications for People with Disabilities and Elderly Persons. In International Conference on Computers Helping People with Special Needs (pp. 101–104). Lecture Notes in Computer Science (LNCS), 13342. ICCHP-AAATE 2022. Retrieved from https://link.springer.com/https://doi.org/10.1007/978-3-031-08645-8_12 Accessed on 17 July 2023
29. Jindal, N., Yadav, N., Nirvan, N., & Kumar, D. (2022). Sign Language Detection using Convolutional Neural Network (CNN). In 2022 IEEE World Conference on Applied Intelligence and Computing (pp. 354–360). https://doi.org/10.1109/AIC55036.2022.9848844. New Delhi, India: Delhi Technological University. Retrieved from https://ieeexplore.ieee.org/document/9848844/ Accessed on 18 July 2023
30. Mubin, S. A., You, J. S. T., Samiraj, E. P. R., & Jaafar, S. P. (2022). A Framework for Supporting Deaf and Mute Learning Experience Through Extended Reality. In 2022 IEEE International Conference on Computing (ICOCO) (pp. 1–6). Kota Kinabalu, Malaysia: IEEE Retrieved from https://ieeexplore.ieee.org/document/10031865/ Accessed on 17 July 2023
31. Cano, S., Peñeñory, V., Collazos, C.A., Albiol-Pérez, S.: Designing internet of tangible things for children with hearing

impairment. Information **11**(2), 70 (2020). https://doi.org/10.3390/info11020070
32. Uden, L., He, W.: How the internet of things can help knowledge management: a case study from the automotive domain. J. Knowl. Manag. **21**(1), 57–70 (2017). https://doi.org/10.1108/JKM-07-2015-0291
33. Apthorpe, N.J., Emami-Naeini, P., Mathur, A., Chetty, M., Feamster, N.: You, me, and IoT: how internet-connected consumer devices affect interpersonal relationships. ACM Transa Int Things **3**(4), 1–29 (2022). https://doi.org/10.1145/3539737
34. Tough, H., Siegrist, J., Fekete, C.: Social relationships, mental health and well-being in physical disability: a systematic review. BMC Public Health (2017). https://doi.org/10.1186/s12889-017-4308-6
35. McNaughton, D., Light, J.: The iPad and mobile technology revolution: benefits and challenges for individuals who require augmentative and alternative communication. Augment. Altern. Commun. **29**, 107–116 (2013). https://doi.org/10.3109/07434618.2013.784930
36. Rodić-Trmčić, B., Labus, A., Bogdanović, Z., Despotović-Zrakić, M., Radenković, B.: Development of an IoT System for students' stress management. Facta Universitatis, Series: Electronics Energetics **31**(3), 329–342 (2018). https://doi.org/10.2298/FUEE1803329R
37. dos Santos Paula, L., Salomão Dias, L.P., Francisco, R., Victória Barbosa, J.L.: Analysing IoT data for anxiety and stress monitoring: a systematic mapping study and taxonomy. Int J Human Computer Interaction (2022). https://doi.org/10.1080/10447318.2022.2132361
38. Polonetsky, J., & Gray, S. (2017).The Internet of Things as a Tool for Inclusion and Equality, Journal: Federal Communications Law Journal, Volume: 69, Issue: 2, Publication Date: August 2017, Pages: 103–118
39. Dos Santos, G.B., Damasceno, J.P.L., Cunha, N.B.A., Pachú, C.O.: Accessibility and social inclusion of persons with physical disabilities: an integrative review. Int J Dev Res. **11**(04), 46343–46347 (2021)
40. Ravi Chandra, M. L., Varun Kumar, B., & Suresh Babu, B.Babu (2017). IoT enabled home with smart security. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 1–2 August 2017. IEEE. https://doi.org/10.1109/ICECDS.2017.8389630.
41. World Health Organization. (2023, 27 February). Deafness and hearing loss. Retrieved from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss Accessed on 20 July 2023
42. Holman, J.A., Drummond, A., Hughes, S.E., Naylor, G.: Hearing impairment and daily-life fatigue: a qualitative study. Int. J. Audiol. **58**(7), 408–416 (2019). https://doi.org/10.1080/14992027.2019.1597284
43. Trujillo, M., & Viviana, M. (2020). Estrategias de afrontamiento y calidad de vida en personas con discapacidad auditiva [Coping Strategies and Quality of Life in Individuals with Auditory Disabilities]. Clinical psychology research project with a mental health focus. Universidad Central del Ecuador, pp 56–57. Retrieved from http://www.dspace.uce.edu.ec/handle/25000/22255 Accessed 23 April, 2023.
44. Verdugo Alonso, M. Á., Arias Martinez, B., Gómez Sánchez, L. E., & Schalock, R. L. (2009). Escala de Calidad de Vida - GENCAT. Retrieved from https://inico.usal.es/escala-de-calidad-de-vida-gencat/ Accessed on 23 April 2023.
45. McKeown, C., McKeown, J.: Accessibility in online courses: understanding the deaf learner. TechTrends **63**(4), 506–513 (2019). https://doi.org/10.1007/s11528-019-00385-3
46. MoSIoT Backend GitHub Repository https://github.com/santiagomelia/MoSIoT-Backend
47. Meliá, S., Martínez Domene, J. J., Pazmiño Pérez, A. M., & Gomez, J. (2009). OOH4RIA Tool: Una Herramienta basada en el Desarrollo Dirigido por Modelos para las RIAs. In XIV Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2009), San Sebastián, Spain, September 8–11, 2009. Source: DBLP.
48. MoSIoT Admin App GitHub Repository https://github.com/JavierJH99/MoSIoT
49. Domain Expert App GitHub Repository https://github.com/hichman662/medico
50. Patient with Disability App GitHub Repository https://github.com/hichman662/paciente
51. Stackowiak R (2019). IoT Central and Solution Accelerators. In: Azure Internet of Things Revealed. Apress, Berkeley, CA.
52. Google Cloud Speech-to-Text. https://cloud.google.com/speech-to-text Accessed on 23 April 2023.
53. Google Cloud Text-to-Speech. Retrieved from https://cloud.google.com/text-to-speech Accessed on 23 April 2023.
54. Google LLC. MediaPipe. Retrieved from https://developers.google.com/mediapipe Accessed on 23 April 2023.
55. Fletcher, P. (2020). Introduction to SignalR. Retrieved from Microsoft Learn website: https://learn.microsoft.com/en-us/aspnet/signalr/overview/getting-started/introduction-to-signalr Accessed on 23 April 2023.

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

From Quantum Computing to Condensed Matter : A Machine Learning Perspective

by

Pranav Kairon

B.Tech, Delhi Technological University, 2021

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Physics)

The University of British Columbia
(Vancouver)

August 2024

© Pranav Kairon, 2024

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

From Quantum Computing to Condensed Matter : A Machine Learning Perspective

submitted by **Pranav Kairon** in partial fulfillment of the requirements for the degree of **Master of Science** in **Physics**.

Examining Committee:

Roman V. Krems, Professor, Chemistry, UBC
Supervisor

Mona Berciu, Professor, Physics & Astronomy, UBC
Supervisory Committee Member

Abstract

This thesis investigates the intersection of physics and machine learning, with a primary focus on applications in condensed matter and quantum computing. In the first half, we demonstrate Bayesian machine learning (ML) models that use quantum properties in an effectively lower-dimensional Hilbert space to make predictions for the Hamiltonian parameters that require a larger basis set as applied to a well-known problem in quantum statistical mechanics, the polaron problem. Two models of interest include the Su-Schrieffer-Heeger (SSH) model and the mixed SSH-Holstein model. We demonstrate ML models that can extrapolate polaron properties in phonon frequency. We consider the sharp transition in the ground-state momentum of the SSH polaron and examine the evolution of this transition from the anti-adiabatic regime to the adiabatic regime. We also demonstrate Bayesian models that use the posterior distributions of highly approximate quantum calculations as the prior distribution for models of more accurate quantum results. This drastically reduces the number of fully converged quantum calculations required to map out the polaron dispersion relations for the full range of Hamiltonian parameters of interest. In the second half, we show that a parametrized quantum neural network can be used to build a parameter-free quantum kernel for machine learning that inherits the concentration properties of the neural network cost function. This establishes a rigorous connection between barren plateaus in variational quantum algorithms and exponential concentration of quantum kernels for machine learning. Our results imply that recently proposed algorithms for building barren plateau-free quantum circuits can be applied to construct useful quantum kernels for machine learning without inductive bias. Additionally, we investigate how leveraging the group-theoretic structure of datasets enables us to exploit the block diagonal structure of the unitary representation of the symmetry group. We derive new bounds on the variance of the kernel matrix, enhancing the stability and performance of quantum kernel-based machine learning algorithms.

Lay Summary

We explore the connection between machine learning and quantum physics through two key examples. Firstly, we focus on the study of electron-phonon physics, crucial for understanding high-temperature superconductivity—a phenomenon that enables electricity to flow without resistance. Machine learning allows us to create computational models that uncover hidden relationships in raw data. Using machine learning, we can predict electron behavior in these materials in regimes not accessible by traditional computational methods. Secondly, we examine machine learning on quantum computers, the next generation of devices capable of solving incredibly complex problems beyond the capabilities of current computers. Quantum computing holds promise for advances in fields such as chemistry, cryptography, and materials science. However, challenges in training quantum machine learning models currently hinder progress. Overcoming these obstacles will bring us closer to practical applications of quantum computing in the real world. Both examples illustrate how combining machine learning with quantum physics leads to exciting and novel discoveries.

Preface

The material in Chapter 3 was published in the article: Kairon, P., Sous, J., Berciu, M., & Krems, R. V. (2024). Extrapolation of polaron properties to low phonon frequencies by Bayesian machine learning. Physical Review B, 109(14), 144523. The project was identified and designed by Roman V. Krems, Mona Berciu, John Sous, and the author. The author developed the code, performed the calculations, and the data analysis.

The preparation for the manuscript consisting of the material in Chapter 4 is currently in progress. The project was identified and designed by Roman V. Krems, Jonas Jäger, and the author. The author performed the calculations in collaboration with Jonas Jäger.

Table of Contents

Abstract . . . . . **iii**

Lay Summary . . . . . **iv**

Preface . . . . . **v**

Table of Contents . . . . . **vi**

List of Figures . . . . . **viii**

List of Abbreviations . . . . . **xiv**

Acknowledgments . . . . . **xvi**

Dedication . . . . . **xix**

1 Introduction **1**

1.1 Organization of this thesis 4

2 Machine learning techniques **6**

2.1 Gaussian Processes 6

2.2 RKHS and Bayesian Machine learning 7

2.3 Gaussian Process Regression 9

2.4 Kernels and Automatic Model construction 9

2.5 Example: Ginzburg-Landau phase transitions 12

2.6 Multi-Fidelity Gaussian Processes 15

3 Bayesian extrapolation of quantum observables in Hamiltonian parameter space **19**

3.1 Electron phonon coupling and the Polaron problem 19
3.1.1 Bipolaronic superconductivity and Peierls model 23
3.2 Electron phonon models 25
3.2.1 Holstein Model 26
3.2.2 Peierls Model 27
3.3 Numerical method for Green's function calculation 28
3.4 Results 31
3.4.1 SSH polaron 31
3.4.2 Mixed Holstein – SSH Model 33
3.4.3 Predictions by extrapolation to extreme adiabatic regime 33
3.4.4 Interpolation in variational space 40
3.5 Discussion and Outlook 42

4 Equivalence between Exponential Concentration and Barren Plateaus 44
4.1 Quantum Machine learning models 45
4.1.1 Quantum Kernels 46
4.1.2 Quantum Neural Networks 50
4.2 Relationship between BP and EC 54
4.3 Invariance as an inductive bias 59
4.3.1 A new bound for covariant quantum kernels 60
4.4 Discussion and Outlook 63

5 Conclusion 65

Bibliography 68

List of Figures

Figure 2.1 Illustration of Gaussian process regression in one dimension, for the target test function $f : x\sin(x)$ (dashed line). Training data are acquired (blue dots). The blue line shows the mean prediction of the Gaussian process regression and the shaded region shows the corresponding 95% confidence interval. Gaussian processes allow interpolating and extrapolating predictions in regions of parameter space where training data are absent 7

Figure 2.2 Schematic diagram of the automatic model construction method employed to develop kernels that can extrapolate well. At each layer (L), the kernel with the highest Bayesian information criterion (Equation 2.15) is selected (highlighted in red). In this particular diagram, the algorithm is shown for two layers. The search tree can be extended to arbitrary L, but it is a good idea to stop when $|\mathrm{BIC_L} - \mathrm{BIC_{L-1}}| \leq 0.05 \times \mathrm{BIC_L}$ 11

Figure 2.3 **(a)** Theoretically calculated free energy(solid lines) using Eq. (2.18). GP prediction(dashed lines). Each curve corresponds to a unique value of Θ **(b)** Black line is theoretically calculated Order parameter ϕ_0 as a function of twisting angle Θ. Red dots are the corresponding GP prediction. The pink region shows the training regime, such that 300 points were taken from $0 < \Theta < 17^{\circ}$ and $-\pi < \phi < \pi$. 14

Figure 2.4 **(a)** Band structure plots $\pm E_{ka}(a = 1,2)$ in the two phases. **(b)** Black line is theoretically calculated band gap δ_{BG} as a function of twisting angle Θ. Red dots are the corresponding GP prediction. δ_{BG} is calculated by subtracting the maximum of the lower band from the minimum of the upper band. E_{ka} calculation requires knowing the order parameter value for that particular twisting angle which is obtained using GP in Fig. 2.3b 16

Figure 2.5 **(a)** Theoretically calculated free energy for multiple order parameter case (Left) Free energy in the trivial phase with one minimum (Right) Free energy in the topological phase with two minimas. **(b)** Solid lines are theoretically calculated Order parameter ϕ_1, ϕ_2 as a function of twisting angle Θ. Solid lines with triangles are the corresponding GP prediction. The pink region shows the training regime, such that 600 points were taken from $0 < \Theta < 25^{\circ}$, $-\pi < \phi_1 < \pi$ and $-\pi < \phi_2 < \pi$. . 17

Figure 3.1 Electrons scattered by phonons 22

Figure 3.2 Electron self-interaction through phonons 22

Figure 3.3 Two electrons interacting by an exchange of phonons 23

Figure 3.4 Holstein model 27

Figure 3.5 Depiction of GGCE variational algorithm: In this example, only $N = 6$ phonons can be created inside the dotted box of size $M = 4$. The bare particle can be anywhere on the lattice 30

Figure 3.6 (Left) Spectral function $A(k,\omega)$ for $\Omega = 4.0$ and coupling strength $\lambda_{\mathrm{SSH}} = 1.75$ and $\alpha_{\mathrm{H}} = 0$; (Center) Polaron energy $E(k)$ for several coupling strengths: dotted curves are polaron energies for $\lambda < \lambda^{\mathrm{c}}$ with the energy minimum at $k = 0$ (black circles) and solid curves are for $\lambda > \lambda^{\mathrm{c}}$ with the ground state occurring at $k = \pm K_{\mathrm{GS}}$ (black cirlces). (Right) SSH polaron ground state momentum K_{GS} and effective polaron mass $1/m^*$ as functions of λ_{SSH}. The sharp transition occurs at $\lambda^{\mathrm{c}}_{\mathrm{SSH}} = 0.644$. 31

Figure 3.7 Critical coupling strength $\lambda^c_{\rm SSH}$ of the SSH polaron as a function of phonon frequency (Ω): squares – ML predictions of the sharp transition; triangles – location of the sharp transition from the GGCE calculations. The ML models are trained by 600 polaron energies randomly sampled from the three-dimensional space $[\Omega, \lambda_{\rm SSH}, k]$ at $[\Omega, \lambda_{\rm SSH}]$ indicated by the symbols in the shaded region. 32

Figure 3.8 Critical coupling strength $\lambda^c_{\rm SSH}$ of the SSH polaron as a function of phonon frequency (Ω): squares – predictions of the sharp transitions by the ML models trained by polaron energies at phonon frequencies $\Omega \geq 1$; circles – predictions of ML models trained by polaron energies at phonon frequencies $\Omega \geq 3.5$; triangles – GGCE calculations. The ML models are trained by 415 polaron energies randomly sampled from the three-dimensional space $[\Omega, \lambda_{\rm SSH}, k]$. The training range for predictions represented by squares is shown by the shaded region. . . . 34

Figure 3.9 Critical coupling strength $\lambda^c_{\rm SSH}$ of the polaron for the mixed Holstein - SSH model as a function of phonon frequency (Ω): squares – ML predictions of the sharp transition; triangles – GGCE calculations. The ML models are trained by 600 polaron energies randomly sampled from the three-dimensional space $[\Omega, \lambda_{\rm SSH}, k]$ at $[\Omega, \lambda_{\rm SSH}]$ indicated by the symbols in the shaded region. The Holstein coupling strength is fixed to $\lambda_H = 0.5$. 35

Figure 3.10 (Left) SSH polaron dispersions for $\Omega = 0.4$: the symbols represent the results of the GGCE calculations with $M = 5, N = 10$; the solid lines are the ML predictions. The ML model is trained by the polaron energies at $\Omega \in [0.5, 0.7]$ and $\lambda_{\rm SSH} \in [0.25, 1.7]$. (Right) The sharp $K_{\rm GS} = 0 \leftrightarrow K_{\rm GS} > 0$ transition observed in the ground state momentum at $\Omega = 0.4$: circles – the ML predictions; squares – the GGCE calculations. 37

Figure 3.11 Extrapolation of the polaron dispersion relation E in phonon frequency space (Ω) for the pure SSH model. The dispersion curves are predicted by a ML model trained by polaron energies at phonon frequencies $\Omega \in [0.4, 0.9]$. The solid lines are the predictions of the model for various coupling strengths. The horizontal dashed lines show the limit of the dispersion bandwidth equal to Ω. Lower right: The dependence of K_{GS} of the SSH polaron on λ_{SSH} at three phonon frequencies corresponding to the dispersion curves shown in the other panels. The symbols represent different values of Ω: triangles for $\Omega = 0.4$, circles for $\Omega = 0.3$, and squares for $\Omega = 0.2$. 38

Figure 3.12 Extrapolation of the polaron dispersion relation E in phonon frequency space (Ω) for the pure SSH model. The dispersion curves are predicted by a ML model trained by polaron energies at phonon frequencies $\Omega \in [0.4, 0.9]$. The solid lines are the predictions of the model for various coupling strengths. The horizontal dashed lines show the limit of the dispersion bandwidth equal to Ω. Lower right: The dependence of K_{GS} of the SSH polaron on λ_{SSH} at three phonon frequencies corresponding to the dispersion curves shown in the other panels. The symbols represent different values of Ω: triangles for $\Omega = 0.1$, circles for $\Omega = 0.05$, and squares for $\Omega = 0.01$. 39

Figure 3.13 Top Left: SSH polaron dispersions: for $\Omega = 0.5$, the green squares (red circles) show the total of 100 (15) points sampled from the lower (higher) fidelity quantum calculations and the curves are the NARGP model predictions trained on this combination of points. The circles represent expensive quantum calculations. The solid lines are evaluated for $\lambda_{\mathrm{SSH}} = [0.25, 0.5, 0.75, 1.0, 1.1, 1.25, 1.5, 1.75]$ and 20 uniformly distributed points $k \in [0, 0.8\ \pi]$. Top Right: Ground-state momentum K_{GS} of the SSH polaron as a function of λ_{SSH}. The red circles and the green squares are the NARGP model predictions for high fidelity data with the sharp transition at $\lambda^{c}_{\mathrm{SSH}} = 1.117$ and low fidelity data with the transition at $\lambda^{c}_{\mathrm{SSH}} = 1.121$, respectively. The black circles are the GGCE calculations presented for reference. Bottom Left: Red circles – rigorous calculations (reference); broken curves – predictions of ML models trained directly by 15 high fidelity calculations Bottom Right: Red circles – rigorous calculations (same as in left); solid curves – NARGP model predictions. The difference between the dashed green curves and the solid red curves represents the improvement of accuracy due to combination of low-fidelity and high-fidelity data. 41

Figure 4.1 Exponential concentration of a quantum kernel: comparison of the kernel matrices constructed using an ansatz proposed in Havelick *et al.* [40] for binary classification on MNIST dataset [148]. The matrix is constructed for 100 data points with the number of dimensions reduced to the corresponding number of qubits by principal component analysis. Each dimension is encoded into a separate, one qubit. The kernel matrices shown are for qubit numbers $n = [2, 5, 10]$. 48

Figure 4.2 Quantum circuit for QNN (left) and a quantum kernel (right). The final output is the expectation value of $|0\rangle^{\otimes n}\langle 0|^{\otimes n}$. 51

Figure 4.3 Cost Concentration around the mean: Two-dimensional cross-section through the landscape of (negative of) cost function $C = 1 - \prod_{j=1}^{n} \cos^2 \frac{\theta^j}{2}$ for $n = 3$ (Left) and $n = 20$ (Right), where n remains the number of qubits. The cost function landscape flattens as we increase the number of qubits, with only a few minima appearing which require an exponential amount of resources to traverse. 53

Figure 4.4 Kernel matrices computed with the BP-free ansatz [144] for binary classification on the MNIST dataset [148]. The matrices are constructed for 100 data points with the number of dimensions reduced to the corresponding number of qubits (n) by principal component analysis. Each dimension is encoded into a single independent qubit. (Right) Variances of the kernel matrix elements computed with the BP-free ansatz (black) as a function of number of qubits in the range $n \in [2, 20]$. The depth of the circuit is equal to the number of qubits at each point. The purple line corresponds to the Havlíček *et al.* [40] considered in Fig. 4.1. The depth of the circuit is kept constant. 59

List of Abbreviations

ARPES	Angle-resolved photoemission spectroscopy
BCS	Bardeen-Cooper-Schrieffer
BIC	Bayesian Information Criterion
BDG	Bogoliubov-de Gennes
BOA	Born-Oppenheimer Approximation
BP	Barren Plateaus
EC	Exponential Concentration
EOM	Equations of Motion
EPC	Electron phonon coupling
GGCE	Green's function Cluster State Expansion
GL	Ginzburg-Landau
GP	Gaussian Process
GPR	Gaussian Process Regression
LML	Log Marginal Likelihood
MA	Momentum Approximation
ME	Migdal-Elishaberg
NARGP	Non-linear Auto-Regressive multi-fidelity Gaussian process

NP	Non Polynomial
PQC	Parametrized Quantum Circuit
PSD	Positive Semi Definite
QML	Quantum Machine Learning
QNN	Quantum Neural Network
RBF	Radial Basis Function
RKHS	Reproducing Kernel Hilbert Space
SSH	Su-Schrieffer-Heeger

Acknowledgments

I would like to express my heartfelt gratitude to the individuals and institutions that have played an instrumental role in the completion of this thesis. First and foremost, I am deeply indebted to my advisor Dr. Roman Krems, whose guidance and mentorship have been invaluable throughout this journey. Besides his expertise in physics and machine learning, I enjoyed our discussions on various other subjects like finance, culture, and history; be it during the coffee hour or the strenuous hikes. I have learned a lot from his perspective on research and academia. I will always cherish the memories of our (dangerous) excursion to St. Mark's. I really look forward to my future visits to Vancouver and having more such experiences. I also thank Dr. Mona Berciu, who has been incredibly helpful and supportive throughout my Master's. I have been fortunate enough to sit in on some of her lectures on condensed matter physics. Throughout my Master's degree, I have benefited greatly from discussions with Dr. Joe Salfi, Dr. Marcel Franz, Dr. Ludmila Szulakowska, Dr. Polina Feldmann, and especially Dr. Dmytro Bondarenko. His guidance and patience with my inquiries and blackboard sessions were immensely appreciated. I extend my best wishes to Dmytro and his family. I would like to thank my professors in India, Dr. Anirban Pathak and Dr. Mukhtiyar Singh, whose foundational support paved the way for my current achievements. In addition, I extend my thanks to my alma mater, St. Xavier's School, and Delhi Technological University, where I had the privilege of meeting many brilliant individuals. Particularly, I want to thank Mrs. Prabha George, my sixth grade science teacher for inculcating a passion for learning in a chaotic child.

I would like to thank all my friends at UBC. Dhruv Pahadi, thanks for helping me perfect my lat pulldowns and cover drive. Though these things represent just a tiny subspace in the Hilbert space of our friendship (with memes as the inner product), I am grateful that they exist. I look forward to spending the next 5 years with you. Rhea, thanks for all the phone calls and the Rajma Chawal. I hope the tradition of giving you terrible gifts continues

forever. My worries seem to fade when I know I'll be sharing them with you. Mr. Mukherjee, I really admire the wealth of knowledge you have and I'm glad I got to absorb some of it. Thanks again, for making me part of the chosen few, it meant a lot. Samarth, thanks for all the advice on academia or otherwise, be it hinged or unhinged. Your perspective is very unique and something I admire a lot. I wish I could write a blog about our friendship. Saran, thanks for all the bluechip conversations and for helping me practice how to slog left arm pacers. Sudhang Varshney, among many other things, thanks for taking my photo at the Wreck beach and for opening the door to your house (the few times you did). I really look forward to the publishing of your book, A physicist's guide to judgement in the modern world, and my favourite, The Art of *cheezon ke baare mein sochna*. Hyderbadi Biryani is a lot more fun, with you all around the table.

I want to express my gratitude to my labmates, the potluck group: Krish Dosa, thank you for being my partner in crime and for always being up for spontaneous coffee breaks. I wish you and Avinash a fulfilling journey through your PhD and beyond. Dawn Roomie, I'll always cherish our time as roommates and those late-night conversations. Thank you for looking after me like an elder sister during our excursions. Elham, you are one of the kindest and most genuine people I have ever met. I hope I have absorbed some of your Rumi-esque wisdom. I wish you and Ramin endless happiness and joy. Jonas, thank you for all the blackboard discussions and for introducing me to Irish folk music. Jun, the code you shared was invaluable to this thesis, and thanks for mentoring me as a young master's student. Kasra, thank you for making me feel like I wasn't the only person with beer in the group. Finally, Xuyang, thanks for letting me win all those pool games. You made my "only and you" trips a lot more fun. I will definitely cherish my memories with you all, especially the last few months. You are all incredible and so talented—I had an amazing time working with you. I am thankful for the support and camaraderie of my peers and colleagues within the physics department and the entire Glueballs team. I would also like to thank the cricket group especially Biswajeet, Nihar, and Mussanna for all the fun friday evenings. I want to thank my roommates Joseph and Rushil for all their help. I also want to thank all the people from CREATE program especially Bahiyyih, Luis and Saasha. I thank my friend Rashmi for helping me with my first TA course. I want to express my gratitude to my friends back home, the 10-C gang, Katla & Bros and the Timarpur crew. As the saying goes, sometimes to meet yourself, you have to meet with an old friend.

I am deeply grateful to my amazing parents and my sister Kitoo, whose unwavering

support has been the backbone of my journey. Being away from them has made me value the things that I would take for granted. Their constant care and encouragement have empowered me to chase every ambition. I also want to express my heartfelt thanks to my family in Canada—Megha & Karan, Pooja & Rahul—whose support and encouragement were invaluable during my studies abroad. Lastly, a special shout out to my headphones for keeping me company on all those late-night walks.

Dedication

To Chaiji.

Chapter 1

Introduction

In recent years, the field of physics has witnessed a significant integration of machine learning (ML) techniques. ML models have emerged as powerful tools for recognizing patterns in data and providing valuable insights, particularly in scenarios where analytical solutions are unavailable or computationally expensive [1]. Quantum systems, in particular, present a plethora of challenges due to the absence of exact solutions and exponential increases in computational complexity with system size. The ability to accurately and efficiently interpolate and extrapolate from a limited set of observations is highly desirable, for a wide class of problems in quantum mechanics and condensed matter physics [2]. Even though statistical learning theory has been using ML for the past few decades [3], ML has only recently become a trend in condensed matter physics [4]. Generally, ML has been utilized in quantum physics for six different types of applications: Neural networks for phase classification [5–9], acceleration of Monte-Carlo/Density-Functional Theory calculations [10, 11], ML on quantum architectures [12, 13], ML for experimental protocols and optimal control [14–16] and material design/search [17, 18] etc. To a physicist, increasing the accuracy by a few percent, at the cost of designing and training a model which takes hours of computational resources might seem trivial. But there is a need to look beyond the "black-box" stigma that ML carries, especially since deeper relations between fundamental physics concepts and ML have been revealed in recent years. Besides regression, classification, and speeding up computation, ML models can also be used to guide the theory behind the observable physics. For example, *(i)* it was recently shown that finite-width neural networks can be thought of as interacting field theories, whereas a Gaussian process is equivalent to an interacting free field theory [19, 20] or that *(ii)* neural networks can be made interpretable by using concepts from renormalization group [21] or that *(iii)* the competition between pre-

diction accuracy and model complexity can be modeled as a second-order phase transition [22]. At the same time, it should be noted that ML is a highly result-driven field as it is dominated by approaches that "work" for specific datasets. This is due to a fundamental theorem in Machine learning i.e. the *no-free lunch theorem*, which states that there is no universal algorithm that performs optimally for every possible dataset [23]. This implies that there is an inherent trade-off in ML known as the bias-variance trade-off which refers to balancing model simplicity (bias) and flexibility (variance) to achieve optimal prediction accuracy and generalization. Increasing model complexity reduces bias but may lead to overfitting and higher variance. Neural networks have proven to be exceptional when it comes to tasks such as image classification and functional interpolation [24]. However, for regimes where no training data is available, the predictions get worse [25]. (This is mostly true for feed-forwards neural networks, so recently people have started using Recurrent Neural Network models for better predictions.) Linear ML models such as kernel methods and Gaussian Processes (GP) are quite good at extrapolation [26, 27]. The latter is a Bayesian machine learning model and can incorporate our prior beliefs about the dataset such as continuity, scale, periodicity, etc [28]. These beliefs in conjunction with training data can be used to extrapolate to different data regimes. One way to ensure that a specific model in practice is robust to information-theoretic limitations of ML models is to validate it by training using datasets sampled from different regions of input space. This eliminates the biases coming from specific datasets and hyper-parameter (human-controlled parameters of the model) choices. While the problem is not solved completely, this method can provide a degree of confidence in one's predictions [29]. In Chapter 3, we will use ML to make accurate and physically motivated predictions for a condensed matter system. Specifically, we will consider the Peierls model of electron-phonon coupling, which is motivated by recent works highlighting the role it plays in high-temperature superconductivity [30, 31]. We thoroughly explore this motivation in Section 3.1.1. We show that machine learning models are capable of extrapolating polaron properties in phonon frequencies. Additionally, Bayesian models are demonstrated, leveraging posterior distributions from approximate quantum calculations as priors for more accurate models. This significantly minimizes the need for fully converged quantum calculations to map out polaron dispersion relations across a wide range of Hamiltonian parameters.

Quantum computers hold great potential to provide a significant speedup over existing computational models. At this moment, we are in the noisy intermediate scale quantum (NISQ) era of quantum computing, which means that the present-day quantum computers

suffer from many problems such as small number of qubits (< 100), lack of all-to-all qubit connectivity, short coherence times, noisy quantum operations such as poor gate fidelity, readout error, *et cetera* [32, 33]. To perform quantum operations in a noiseless manner, we need to achieve *fault tolerance* by encoding physical qubits into logically error-corrected qubits. In the future, this can be done by the development of efficient error-correcting codes [34]. Until this is achieved, the big question is, Can we extract any quantum advantage from the currently available quantum computers? [35] To answer that question, there has been an emergence of a novel field of quantum machine learning (QML). QML seeks to perform ML operations on a quantum information processing unit, through the realization of quantum circuits. The goal is to take advantage of phenomena inherent to quantum physics such as entanglement, and exponentially large Hilbert spaces, to develop algorithms that can possibly outperform classical machine learning algorithms in certain applications. Various quantum machine learning algorithms have been proposed, surpassing their classical counterparts in specific problem classes [12, 36–38]. The quantum kernel method is one of the prime examples of the QML algorithm which incorporates quantum computing into classical kernel methods [39, 40]. When combined with classical linear classifiers like support vector machines, this approach effectively classifies data that classical models struggle with [37]. Recently, the discrete logarithmic problem which is proven to be NP-hard has been successfully classified using the quantum kernel method [38]. Quantum Kernels have also been applied to real-world datasets [41, 42]. There exists another class of QML models known as Quantum Neural Networks (QNN), developed as an analog to classical deep learning models [43, 44]. Schuld *et al.* [45] proved a fundamental connection between quantum kernel methods and quantum neural networks highlighting the generality of the former for supervised learning tasks. However, the exponential Hilbert space which was thought to be the source of the speedup, also turns out to be a disadvantage during training. The resources required to extract the global information embedded into this exponentially large Hilbert space scale exponentially with the system size. Hence, both QNNs and quantum kernels suffer from trainability problems that have similar origins. The former suffers from Barren Plateaus (BP) [46], whereas the latter suffers from Exponential Concentration (EC) [47]. BPs have been known to exist since 2018, whereas EC was discovered quite recently. Hence, people have suggested various methods to avoid the problem of vanishing gradients in QNN [48–52]. However, only a few methods are known that avoid the vanishing similarity issue [53, 54]. In the second part of the thesis, we will derive a concrete relation between the problem of barren plateaus and exponential concentration, such that we can transfer some of the results obtained for BPs to mitigate the trainability issues that

arise in quantum kernels (given certain conditions hold true). Kübler *et al.* [55] proved that to obtain a quantum advantage, we don't need an exponentially large Hilbert space, but rather an ability to encode good inductive biases pertaining to our data in our QML model. To have a quantum advantage, one needs to show that such inductive bias cannot be encoded into an analogous classical model. Here, we also provide a way to encode to inductive biases to mitigate the problem of exponential concentration in quantum kernels.

1.1 Organization of this thesis

In Chapter 2, we provide a brief overview of the classical machine-learning techniques which will be used in Chapter 3 for the electron-phonon problem, followed by a demonstration of the techniques on a simpler problem of Ginzburg-Landau second order phase transitions. In Chapter 3 we begin by briefly going over the theory of electron-phonon interactions in Section 3.1. We describe the motivation for our work in Section 3.1.1. Specifically, we discuss how an electron moving through the lattice, interacting with phonons via a Peierls type coupling [56] can be a precursor to high-temperature superconductivity. This problem becomes quite challenging for numerical methods in lower phonon frequency regimes, such that it is impossible to obtain an exact solution in the desired time frame [57]. In Section 3.2 and 3.3 we describe the specific electron-phonon models used in the present work and the numerical method used to probe such systems. Then, we move on to Section 3.4 where we describe our results. We will use Gaussian process regression (GPR) to train our models on the available dataset and extrapolate to parameter regimes where simulations become difficult. We will use a kernel composition algorithm [58], which has previously been shown to perform quite well for similar extrapolation problems [59]. We obtain the dataset by performing an exact Green's function calculation on the Peierls electron-phonon coupling model using Generalized Green's function Cluster Expansions (GGCE) method [60] described in Section 3.3. Finally in Section 3.4.3, we will comment on the existence of a sharp transition in the low-phonon frequency regime. In Section 3.4.4, we demonstrate how multi-fidelity Gaussian processes [61] can be used to interpolate in the variational space of GGCE, thus reducing the compute time for computationally expensive calculations. In Chapter 4, we begin with Section 4.1 which provides a brief overview of the theory behind quantum kernels and quantum neural networks. We also define the trainability problems in quantum machine learning models such as Barren plateaus and exponential concentration concretely since these will be used to explain the theorems later on. In Section 4.2, we derive results to transfer the bounds on the trainability of a quantum neural

network to an effective quantum kernel constructed in a similar manner. More concretely, we draw parallels between the trainability issue in quantum neural networks i.e. barren plateaus [46] and similar issues faced during the training of quantum kernel i.e. exponential concentration [47]. Using our results, one can use a barren plateau-free quantum ansatz in a quantum kernel setting, given it satisfies some conditions. Lastly in Section 4.3, we provide a rigorous mathematical proof demonstrating how encoding good inductive biases such as symmetries of the dataset can improve the exponential concentration problem in quantum kernels. We show how utilizing the group-theoretic structure of datasets enables us to harness the block diagonal structure of symmetry's unitary representation. This approach allows us to derive new bounds on variance of kernel matrix elements, thereby improving the stability and performance of quantum kernel-based machine learning algorithms.

Chapter 2

Machine learning techniques

In this chapter, a concise overview of the Gaussian Process (GP) theory will be provided, accompanied by a detailed investigation into methodologies aimed at optimizing the construction of efficient extrapolation models. In the end, we will highlight applications of such methodologies to extrapolate quantum observables.

2.1 Gaussian Processes

Machine learning practitioners commonly follow two steps for model construction: (1) selecting the functional structure, e.g., a neural network, and (2) concentrating learning endeavors on refining the model's parameters. This requires a lot of computational effort and yet its interpretability is limited. Gaussian processes (GPs) offer an alternative paradigm, by shifting to a *function space* view and directly endowing distributions upon functions that can conform to data. In the *function space* view, we can assume a prior distribution over the space of functions. After that, we can iteratively update the probability using a closed-form expression, by observing a finite number of points of our target function. Consider a dataset in the real domain, $\mathscr{D} = \{\mathbf{X}, \mathbf{y}\}_{i=1}^{N}$, consisting of N input vectors denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^\intercal$, where each input data point is a p-dimensional vector $\mathbf{x}_i$. Since we are in the framework of supervised learning for each input data point $\mathbf{x}_i$ we have a corresponding output $\mathbf{y}_i$. Our goal is to learn the underlying generating function $f : \mathbf{X} \to \mathbf{y}$ such that $y = f(x)$. A *stochastic process* can be defined as a collection of random variables indexed by time or space. Strictly speaking, a Gaussian process is a *stochastic process*, such that if we take any finite subset of them, the marginal distribution would be a multivariate Gaussian. The reason we say the Gaussian process can be thought of as an infinite dimensional Gaussian random variable is because it is defined using a mean and a covariance function

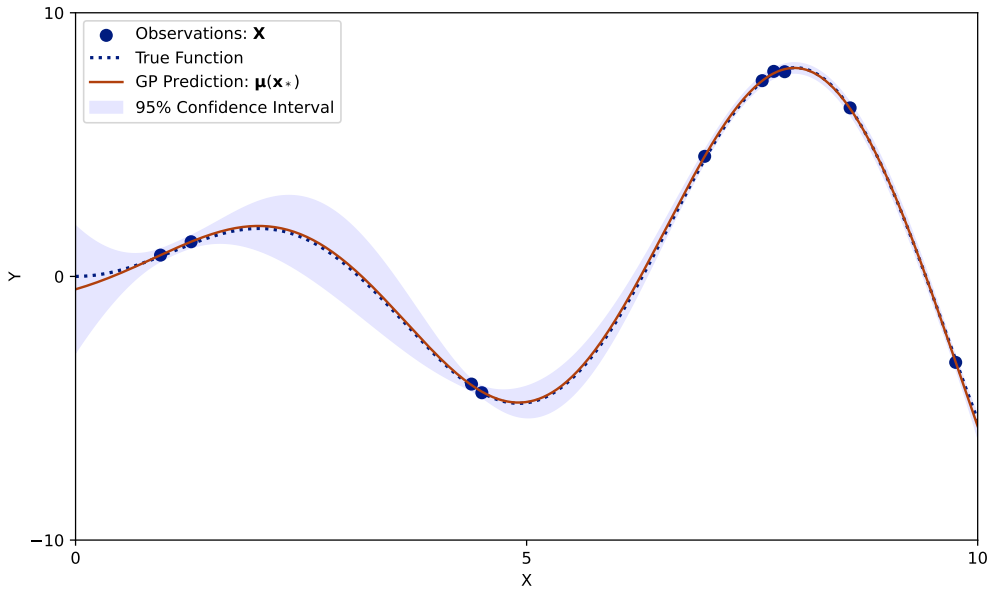


Figure 2.1: Illustration of Gaussian process regression in one dimension, for the target test function $f : x\sin(x)$ (dashed line). Training data are acquired (blue dots). The blue line shows the mean prediction of the Gaussian process regression and the shaded region shows the corresponding 95% confidence interval. Gaussian processes allow interpolating and extrapolating predictions in regions of parameter space where training data are absent

(as we will see shortly) which can have an infinite input domain. In the case of Gaussian process regression, those random variables are the data points/observed values of the target function. A Gaussian process can be completely described using its mean $\mu(\mathbf{x})$ and covariance/kernel function $k(\mathbf{x},\mathbf{x}')$ as

$$f(\mathbf{x}) \sim \mathcal{GP}[\mu(\mathbf{x}),k(\mathbf{x},\mathbf{x}')] \tag{2.1}$$

$$\mu(\mathbf{x}) = \mathrm{E}[f(\mathbf{x})] \tag{2.2}$$

$$k(\mathbf{x},\mathbf{x}') = \mathrm{Cov}[f(\mathbf{x}),f(\mathbf{x}')] = \mathrm{E}[(f(\mathbf{x})-\mu(\mathbf{x}))(f(\mathbf{x}')-\mu(\mathbf{x}'))] \tag{2.3}$$

2.2 RKHS and Bayesian Machine learning

Gaussian process models are Bayesian non-parametric linear models that can be used for regression or classification, here we will focus on the former. GPs are mainly characterized by a symmetric positive definite kernel function k. Any symmetric function k is said to be

positive definite if the matrix $K_{\mathbf{X},\mathbf{X}} \in \mathrm{R}^{N\times N}$ with elements $[K_{\mathbf{X},\mathbf{X}}]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, is a positive semi-definite matrix (has non-negative eigenvalues). $K_{\mathbf{X},\mathbf{X}}$ is referred to as *kernel matrix* or *Gram matrix* [62]. Now we can define Reproducing Kernel Hilbert Spaces,

Definition 2.1. *Reproducing Kernel Hilbert Space (RKHS) is a specialized type of Hilbert space $\mathscr{H}_k$ equipped with an inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}_k}$ and associated with a positive definite kernel function/reproducing kernel such that the following statements hold true:*

$$\forall x \in \mathscr{X} \text{ we have } k(\cdot, x) \in \mathscr{H}_k$$

$$\forall x \in \mathscr{X} \text{ and } \forall f \in \mathscr{H}_k \quad \rightarrow f(x) = \langle f, k(\cdot, x)\rangle_{\mathscr{H}_k}$$

The Moore-Aronszajn theorem [63] establishes a fundamental result, showing that for any positive definite kernel k, there exists a distinct RKHS, such that k serves as its corresponding reproducing kernel. Hence, k possesses a remarkable property that taking the inner product of the function by evaluating it at two different points $\mathbf{x}$ and $\mathbf{x}'$ is the same as evaluating the kernel function at the pair $\mathbf{x}, \mathbf{x}'$ and hence it "reproduces" the values of the kernel function.

$$\langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{x}')\rangle = k(\mathbf{x}, \mathbf{x}') \tag{2.4}$$

This is the idea of the kernel trick, i.e. mapping features to higher dimensions doesn't incur any computational overhead since one can directly evaluate the inner product in higher dimensions using the kernel function. Therefore, $k(\cdot, x)$ is a feature representation of x that forms a complete basis for the RKHS, and hence any function inside the RKHS can be represented as a linear combination of these basis $f(\cdot) = \sum_{j=1}^{N} \alpha_j k(\cdot, \mathbf{x})$. We can also write our model in matrix notation. Let's define a vector $\Phi(\mathbf{x})$ of length N such that $\Phi(\mathbf{x}) = [\phi(\mathbf{x}_1), ..\phi(\mathbf{x}_N)]$ such that $\phi(\mathbf{x}_i) = k(., \mathbf{x}_i)$. We can write any linear model as $f(\mathbf{x}) = \Phi(\mathbf{x})\mathbf{w}$ where $\mathbf{w}$ is a set of weights. As we can see the functions inherit many properties of the kernel function for eg. if the kernel is s-times differentiable so is the function f. The choice of different kernel functions will determine the characteristic properties of the functions it reproduces such as continuity, smoothness, scale, etc. Additionally, RKHS facilitates efficient inference over Gaussian processes by making use of the kernel trick. The goal of any supervised machine-learning algorithm is to learn the joint probability distribution $p(\mathbf{X}, \mathbf{y})$ over the inputs and the outputs. In the Bayesian machine learning scheme, we model the parameters, which are weights $\mathbf{w}$ in this setting, as random variables. Bayes' rule says

$$p(\mathbf{w}|\mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y}|\mathbf{w}, \mathbf{X})p(\mathbf{w})}{p(\mathbf{y}|\mathbf{X})} \tag{2.5}$$

where $p(\mathbf{w}|\mathbf{y},\mathbf{X})$ is the posterior distribution of the weights given the observations $\mathbf{X},\mathbf{y}$. $p(\mathbf{w})$ is the prior distribution guided by our belief in the true underlying modeling function. Since our prior is a multivariate Gaussian distribution, our posterior will be a Gaussian too. $p(\mathbf{y}|\mathbf{w},\mathbf{X})$ is our likelihood i.e. probability of observing $\mathbf{y}$ at given parameter $\mathbf{w}$. The denominator which acts as a normalizing constant is defined as marginal likelihood and is obtained by integrating the parameters out $p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{w},\mathbf{X})p(\mathbf{w})dw$

2.3 Gaussian Process Regression

Now that we have the recipe, we can start to put all the ingredients together. Our goal is to make a prediction at an unknown set of points $\mathbf{x}^*$ given the training dataset $\mathscr{D}$. We pick a kernel function and evaluate the $N \times N$ square kernel matrix between input observations $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$ and similarly define the training-test kernel matrix $K^i_* = k(\mathbf{x}^*,\mathbf{x}_i)$. Now the predictive posterior is given by:

$$p(\boldsymbol{f}_* \mid \boldsymbol{x}^*,\boldsymbol{X},\boldsymbol{f}) = \mathscr{N}(\boldsymbol{f}_* \mid \boldsymbol{\mu}_*,\boldsymbol{\Sigma}_*), \tag{2.6}$$

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}(\boldsymbol{x}^*) + \mathbf{K}_*^T(\mathbf{K}+\sigma^2\mathscr{I})^{-1}(\mathbf{f}-\boldsymbol{\mu}(\boldsymbol{X})), \tag{2.7}$$

$$\boldsymbol{\Sigma}_* = \mathbf{K}_{**} - \mathbf{K}_*^T(\mathbf{K}+\sigma^2\mathscr{I})^{-1}\mathbf{K}_* \tag{2.8}$$

Here $\boldsymbol{\mu}_*$ is the model prediction, $\mathscr{I}$ is the identity matrix and $\boldsymbol{\Sigma}_*$ gives us a measure of the confidence/uncertainty in the predictions of our model. In case our training data is noisy such that $y = f + \varepsilon$, we can model the noise as a Gaussian distribution with a noise parameter σ^2 such that $\varepsilon \sim \mathscr{N}(0,\sigma^2)$, and compute it together with other hyper-parameters of the model. Even if our observations are noise free, it helps us to add a really small value of $\sigma^2 \sim 10^{-6}$ to obtain a numerically stable inverse of K, in case it is near singular. The objective of training a GP model is to find the parameters $\boldsymbol{\theta}$ of a function chosen to represent $\kappa(\boldsymbol{x},\boldsymbol{x}')$. A GP model is trained by maximizing the logarithm of the marginal likelihood (LML), which can be written in closed form in terms of $\mathbf{K}$ and $\boldsymbol{y}$ as follows:

$$\log \mathscr{L}(\theta) = -\frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{N}{2}\log 2\pi \tag{2.9}$$

2.4 Kernels and Automatic Model construction

Kernel functions are the workhorse of any GP model. To reiterate, a kernel matrix is a $N \times N$ square covariance matrix, where N is the number of training data entries. In the limit

of large N, any universal kernel such as the Radial Basis Function (RBF) kernel, would give accurate predictions. The problem becomes interesting when you have a small number of data points. In that limit, the choice of kernel serves as a measure of your prior belief about the dataset. For example, if you have periodic data with a linear trend, rather than going for standard RBF, it would be better to use a combination of linear and a periodic kernel. Thus the choice of the kernel is a measure of inductive bias in your model. It is not trivial to design custom kernels for your datasets, as the trends in datasets may be more convoluted than periodic in many cases, especially for multi-dimensional problems. We can construct more robust kernels simply by adding or multiplying any two kernels. This comes from the property of positive definite matrices (PSD) such that the addition/multiplication of two PSD matrices is a PSD matrix.

$$k_\alpha + k_\beta = k_\alpha(\cdot,\cdot) + k_\beta(\cdot,\cdot) \tag{2.10}$$

$$k_\alpha \times k_\beta = k_\alpha(\cdot,\cdot) \times k_\beta(\cdot,\cdot) \tag{2.11}$$

where $k_i(\cdot,\cdot)$ is a simple base kernel. The following set of kernel functions (base kernels) serve as a good example:

$$k\left(\boldsymbol{x},\boldsymbol{x}'\right) = \sigma \exp\left(-\frac{1}{2} r^2\left(\boldsymbol{x},\boldsymbol{x}'\right)\right) \tag{2.12}$$

$$k(\boldsymbol{x},\boldsymbol{x}') = \sigma\left(1+\sqrt{5}\cdot r(\boldsymbol{x},\boldsymbol{x}') + \frac{5}{3}\cdot r^2(\boldsymbol{x},\boldsymbol{x}')\right) \times \exp\left(-\sqrt{5}\cdot r(\boldsymbol{x},\boldsymbol{x}')\right) \tag{2.13}$$

$$k\left(\boldsymbol{x},\boldsymbol{x}'\right) = \sigma\left(1+\frac{1}{2\alpha} r^2\left(\boldsymbol{x},\boldsymbol{x}'\right)\right)^{-\alpha} \tag{2.14}$$

where $r^2\left(\boldsymbol{x},\boldsymbol{x}'\right) = \left(\boldsymbol{x}-\boldsymbol{x}'\right)^T \times \blacksquare \times \left(\boldsymbol{x}-\boldsymbol{x}'\right)$, $\blacksquare$ is a diagonal matrix of shape $|\mathscr{M}_i| \times |\mathscr{M}_i|$, and $|\mathscr{M}_i|$ is the number of kernel parameters for model $\mathscr{M}_i$.

So we can construct more complex kernels, but how do we know that they are *necessarily better*? Here we will use a model selection metric that will guide the construction of more complex kernels. In this thesis, complex kernels are built by a greedy search algorithm that uses Bayesian Information Criterion (BIC) as the model selection metric. This approach was first formulated by Duvenaud *et al.* [58] and is referred to as Automatic model construction in the literature. The BIC is given by:

$$\text{BIC} = \log \mathscr{L}(\boldsymbol{\theta}^*) - \frac{1}{2}\gamma \log N \tag{2.15}$$

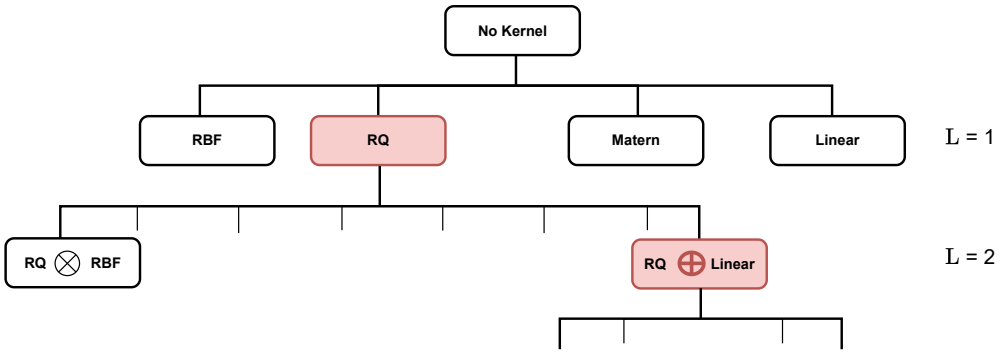


Figure 2.2: Schematic diagram of the automatic model construction method employed to develop kernels that can extrapolate well. At each layer (L), the kernel with the highest Bayesian information criterion (Equation 2.15) is selected (highlighted in red). In this particular diagram, the algorithm is shown for two layers. The search tree can be extended to arbitrary L, but it is a good idea to stop when $|\mathrm{BIC_L} - \mathrm{BIC_{L-1}}| \leq 0.05 \times \mathrm{BIC_L}$

Here γ is the number of hyper-parameters in the model. The second term serves to punish the more complex models, whereas the first term can be thought of as an estimate of the squared loss function. The algorithm works like a search tree as described in Fig. 2.2. Basically, we start with some simple kernel functions which we will call the base kernels. One evaluates BIC for each kernel, selects the highest one, and ascribes that as the new base kernel. Multiply and add the new base kernel with all the simple kernels and then evaluate BIC for each kernel, and again select the one with the highest BIC and make it the new base kernel. One can perform these steps for an arbitrary number of layers. In this case, we stop adding more layers once the incremental increase in BIC is less than 5%. Automatic model construction has been used in various applications such as fitting potential energy surfaces [58, 64–66] and extrapolating quantum observables [59]. It is worth pointing out that there are new ways to go beyond the greedy search algorithm and use gradient descent to optimize kernels for building GP models as described in Ref [67] or to use evolutionary algorithms to construct kernels [68]. Another route is to construct kernels out of neural networks, which is the field of deep kernel learning [69]. In this thesis, we will use the simpler greedy search algorithm since we are working with small datasets. In practice, we parallelize the evaluation of BIC over different cores of the computer, to save some time.

2.5 Example: Ginzburg-Landau phase transitions

We can demonstrate the efficacy of this approach for our main results (Chapter 3), by demonstrating it using a simpler example. Previously, Vargas-Hernández *et al.* [59] used this method to extrapolate the free energy of an Ising model by training on data from one side of the phase transition. We will consider a similar albeit slightly different example to demonstrate the power of this approach.

The Landau theory [70] is a concise and effective approach to describe phase transitions that arise due to symmetry breaking. In this theory, a key concept is the order parameter ψ, which characterizes the transition from a high symmetry phase ($\psi = 0$) to a low symmetry phase ($\psi \neq 0$). This order parameter often corresponds to physically measurable quantities, such as distortions in crystal lattice structures or spontaneous magnetization. The Ginzburg-Landau (GL) theory is a localized version of the Landau theory, where the order parameter becomes dependent on spatial coordinates. This variation makes it a fundamental tool in the phenomenological study of superconductivity [71, 72]. In the context of superconductivity, the GL theory accounts for how the order parameter varies throughout the material, giving insights into the behavior of superconducting currents and magnetic fields at different points. To determine the phase of the system, we describe a system by some order parameters and use them to construct a Landau potential. There exist many systems with two or more competing order parameters, coexisting with each other [73–76]. A simple Landau potential for a system with two order parameters can be given by

$$V(\psi_1, \psi_2) = -a_{ij}(\psi_i^* \psi_j) + \frac{1}{2} b_{ijkl}(\psi_i^* \psi_j)(\psi_k^* \psi_l) \quad i, j, k, l = 1, 2 \tag{2.16}$$

The coefficients are generally independent even though they can be related, depending upon the application. The Eq. (2.16) shows a general Landau potential for two-order parameters consisting of quadratic and quartic terms. To find the minimum of V one needs to solve static homogeneous Ginzburg-Landau equations. The complexity of solving them grows with the number of order parameters, as you need to solve six coupled algebraic equations for two order parameters, which is difficult already [77]. Moreover depending upon the problem being considered it might be easier to obtain the free energy density ($V(\psi_1, \psi_2)$) in one phase than the other. Hence ML models that can extrapolate information by only considering knowledge from one side of the phase transition can be really helpful in these situations. The way the calculation presented in the present work, differs from the previous application to the Ising model [59] is that we show how to calculate some physically relevant parameters (such as bandgap) of a model based on our ML predictions. Moreover, we

explore the extent to which we can push the boundaries of our ML framework with multiple order parameters.

The exact model we are considering is taken from the paper by Can *et al.* [78], where they construct a phenomenological Ginzburg Landau Hamiltonian to explain a topological phase transition in bilayer cuprates. It should be noted that topological phase transitions cannot be studied under the GL framework in general, due to the absence of a local order parameter like magnetization, superconducting order parameter, *etc*. The topological phase transitions depend upon the nontrivial topology of the system's ground state, which cannot be fully described by such local quantities [79, 80]. However in certain situations, one can construct a phenomenological Hamiltonian which captures the essential physics such that the phase transition occurring in the symmetry-breaking framework in our toy model corresponds to the topological phase transition of the system being modeled. This allows us to study key properties of our system marred by non-locality and global properties in a simplified framework as shown in Ref. [78].

In this thesis, we will not go deep into the explanation of the model that was used but just a brief overview of the problem, since these results can be extended to any GL model of this form. We directly consider the Bogoliubov-de Gennes (BdG) Hamiltonian considered in the paper of the form $\mathscr{H} = \sum_{\mathbf{k}} \Psi_{\mathbf{k}}^{\dagger} h_{\mathbf{k}} \Psi_{\mathbf{k}}$ such that $\xi_{\boldsymbol{k}a} = \hbar^2\mathbf{k}^2/2m$

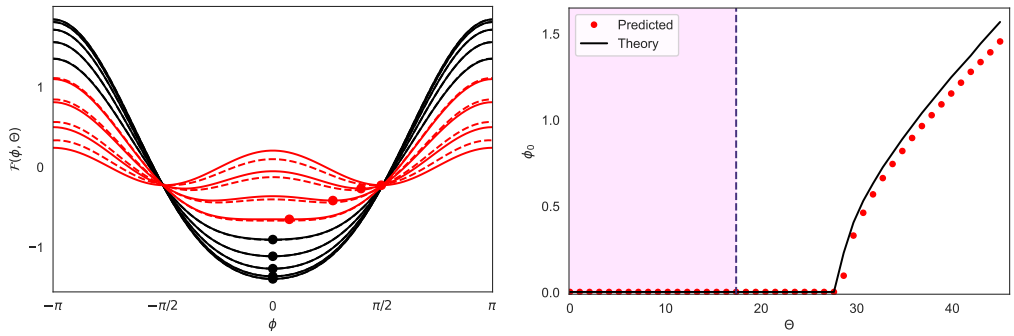
$$h_k = \begin{pmatrix} \xi_{\mathbf{k}} & \Delta_{\mathbf{k}1} & g & 0 \\ \Delta_{\mathbf{k}1}^* & -\xi_{\mathbf{k}} & 0 & -g \\ g & 0 & \xi_{\mathbf{k}} & \Delta_{\mathbf{k}2} \\ 0 & -g & \Delta_{\mathbf{k}2}^* & -\xi_{\mathbf{k}} \end{pmatrix} \tag{2.17}$$

Here $\Delta_{\mathbf{k}1} = \psi\cos(2\alpha_{\mathbf{k}})$, $\Delta_{\mathbf{k}2} = \psi\cos(2\alpha_{\mathbf{k}} - 2\Theta)$ are the superconducting order parameters for layers 1 and 2, which depend on twisting angle Θ between the two layers and g is the in-plane kinetic energy parameter. If we diagonalize $h_{\mathbf{k}}$, we obtain two values of the eigenenergies $E_{\mathbf{k}\alpha} = \pm\sqrt{(\Delta_{\mathbf{k}1}^2 + \Delta_{\mathbf{k}2}^2)/2 + \xi_{\mathbf{k}}^2 + g^2 + (-)^{\alpha} D_{\mathbf{k}}}$, where $D_k^2 = (\Delta_{\mathbf{k}1}^2 - \Delta_{\mathbf{k}2}^2)^2/4 + g^2(\Delta_{\mathbf{k}1}^2 + \Delta_{\mathbf{k}2}^2 + 4\xi_{\mathbf{k}}^2) - 2g^2\Delta_{\mathbf{k}1}\Delta_{\mathbf{k}2}\cos\varphi$, for each value of $\mathbf{k}$, where $\alpha_{\mathbf{k}}$ represents the polar angle of the vector $\mathbf{k}$. and $\alpha = 1, 2$. We can calculate the free energy using the expression

$$\mathscr{F}_{\mathrm{BdG}} = -2k_BT \sum_{\boldsymbol{k}\alpha} \ln\left[2\cosh\left(E_{\boldsymbol{k}\alpha}/2k_BT\right)\right] \tag{2.18}$$

Ref. [78] also provides a simpler phenomenonlogical expression for free energy give by:

$$\mathscr{F}(\Theta, \phi) \propto [-\cos(2\theta)\cos\varphi + \mathscr{K}\cos(2\varphi)] \tag{2.19}$$



(a) **(b)**

Figure 2.3: **(a)** Theoretically calculated free energy(solid lines) using Eq. (2.18). GP prediction(dashed lines). Each curve corresponds to a unique value of Θ **(b)** Black line is theoretically calculated Order parameter ϕ_0 as a function of twisting angle Θ. Red dots are the corresponding GP prediction. The pink region shows the training regime, such that 300 points were taken from $0 < \Theta < 17^{\circ}$ and $-\pi < \phi < \pi$.

where $\mathscr{K}$ has been kept as a constant in the present calculation. $\mathscr{F}_{\text{BdG}}$ and $\mathscr{F}(\Theta,\phi)$ can be shown to be equivalent [78]. The free energy $\mathscr{F}_{\text{BdG}}$ depends on two parameters, ϕ which is the order parameter, and the twisting angle between two cuprate layers, Θ. The system is in a gapless trivial phase for $\Theta < \Theta_c$ and transitions to a gapped topological phase for $\Theta > \Theta_c$. For the parameters, we chose for this calculation $\Theta_c = 28^{\circ}$. We train our GP model using the automatic model construction method with our $\mathbf{X} = \{\Theta,\phi\}$ and $\mathbf{y} = \{\mathscr{F}_{\text{BdG}}\}$. We only consider data from the trivial phase far away from the phase transition, as highlighted in the pink region in Fig. 2.3b. We then extrapolate our $\mathscr{F}_{\text{BdG}}$ and compute the order parameter ϕ_0 which minimizes the free energy obtained using GPR. Fig. 2.3a demonstrates the accuracy of the extrapolated free energy values and we see a good agreement between the GP prediction and the theoretically obtained order parameter as seen in Fig. 2.3b. The kernel for this dataset was found by to be $k = \text{MAT}(5/2) \times \text{RQ}$ by running the Automatic model construction for two layers. A similar analysis was done for the Ising model in Ref. [59]. We take this a step further by predicting the bandgap that arises due to the phase transition. The calculation of the bandgap requires accurate prediction of ϕ_0, based on which we calculate the band structures. We can see that the ML model accurately predicts the bandgap as shown in Fig. 2.4b by having no knowledge about the existence of any gap in the system. The bandgap is calculated by taking the difference between the two energy bands $|E_{\mathbf{k}1} - E_{\mathbf{k}1}|$. To take a broader view, we can also see that the band structure exhibits Dirac Cones in the trivial phase [81]. In quantum chemistry, similar structures exist when

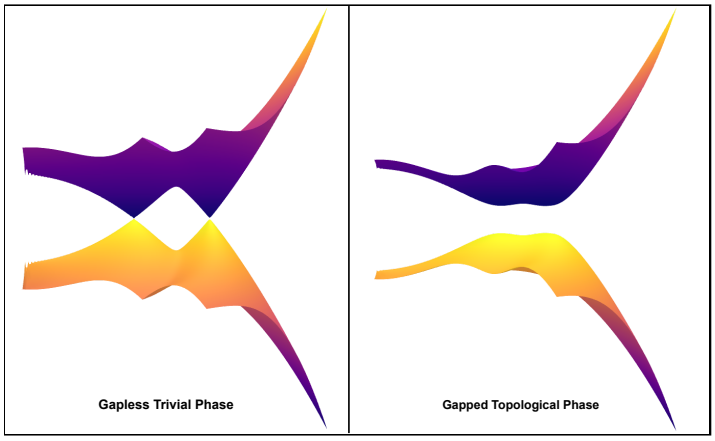
two potential energy surfaces touch at one point, they are known as Conical Intersections [81]. During the course of this project, we tried using similar ML models to interpolate and extrapolate conical intersections. We were only partially successful and hence the results are not recorded here. But Fig. 2.4a shows the evolution of band structure from trivially gapless to gapped topological phase. This suggests that similar ML models can be used to explore the nature of conical intersections.

We can take this model a step further by expanding it to models with multiple order parameters [77]. Here we will take a similar model provided in [82]. This is the same model but with 3 layers now, such that we have two order parameters ϕ_1, ϕ_2. The results are shown in Fig. 2.5. We can see that the GP model accurately predicts the phase transition even though it is trained from data only from the trivial phase. The model also predicts both order parameters successfully. The kernel for this dataset was found by to be $k = \mathrm{MAT}(5/2) \times \mathrm{RQ}$ by running the automatic model construction for two layers.

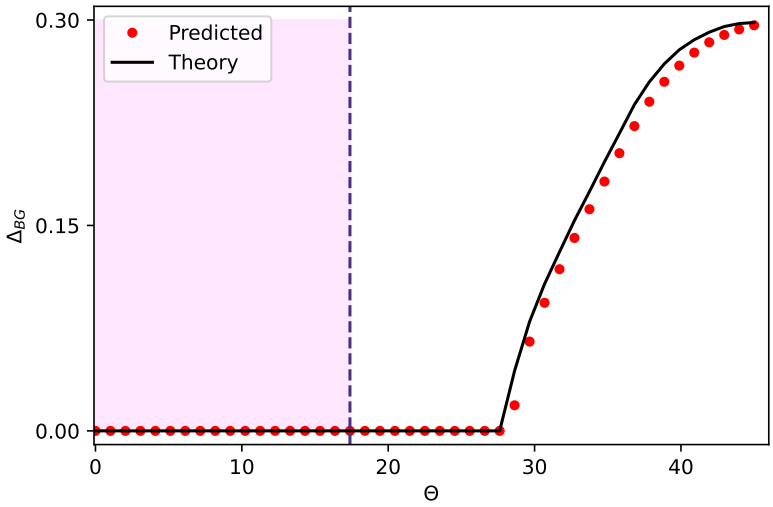
2.6 Multi-Fidelity Gaussian Processes

One of the main challenges in machine learning is to construct models that can generalize well by training on a few data points since data is often scarce or really expensive. For example in applications such as aerospace engineering or climate modeling, data acquisition can be challenging from a computational perspective or otherwise. A common solution is to develop simulators that can approximate original behavior instead of running experiments on actual systems. But often times one can expect these simulators to be noisy, biased/confined to a certain parameter regime. Hence, *multi-fidelity* models are designed in such a way that they can combine information from multiple sources i.e. the scarce true observations and the cheap simulations of the same. We will denote the cheaply available, noisy data samples as the *low-fidelity* observations whereas the scarcely available exact data will be referred to as *high-fidelity* observations. The goal is to systematically combine the high-fidelity and low-fidelity data to construct accurate models that can generalize better compared to single-fidelity ones.

In the present work, we follow the approach described by Perdikaris et al. [83] based on non-linear auto-regressive multi-fidelity Gaussian processes (NARGP). The goal is to build a model of a target function f_t, which is expensive to evaluate, by exploiting cheap auxiliary functions $f_0, f_1 ..., f_{t-1}$, which approximate f_t in the increasing order of fidelity. The relationship between the target function and the auxiliary functions can be generally

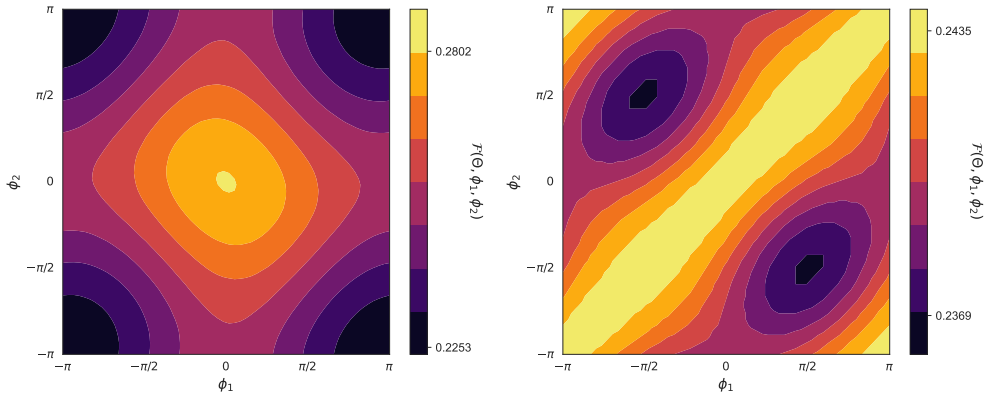


(a)

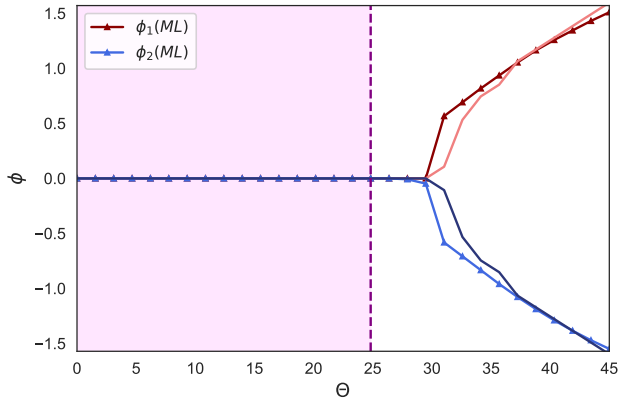


(b)

Figure 2.4: **(a)** Band structure plots $\pm E_{ka}(a = 1, 2)$ in the two phases. **(b)** Black line is theoretically calculated band gap δ_{BG} as a function of twisting angle Θ. Red dots are the corresponding GP prediction. δ_{BG} is calculated by subtracting the maximum of the lower band from the minimum of the upper band. E_{ka} calculation requires knowing the order parameter value for that particular twisting angle which is obtained using GP in Fig. 2.3b



(a)



(b)

Figure 2.5: (a) Theoretically calculated free energy for multiple order parameter case (Left) Free energy in the trivial phase with one minimum (Right) Free energy in the topological phase with two minimas. **(b)** Solid lines are theoretically calculated Order parameter ϕ_1, ϕ_2 as a function of twisting angle Θ. Solid lines with triangles are the corresponding GP prediction. The pink region shows the training regime, such that 600 points were taken from $0 < \Theta < 25^{\circ}$, $-\pi < \phi_1 < \pi$ and $-\pi < \phi_2 < \pi$.

represented as

$$f_t(\boldsymbol{x}) = \rho_{t-1}(f_{t-1}(\boldsymbol{x})) + \gamma_t(\boldsymbol{x}) \tag{2.20}$$

where $\rho_{t-1}(\cdot)$ is a nonlinear mapping between two successive auxiliary functions, which is independent of $\gamma_t(\boldsymbol{x})$. This non-linear mapping and γ_t are provided by independent GPs.

NARGP model can thus be viewed as a GP process with the following kernel function

$$\begin{aligned} k_t(\boldsymbol{x},\boldsymbol{x}') = k_{\rho_{t-1}}(\boldsymbol{x},\boldsymbol{x}') \times k_{f_t}(f^*_{t-1}(\boldsymbol{x}), f^*_{t-1}(\boldsymbol{x}')) \\ + k_{\gamma_t}(\boldsymbol{x},\boldsymbol{x}'), \end{aligned} \tag{2.21}$$

where each of the kernel functions k_ρ, k_f and k_γ is parametrized by an independent set of parameters to be learned and f^*_{t-1} denotes the posterior at fidelity level $t-1$.

When training the GP model by LML maximization, the GP at level t thus learns simultaneously from the data of fidelity t and the GP at level $t-1$ (which is trained by data of lower fidelity). The expectation that the resulting Bayesian models must be efficient is based on the assumption that low-level approximate calculations provide a better description of the physical process than random guessing. Using NARGP, one can propagate the uncertainty fully from a lower level of fidelity to higher levels of fidelity. Hence, training a NARGP model is akin to training a regular GP model. For a more detailed discussion of multi-fidelity Gaussian processes see Ref. [84]. For a more detailed discussion on multi-fidelity Gaussian processes see Ref. [84]. Multi-Fidelity models have found a lot of applications in chemistry for potential energy surface calculations [85]. Here we will apply them to some condensed matter models in Section 3.4.4 .

Chapter 3

Bayesian extrapolation of quantum observables in Hamiltonian parameter space

The electron-phonon coupling i.e., the interaction between electrons and crystal lattice plays a vital role in determining electronic properties of materials at various scales. Most importantly, the electron-phonon interaction at microscopic scales gives rise to exciting phenomena such as charge density waves, BCS superconductivity, etc. The term "Polaron" refers to the state of the quasiparticle dressed by its interaction with crystalline lattice [86, 87]. In this section, I aim to give an overview of the electron-phonon coupling phenomena and how it leads to the formation of polarons and bipolarons. I further discuss the role polaronic effects play in explaining conventional (BCS) superconductivity, especially through the bipolaron mechanism, and how it relates to high-temperature superconductors. I will give a review of the recent results in this direction and discuss the motivations for using machine learning techniques for this problem. Finally, I will present our results in Section 3.4.

3.1 Electron phonon coupling and the Polaron problem

In order to study condensed matter systems, we need to write a Hamiltonian describing the state of $\sim 10^{23}$ number of atoms and electrons. Since the time/space complexity of this problem grows exponentially in the number of electrons, solving this system using direct diagonalization would be infeasible. We can simplify the problem by imposing certain as-

sumptions. The first approximation is known as Born-Oppenheimer Approximation (BOA). Since electrons are much lighter compared to the nucleus, we can decouple electronic motion from nuclear motion such that nuclear coordinates are just parameters in the electronic Hamiltonian. The decoupling is only partial since the dynamics of the electrons are expanded in terms of the parameter κ, which is the ratio of electron mass to ion(nucleus+core electron) mass $\kappa = (m/M)^{1/4}$. In cases where the ions are light, this approximation may fail. Secondly, we will only consider the valence electrons i.e. electrons lying in the outermost shell because they are shielded from nuclear attraction due to Coulombic screening by core electrons. This screening allows them to have higher degrees of freedom and spatially extended wave functions. In general, we should note that we make many more approximations like these. For example the Relativistic approximation, which states that the electron velocities being considered are small compared to the speed of light. We have highlighted the most common assumptions here. These assumptions can be the source of error when we are considering specific systems, but for most of the physics we will consider here, these approximations hold true unless stated otherwise. [88]

Under BOA we can write total wavefunction as $\Psi(\mathbf{r},\mathbf{R}) = \chi(\mathbf{R})\psi(\mathbf{r};\mathbf{R})$ where $\mathbf{r}$ and $\mathbf{R}$ are electron and ion coordinates, respectively. The electronic wavefunction obeys the Schrodinger equation

$$\underbrace{\left[\underbrace{-\frac{\hbar^2}{2}\sum_i \frac{\nabla^2_{r_i}}{m}}_{T_e} + \underbrace{\frac{1}{4\pi\varepsilon_0}\sum_{ij}\frac{e^2 Z}{|R_i - r_j|}}_{V_{e-ion}} - \underbrace{\frac{1}{8\pi\varepsilon_0}\sum_{i\neq j}\frac{e^2}{|r_i - r_j|}}_{V_{e-ion}}\right]}_{H_e}\psi(\mathbf{r};\mathbf{R}) = E_n(\mathbf{R})\psi(\mathbf{r};\mathbf{R}) \qquad (3.1)$$

in which M is the mass of an ion, m is the mass of an electron, r_i is the position of the i^{th} electron and R_i is the position of the i^{th} ion. In the expression of electronic Hamiltonian H_e, the first term is the kinetic energy of electrons T_e, the second term is the Coulomb attraction felt by electrons in the periodic lattice of ions V_{e-ion}, and the third term is the Coulomb repulsion between electrons V_{e-e}. The non-interacting part of the electronic Hamiltonian i.e. $V_{e-ion} + T_e$ can be solved exactly. The interacting problem V_{e-e} is dealt with in various different ways. One way is to use first principle tools such as perturbation theory or Density Functional theory [87]. The alternative approach involves seeking a simpler and more effective Hamiltonian that can accurately represent the essential physics of the system. This approach makes the original problem more tractable such that it can be solved by quantum Monte Carlo, exact diagonalization, or other numerical methods [89, 90].

The ions arranged in the lattice aren't stationary (otherwise this would violate Heisenberg's uncertainty principle) but they vibrate about their mean position. In most cases, these oscillations can be modeled as a harmonic oscillator. Considering the simple case of one ion in each of the N unit cells, there are $3N$ harmonic oscillators. The phonon modes are usually of two types, acoustic modes which are low in energy and oscillate in phase with the motion of atoms. There are also optical modes that are high in energy and out-of-phase with the motion of atoms. These out-of-phase motions produce electric dipoles which strongly couples to the electromagnetic field of electrons. In case there are anharmonic terms, this may lead to three body processes such as phonon-phonon scattering. To formulate the electron-phonon Hamiltonian we can Taylor expand the electron-ion potential about its mean position R_i^0. If the i^{th} ion has a small displacement of u_i about its mean position, we assume that $u_i = R_i - R_i^0 \ll a$ where a is the lattice constant. Assuming linear dependence on lattice distortion, we can write the interaction potential between electron and ion, as

$$V_{e-ion}(r-R_i) = V_{e-ion}\left(r-R_i^0\right) - u_i \cdot \nabla_r V_{e-ion}\left(r-R_i^0\right) \tag{3.2}$$

$$H = H_e + H_{ph} - \sum_i u_i \cdot \nabla_r V_{e-ion}\left(r-R_i^0\right) \tag{3.3}$$

Here we will drop the first term in the Eq. (3.2) in the final equation since that represents electron interaction with the frozen lattice, and is already accounted for in H_e. The last term represents our electron-phonon interaction. Now we shift to the second quantization picture, under the tight-binding approximation. This approximation replaces the Bloch state with a localized state centered on each ion. The electron can hop to neighboring sites, and the hopping is controlled by a hopping parameter t which is proportional to the wave function overlap between the sites. This overlap decreases rapidly such that higher-order hoppings are exponentially suppressed in most systems. Given that the total momentum is conserved, it is easy to work in Fourier (momentum) space where we only need to specify the coupling g. In the momentum space picture, the electronic part of the Hamiltonian H_e transforms as $H_e = \sum_k \varepsilon(k) c_k^\dagger c_k$, where $\varepsilon(k)$ is eigenenergy/electron dispersion of free electron, given by $\varepsilon(k) = -2t\cos(k)$. $c_k(c_k^\dagger)$ is the electron annihilation(creation) operator, and k is electron crystal momentum. This is still an approximation, but it holds well when we work with the quasiparticle picture and weak electron-electron coupling. Here we use the harmonic approximation to write $H_{ph} = \sum_q \omega(q) b_q^\dagger b_q$ where $\omega(q)$ is phonon dispersion relation of H_{ph}, $b_q(b_q^\dagger)$ is the phonon annihilation(creation) operator, and q is phonon momentum. Under the Einstein model of optical phonons $\omega(q) = \Omega$ i.e. phonons are dispersionless. These are the kinds of phonons we will consider throughout this thesis. Here I will write the final

Hamiltonian [88]:

$$H = \sum_k \varepsilon(k) c_k^\dagger c_k + \sum_q \Omega b_q^\dagger b_q + \sum_{k,q} \frac{g(k,q)}{\sqrt{N}} c_{k+q}^\dagger c_k \left(\hat{b}_q^\dagger + \hat{b}_q\right) \tag{3.4}$$

These electron-phonon interactions play important roles in many different situations. Electrons can interact with each other through an exchange of virtual phonons. For electrons near the Fermi surface, this interaction is attractive such that electrons bind together to form cooper pairs, which is the microscopic origin of the superconductivity. We can understand the main electron-phonon interaction through Feynman diagrams.

1. Electrons with momentum k being scattered by phonons carry a momentum q. This form of electron-phonon scattering is a dominant source of resistivity in many materials.

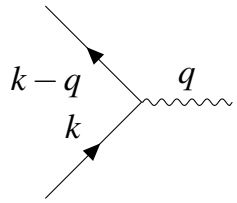


Figure 3.1: Electrons scattered by phonons

2. An electron can emit a virtual phonon and reabsorb it. These types of diagrams are important while calculating a single-particle Green's function. This process leads to the renormalization of the properties of the bare particle.

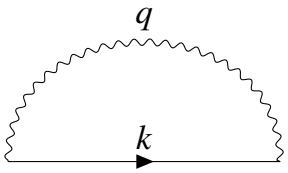


Figure 3.2: Electron self-interaction through phonons

3. An exchange of virtual phonons by two electrons gives rise to an effective attractive potential. This leads to the formation of Cooper pairs as explained by BCS theory of superconductivity.

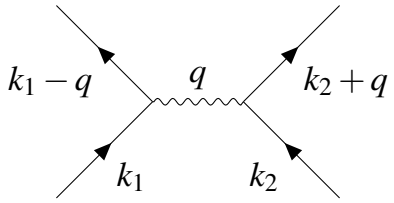


Figure 3.3: Two electrons interacting by an exchange of phonons

3.1.1 Bipolaronic superconductivity and Peierls model

The Bardeen-Cooper-Schrieffer (BCS) theory presents a formalism to explain conventional superconductivity [91, 92]. It states that Coulomb repulsion between the electrons is screened by an arbitrarily weak effective phonon-mediated electron-electron attraction. An electron present near the Fermi surface can use this attractive interaction to pair up with a nearby electron with opposite momenta and spin, thus lowering its energy and forming Cooper pairs. The wavefunctions of these pairs are delocalized over multiple lattice sites, this allows for the coherent transport of electrons without scattering. The emergence of this superconducting state occurs below a critical temperature (T_c) where thermal fluctuations are dominated by attractive interactions. One can observe the emergence of Cooper pairs by solving the BCS Hamiltonian (H_{BCS}) under the mean-field approximation.

$$H_{BCS} = \sum_{k,\sigma} (\varepsilon_k - \mu) c^\dagger_{k\sigma} c_{k\sigma} + \sum_k \Delta_k (c^\dagger_{k\uparrow} c^\dagger_{-k\downarrow} + c_{-k\downarrow} c_{k\uparrow}) \tag{3.5}$$

Here ε_k is the electron dispersion, μ is the chemical potential and $\Delta_k = \sum_{k'} V_{k,k'} \langle c_{-k'\downarrow} c_{k'\uparrow} \rangle$, where $|\Delta_{\mathbf{k}}|$ plays the role of an energy gap in the superconducting spectrum. This phonon-induced interaction was first derived in Refs. [93, 94]. An interesting observation one can make is that increasing Δ_k doesn't lead to high T_c superconductivity. This is because the phonon dressing around the electron leads to an increased effective mass, thus creating highly localized Cooper pairs. Hence the persistent interest in the polaron problem stems from its applications in understanding high-temperature/T_c superconductivity [95, 96], especially through the bipolaron mechanism. A bipolaron is a localized self-trapped state for pair of electrons and their collective interaction with the lattice (lattice distortion by the charges is also called self-trapping). [97, 98] Generally, bipolarons are considered bad for superconductivity as put very concretely by Chakraverty *et al.* in Ref. [99]. This comes from the conventional belief that when a single electron dressed by the cloud of phonons moves through the lattice, the phonon cloud acts as extra weight contributing heavily to the

effective mass of the quasiparticle. When we have two polarons i.e. two electrons in the system with their respective phonon clouds, these two clouds can exchange phonons leading to the formation of a larger phonon glue which induces binding between two polarons to form a bipolaron. Hence, in the conventional models where phonons couple to electron density and modulate the on-site energy of the electrons, the effective mass of the quasiparticles is often exponential in the coupling strength. Since the effective mass of cooper pairs (m_{cp}) is inversely related to critical temperature as in Eq. (3.6), this presents a problem for High-T_c superconductors.

$$k_B T_c \propto \frac{1}{m_{cp}} \tag{3.6}$$

A natural question to ask is whether we can adjust this weak coupling theory to understand the physics of superconductors in a strong coupling regime. One answer to that question might be the Migdal-Elisaberg (ME) theory [100], which is a theory based upon Migdal theorem [101] and adiabatic approximations. ME theory allows us to ignore the vertex corrections perturbatively. Here the small perturbation theory parameter is $\lambda \frac{\Omega}{E_F} T \ll 1$ where λ is the dimensionless electron-phonon coupling strength multiplied by a small parameter which is the ratio of phonon frequency (Ω) to Fermi energy (E_F). But this approximation breaks down at intermediate coupling λ [100] because at this point bipolarons start to emerge in the ME theory. Moreover, it is not applicable in the low-density limit where E_F is quite small.

Recently, researchers have started exploring a different type of electron phonon coupling, where phonons modulate the hopping amplitude of the electron, such as the Su-Schieffer-Heeger (SSH) model, commonly known as the Peierls model. Sous *et al.* [102] showed that a breathing mode Peierls model leads to polarons whose effective mass stays roughly constant $\mathcal{O}(1)$ as we transition to the strong coupling regime for a one-dimensional model. However, this result was proven to be true for a one-dimensional model in the adiabatic regime ($\Omega \sim 4t$). Hence, the authors introduced the Bond-Peierls model [103, 104] to study this phenomenon in two dimensions, which is akin to the breathing mode Peierls model, but the phonons dangle on the bonds rather than the sites to modulate the electron hopping. The authors still found the signatures of light polarons in two dimensions. Recently, this result has been extended to the case of bipolarons in the presence of screened [31] and unscreened Coloumb repulsion [30], such that phonons form strongly bound yet light bipolarons and are resilient to local repulsive interactions of Hubbard type. Specifically, in [31] Sous *et al.* demonstrated a mechanism to obtain light bipolarons in Peierls model that have T_c higher than what is predicted by the Migdal-Elishaberg theory [100]. Thus,

exploring the Peierls model opens up the possibility of understanding phonon-mediated high-temperature superconductivity. However, most of these results focus on the antiadiabatic regime i.e. when phonon dynamics are fairly fast compared to electron motion. To gain insight into real physical phenomena, we need to solve the same problem in the extreme adiabatic regime ($\Omega \ll 4t$). The simulations in this regime become computationally intractable since phonons become quite cheap at lower phonon frequencies, and hence the bosonic Hilbert space becomes too big to simulate using available numerical/quantum Monte Carlo methods [105, 106]. There is also a debate surrounding the survival of the sharp transition observed in the electron ground state momentum for the Peierls polaron [57] as one goes to the extreme adiabatic regime. Data-driven machine learning models provide a promising framework to help address these concerns. We will describe how to build efficient and accurate ML models for this problem in Section 3.4 of this thesis.

3.2 Electron phonon models

In this thesis, we will consider polarons formed due to electrons moving through a crystalline lattice, dressed by the lattice vibrational modes called phonons. The electron-phonon interaction on the lattice can be broadly described by two types of models, $g(q)$ and $g(k,q)$ models, with the Holstein and Peierls models being examples of each type, respectively. We consider a one-dimensional lattice electron-phonon model with infinite sites ($N \to \infty$) with generalized coupling between the fermionic (electron) and bosonic (phonon) parts of Hamiltonian. The most general Hamiltonian considered in this work can be written as follows:

$$\begin{aligned}\hat{H} = & \underbrace{-t\sum_{\langle ij\rangle}\hat{c}_i^\dagger\hat{c}_j}_{H_e} + \underbrace{\Omega_{\mathrm{H}}\sum_i \hat{b}_i^\dagger\hat{b}_i + \Omega_{\mathrm{SSH}}\sum_i \hat{b}_i^\dagger\hat{b}_i}_{H_{ph}} \\ & \underbrace{+\alpha_{\mathrm{H}}\sum_i \hat{c}_i^\dagger\hat{c}_i\left(\hat{b}_i^\dagger+\hat{b}_i\right)}_{V_{\mathrm{H}}} \\ & \underbrace{+\alpha_{\mathrm{SSH}}\sum_{\langle ij\rangle}\left(\hat{c}_i^\dagger\hat{c}_j+\hat{c}_j^\dagger\hat{c}_i\right)\left(\hat{b}_i^\dagger+\hat{b}_i-\hat{b}_j^\dagger-\hat{b}_j\right)}_{V_{\mathrm{SSH}}}\end{aligned} \tag{3.7}$$

where t is the electron hopping amplitude and will be set to $t = 1$ throughout our calculations, Ω is the dispersionless frequency of phonons, α_{SSH} and α_{H} are the corresponding strengths of the electron-phonon couplings for the SSH and Holstein models, respectively, and the operators $\hat{c}_i$ and $\hat{b}_i$ create the electron and phonon at site i of an infinite lattice. In the

present work, we consider a single phonon branch and perform calculations for two types of models: (i) the pure SSH model that corresponds to the $\alpha_{\mathrm{H}} = 0$ limit of Eq. (3.7); and (ii) the mixed model (3.7) with both $\alpha_{\mathrm{SSH}} \neq$ and $\alpha_{\mathrm{H}} \neq 0$. The electron - phonon coupling is quantified by $\lambda_{\mathrm{SSH}} = 2\alpha^2_{\mathrm{SSH}}/\Omega_{\mathrm{SSH}}t$ for the pure SSH model and by a combination of λ_{SSH} and $\lambda_{\mathrm{H}} = \alpha^2_{\mathrm{H}}/2\Omega_{\mathrm{H}}t$ for the mixed model. We also use the adiabaticity ratio

$$\Lambda = \Omega_{\mathrm{H/SSH}}/4t, \tag{3.8}$$

to quantify the regime of electron - phonon interactions. If $\Lambda >> 1$, we are in the anti-adiabatic regime such that phonons are moving very fast compared to electrons. This regime is easier to compute since the timescales for the two systems are quite different, hence one can apply perturbation theory quite successfully. For $\Lambda << 1$, we are in the extreme adiabatic regime, where phonons move quite slowly. This means that this regime is quite hard to probe with traditional numerical methods [60]. The effective mass is another standard way of measuring how strongly the interaction modifies the properties of the dressed particle compared to the bare electron. It is defined as :

$$\frac{1}{m^*} = \left[\frac{d^2E_P(k)}{dk^2}\right]_{k=K_{\mathrm{GS}}} \tag{3.9}$$

3.2.1 Holstein Model

The $g(q)$ coupling has its origins in the modulation of on-site energy due to neighboring lattice vibrations. Due to the Coulombic nature of the coupling, it is diagonal in electron operators in real space and hence only depends on phonon momentum q and not on electron momentum k. Holstein's model is a paradigmatic example of $g(q)$ coupling [107, 108]. The $g(k,q)$ models arise when hopping amplitudes of electrons are affected by lattice distortions, and since the electronic part of Hamiltonian is off-diagonal in real space, this brings k back into the picture. The Holstein model is obtained by setting Ω_{SSH} and α_{SSH} equal to zero in the generalized Hamiltonian in the Eq. (3.7)

$$\hat{H}_H = -t\sum_{\langle ij\rangle}\hat{c}_i^\dagger\hat{c}_j + \Omega_{\mathrm{H}}\sum_i \hat{b}_i^\dagger\hat{b}_i + \alpha_{\mathrm{H}}\sum_i \hat{c}_i^\dagger\hat{c}_i\left(\hat{b}_i^\dagger + \hat{b}_i\right) \tag{3.10}$$

It is a kind of $g(q)$ model, such that electron-phonon interaction modulates the potential(on-site) energy of the electron as shown in Fig. 3.4. The free electron dispersion is given by $-2t\cos(k)$. For a single polaron, in the weak coupling regime, the polaron is large and

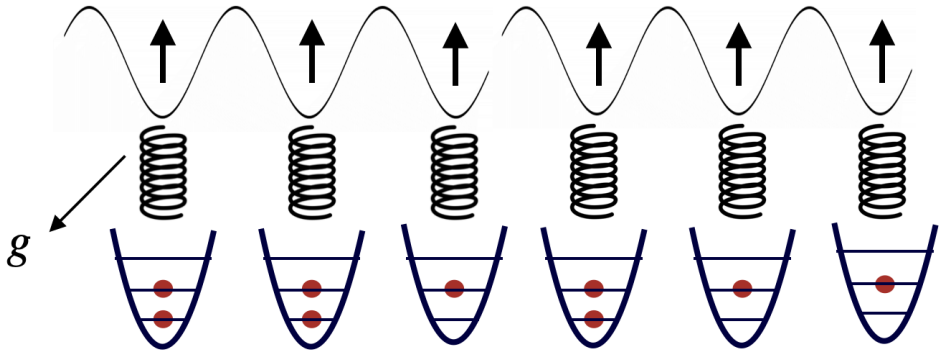


Figure 3.4: Holstein model

light and behaves like a continuous large polaron, at least at small momentum. In the strong coupling regime, $\lambda > 1$, we can calculate the renormalized dispersion relation using perturbation theory which is, $-2te^{-2t\lambda/\Omega}\cos(k)$. This results in an exponentially enhanced polaron mass given by $(2te^{-2t\lambda/\Omega})^{-1}$ [99]. The ground state of a single polaron smoothly evolves under the transition from a weak to a strong coupling regime. Similar statements can be made for the Holstein bipolaron, i.e. bipolaron mass is exponentially enhanced and there is a smooth crossover in bipolaronic ground state properties. This follows from the Gerlach-Lőwen theorem which states that there exist no non-analyticities in ground state properties of $g(q)$ type electron-phonon interaction models [109].

3.2.2 Peierls Model

The $g(k,q)$ models arise when hopping amplitudes of electrons are affected by lattice distortions, and since the electronic part of Hamiltonian is off-diagonal in real space, this brings k back into the picture. The Su-Schrieffer-Heeger (SSH)/Peierls model [56, 110, 111] is a well-studied $g(k,q)$ model which has starkly different properties compared to widely used Holstein models such as low effective mass in strong coupling regimes and sharp transitions of ground state momentum as we go from weak to strong coupling regimes [57]. We can write the Peierls model by setting Ω_{H} and α_{H} equal to zero in the generalized Hamiltonian in Eq 3.7

$$\hat{H}_P = -t\sum_{\langle ij\rangle}\hat{c}_i^\dagger\hat{c}_j + \Omega_{\mathrm{SSH}}\sum_i \hat{b}_i^\dagger\hat{b}_i + \alpha_{\mathrm{SSH}}\sum_{\langle ij\rangle}\left(\hat{c}_i^\dagger\hat{c}_j + \hat{c}_j^\dagger\hat{c}_i\right)\left(\hat{b}_i^\dagger + \hat{b}_i - \hat{b}_j^\dagger - \hat{b}_j\right) \tag{3.11}$$

It is a kind of $g(k,q)$ model, such that electron-phonon interaction modulates the kinetic energy (hopping parameter) of the electron. The k dependence of the coupling implies that we cannot apply the Gerlach-Lőwen theorem, which forbids non-analytic behavior in the polaron ground state properties. The SSH coupling opens up another channel of delocalization for the electron. For example, the particle can hop from one site creating a phonon there and then absorbing it as it hops further. All these self-interactions result in an effective nearest neighbor hopping and polaron dispersion goes from $-2t\cos k$ to $\approx -2t^*\cos k + 2t_2\cos 2k$, where t^* is the normalized nearest-neighbor hopping. (Note that this ansatz is only true for higher frequency values). For single Peierls polaron, there exists a sharp transition(due to the dominant t_2 term) in polaron ground state momentum and effective quasiparticle mass, as we tune the coupling strength λ [57]. The transition occurs at a critical coupling strength λ_c below which the quasimomentum of the ground state is zero. This is calculated by locating the quasimomentum k value corresponding to the minimum of the polaron dispersion $E(k)$. In other words, if we define the total quasimomentum operator K_{total} as

$$K_{total} = \sum_k k\hat{c}_k^\dagger \hat{c}_k + \sum_q q\hat{h}_q^\dagger \hat{h}_q \tag{3.12}$$

then the minimum of the energy corresponds to eigenvalue $K = 0$ of K_{total}. Above λ_c the ground state of $\hat{H}_P$ is two-fold degenerate corresponding to $\pm K_{\mathrm{GS}}$ quasimomenta values, on each side of the Brillouin zone. For a sufficiently strong coupling strength, K_{GS} becomes $\pi/2$.

3.3 Numerical method for Green's function calculation

Green's functions are the workhorse of quantum many-body calculations especially when it comes to dynamics of condensed matter systems. In cases where the exact solutions to wavefunctions are not known, Green's functions are often used to calculate the eigenstates and response functions of complex Hamiltonians. The single particle retarded Green's function which is a type of retarded two-time correlation function can be used to probe the dynamics of a single electron moving through a crystalline lattice whilst interacting with bosonic excitations. One can compute the dynamical response function by taking its Fourier transform, and hence study various magnetic, transport, and even optical properties [112]. In the present context, we define the Green's function as the matrix element of the resolvent $G(\omega) = \frac{1}{\omega - H + \imath\eta}$. Assuming we denote the complete set of eigenstates of the Hamiltonian under consideration H by E_α, indexed by some quantum number α. Then the elements of

$G(\omega)$ can be decomposed using Lehmann's representation:

$$G^{fi} = \langle f|G(\omega)|i\rangle = \sum_{\alpha} \frac{\langle f|\alpha\rangle\langle\alpha|i\rangle}{\omega - E_\alpha + \imath\eta} \tag{3.13}$$

Here η is an artificial broadening parameter. We can see that in the limit $\eta \to 0$, the imaginary part of G^{fi} is non-zero only at eigenvalues E_α. Hence, we can represent it using a delta function such that $-\frac{1}{\pi}\text{Im}\frac{1}{\omega - H + \imath\eta} \to \delta(\omega - E_\alpha)$. If we assume the initial state of time $t = 0$ to be single excitation of the bare particle i.e. $c_k^\dagger|0\rangle$ we can write the Green's function as a sum of multiple Lorentzian peaks with a certain peak height.

$$A(k,\omega) \equiv -\frac{1}{\pi}\text{Im } G(k,\omega) = \sum_{\alpha} Z_{k,\alpha}\delta(\omega - E_\alpha) \tag{3.14}$$

If we have an outgoing photo-electron with momentum k, using a current detector in a spectroscopic experiment, can give us an accurate measure of the quantity $A(k,\omega)$ known as the *momentum-frequency resolved spectral function*, which is the case in ARPES [87]. The imaginary part of Green's function can give us many quantities of interest such as the quasiparticle weight given by $Z_{k,\alpha} = |\langle\alpha|c_k^\dagger|0\rangle|^2$. $Z_{k,\alpha}$ contains information about renormalization of eigenstate of H. Additionally, the poles in the spectra provide information on the eigen-energies of the particle. For each quasimomentum k, $A(k,\omega)$ satisfies the sum rule i.e. $\int_{-\infty}^{\infty} A(k,\omega)d\omega = 1$. We can also obtain the total density of states $\rho(\omega)$ by integrating out the momentum part from the spectral function, $\rho(\omega) = \int A(k,\omega)dk$. If we perform an inverse Fourier transform on $A(k,\omega)$, we can obtain the quantity $\langle\psi(t)|c_k^\dagger|0\rangle$ which is the overlap between the state $\psi(t)$ of the system at time t to its initial state at $t = 0$. If we take the square modulus of it, we can obtain Loschimdt echo [113] defined as

$$\mathscr{L}_k(t) \equiv |\langle\psi(t)|c_k^\dagger|0\rangle|^2 \tag{3.15}$$

Loschimdt echo is an important quantity in many-body physics since it measures the sensitivity of the system to perturbations [114]. It can be used to study dynamical quantum phase transitions [115], quantum chaos [116], decay of entanglement [117], etc. In this thesis, we will use Generalized Green's function Cluster state Expansion (GGCE) method [60, 118] which comes from the Momentum Approximation (MA) [119, 120] family of methods to calculate Green's functions for electron-phonon Hamiltonian. MA provides a variational approximation to the exact Green's function and has been proven to be quite efficient compared to many quantum monte carlo approaches for specific systems. The goal

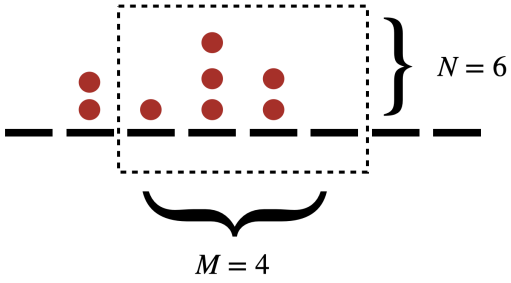


Figure 3.5: Depiction of GGCE variational algorithm: In this example, only $N = 6$ phonons can be created inside the dotted box of size $M = 4$. The bare particle can be anywhere on the lattice

of the GGCE method [60] is to calculate Equations of Motion (EOM) for a single electron Green's function for models of the form in Eq. (3.7) which is given by

$$G(\mathbf{k},\omega) = \langle 0|\hat{c}_{\mathbf{k}}\hat{G}(\omega)\hat{c}^{\dagger}_{\mathbf{k}}|0\rangle \tag{3.16}$$

where $\hat{G}(\omega) = [\omega - \hat{H} + \imath\eta]^{-1}$. Here η is an artificial hyperparameter that controls the broadening of the peaks in the spectrum. By applying Dyson's identity iteratively, we can derive an infinite hierarchy of EOMs. The Momentum Approximation (MA) [119–122] family of methods takes advantage of the structure of EOMs in Green's function space. Specifically, GGCE approximates the EOMs by a variational ansatz parameterized by, M which represents the spatial extent of the phonon cloud, and N which defines the number of bosons in one phonon cloud. The carrier is free to move such that the N bosons are confined in a cloud of length M around it. For our purposes, GGCE is a numerical method that implements the evaluation of Green's function for a single charge carrier based on the two variational/convergence parameters M,N. The number of EOMs we need to solve scales combinatorially in M and N. For a single (k,ω) grid point if the number of equations we need to solve for $(M,N) = (5,10)$ is roughly $\sim 2k$. This number scales to $\sim 12k$ for $(M,N) = (6,12)$ and hence one can see why this quickly becomes intractable for present-day supercomputing clusters for even a small (k,ω) grid of a thousand points. In the present work, we use GGCE to calculate electron dispersions for the Peierls model. We use a single branch MA code, to calculate the electron dispersion for Peierls+Holstein mixed model. We provide a solution to this problem in Section 3.3 to ease the computational expense of these calculations.

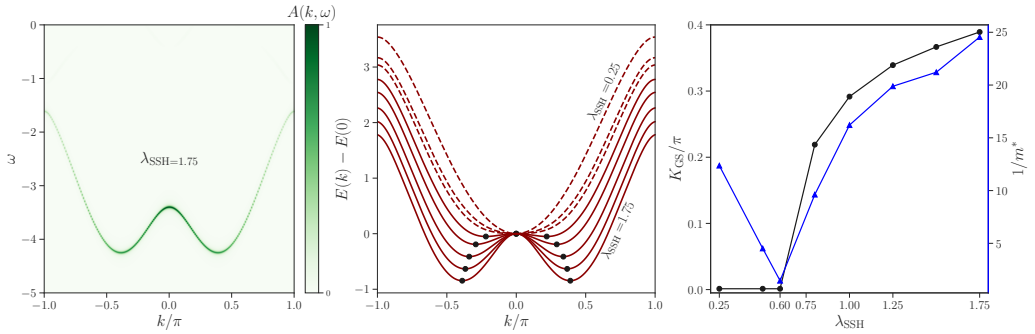


Figure 3.6: (Left) Spectral function $A(k,\omega)$ for $\Omega = 4.0$ and coupling strength $\lambda_{\mathrm{SSH}} = 1.75$ and $\alpha_{\mathrm{H}} = 0$; (Center) Polaron energy $E(k)$ for several coupling strengths: dotted curves are polaron energies for $\lambda < \lambda^{\mathrm{c}}$ with the energy minimum at $k = 0$ (black circles) and solid curves are for $\lambda > \lambda^{\mathrm{c}}$ with the ground state occurring at $k = \pm K_{\mathrm{GS}}$ (black cirlces). (Right) SSH polaron ground state momentum K_{GS} and effective polaron mass $1/m^*$ as functions of λ_{SSH}. The sharp transition occurs at $\lambda^{\mathrm{c}}_{\mathrm{SSH}} = 0.644$.

3.4 Results

We present the main results of this work as follows. First, we consider a pure SSH model and demonstrate ML models that extrapolate the ground-state momentum of the SSH polaron in phonon frequency. We build several ML models trained by the polaron energy dispersions in different regimes of adiabaticity. We then apply a similar analysis to the combined SSH - Holstein model. We obtain ML predictions of the sharp transition in the polaron ground-state momentum down to the extreme adiabatic limit. Finally, we demonstrate the application of multi-fidelity models to obtain accurate predictions of polaron properties from low-level GGCE calculations, thus reducing the overhead of computationally expensive numerical methods.

3.4.1 SSH polaron

We begin by training Gaussian process regression models with composite kernels by polaron energies $E(k,\Omega,\lambda_{\mathrm{SSH}})$ calculated using GGCE as described in Section 3.3. The resulting ML models predict the polaron energy for given values of the Hamiltonian parameters $\Omega, \lambda_{\mathrm{SSH}}$ and a given value of k. Our ML models can thus be viewed as surrogate models of GGCE calculations, with inputs given by three-dimensional vectors of $[\Omega, \lambda_{\mathrm{SSH}}, k]$ and outputs representing the polaron energy. In Section 3.4.4 below, we show how to extend these models to include the parameters N and M of the GGCE calculations on input, in order to

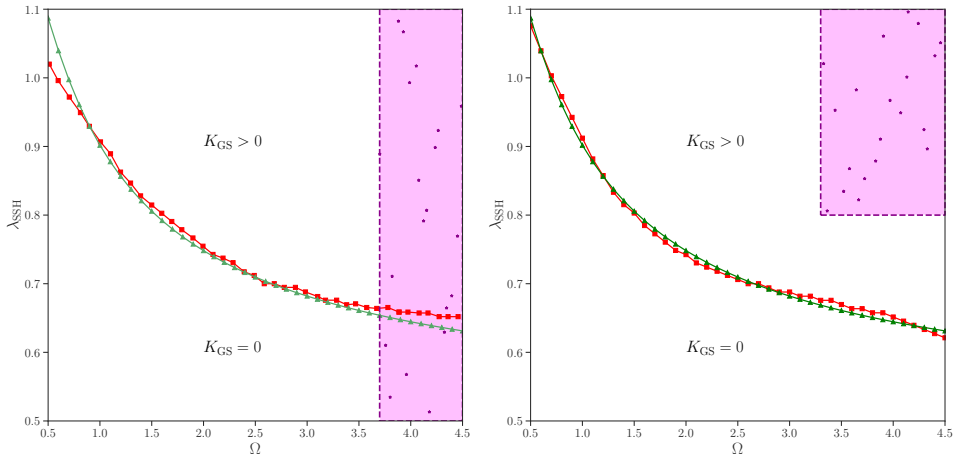


Figure 3.7: Critical coupling strength λ^c_{SSH} of the SSH polaron as a function of phonon frequency (Ω): squares – ML predictions of the sharp transition; triangles – location of the sharp transition from the GGCE calculations. The ML models are trained by 600 polaron energies randomly sampled from the three-dimensional space $[\Omega, \lambda_{SSH}, k]$ at $[\Omega, \lambda_{SSH}]$ indicated by the symbols in the shaded region.

improve the accuracy of predictions based on low-level GGCE calculations. In Ref. [59], the ML models were trained by full polaron dispersions at fixed values of phonon frequency and particle - phonon couplings. We note that we use a different approach in the present work. In particular, we do not discriminate between the input variables $[\Omega, \lambda_{SSH}, k]$. Instead, the polaron energies are computed at randomly selected combinations of the three variables $[\Omega, \lambda_{SSH}, k]$. We use a Latin Hypercube sampling strategy [123] to avoid accidental clustering of the training data and train ML models by energy points in the three-dimensional space thus sampled.

Figure 3.7 illustrates the ML predictions of K_{GS} for the SSH polaron as a function of Ω and λ_{SSH} in the range up to 1.1. Our particular focus is on the location of the sharp transition $K_{GS} = 0 \leftrightarrow K_{GS} > 0$ in the $[\lambda_{SSH}, \Omega]$ diagram. The models are trained by polaron energies computed for the Hamiltonian parameters in the anti-adiabatic limit $\Lambda > 1$ indicated by the symbols in the shaded region. The predictions are tested at different phonon frequencies, extending to the adiabatic regime. Figure 3.7 (left) illustrates ML extrapolation of the sharp transition in the frequency domain. It can be observed that ML models capture the evolution of the combination of λ_{SSH} and Ω that gives rise to the sharp transition for the SSH polaron down to low values of frequency $\Omega = 0.5$.

Figure 3.7 (right) demonstrates the accuracy of the ML predictions obtained by extrap-

olation in both Ω and $\lambda_{\rm SSH}$. As illustrated by the shaded region of Figure 3.7 (right), the polaron energies used for training ML models are entirely in the high-frequency, $K_{\rm GS} > 0$ region of the $[\lambda_{\rm SSH}, K_{\rm GS}]$ diagram, far removed from the location of the sharp transition. The evolution of the SSH polaron dispersion in the $[\lambda_{\rm SSH}, K_{\rm GS}]$ space is analogous to the evolution of free energy across second order phase transitions for some thermodynamic systems, such as the many-body spin system described by the Heisenberg model. Figure 3.7 (right) thus illustrates that ML models can be used to identify a phase transition by extrapolation of smooth functions from a single phase.

The accuracy of ML predictions obtained by extrapolation from a given phase is expected to improve as the training data approach the transition. This is illustrated in Figure 3.8, where the polaron energies used to train the ML models are calculated in the phonon frequency range $\Omega \in [1, 1.3]$. The circles in this figure are the predictions of the ML model represented by squares in Figure 3.7 (left).

3.4.2 Mixed Holstein – SSH Model

We now perform the same analysis as in the previous subsection but for a mixed coupling model given by Eq. (3.7). This model has been used for the description of excitations in molecular complexes and organic crystals [124–126]. It can be viewed as a generalization of both the SSH and Holstein models to include competition between phonon-induced interactions that modulate the potential and kinetic energy of the bare particle.

It was previously shown that the ground-state momentum of the polaron described by the mixed model also exhibits a sharp transition. However, the critical coupling strength for the generalized model is different. The dual coupling model gives rise to a critical coupling surface $\lambda^{\mathbf{c}} = [\lambda^{c}_{\rm SSH}, \lambda_{\rm H}]$, where $\lambda^{\mathbf{c}}$ is a function of Ω. We examine cuts through this surfaces that correspond to fixed Holstein coupling. For the case illustrated in Fig. 3.9, we fix $\lambda_{\rm H} = 0.5$. We train ML models using data at $\Omega = [3.7, 4.5]$ and predict the shift of this new critical coupling strength with the phonon frequency. Here, the polaron energies are calculated as in Ref. [119].

3.4.3 Predictions by extrapolation to extreme adiabatic regime

The ML models illustrated in the previous section are here used to predict the evolution of the sharp transition to the extreme adiabatic regime ($\Lambda \ll 1$). As Λ decreases, accurate calculations of the polaron energy become exceedingly difficult. As a result, $\Omega = 0.5$ is the lowest phonon frequency, for which polaron dispersions of the SSH polaron have been

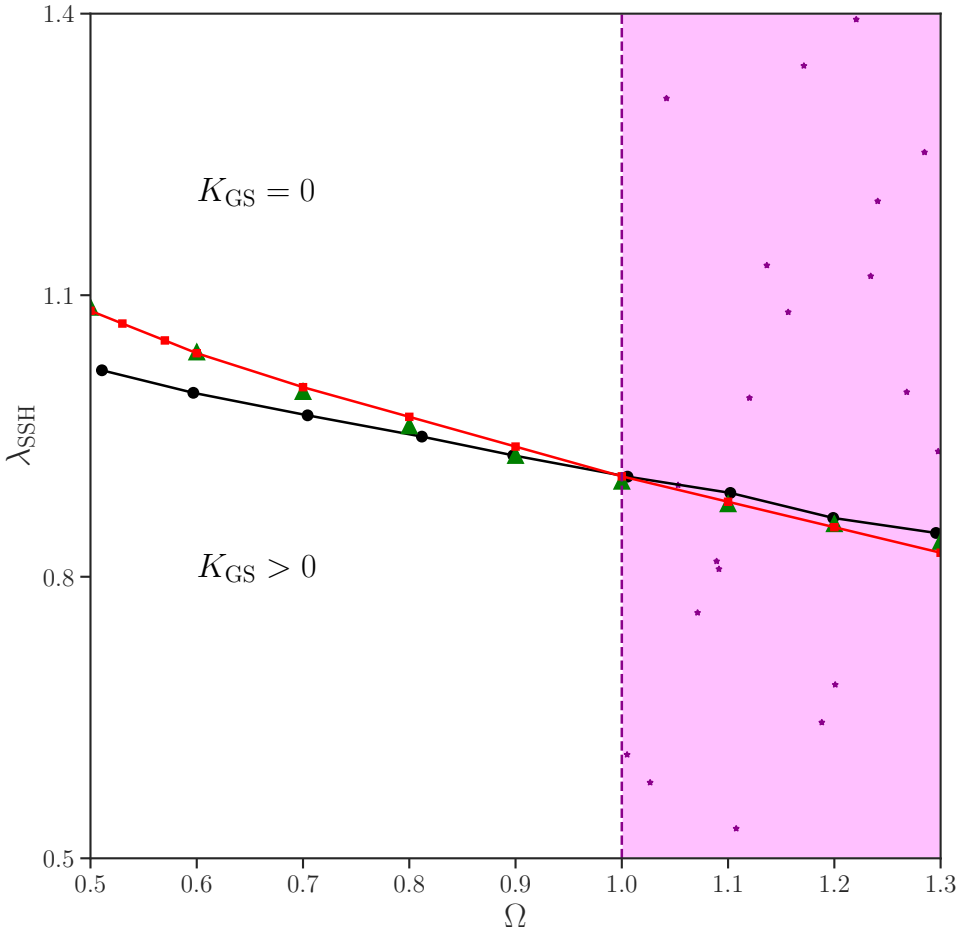


Figure 3.8: Critical coupling strength $\lambda^{c}_{\mathrm{SSH}}$ of the SSH polaron as a function of phonon frequency (Ω): squares – predictions of the sharp transitions by the ML models trained by polaron energies at phonon frequencies $\Omega \geq 1$; circles – predictions of ML models trained by polaron energies at phonon frequencies $\Omega \geq 3.5$; triangles – GGCE calculations. The ML models are trained by 415 polaron energies randomly sampled from the three-dimensional space $[\Omega, \lambda_{\mathrm{SSH}}, k]$. The training range for predictions represented by squares is shown by the shaded region.

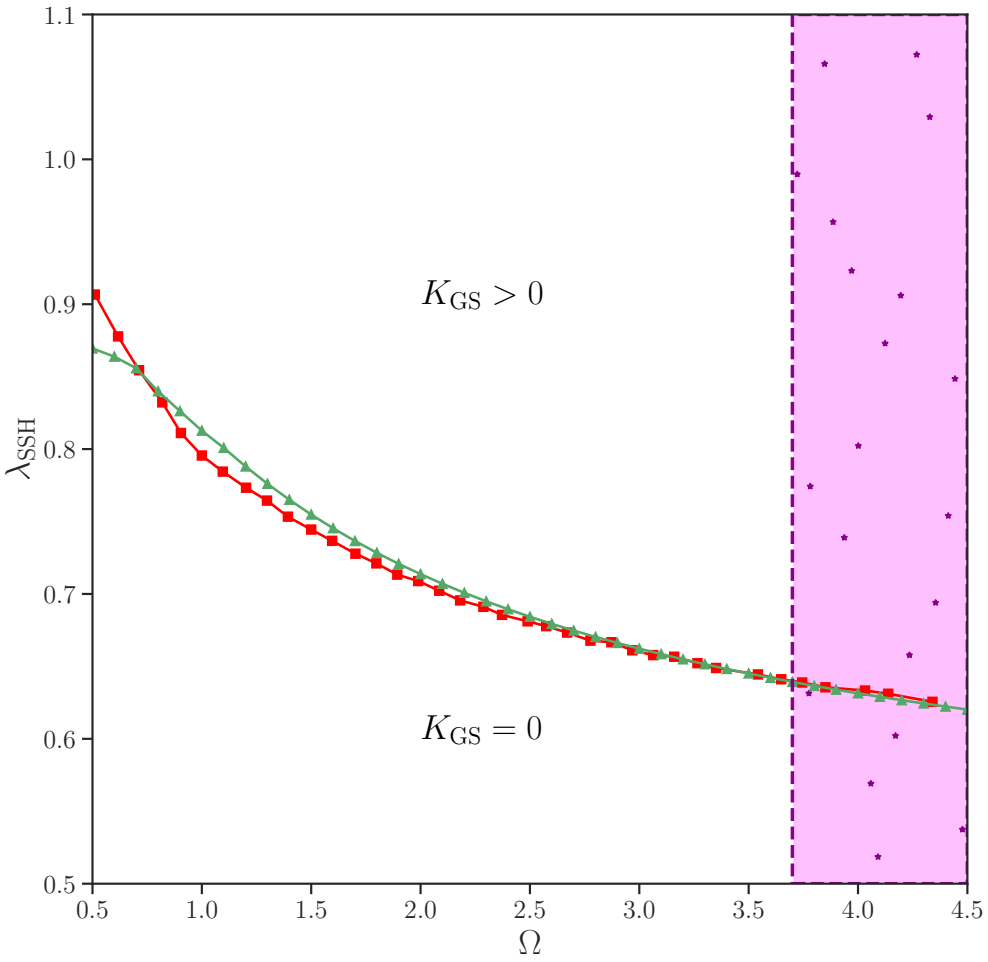


Figure 3.9: Critical coupling strength $\lambda^{c}_{\mathrm{SSH}}$ of the polaron for the mixed Holstein - SSH model as a function of phonon frequency (Ω): squares – ML predictions of the sharp transition; triangles – GGCE calculations. The ML models are trained by 600 polaron energies randomly sampled from the three-dimensional space $[\Omega, \lambda_{\mathrm{SSH}}, k]$ at $[\Omega, \lambda_{\mathrm{SSH}}]$ indicated by the symbols in the shaded region. The Holstein coupling strength is fixed to $\lambda_H = 0.5$.

reported to date in the literature [57]. Very recently, Grundner *et al.* performed DMRG calculations for a lattice with 256 sites and the SSH particle - phonon coupling for the phonon frequency 0.2 [127]. Interestingly, the results of their work suggest the absence of the sharp transition at this phonon frequency. We attempt to explore the persistence of the sharp transition in the SSH polaron diagram by extrapolation with ML to $\Omega \leq 0.4$. While large values of Ω are relevant for some materials, such as $SrTiO_3$ [128], the focus of the present work is on using the results at high Ω to make predictions of polaron properties at low values of Ω. Thus, polaron energies at large values of Ω should be viewed as easy-to-compute inputs into ML models.

First, we consider the case of $\Omega = 0.4$. Our models are trained by the polaron dispersions at $\Omega \in [0.5, 0.7]$ and $\lambda_{\mathrm{SSH}} \in [0.25, 1.7]$, down to the lowest phonon frequencies probed by the previous calculations, and predict the polaron dispersions and transition curve at $\Omega = 0.4$. The results reported in Fig. 3.10 show that ML models predict the expected trend for both the polaron dispersion and the evolution of the transition. The predicted value of $\lambda^{c}_{\mathrm{SSH}}$ at $\Omega = 0.4$ is 1.19.

To verify this prediction, we perform selected GGCE calculations with $M = 5$ and $N = 10$. The results are compared with the ML predictions in Fig. 3.10. The value of $\lambda^{c}_{\mathrm{SSH}}$ inferred from the GGCE calculations is between 1.1 and 1.2. A more accurate prediction would require a large number of GGCE calculations on a dense grid of $[\omega, k, \lambda_{\mathrm{SSH}}]$, which is currently out of reach of our computation resources.

Given that ML predictions are accurate for $\Omega = 0.4$, we now test the limits of ML models for predicting polaron dispersions in the extreme adiabatic regime. Previous work showed that K_{GS} undergoes the transition to $\pi/2$ for all shown frequencies above 0.5 [57]. However, the transition at the lowest considered phonon frequencies is predicted to occur at a very large coupling strength $\lambda_{\mathrm{SSH}} \sim 100$ [57]. Our results of Figure 3.10 show that the transition may happen at much lower values of λ_{SSH}.

We include the GGCE results in the training set for another set of ML models. In the process of training these ML models, we have found that including data from a narrower range of frequencies, just above the prediction range, leads to better quantitative predictions. This happens because the resulting models are less biased by the different physics at high phonon frequencies. Therefore, we produce a new set of ML models by training with the polaron energies in the phonon frequency range $\Omega \in [0.4, 0.7]$. We predict the polaron dispersions at $\Omega = [0.4, 0.3, 0.2, 0.1, 0.01]$ as shown in Figs. 3.11 and 3.12.

These predictions are in the phonon frequency regime that is currently out of reach of rigorous quantum calculations. They can therefore not be independently tested. However,

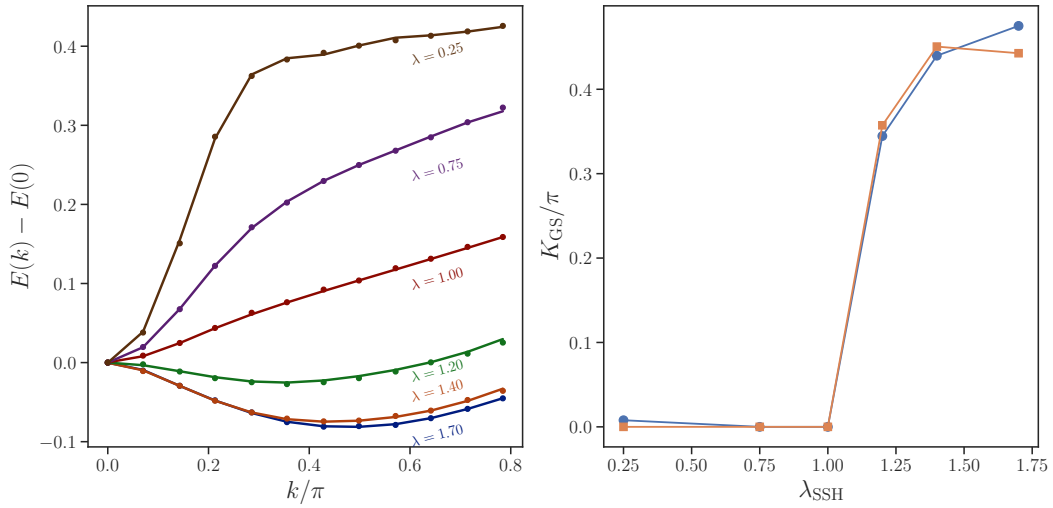


Figure 3.10: (Left) SSH polaron dispersions for $\Omega = 0.4$: the symbols represent the results of the GGCE calculations with $M = 5, N = 10$; the solid lines are the ML predictions. The ML model is trained by the polaron energies at $\Omega \in [0.5, 0.7]$ and $\lambda_{SSH} \in [0.25, 1.7]$. (Right) The sharp $K_{GS} = 0 \leftrightarrow K_{GS} > 0$ transition observed in the ground state momentum at $\Omega = 0.4$: circles – the ML predictions; squares – the GGCE calculations.

there are some general physical features that one can verify to confirm the validity of the ML predictions. As described in Ref. [57], the polaron ground state is determined by the lowest peak in the spectral function at a particular value of momentum k. The onset of the continuum happens at Ω above the ground-state energy. For Einstein phonons, a state with energy $E' > E(0) + \Omega$ can therefore dissociate into a ground state polaron + one phonon, so a state with such energy cannot be bound. Thus, the energy of a stable polaron must be lower than this bound. We observe that the dispersion curves predcited by ML for $\Omega = 0.4$ and $\Omega = 0.3$ satisfy $|E(k) - E(0)| < \Omega$. This condition is also approximately satisfied for $\Omega = 0.2$. We further observe that the dispersion curves predicted for $\Omega < 0.2$ do not follow this expected behaviour.

Another way to confirm the validity of ML predictions is by training multiple models with different distributions of training data (e.g., data from different regions of the $[\lambda, \Omega]$ diagram). All predictions in Figs. 3.11 and 3.12 are qualitatively robust to variations in training data. Thus, we attach a high level of confidence to the ML predictions for $\Omega \geq 0.3$. The dispersion curves predicted by ML at $\Omega \geq 0.3$ exhibit the sharp $K_{GS} = 0 \leftrightarrow K_{GS} > 0$ transition. The ML predictions at $\Omega < 0.3$ cannot be regarded as reliable. Interestingly, our

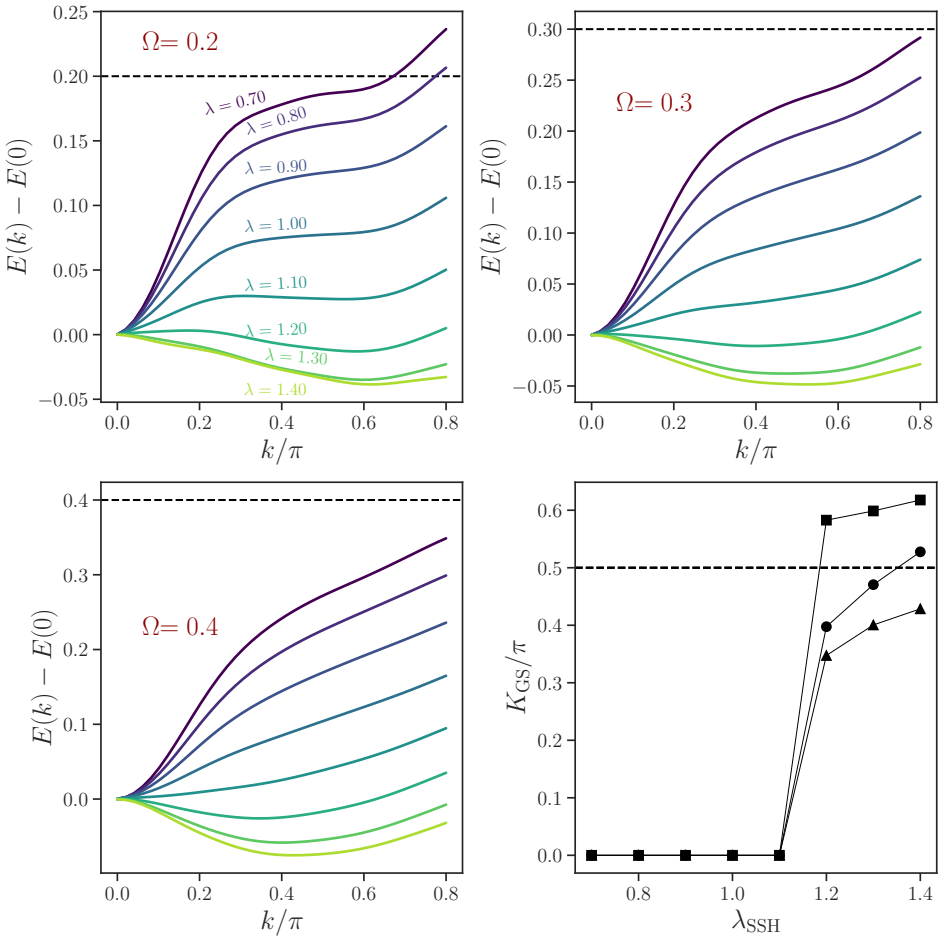


Figure 3.11: Extrapolation of the polaron dispersion relation E in phonon frequency space (Ω) for the pure SSH model. The dispersion curves are predicted by a ML model trained by polaron energies at phonon frequencies $\Omega \in [0.4, 0.9]$. The solid lines are the predictions of the model for various coupling strengths. The horizontal dashed lines show the limit of the dispersion bandwidth equal to Ω. Lower right: The dependence of K_{GS} of the SSH polaron on λ_{SSH} at three phonon frequencies corresponding to the dispersion curves shown in the other panels. The symbols represent different values of Ω: triangles for $\Omega = 0.4$, circles for $\Omega = 0.3$, and squares for $\Omega = 0.2$.

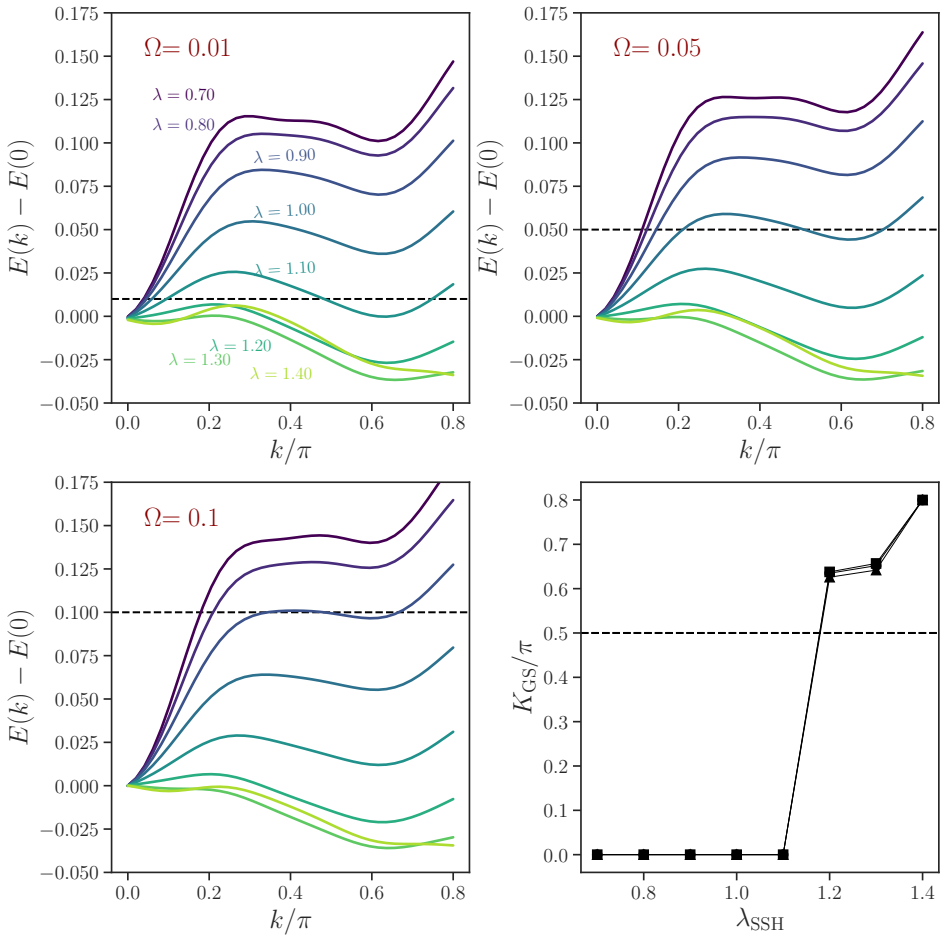


Figure 3.12: Extrapolation of the polaron dispersion relation E in phonon frequency space (Ω) for the pure SSH model. The dispersion curves are predicted by a ML model trained by polaron energies at phonon frequencies $\Omega \in [0.4, 0.9]$. The solid lines are the predictions of the model for various coupling strengths. The horizontal dashed lines show the limit of the dispersion bandwidth equal to Ω. Lower right: The dependence of K_{GS} of the SSH polaron on λ_{SSH} at three phonon frequencies corresponding to the dispersion curves shown in the other panels. The symbols represent different values of Ω: triangles for $\Omega = 0.1$, circles for $\Omega = 0.05$, and squares for $\Omega = 0.01$.

results indicate a qualitative change in the ML predictions at the phonon frequency around 0.3. The breakdown of ML predictions below $\Omega = 0.3$ may be indicative of a qualitative change in polaron physics. This is consistent with the DMRG results of Grundner *et al.* [127] at $\Omega = 0.2$.

3.4.4 Interpolation in variational space

In this final section, we demonstrate multi-fidelity models that can be used to enhance the efficiency and accuracy of the GGCE calculations described in Section 3.3. As mentioned previously, we generalize the inputs to our machine learning models to include N and M of the GGCE calculations, and the outputs to describe the results of quantum calculations at different levels of theory. One can view the resulting ML models as surrogate models of the polaron energy in the space of $[\Omega, \lambda_{\mathrm{SSH}}, k]$ as well as the parameters of the GGCE calculations $[N, M]$.

We chose Eq. (2.13) as our kernel function in the NARGP scheme. NARGP models are first trained on 100 points sampled across the $[\lambda_{\mathrm{SSH}}, k]$ grid, with the polaron dispersions calculated using $(M,N) = (2,4)$, far smaller values than required for convergence. We limit our calculations to $k \leq 0.8\ \pi$. The reasons for this choice are twofold. First, for high momentum states $k > 0.8\ \pi$, the peaks corresponding to ground state energy in the spectral function tend to be small. This poses a challenge for accurate peak detection algorithms, potentially leading to less precise data for high momentum states compared to the well-defined peaks observed for low-lying energy eigenstates. Second, the sharp transition is entirely captured by $k \leq 0.8\ \pi$. With this set of parameters, the full diagram of the polaron energies in $[\lambda_{\mathrm{SSH}}, k]$ can be computed on an 8-core iMac with 16 GB RAM in less than ten minutes. The poor accuracy of this fidelity level can be seen in the green sharp transition curve in Fig. 3.13, as the critical coupling strength $\lambda^{\mathrm{c}}_{\mathrm{SSH}}$ at this level is shifted to 1.21 from the exact value $\lambda^{\mathrm{c}}_{\mathrm{SSH}} = 1.117$. The poor accuracy of this calculation is also evident from the departure of green symbols (squares) from the fully converged calculations represented by red circles in Fig. 3.13. However, GP models of these approximate calculations can serve as a very good prior into models of the more accurate results.

We show this by subsequently sampling 15 points from the high-fidelity model with the polaron energies calculated using $(M,N) = (3,9)$, which correspond to fully converged calculations. With this choice of parameters, full GGCE calculations of the entire diagram would require approximately 10 hours to generate the illustration shown in Fig. 3.13 on the same computing device. The difficulty of this calculation quickly scales up with $M \to M+1, N \to N+1$ and becomes intractable at $\sim (10,8)$ for most supercomputers [60, 118]. The

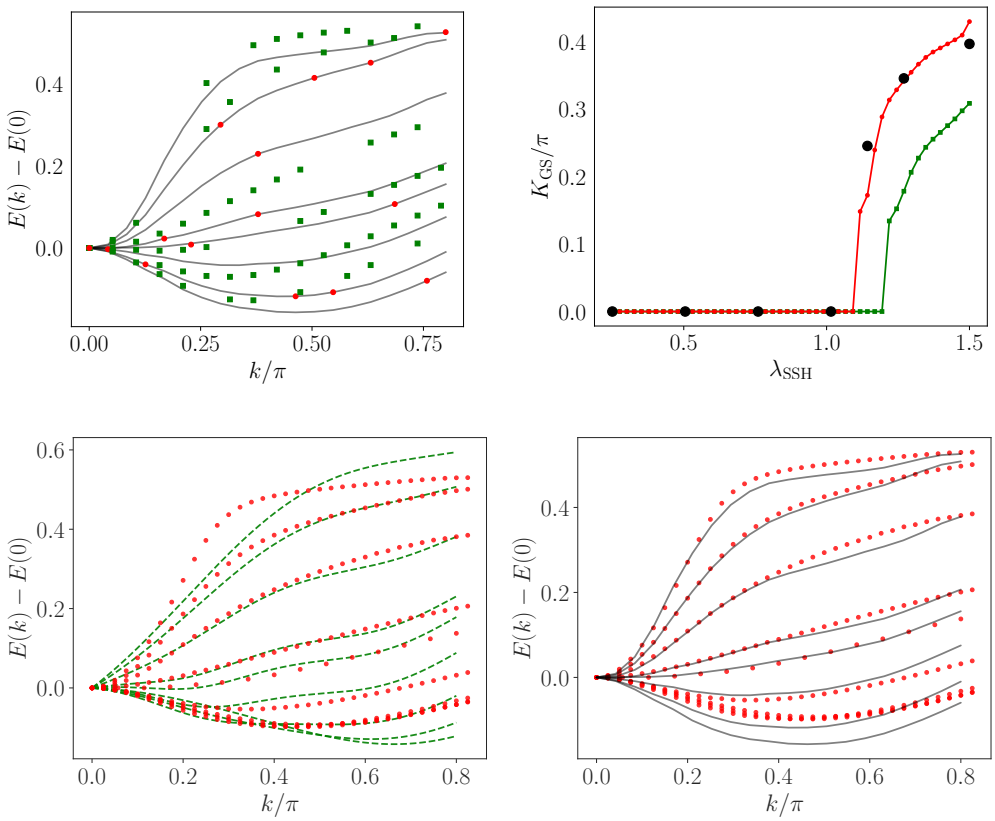


Figure 3.13: Top Left: SSH polaron dispersions: for $\Omega = 0.5$, the green squares (red circles) show the total of 100 (15) points sampled from the lower (higher) fidelity quantum calculations and the curves are the NARGP model predictions trained on this combination of points. The circles represent expensive quantum calculations. The solid lines are evaluated for $\lambda_{SSH} = [0.25, 0.5, 0.75, 1.0, 1.1, 1.25, 1.5, 1.75]$ and 20 uniformly distributed points $k \in [0, 0.8\,\pi]$. Top Right: Ground-state momentum K_{GS} of the SSH polaron as a function of λ_{SSH}. The red circles and the green squares are the NARGP model predictions for high fidelity data with the sharp transition at $\lambda^c_{SSH} = 1.117$ and low fidelity data with the transition at $\lambda^c_{SSH} = 1.121$, respectively. The black circles are the GGCE calculations presented for reference. Bottom Left: Red circles – rigorous calculations (reference); broken curves – predictions of ML models trained directly by 15 high fidelity calculations Bottom Right: Red circles – rigorous calculations (same as in left); solid curves – NARGP model predictions. The difference between the dashed green curves and the solid red curves represents the improvement of accuracy due to combination of low-fidelity and high-fidelity data.

results of Figure 3.13 show that these calculations can be combined to produce predictions for the polaron energies at the high level of theory with a small fraction of the computation cost.

The improvement of the prediction accuracy by the inclusion of cheap, low-level calculations is seen in the comparison of the lower panels in Figure 3.13. The left panel illustrates the ML predictions from direct interpolation of the small number of exact polaron energies, whereas the right panel illustrates the predictions of multi-fidelity NARGP models. The computation time to produce the dispersion curves in both panels is very similar, as it is determined by the computation of the 15 polaron energies at the $(M,N) = (3,9)$ level.

3.5 Discussion and Outlook

The present work demonstrates that it is possible to obtain accurate predictions of polaron properties at low phonon frequencies by machine learning models trained on polaron properties at high phonon frequencies. This represents an example of machine learning for a quantum problem, where data from an effectively lower-dimensional Hilbert space are used to make predictions for an effectively higher-dimensional Hilbert space. There are numerous examples in quantum theory, where this strategy can be used to extend the range of quantum predictions beyond the limitations of numerical calculations. In the present work, we consider the sharp transition in the ground-state momentum of the SSH polaron and examine the evolution of this transition from the anti-adiabatic regime to the adiabatic regime. Whether the sharp transition observed in the anti-adiabatic regime of high phonon frequencies occurs at reasonable electron-phonon coupling strength in the adiabatic limit has been debated in the recent literature. All of the ML models in the present work predict the sharp transition, even at very low phonon frequencies.

Extrapolation, by definition, cannot be rigorously tested, until new results become available, whether from theory or experiment. Our conclusions rely on three results. First, we test the prediction accuracy of our ML models in the range of phonon frequencies accessible by quantum calculations. The calculations for this range of phonons show that Bayesian ML models proposed here can produce quantitative predictions by extrapolation in the phonon frequency domain.

For the range of phonon frequencies outside of available rigorous results, we discuss the validity of our ML predictions based on two criteria: criterion (i) – ML predictions must satisfy basic physical principles; criterion (ii) – models trained with different distributions of data must produce consistent predictions. The predictions of the sharp transition for the

SSH polaron at $\Omega > 0.3$ presented in this work are robust to changes of the training data distributions and should thus be viewed as qualitatively, if not quantitatively, correct. This is best exemplified by the two panels of Figure 2 that present ML predictions with training data sampled from two very different parts of the Hamiltonian parameter space.

We observe that ML predictions of polaron dispersions yield energies within the expected band limits for phonon frequencies just beyond the range of the training data. However, as the prediction range of phonon frequencies is reduced to below $\Omega = 0.3$, polaron dispersions predicted by ML become unphysical and exceed the allowed energy range. This signals the onset of the break-down of the ML predictions. This may be the result of the qualitative change in polaron physics, as suggested by the work in Ref. [127]. While the present article was in preparation, the authors of Ref. [127] reported the DMRG calculations of the SSH bipolaron energy for a lattice of 256 sites with phonon frequency $\Omega = 0.2$. The authors find that the sharp transition does not occur at this phonon frequency in the range of considered values of $\lambda_{\mathrm{SSH}} \in [0, 1.5]$. Their results suggest the formation of a weakly bound bipolaron or unstable bipolaron-like state. This suggests departure from the trends observed in Refs. [57, 102, 129, 130], where stable bipolarons exist at strong coupling.

We have also demonstrated Bayesian models that use the posterior distributions of Gaussian processes based on highly approximate quantum calculations as the prior distribution for models of more accurate quantum results. This drastically reduces the number of fully converged calculations required to map out the full dispersion relations for the full range of Hamiltonian parameters of interest. For example, we have demonstrated that full dispersion relations for the SSH polarons in the range of electron - phonon couplings from $\lambda_{\mathrm{SSH}} = 0.25$ to 1.75 producing the correct evolution of the sharp transition in polaron ground-state momentum for $\Omega = 0.5$ can be obtained with only 15 polaron energy calculations. This strategy can be employed to build accurate surrogate models of quantum results for problems with a hierarchy of approximate methods, where the number of required quantum calculations decreases with the numerical complexity of the method. The loss of accuracy is fully controlled and can potentially be reduced to negligible errors by increasing the number of calculations as well as by aligning the kernels of the underlying models with the most accurate results.

Chapter 4

Equivalence between Exponential Concentration and Barren Plateaus

Machine learning with quantum kernels [39, 41, 45, 131, 132] and variational quantum algorithms (VQA) [133] are two distinct paradigms currently explored for computation with noisy quantum devices [12, 13, 32, 33, 35]. VQAs are restricted by barren plateaus (BPs) [46, 134]. BPs are regions in the variational parameter space where gradients of the cost function decrease exponentially with the number of qubits, leading to vanishingly narrow gorges of optimal solutions. This implies that an exponentially large number of measurements is required to estimate cost functions for VQAs with BP. The origin of BPs can be traced to vanishing of inner products between quantum states in exponentially large Hilbert spaces. Exponential vanishing of inner products also hinders applications of quantum computing for estimating kernels for machine learning models [39, 40]. In kernel methods of machine learning, models are based on kernel matrices [40], which in quantum machine learning (QML) are estimated by measurements on a quantum computer [41, 131, 132]. Refs. [47, 55] show that both the expected value and variance of the off-diagonal elements of quantum kernel matrices decrease exponentially as the number of qubits used for building quantum kernels grows. This leads to exponential concentration (EC) of the kernel matrix around a diagonal matrix, which implies that quantum kernels reduce to δ functions in a large Hilbert space and make machine learning impractical.

It has been assumed that BPs and EC of quantum kernels are different manifestations of the same problem [47, 134–138]. However, to the best of our knowledge, no rigorous proof of the connection, or limitations thereof, between BPs in VQA and EC in quantum kernels has been presented. While training a machine learning model with parametrized kernels

can be formulated directly as a variational problem [132, 139], it is not clear whether the equivalence between BPs and EC remains for learning with parameter-free kernels. Here, we formalize the relation between BPs in VQA and EC of quantum kernels by transferring training bounds derived for VQA to ML with quantum kernels.

Since BPs preclude the quantum advantage of VQA [46], a significant focus of recent work has been on how to engineer quantum circuits for VQA without BP. It has been shown that BPs can be avoided by restricting the depth of quantum circuits [140], by building symmetries of the data into the quantum ansatz [141], as well by employing non-random intialization strategies for variational optimization [48–52]. Previous work on how to prevent EC of quantum kernels is much more limited. The proposed approaches include models based on covariant quantum kernels [142], quantum kernel bandwidth methods [53], quantum Fisher kernels [54], and projected quantum kernel methods [143]. All these methods encode inductive bias into the quantum models, thereby restricting learning to a particular subspace of the Hilbert space. A rigorously proven relation between BPs in VQA and quantum kernels in machine learning is required in order to apply the results from the BP literature to quantum machine learning.

Schuld *et al.* [45] established connections between quantum kernel methods and VQAs. Here, we use a similar approach to demonstrate how training bounds derived for BPs specific to a particular circuit ansatz can, under specific conditions, be extended to overcome the EC of quantum kernels. This is achieved by analyzing the upper and lower bounds on the cost function of a quantum neural network and constructing an equivalent quantum kernel using the variational ansatz. We prove that the trainability bounds of this newly constructed kernel can be related to the variational ansatz used in the VQA framework. In particular, this proves that if an ansatz leads to a BP-free VQA [144], then there exists an equivalent quantum kernel that does not suffer from EC.

4.1 Quantum Machine learning models

For a supervised learning task, the dataset is given by $\mathscr{D} := (\mathbf{x}_i, \mathbf{y}_i)_{i=1}^{N_s}$, where $\mathbf{x}_i \in \mathscr{X}$ are input vectors, $\mathbf{y}_i \in \mathscr{Y}$ are the labels, and N_s is the number of data points. The inputs are related to the labels via a black box function $f : \mathscr{X} \to \mathscr{Y}$, and the task is to determine a QML model h_θ parameterized by θ such that $h_\theta : \mathscr{X} \to \mathscr{Y}$ approximates f. We introduce a distance metric ℓ and use that to define $\widehat{\mathscr{L}_\theta}$ called the empirical loss function, which quantifies how close h_θ is to the true function f given that we have access to the dataset $\mathscr{D}$.

We can obtain the optimal parameters of the model by

$$\widehat{\mathscr{L}_\theta} := \frac{1}{N_s}\sum_{i=1}^{N_s} \ell(h_\theta(\mathbf{x}_i), \mathbf{y}_i) \tag{4.1}$$

$$\theta_{\text{opt}} := \underset{\theta}{\text{argmin}}\ \mathscr{L}_\theta \tag{4.2}$$

Given the fact that we only have access to N_s datapoints, in reality, we only approximate the true loss function, which is defined over the probability distribution of our data $P(\mathscr{X})$, defined as

$$\mathscr{L}_\theta = \mathbb{E}_{x_i \sim P(x)}[\ell(h_\theta(\mathbf{x}_i), \mathbf{y}_i)] \tag{4.3}$$

If we assume that N_s datapoints were sampled independently and identically from $P(\mathscr{X})$, we can say that $\widehat{\mathscr{L}_\theta}$ is a true and an unbiased estimator of true loss $\mathscr{L}_\theta$. The difference between the empirical loss and the true loss is defined as the generalization error of the model given by:

$$\text{gen}(\theta) = |\mathscr{L}_\theta - \widehat{\mathscr{L}_\theta}| \tag{4.4}$$

4.1.1 Quantum Kernels

In this thesis, we will concern ourselves with a specific subset of algorithms known as *kernel methods*. Here, a kernel function k maps our input data lying in a vector space $\mathscr{V}$ to a higher dimensional feature space $\mathscr{V}'$ such that the classification/regression problem being solved becomes easier [145]. One might think that calculations in higher dimensional space would only add to the complexity of the problem, but this is not the case. The *kernel trick* makes sure that the explicit calculations of feature vectors are never performed in higher dimensional space, it's only their inner products that enter the calculation which are evaluated using the kernel function κ [145]. Mathematically we can describe the kernel as a function $\kappa : \mathscr{V}' \times \mathscr{V}' \to \mathbb{R}$ such that

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) \equiv \left\langle \phi(\mathbf{x_i}), \phi(\mathbf{x_j}) \right\rangle, \tag{4.5}$$

where $\phi : \mathscr{V} \to \mathscr{V}'$ is map connecting the two vector spaces and $\mathbf{x}_i, \mathbf{x}_j$ are input data points living in the $\mathscr{V}$ space. The brackets $\langle \cdot, \cdot \rangle$ specify the inner product between the two vectors. The kernel function satisfies the Mercer condition [146] such that it is always positive semi-definite i.e. $\sum_{i=1}^{N}\sum_{j=1}^{N} \kappa(\mathbf{x}_i, \mathbf{x}_j) c_i c_j \geq 0 \quad \forall c_{i,j}$ where $\{c_{i,j}\}$ is a set of N real numbers.

In classical machine learning, we know that Support Vector Machines (SVMs) [147] can efficiently train a linear model using the inner products of feature vectors. Training occurs in the higher-dimensional feature space, although feature vectors are never evaluated themselves, but only their inner product at two data points $\mathbf{x}$ and $\mathbf{x}'$. Kernel theory utilizes relation in functional analysis and convex optimization to make statements about training and generalization error of linear models in feature space. Quantum computation enters the picture through the feature map implemented by applying unitary circuits such that the input x, is mapped from the dataspace to the feature space:

$$x \to |\phi(x)\rangle\langle\phi(x)| \tag{4.6}$$

The feature space for quantum machine learning models is the exponentially large space of density matrices $\mathbb{S}(2^n)$ spanned by n qubits. To access this space, a kernel function known as the fidelity quantum kernel is used. This kernel function is defined using a fidelity-based measure and is constructed with respect to the Hilbert-Schmidt inner product of matrices.

$$\kappa(\mathbf{x},\mathbf{x}') = |\langle\phi(\mathbf{x})|\phi(\mathbf{x}')\rangle|^2 = |\langle 0^{\otimes n} \mid U^\dagger(\mathbf{x})U(\mathbf{x}') \mid 0^{\otimes n}\rangle|^2 = \mathrm{Tr}(\rho(\mathbf{x})\rho(\mathbf{x}')) \tag{4.7}$$

$$\rho(\mathbf{x}) = U(\mathbf{x})\rho_0 U^\dagger(\mathbf{x}) \tag{4.8}$$

where $\rho(\mathbf{x})$ and $\rho(\mathbf{x}')$ are the density matrices corresponding to the quantum states represented by the feature vectors $\mathbf{x}$ and $\mathbf{x}'$, respectively. Throughout the thesis, we use the notation $\rho_0 = |0\rangle^{\otimes n}\langle 0|^{\otimes n}$ to define our initial state. The feature map as represented in Eq. (4.6) can be implemented through a class of quantum circuits $U(\mathbf{x})$ defined by the input vector $\mathbf{x}$. The data embedding $U(\mathbf{x})$ acts as a generator of a unitary ensemble, defined over all possible input data vectors $x_i \in \mathscr{X}$. From a functional analysis perspective, $U : \mathscr{X} \to \mathbb{U}_x \subseteq \mathscr{U}(2^n)$ where $\mathscr{U}(2^n)$ is the total space of unitaries, and $\mathbb{U}_x = \{U(\mathbf{x})|\mathbf{x} \in \mathscr{X}\}$.

Also, it is worth considering the Reproducing Kernel Hilbert Space (RKHS) defined by our feature map $U(x)$ [145]. The feature vectors $\phi(x) = k(x,.)$ form a basis in RKHS and we can reproduce any function in RKHS as a linear combination of these basis functions. We will show later that these functions are identical to linear quantum models in the same feature space. The Representer theorem uses this property of RKHS to show that, we can write a function f_{opt} which minimizes the loss landscape, which is convex with respect to

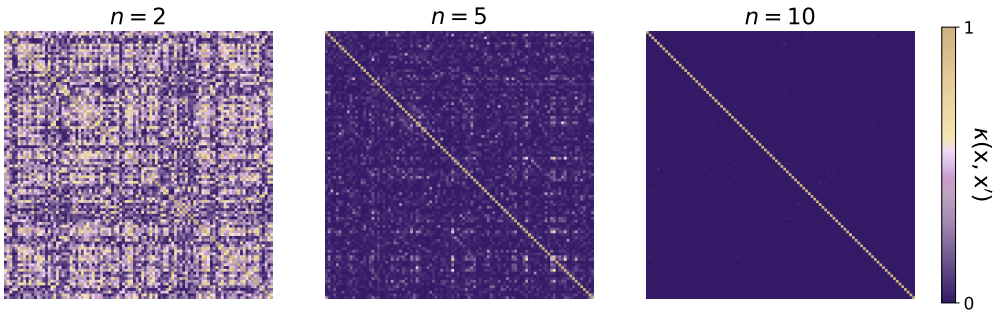


Figure 4.1: Exponential concentration of a quantum kernel: comparison of the kernel matrices constructed using an ansatz proposed in Havelick *et al.* [40] for binary classification on MNIST dataset [148]. The matrix is constructed for 100 data points with the number of dimensions reduced to the corresponding number of qubits by principal component analysis. Each dimension is encoded into a separate, one qubit. The kernel matrices shown are for qubit numbers $n = [2, 5, 10]$.

parameters α_i having the following form [45]:

$$f_{\text{opt}}(x) = \sum_{m=1}^{N_s} \alpha_m \,\text{Tr}[\rho(x_m)\rho(x)] = \text{Tr}\left[\left(\sum_{m=1}^{N_s} \alpha_m \rho(x_m)\right)\rho(x)\right] \tag{4.9}$$

Exponential Concentration

Kernel methods are favored in situations with limited data because the *representer theorem* ensures a convex training landscape, which means that the optimization problem has a single globally optimal solution that can be efficiently found. Support vector machine, kernel ridge regression would fall into this category. (Note: optimization of the log marginal likelihood for a gaussian process is not convex). But in the case of the quantum kernel, kernel matrix elements are only estimated by finite sampling from the Born distribution. This allows statistical uncertainty to exacerbate the predictions. Moreover, suppose that the feature vectors are too spread out in the exponentially large Hilbert space. In that case, the cost of taking the overlap (inner product for kernels) between the two states exponentially increases in the number of qubits, and all the kernel values start to concentrate about the mean. This phenomenon is known as the exponential concentration (EC) of the kernel values or the vanishing similarity issue in quantum kernels [47]. Formally, this is defined as:

Definition 4.1. (Deterministic) Deterministic exponential concentration of quantum kernel $\kappa(\mathbf{x}, \mathbf{x}')$ in number of qubits n is defined as

$$
\begin{aligned}
&| \kappa(\mathbf{x}, \mathbf{x}') - \mu | \leq \beta, \quad \beta \in O(1/b^n) \\
&\mu = \mathbb{E}_{\mathbf{x},\mathbf{x}'}[\kappa(\mathbf{x}, \mathbf{x}')]
\end{aligned}
$$

for some $b > 1$ and all pairs of $(\mathbf{x}, \mathbf{x}')$.

Definition 4.2. (Probabilistic) Probabilistic concentration of quantum kernel $\kappa(\mathbf{x}, \mathbf{x}')$ is

$$
\Pr_{\mathbf{x},\mathbf{x}'}[| \kappa(\mathbf{x}, \mathbf{x}') - \mu | \geq \delta] \leq \frac{\beta^2}{\delta^2}, \quad \beta \in O(1/b^n) \tag{4.10}
$$

Here we have made use of Chebyshev's inequality:

$$
\Pr\big(|\kappa(\mathbf{x}, \mathbf{x}') - \mu| \geq \delta\big) \leq \frac{\mathrm{Var}_{\mathbf{x},\mathbf{x}'}[\kappa(\mathbf{x}, \mathbf{x}')]}{\delta^2} \tag{4.11}
$$

EC can also be observed in the eigenspectrum of the kernel matrix. The framework proposed in Ref. [149] considers the number of samples necessary for a model to learn a target function. The learning requirement for each eigenmode of the kernel is determined by its associated eigenvalue. Consequently, a "flat" spectrum indicates a requirement of an exponential amount of data to learn each eigenmode, which leads to a high generalization error. A numerical demonstration of exponential concentration is shown in Fig. 4.1. In summary, the vanishing similarity problem associated with the fidelity-based quantum kernel as defined in Eq. (4.7) indicates that both the expected value and the variance of the QK decrease at an exponential rate as the number of qubits increases. Therefore, the estimation of each kernel matrix element on an actual quantum device necessitates a number of measurement shots that increase exponentially with the number of qubits as well. In other words, as the quantity of qubits becomes very large, the Kernel matrix approaches the identity matrix. This leads to overfitting and subpar generalization performance.

Expressibility of a quantum circuit

We know that unitaries are norm-preserving operations, hence taking expectation over all input elements $\mathbf{x} \in \mathscr{X}$ should be the same as average over the ensemble of data-encoded unitaries $\mathbb{U}_x$, or the data-encoded states defined as $\mathbb{S}_x = U(\mathbf{x}_i)\rho_0 U^\dagger(\mathbf{x}_i)$. We can estimate the expressibility of an ensemble $\mathbb{U}_x$ by quantifying how close it is to a t-design. A set of unitary circuits is said to follow a t-design when the statistical properties of those unitaries

are the same as those of truly random Haar unitaries up to the t^{th} moment. More concretely, we can define the expressibility of ensemble $\mathbb{U}_x$ as $\varepsilon_{\mathbb{U}_x}$ for a given input state ρ_0 given by [138]:

$$\mathscr{A}_{\mathbb{U}_x}(\cdot) = \int_{\mathscr{U}(2^n)} dV V^{\otimes 2}(\cdot)(V^\dagger)^{\otimes 2} - \int_{\mathbb{U}_x} dU U^{\otimes 2}(\cdot)(U^\dagger)^{\otimes 2} \tag{4.12}$$

$$\varepsilon_{\mathbb{U}_x} := \|\mathscr{A}_{\mathbb{U}_x}(\rho_0)\|_1 \tag{4.13}$$

where we take a trace norm in the last equation. Here, we are considering circuits that follow a 2-design. In the integral, the first term is a sum over the Haar ensemble and will be denoted by $\mathscr{V}_{Haar}(\cdot)$ and the second term is integral over our ensemble $\mathbb{U}_x$. If $\varepsilon_{\mathbb{U}_x} = 0$ that means our encoding is maximally expressive.

4.1.2 Quantum Neural Networks

Quantum Neural Networks (QNN) are a family of parameterized circuits that have been used for regression and classification tasks in quantum machine learning [43, 44]. As shown in Fig. 4.2, they can be defined using a data-based embedding $U_{\mathbf{x}}$ for $\mathbf{x} \in \mathscr{X}$ followed by variational embedding $W(\boldsymbol{\theta})$ with m parameters $\boldsymbol{\theta} \in [-\pi, \pi]^m$. The most general cost function for QNNs can be defined as follows:

$$\begin{aligned} \mathscr{C}(\boldsymbol{\theta}, \mathscr{X}) &= \sum_{i=1}^{N_s} c_i \mathrm{Tr}[W(\boldsymbol{\theta})\rho(\mathbf{x}_i)W^\dagger(\boldsymbol{\theta})O] \\ &= \sum_{i=1}^{N_s} c_i \mathrm{Tr}[W(\boldsymbol{\theta})U(\mathbf{x}_i)\rho_0 U^\dagger(\mathbf{x}_i)W^\dagger(\boldsymbol{\theta})O] = \sum_{i=1}^{N_s} c_i\, C(\boldsymbol{\theta}, \mathbf{x}_i) \end{aligned} \tag{4.14}$$

where O is a measurement operator. In this work, O is a global measurement operator that acts on all qubits in the circuit to relate the cost function to the quantum state overlaps serving as the kernel function.

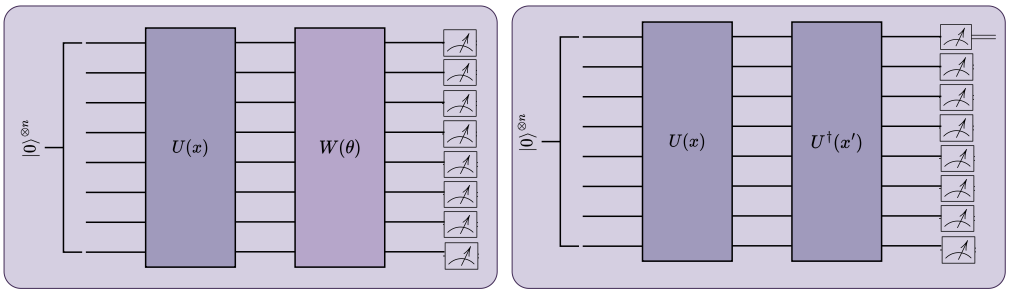


Figure 4.2: Quantum circuit for QNN (left) and a quantum kernel (right). The final output is the expectation value of $|0\rangle^{\otimes^n}\langle 0|^{\otimes^n}$.

Barren Plateaus

Intuitively, one can understand barren plateaus as the phenomenon where the average value of the cost function gradient with respect to variational parameters of the QNN is exponentially suppressed. Mathematically, we can formulate the problem in terms of the variance and means of the partial derivatives of the cost function.

Definition 4.3. (Barren Plateaus). The cost function defined for QNN in Eq. (4.14) exhibits a BP if the variance of the derivative of the cost function with respect to $\theta_\mu \in \boldsymbol{\theta}$ is exponentially decaying in the number of qubits:

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta}, \mathbf{x}_i)] \leq F(n), \quad F(n) \in O(1/b^n) \tag{4.15}$$

We can prove the mean of the derivative of the cost function i.e. $\mathbb{E}_\theta[\partial_\mu C(\theta)]$ to be equal to zero, since the variational space is periodic and the integral of $\partial_\mu C(\theta)$ must vanish along a closed loop. BPs were first introduced in Ref. [46], where the source of exponential suppression is the high expressivity of PQCs, hence they are also called randomness-induced Barren Plateaus. This affects both gradient-free [150] and gradient-based [151] optimization approaches. Concretely, their argument is that if we randomly initialize the θ for a given PQC $W(\theta)$ such that it forms a 2-design [46, 152, 153], then QNN will suffer from a BP. Cerezo *et al.* [154] also studied BPs in shallow circuits and concluded that PQC with global measurement-based cost functions (O acting on all qubits) can suffer from BPs regardless of depth. The reason behind this suppression is that a global cost function compares two states in an exponentially large Hilbert space which always leads to an exponentially small overlap given the number of shots available is only $poly(n)$. These kinds of BPs can be mitigated by using local cost functions and shallow-depth circuits. However, the quan-

tum advantage obtained from such models is questionable. The phenomenon of BPs has also been extended to machine learning models based on tensor networks [155] and PQCs with high amounts of entanglement [137]. A recent review [134] summarizes all the major developements in BPs over the past few years. We can demonstrate BPs through a simple example considered in [154]. Consider an ansatz of the form $W(\boldsymbol{\theta}) = \bigotimes_{j=1}^{n} e^{-i\theta^j \sigma_x^{(j)}/2}$ and we want to find the optimal parameters θ_j^* such that $W(\boldsymbol{\theta})|\mathbf{0}\rangle = |\mathbf{0}\rangle$. Here the global measurement operator takes the form $O = \mathbb{I} - |\mathbf{0}\rangle\ \langle\mathbf{0}|$, such that cost function C has a nice analytical form $C = 1 - \prod_{j=1}^{n} \cos^2\frac{\theta^j}{2}$. One can easily calculate the variance of the cost function which comes out to be $\mathrm{Var}[\frac{\partial C}{\partial \theta^j}] = \frac{1}{8}(\frac{3}{8})^{n-1}$, which exponentially goes to zero as we increase the number of qubits n. This result is illustrated in Fig. 4.3, showing a two-dimensional slice of the cost function's multi-dimensional parameter space for qubit counts of $n = 3$ and $n = 20$. The apparent randomness in the figure arises from finite sampling of the parameter space, though the curve is theoretically smooth.

Cost concentration

The phenomenon of narrow gorges is adjacent to the Barren Plateaus. Narrow gorges occur in a cost function landscape when the volumetric fraction of the parameter space below a certain value is exponentially suppressed. Rigorously we can define the idea of cost concentration as [156]:

Definition 4.4. (Cost Concentration). The cost function defined for QNN in Eq (4.14) will exhibit a narrow gorge if there exists a local minimum lower by the mean cost by at least $\delta(n) > 0$, with $\delta(n) \in \Omega(1/poly(n))$. The probability that the cost function value deviates from its mean by at least δ at a randomly chosen point from a uniform distribution over the parameter space is bounded, for some $b > 1$ and θ_A which is a random draw from the uniform probability distribution over the parameter space.

$$\mathrm{Var}_{\boldsymbol{\theta}_A}[\mathbb{E}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta}, \mathbf{x}_i)] - C(\boldsymbol{\theta}_A, \mathbf{x}_i)] \leq G(n), \quad G(n) \in O(1/b^n), \quad b > 1. \tag{4.16}$$

where $\boldsymbol{\theta}_A$ is a random draw from the uniform probability distribution over the parameter space. Furthermore, the cost function defined for QNNs in Eq. (4.14) exhibits a narrow gorge if there exists a local minimum lower than the mean cost by at least $\delta(n) > 0$ with $\delta(n) \in \Omega(1/\mathrm{poly}(n))$. The probability that the cost function value deviates from its mean by at least δ at a randomly chosen point from a uniform distribution over the parameter

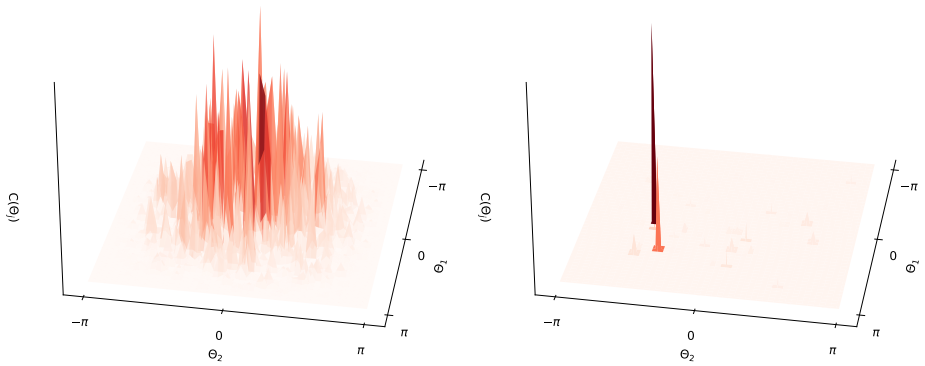


Figure 4.3: Cost Concentration around the mean: Two-dimensional cross-section through the landscape of (negative of) cost function $C = 1 - \prod_{j=1}^{n} \cos^2 \frac{\theta^j}{2}$ for $n = 3$ (Left) and $n = 20$ (Right), where n remains the number of qubits. The cost function landscape flattens as we increase the number of qubits, with only a few minima appearing which require an exponential amount of resources to traverse.

space is bounded and

$$\Pr_{\boldsymbol{\theta}}[|\, \mathbb{E}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta}, \mathbf{x}_i)] - C(\boldsymbol{\theta}_A, \mathbf{x}_i)\, | \geq \delta] \leq \frac{G(n)}{\delta^2}, \quad G(n) \in O(1/b^n), \quad b > 1 \tag{4.17}$$

Importantly, these two formulations of narrow gorges and cost concentration occur simultaneously, and, moreover, if a parametrized quantum circuit exhibits a BP then there exists a cost concentration (and narrow gorges) satisfying the above condition and vice versa, as proven by Arrasmith *et al.* [156]. Hence, instead of considering cost function gradients, one can examine concentration of Var[ΔC].

4.2 Relationship between BP and EC

In what follows, we relate the exponential concentration problem in un-parameterized (no variational parameters) quantum kernels with the barren plateaus problem that impedes the training of quantum neural networks. For the case of the parameterized quantum kernels, there exists an equivalent cost function namely, kernel target alignment [42]. It was shown in Ref. [47], that the exponential upper bounds on the variance of the QNN cost function can be transferred to parameterized quantum kernels directly. This relationship is straightforward since barren plateaus is a phenomenon observed in the variational space, and kernel target alignment depends on variational parameters. Here, we want to verify whether the quantum ansatz proposed to mitigate barren plateaus can be used in a quantum kernel setting. Secondly, we show that an ansatz suffering from barren plateaus also suffers from exponential concentration if we were to build an equivalent kernel out of it. We begin with a QNN with a BP. We consider the cost function given by Eq. (4.14) with fixed $\mathbf{x}_i$. The cost concentration for such an ansatz is given by

$$\forall \mathbf{x}_i \in \mathscr{X}: \quad \mathrm{Var}_{\boldsymbol{\theta}_A}[\mu_{\mathbf{x}_i} - C(\boldsymbol{\theta}_A|\mathbf{x}_i)] \leq F(n), \quad F(n) \in \mathscr{O}\left(\frac{1}{b^n}\right) \tag{4.18}$$

where $\mu_{\mathbf{x}_i} = \mathbb{E}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta}|\mathbf{x}_i)]$. We construct a kernel from the variational part of the circuit $W(\boldsymbol{\theta})$ with the data embedding $\rho(\mathbf{x}_i) = W(\mathbf{x}_i)\rho_0 W^\dagger(\mathbf{x}_i)$, where $\mathbf{x}_i \in \mathscr{X}$. The dataset $\mathscr{X}$ is assumed to be normalized and standardized to the range $[-\pi, \pi]$, which is also the range of $\boldsymbol{\theta}$. This leads to the kernel

$$\kappa(\mathbf{x}, \mathbf{x}') = \mathrm{Tr}[W(\mathbf{x})\rho_0 W^\dagger(\mathbf{x}) W(\mathbf{x}')\rho_0 W^\dagger(\mathbf{x}')] \tag{4.19}$$

We next relate the bounds as determined by Eq. (4.18) to EC of the kernel as defined in Eq. (4.19). First, we choose the measurement operator O to be ρ_0. Second, we notice that each column in a quantum kernel matrix (corresponding to the datapoint $\mathbf{x}_j$) is a cost function of QNN considered above, concentrating around $\mu_{\mathbf{x}_j}$. A transition from QNN to a quantum kernel setting requires $\mu_{\mathbf{x}_i} = \mu \ \forall \mathbf{x}_i$. We note that the mean of each column in the kernel matrix can be written as $\mu_{\mathbf{x}_i} = \frac{1}{N_s} + \varepsilon(\mathbf{x}_i)$, where the first term comes from the diagonal element being one, and the second term decays exponentially in the number of qubits, i.e. $\varepsilon \sim O(1/c^n)$ for $c > 1$. Hence, the mean of the kernel matrix is constant up to ε. This

allows us to write Eq. (4.18) as

$$\mathrm{Var}_{\boldsymbol{\theta}_A}[C(\boldsymbol{\theta}_A|\mathbf{x}_i)] \leq F(n) \tag{4.20}$$

where we use the property of the variance $\mathrm{Var}[f(x)+c] = \mathrm{Var}[f(x)]$, where $c = \text{const}$, and include ε into the error of $F(n)$.

We then consider a BP-free QNN, whose cost function is bounded as follows [144]:

$$\mathrm{Var}_{\boldsymbol{\theta}_A}[C(\boldsymbol{\theta}_A|\mathbf{x}_i)] \geq G(n), \quad G(n) \in \Omega\left(\frac{1}{\mathrm{poly}(n)}\right) \tag{4.21}$$

Given the exponential (polynomial) upper (lower) bound on the cost function of the QNN, we relate these bounds to EC of the corresponding kernels. To do this, we need to prove the following lemmas.

Lemma 4.1. If there exists an upper bound under variation of each single random variable $\mathbf{x}_j$,

$$\forall \mathbf{x}_i: \quad \mathrm{Var}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i, \mathbf{x}_j\right) \mid \mathbf{x}_i\right] \leq F(n)$$

and the additional condition holds

$$\mathrm{Var}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i, \mathbf{x}_j\right)|\mathbf{x}_i\right]\right] \leq \bar{F}(n),$$

then this upper bound holds for the total variance: $\mathrm{Var}_{\mathbf{x}_i,\mathbf{x}_j}\left[f\left(\mathbf{x}_i, \mathbf{x}_j\right)\right] \leq F(n) + \bar{F}(n)$

Proof.

$$\begin{aligned}
\mathrm{Var}_{\mathbf{x}_i,\mathbf{x}_j}\left[f\left(\mathbf{x}_i, \mathbf{x}_j\right)\right] &= \mathbb{E}_{\mathbf{x}_i}\left[\mathrm{Var}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i, \mathbf{x}_j\right)|\mathbf{x}_i\right]\right] + \mathrm{Var}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i, \mathbf{x}_j\right)|\mathbf{x}_i\right]\right] && (4.22)\\
&\leq \mathbb{E}_{\mathbf{x}_i}[F(n)] + \bar{F}(n) && (4.23)\\
&= F(n) + \bar{F}(n) && (4.24)
\end{aligned}$$

where independence between $\mathbf{x}_i$ and $\mathbf{x}_j$ is used to apply the law of total variance, and the additional condition is substituted. □

Lemma 4.2. If there exists a lower bound under variation of each single variable $\mathbf{x}_j$,

$$\forall \mathbf{x}_i: \quad \mathrm{Var}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i, \mathbf{x}_j\right) \mid \mathbf{x}_i\right] \geq G(n)$$

then this lower bound holds for the total variance:

$$\mathrm{Var}_{\mathbf{x}_i,\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right)\right] \geq G(n)$$

Proof.

$$\begin{aligned}
\mathrm{Var}_{\mathbf{x}_i,\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right)\right] &= \mathbb{E}_{\mathbf{x}_i,\mathbf{x}_j}\left[f\left(\mathbf{x}_i,,\mathbf{x}_j\right)^2\right] - \mathbb{E}_{\mathbf{x}_i,\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right)\right]^2 && (4.25)\\
&= \mathbb{E}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right)^2 \mid \mathbf{x}_i\right]\right] - \mathbb{E}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right) \mid \mathbf{x}_i\right]\right]^2 && (4.26)\\
&\geq \mathbb{E}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right)^2 \mid \mathbf{x}_i\right]\right] - \mathbb{E}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right) \mid \mathbf{x}_i\right]^2\right] && (4.27)\\
&= \mathbb{E}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right)^2 \mid \mathbf{x}_i\right] - \mathbb{E}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i,,\mathbf{x}_j\right) \mid \mathbf{x}_i\right]^2\right] && (4.28)\\
&= \mathbb{E}_{\mathbf{x}_i}\left[\mathrm{Var}_{\mathbf{x}_j}\left[f\left(\mathbf{x}_i,\mathbf{x}_j\right) \mid \mathbf{x}_i\right]\right] && (4.29)\\
&\geq \mathbb{E}_{\mathbf{x}_i}[G(n)] = G(n). && (4.30)
\end{aligned}$$

The following properties are used to derive Lemma 4.2. The law of total expectation, because of the independence between $\mathbf{x}_i$ and $\mathbf{x}_j$, is used in step 2. Jensen's inequality, which states that the convex transformation of a mean is less than or equal to the mean applied after convex transformation, is applied in step 3 as $x \mapsto x^2$ is a convex function. We also use linearity of expectation. □

Theorem 4.1. Given a QNN as defined in Eq. (4.14) that follows either an exponential upper bound (BP landscape)[1] or polynomial lower bound (BP-free landscape), there exists a quantum kernel as defined by Eq. (4.19) such that the variance bounds from QNN apply to the quantum kernel.
Formally, for the QNN (Eq. (4.14)) exhibiting a BP landscape, i.e.,

$$\exists F(n) \forall \mathbf{x}\colon \quad \mathrm{Var}_{\boldsymbol{\theta}_A}[C(\boldsymbol{\theta}_A,\mathbf{x})] \leq F(n) \quad \text{with} \quad F(n) \in \mathcal{O}\left(\frac{1}{b^n}\right),\ b>1 \tag{4.31}$$

and an additionally exponentially concentrated mean cost under varying input states

$$\mathrm{Var}_{\mathbf{x}}\left[\mathbb{E}_{\boldsymbol{\theta}}\left[f\left(\mathbf{x},\boldsymbol{\theta}\right)|\mathbf{x}\right]\right] \leq \bar{F}(n) \quad \text{with} \quad \bar{F}(n) \in \mathcal{O}\left(\tfrac{1}{\bar{b}^n}\right),\ \bar{b}>1, \tag{4.32}$$

[1] Assuming an exponentially concentrated mean cost under varying input states, i.e., $\mathrm{Var}_{\mathbf{x}}\left[\mathbb{E}_{\boldsymbol{\theta}}\left[f\left(\mathbf{x},\boldsymbol{\theta}\right)|\mathbf{x}\right]\right] \in \mathcal{O}\left(\frac{1}{\bar{b}^n}\right)$.

the corresponding quantum kernel (Eq. (4.19)) concentrates exponentially

$$\mathrm{Var}_{\mathbf{x},\mathbf{x}'}[\kappa(\mathbf{x},\mathbf{x}')] \leq F(n) + \bar{F}(n). \tag{4.33}$$

Vice versa, if the QNN cost function (Eq. (4.14)) is BP-free, i.e.,

$$\exists F(n) \forall \mathbf{x}: \quad \mathrm{Var}_{\boldsymbol{\theta}_A}[C(\boldsymbol{\theta}_A,\mathbf{x})] \geq G(n) \quad \text{with} \quad G(n) \in \Omega\left(\frac{1}{\mathrm{poly}(n)}\right), \tag{4.34}$$

the corresponding quantum kernel (Eq. (4.19)) does *not* concentrate exponentially

$$\mathrm{Var}_{\mathbf{x},\mathbf{x}'}[\kappa(\mathbf{x},\mathbf{x}')] \geq G(n). \tag{4.35}$$

Proof. The theorem assumes our construction introduced in the main text in Sec. 4.2, summarized in the following conditions:

- same ansatz is chosen for the variational part of the QNN as for data embedding in the quantum kernel $U(\boldsymbol{x}) \leftrightarrow W^\dagger(\boldsymbol{x})$
- $O = \rho_0$

Under these conditions, Eq. (4.19) applies, i.e., we recover the quantum kernel formulation from the QNN.

Upon substitution $\boldsymbol{\theta} \to \mathbf{x}_j$, Eq. (4.20) yields

$$\forall \mathbf{x}_i: \quad \mathrm{Var}_{\mathbf{x}_j}\left[\kappa(\mathbf{x}_i,\mathbf{x}_j) \mid \mathbf{x}_i\right] \in \mathscr{O}\left(\frac{1}{b^n}\right), \quad b > 1 \tag{4.36}$$

and a concentration of the mean kernel value

$$\mathrm{Var}_{\mathbf{x}_i}\left[\mathbb{E}_{\mathbf{x}_j}\left[\kappa(\mathbf{x}_i,\mathbf{x}_j)\,|\mathbf{x}_i\right]\right] \in \mathscr{O}\left(\tfrac{1}{\bar{b}^n}\right), \quad \bar{b} > 1. \tag{4.37}$$

Conversely, if no BPs exist using Eq. (4.21) this implies a lower bound on the variance

$$\mathbf{x}_i: \quad \mathrm{Var}_{\mathbf{x}_j}\left[\kappa(\mathbf{x}_i,\mathbf{x}_j) \mid \mathbf{x}_i\right] \in \Omega\left(\frac{1}{\mathrm{poly}(n)}\right). \tag{4.38}$$

It directly follows from Lemmas 4.1 and 4.2 that these bounds also apply to the total vari-

ance

$$\mathrm{Var}_{\mathbf{x}_j,\mathbf{x}_i}\left[\kappa\left(\mathbf{x}_i,\mathbf{x}_j\right)\right] \in \mathscr{O}\left(\frac{1}{b^n}\right) \quad \text{with} \quad b>1, \tag{4.39}$$

$$\mathrm{Var}_{\mathbf{x}_j,\mathbf{x}_i}\left[\kappa\left(\mathbf{x}_i,\mathbf{x}_j\right)\right] \in \Omega\left(\frac{1}{\mathrm{poly}(n)}\right), \tag{4.40}$$

, which concludes the proof. □

We also note that each column in a quantum kernel matrix (corresponding to the data point $\mathbf{x}_j$) is a cost function of the QNN considered above, concentrated around $\mu_{\mathbf{x}_j}$. A transition from QNN to a quantum kernel setting requires $\mu_{\mathbf{x}_i} = \mu \ \forall \mathbf{x}_i$. This is generally not the case in the kernel setting. However, we note that the mean of each column in the kernel matrix can be written as $\mu_{\mathbf{x}_i} = \frac{1}{N_s} + \varepsilon(\mathbf{x}_i)$, where the first term comes from the diagonal element being one, and the second term decays exponentially in the number of qubits, i.e. $\varepsilon \sim O(1/c^n)$ for $c > 1$. Hence, the mean of the kernel matrix is constant up to ε. This allows us to write Eq. (4.18) as

$$\mathrm{Var}_{\boldsymbol{\theta}_A}[C(\boldsymbol{\theta}_A, \mathbf{x}_i)] \leq F(n) \tag{4.41}$$

where we use the property of the variance $\mathrm{Var}[f(x)+c] = \mathrm{Var}[f(x)]$, where $c = \text{const}$, and include ε into the error of $F(n)$.

To illustrate the results of Theorem 4.1, we consider the BP-free ansatz proposed in Ref. [144] which is provably BP-free. We use Eq. (4.19) to construct a quantum kernel with this ansatz and evaluate the kernel matrices for the MNIST dataset considered above in Fig. 4.1. Figure 4.4 shows the kernel matrices thus obtained for different numbers of qubits $n \in [5, 20]$. The kernel matrices in Fig. 4.4 exhibit no concentration, while the off-diagonal elements of the kernel matrices in Fig. 4.1 vanish for $n \leq 10$. To provide a quantitative comparison between the results of Figs. (4.1) and (4.4), we compute the variance of $\kappa(\mathbf{x}_i, \mathbf{x}_j)$. As shown in the right panel of Figure 4.4, $\mathrm{Var}[\kappa(\mathbf{x}_i, \mathbf{x}_j)]$ computed with the BP-free ansatz remains relatively constant as the number of qubit increases, while the ansatz used for the results in Fig. 4.1 leads to an exponential collapse of the variance magnitude.

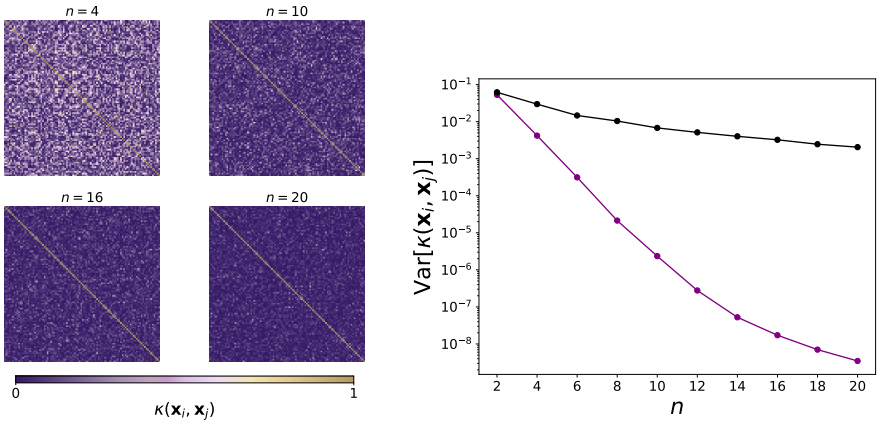


Figure 4.4: Kernel matrices computed with the BP-free ansatz [144] for binary classification on the MNIST dataset [148]. The matrices are constructed for 100 data points with the number of dimensions reduced to the corresponding number of qubits (n) by principal component analysis. Each dimension is encoded into a single independent qubit. (Right) Variances of the kernel matrix elements computed with the BP-free ansatz (black) as a function of number of qubits in the range $n \in [2, 20]$. The depth of the circuit is equal to the number of qubits at each point. The purple line corresponds to the Havlíček *et al.* [40] considered in Fig. 4.1. The depth of the circuit is kept constant.

4.3 Invariance as an inductive bias

If we know that our dataset is symmetric under group $\mathbf{G}$, we can use this information to encode good inductive biases into our quantum kernels [55, 141, 157]. Glick *et al.* [158] introduced a category of feature map circuits known as covariant feature maps, specifically designed for data spaces characterized by a group structure. These quantum feature states are closely connected to covariant measurements [159]. The covariant feature map is defined relative to a unitary representation $D_{\mathbf{x}}$ for the group $\mathbf{G}$ with $\mathbf{x} \in \mathbf{G}$ and a fiducial state $|\psi\rangle \in \mathbb{C}^{2^n}$ on n qubits as $\Phi(x) = D_x|\psi\rangle\langle\psi|D_x^\dagger$. The covariant quantum kernel is then estimated as the fidelity as defined in Ref. [142].

$$K(x,x') = |\langle\psi|D_{x'}^\dagger D_x|\psi\rangle|^2 \tag{4.42}$$

We assume that we have access to a good fiducial state $|\psi\rangle = V|0\rangle$ which will be informed by the choice of dataset or can be variationally prepared by considering a family of fiducial states $|\psi\rangle_\lambda = V_\lambda|0\rangle$. We can find the optimum parameters λ using kernel target alignment or other optimization methods, although we should be mindful that doing this puts us back

in the realm of training problems in variational landscapes. We can see that the kernel function is invariant to the action of any group element $\mathbf{g} \in \mathbf{G}$, and since the kernel function enters our loss function, that means our loss function is also variant to the action of $\mathbf{g}$. Since D is the representation of the group, we have $D_{\mathbf{gx}} = D_{\mathbf{g}} D_{\mathbf{x}}$ and hence

$$\begin{aligned} K(\mathbf{g}x, \mathbf{g}x') &= |\langle \psi | D^{\dagger}_{\mathbf{g}x'} D_{\mathbf{g}x} | \psi \rangle|^2 \\ &= |\langle \psi | D^{\dagger}_{x'} D^{\dagger}_{\mathbf{g}} D_{\mathbf{g}} D_x | \psi \rangle|^2 \\ &= |\langle \psi | D^{\dagger}_{x'} D_x | \psi \rangle|^2 = K(x, x') \end{aligned} \tag{4.43}$$

Another way to build invariance into our model is through the idea of *equivariant* embedding as defined in Ref. [160]. The idea is similar to what is described previously. We design an embedding $U(x)$ such that $U(\mathbf{gx}) = U(\mathbf{g})U(x)U^{\dagger}(\mathbf{g})$ for all $\mathbf{g} \in \mathbf{G}$. We assume our initial fiducial state is chosen such that it is invariant to the action of $U(\mathbf{g})$ such that $U(\mathbf{g})|\psi\rangle = |\psi\rangle$ since $U(\mathbf{g})$ is unitary this implies that state is invariant under its adjoint as well. This leads to:

$$\begin{aligned} K(\mathbf{g}x, \mathbf{g}x') &= |\langle \psi | U^{\dagger}(\mathbf{g}x') U(\mathbf{g}x) | \psi \rangle|^2 \\ &= |\langle \psi | U(\mathbf{g}) U^{\dagger}(x') U^{\dagger}(\mathbf{g}) U(\mathbf{g}) U(x) U^{\dagger}(\mathbf{g}) | \psi \rangle|^2 \\ &= |\langle \psi | U^{\dagger}(x') U(x) | \psi \rangle|^2 = K(x, x') \end{aligned} \tag{4.44}$$

4.3.1 A new bound for covariant quantum kernels

The integration over the manifold $\int_D$ can be factorized into a product of integrals over the Haar measures of the group of unitaries $U(d_\lambda)$ associated with each invariant subspace. This allows us to perform the integration separately for each subspace, which greatly simplifies the calculation of the integral over $\int_D$.

$$\int_D dU f(U(\theta)) = \prod_{\lambda} \int_{U(d_\lambda)} d\mu(U) f(U(\theta))$$

If D is unitary representation of the group G, then the matrix of D is block diagonal since D can be written as direct sum over all irreducible representations (irreps) $D^{\mathscr{J}}$ such that $D_{\mathbf{g}} \simeq \bigoplus_{\mathscr{J}} D^{\mathscr{J}}_{\mathbf{g}}$. Based on this property we can derive a new bound for a covariant kernel with an appropriate fiducial state. The new bound demonstrates that the kernel does concentrate however the rate of concentration is much smaller due to the symmetry encoded into our quantum kernel, which serves as a good inductive bias.

Theorem 4.2. *Consider a covariant quantum kernel as defined in Eq. (4.42). Assume that the input data* $\boldsymbol{x}$ *and* $\boldsymbol{x}'$ *are drawn from the same distribution, such that it is symmetric under the action of a group* $\mathbf{G}$*. The data embedding* U *can be written as* $U(\mathbf{g}) \simeq \bigoplus_{\lambda} U_\lambda(\mathbf{g})$ *such that* d_λ *is the dimension of the Hilbert space corresponding to each irrep* λ*. We also assume that each isotypic block* U_λ *forms a 2-design. Then we have*

$$\mathrm{Var}_{\boldsymbol{x},\boldsymbol{x}'}\left[K\left(\boldsymbol{x},\boldsymbol{x}'\right)\right] \leqslant \sum_{\lambda} \frac{2}{d_\lambda\left(d_\lambda+1\right)} \tag{4.45}$$

Proof. We define our fiducial state density matrix ρ_f as $\rho_f = V\rho_0 V^\dagger$. Now we can derive the Eq. (4.45) by considering a covariant kernel $K(\boldsymbol{x},\boldsymbol{x}')$ as described by Glick *et al.* [142].

$$\begin{aligned}
\mathrm{Var}_{\boldsymbol{x},\boldsymbol{x}'}\left[K\left(\boldsymbol{x},\boldsymbol{x}'\right)\right] &\leqslant \mathbb{E}_{\boldsymbol{x},\boldsymbol{x}'}\left[\left(K\left(\boldsymbol{x},\boldsymbol{x}'\right)\right)^2\right] && (4.46)\\
&\leqslant \int d\boldsymbol{x}\int d\boldsymbol{x}'\,\mathrm{Tr}\left[U(\boldsymbol{x})\rho_f U^\dagger(\boldsymbol{x})U\left(\boldsymbol{x}'\right)\rho_f U^\dagger\left(\boldsymbol{x}'\right)\right]\\
&\quad\times\mathrm{Tr}\left[U(\boldsymbol{x})\rho_f U^\dagger(\boldsymbol{x})U\left(\boldsymbol{x}'\right)\rho_f U^\dagger\left(\boldsymbol{x}'\right)\right] && (4.47)\\
&\leqslant \int dU(\boldsymbol{x})\int dU(\boldsymbol{x}')\,\mathrm{Tr}\left[U(\boldsymbol{x})\rho_f U^\dagger(\boldsymbol{x})U\left(\boldsymbol{x}'\right)\rho_f U^\dagger\left(\boldsymbol{x}'\right)\right]\\
&\quad\times\mathrm{Tr}\left[U(\boldsymbol{x})\rho_f U^\dagger(\boldsymbol{x})U\left(\boldsymbol{x}'\right)\rho_f U^\dagger\left(\boldsymbol{x}'\right)\right] && (4.48)\\
&\leqslant \sum_\lambda\prod_{\lambda,\lambda'}\int_{\mathscr{U}(d_\lambda)} d\mu(U_\lambda(\boldsymbol{x}))\int_{\mathscr{U}(d_{\lambda'})} d\mu(U_{\lambda'}(\boldsymbol{x}'))\\
&\quad\times\mathrm{Tr}\left[U_\lambda(\boldsymbol{x})\rho_f U_\lambda^\dagger(\boldsymbol{x})U_\lambda\left(\boldsymbol{x}'\right)\rho_f U_\lambda^\dagger\left(\boldsymbol{x}'\right)\right]\\
&\quad\times\mathrm{Tr}\left[U_{\lambda'}(\boldsymbol{x})\rho_f U_{\lambda'}^\dagger(\boldsymbol{x})U_{\lambda'}\left(\boldsymbol{x}'\right)\rho_f U_{\lambda'}^\dagger\left(\boldsymbol{x}'\right)\right] && (4.49)\\
&\leqslant \sum_\lambda\int_{\mathscr{U}(d_\lambda)} d\mu(U_\lambda(\boldsymbol{x}))\int_{\mathscr{U}(d_\lambda)} d\mu(U_\lambda(\boldsymbol{x}'))\\
&\quad\times\mathrm{Tr}\left[U_\lambda(\boldsymbol{x})\rho_f U_\lambda^\dagger(\boldsymbol{x})U_\lambda\left(\boldsymbol{x}'\right)\rho_f U_\lambda^\dagger\left(\boldsymbol{x}'\right)\right]\\
&\quad\times\mathrm{Tr}\left[U_\lambda(\boldsymbol{x})\rho_f U_\lambda^\dagger(\boldsymbol{x})U_\lambda\left(\boldsymbol{x}'\right)\rho_f U_\lambda^\dagger\left(\boldsymbol{x}'\right)\right] && (4.50)\\
&\leqslant \sum_\lambda\int_{\mathscr{U}(d_\lambda)} d\mu(U_\lambda(\boldsymbol{x}))\int_{\mathscr{U}(d_\lambda)} d\mu(U_\lambda(\boldsymbol{x}'))\\
&\quad\times\mathrm{Tr}\left[(U_\lambda(\boldsymbol{x}))^{\otimes 2}\rho_f^{\otimes 2}\left(U_\lambda^\dagger(\boldsymbol{x})\right)^{\otimes 2}\left(U_\lambda\left(\boldsymbol{x}'\right)\right)^{\otimes 2}\rho_f^{\otimes 2}\left(U_\lambda^\dagger\left(\boldsymbol{x}'\right)\right)^{\otimes 2}\right] && (4.51)\\
&\leqslant \sum_\lambda\mathrm{Tr}\left[\int_{\mathscr{U}(d_\lambda)} d\mu(U_\lambda(\boldsymbol{x}))(U_\lambda(\boldsymbol{x}))^{\otimes 2}\rho_f^{\otimes 2}\left(U_\lambda^\dagger(\boldsymbol{x})\right)^{\otimes 2}\right.\\
&\quad\left.\times\int_{\mathscr{U}(d_\lambda)} d\mu(U_\lambda(\boldsymbol{x}'))\left(U_\lambda\left(\boldsymbol{x}'\right)\right)^{\otimes 2}\rho_f^{\otimes 2}\left(U_\lambda^\dagger\left(\boldsymbol{x}'\right)\right)^{\otimes 2}\right] && (4.52)\\
&\leqslant \sum_\lambda\mathrm{Tr}\left[\left(\mathscr{V}_{\mathrm{Haar}}^\lambda\left(\rho_f\right)\right)^2\right] && (4.53)\\
&\leqslant \sum_\lambda\frac{2}{d_\lambda(d_\lambda+1)} && (4.54)
\end{aligned}$$

Here, in the second last inequality, we have used the fact that $U_\lambda(x)$ forms a 2-design and $\mathscr{V}_{\mathrm{Haar}}^\lambda\left(\rho_f\right)=\int dU_\lambda(\mathbf{x})(U_\lambda(\mathbf{x}))^{\otimes 2}\rho_f^{\otimes 2}\left(U_\lambda^\dagger(\mathbf{x})\right)^{\otimes 2}$ is the sum over the Haar ensemble as defined previously 4.1.1. We can also consider a more general case where expressivity of

the circuit is lesser compared to an ensemble that forms a 2-design. But here we consider the worst-case scenario. The last inequality comes from Weingarten calculus as described nicely in Refs. [141, 144]. Specifically, the identity is defined as follows:

$$\int_{\mathbb{U}(d)} d\mu(U) \operatorname{Tr}\left[UAU^{\dagger}B\right] \operatorname{Tr}\left[UCU^{\dagger}D\right] = \frac{\operatorname{Tr}[A]\operatorname{Tr}[B]\operatorname{Tr}[C]\operatorname{Tr}[D] + \operatorname{Tr}[AC] + \operatorname{Tr}[BD]}{d^2-1} - \tag{4.55}$$

$$\frac{\operatorname{Tr}[AC]\operatorname{Tr}[B]\operatorname{Tr}[D] + \operatorname{Tr}[A]\operatorname{Tr}[C]\operatorname{Tr}[BD]}{d\,(d^2-1)} \tag{4.56}$$

where A, B, C, and D are linear operators defined on a Hilbert space $\mathscr{H}$. □

This bound doesn't guarantee that the kernel will avoid exponential concentration in every case since one can still, in principle, derive a stricter bound that shows exponential concentration in the number of qubits. In order to prove the absence of EC we need to derive lower bounds as done in [144]. One can see that in the case there is only one irrep, we are back to the original problem. However, the updated upper bound presented in Eq. 4.45 for symmetry encoded kernels surpasses that of the simple kernel.

4.4 Discussion and Outlook

In summary, we have shown that a quantum neural network can be used to build a quantum kernel that inherits the concentration properties of the QNN cost function. More specifically, our results prove that the cost function concentration of a parametrized quantum circuit is equivalent to the concentration of the (non-parametric) kernel matrix. Our results imply that a QNN with a BP leads to quantum kernels that suffer from EC, whereas, more importantly, a BP-free QNN can be used to build a quantum kernel that does not exhibit exponential concentration. We have illustrated this by a numerical example with an ansatz that was previously shown to be BP-free. The present work implies that recently proposed algorithms for building BP-free quantum circuits can be applied to construct performant quantum kernels for machine learning without inductive bias. However, we would like to conclude this work with a word of caution. Constructing a quantum kernel from a BP-free QNN, which does not suffer from exponential concentration, does not automatically guarantee that this quantum kernel is actually practically useful or would show any advantage over a classical kernel. We base our statement on recent findings and evidence that BP-free QNN constructions immediately lead to an efficient classical simulation of the QNN [161–

163], which defeats the use of a quantum computer in the first place, commonly referred to as *de-quantization* of the quantum computing approach. We have also derived new exponential concentration bounds for building symmetry invariant kernels. Our results suggest that symmetry encoded kernels may potentially possess higher variance, leading to a greater probability of non-zero kernel entries when compared to simple kernels.

Chapter 5

Conclusion

In the first part of the thesis, we demonstrated how machine learning techniques can be applied to problems in condensed matter physics. In Chapter 2, we described a supervised machine-learning algorithm called Gaussian process regression. GPR is a Bayesian non-parametric algorithm, that can be used to uncover the hidden relationship $f(\cdot)$ between our input $\mathbf{x}_i$ and the output y_i. The workhorse of a GP is a kernel function, which should encode our prior knowledge about $f(\cdot)$. For example, continuity and translational invariance serve as powerful inductive biases for regression problems in general. However, it is not easy to encode these biases by hand, hence we use an automatic model construction method as described by D. K. Duvenaud *et al.* in Ref. [164]. In this method, we derive complex kernels by combining a few base kernels by maximizing a model selection criterion (BIC). BIC allows us to evaluate different models by only relying on training datasets. This makes it a very useful tool for problems involving extrapolation. It would be an interesting future direction, to find out why certain models selected by BIC perform better than others. Specifically, one can do this by comparing it with other model selection algorithms such as evolutionary algorithms [68], deep kernel learning [69] *etc*. We can also compare the greedy search method to the gradient descent optimization of kernels as described in Ref. [67]. In this thesis, we employ Automatic model selection approach to study a particular problem of extrapolating quantum observables of an electron-phonon system in the Hamiltonian parameter space. Generally, quantum problems are small data, big compute in nature such that computational methods scale very poorly with system size. Here we described how machine learning methods can be used to ease the burden of quantum Monte Carlo/numerical methods. In the main text, we considered an SSH electron-phonon coupling Hamiltonian. This model becomes very challenging for numerical methods such as

GGCE [118] at lower phonon frequencies, since the phonon Hilbert space becomes too big to simulate. We used the BIC algorithm to extrapolate the polaron dispersion relation to lower phonon frequencies, by only using high-frequency data. We showed that these predictions are robust to changes in training datasets. Furthermore, there is a debate surrounding the existence of the sharp transition observed in the SSH model, in the extreme adiabatic regime. We demonstrated how machine learning models can be used to guide the theoretical calculations in this regime. Concretely, we demonstrated that the transition does survive for $\Omega = 0.4$, which is backed by intense GGCE calculations. Beyond this, we don't have validation data, but our model predicts the transition does exist for $\Omega \geq 0.3$. Again, we would like to point out, that these predictions are made by a machine learning model which is generalizing the information it has seen based on the training data and the inductive biases in the kernel function. It wouldn't be fair to treat them in the same footing as we would treat a quantum Monte Carlo algorithm, which is known to perform poorly in certain parameter regimes. However, we can have some confidence in our predictions by validating our model in different training regimes and by checking the output against known physics. Beyond this, our approach is very general and can be applied to different problems such as second-order GL phase transitions as demonstrated in Chapter 2. The BIC approach is quite general and can work for any dataset. Another interesting future direction would be to change the mean function of the Gaussian process, into a parameterized ansatz that encodes our belief about the function. To construct a physics-informed model, we can also use constrained Gaussian processes to impose any domain/derivative constraints that our system respects. This approach would be especially useful to explore the physics near the quantum critical point, which is hard to probe in general using traditional methods.

In the second half of the thesis, we deal with machine learning performed on a quantum processor. As we highlighted previously, we are currently in the NISQ era of quantum computing [32]. We focused our attention on two QML algorithms, namely quantum neural networks and quantum kernel methods. Both fall under the category of hybrid quantum-classical algorithms, which means that these are classical machine learning algorithms with a subroutine being performed on a quantum computer. The initial hope for these algorithms was that the exponentially large Hilbert space might be a source of quantum advantage. But these subroutines suffer from training challenges, mostly because of the exponentially large Hilbert space in which they live. We studied two such problems, namely barren plateaus (in quantum neural networks) and exponential concentration (in quantum kernels). In the second part of the thesis, a concrete relation between the problem of barren plateaus and exponential concentration has been derived. Using analytical tools developed for barren

plateaus, we derive upper and lower bounds on the variance of kernel values. This method enables the adaptation of quantum circuits designed to alleviate barren plateaus directly to the kernel setting, under specified conditions for the circuit ansatz. Lastly, we provided a way to encode inductive biases such as symmetries of the dataset, into quantum kernels. This enabled us to derive new bounds on kernel value variance, enhancing the stability and performance of quantum kernel-based machine learning algorithms. Our work aims to pave the way for tackling future challenges concerning the intricate relationship among barren plateaus, local minima, expressivity constraints, and, crucially, classical simulability [134].

When considering the application of machine learning to quantum physics problems, the issue of replicability and practical utility arises. A crucial question is whether others can easily take an ML model and obtain reliable results in their experiments. The answer to this question is not straightforward because ML models are prone to issues like underfitting and overfitting, particularly when data distributions change. Additionally, hyperparameters are often finely tuned in research papers, making reproducibility challenging due to differences in sampling techniques. There are methods to address hyperparameter tuning, such as Bayesian Optimization [28], DropConnect [165] etc. However, these approaches are not always efficient. Despite these challenges, ML has made significant heuristic advancements over the past two decades, proving more successful than many contemporary approaches. One solution to this problem is making ML more interpretable using tools from quantum physics, such as statistical field theory [19, 20]. ML models can also represent quantum states. We can also study such architectures by analysing their tensor network decomposition and entanglement scaling [166].

Examining the ML-physics relationship from the perspective of quantum machine learning reveals much work remains. Initially, it was assumed that the advantage of QML would stem from the exponentially large Hilbert space it explores. However, many classical models also operate in exponentially large spaces. Therefore, the advantage likely lies not in the size of the space but in how it is navigated. In many methods, gradients are calculated by considering the cost functions in Euclidean space, which may not fully leverage the unique properties of quantum systems [167]. Hence, besides choosing the perfect ansatz for your problem, it is also necessary to look at efficient parameterization and gradient calculation techniques. We believe that the barren plateau study will continue to serve as a guide to identify if and where a practical variational quantum advantage exists. It will also provide insights for developing new variational methods that surpass the standard procedures.

Bibliography

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Rev. Mod. Phys. **91**, 045002 (2019).

[2] M. Rupp, International Journal of Quantum Chemistry **115**, 1058 (2015).

[3] V. Vapnik and R. Izmailov, Machine Learning **108** (2019).

[4] J. Carrasquilla, Advances in Physics: X **5** (2020).

[5] J. Carrasquilla and R. G. Melko, Nature Physics **13**, 431 (2017).

[6] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Phys. Rev. X **7**, 031038 (2017).

[7] A. Glielmo, Y. Rath, G. Csányi, A. De Vita, and G. H. Booth, Phys. Rev. X **10**, 041026 (2020).

[8] K. Inui, Y. Kato, and Y. Motome, Phys. Rev. Res. **3**, 043126 (2021).

[9] S. J. Wetzel and M. Scherzer, Phys. Rev. B **96**, 184410 (2017).

[10] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Phys. Rev. B **95**, 041101 (2017).

[11] L. Fiedler, K. Shah, M. Bussmann, and A. Cangi, Phys. Rev. Mater. **6**, 040301 (2022).

[12] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature **549**, 195 (2017).

[13] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Nature Computational Science **2** (2022).

[14] D. Castaldo, M. Rosa, and S. Corni, Phys. Rev. A **103**, 022613 (2021).

[15] Z. An, H.-J. Song, Q.-K. He, and D. L. Zhou, Phys. Rev. A **103**, 012404 (2021).

[16] I. Khait, J. Carrasquilla, and D. Segal, Phys. Rev. Res. **4**, L012029 (2022).

[17] S. M. Moosavi, K. M. Jablonka, and B. Smit, Journal of the American Chemical Society **142** (2020).

[18] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazzio, Journal of Physics: Materials **2** (2019).

[19] M. Demirtas, J. Halverson, A. Maiti, M. D. Schwartz, and K. Stoner, "Neural Network Field Theories: Non-Gaussianity, Actions, and Locality," (2023).

[20] J. Halverson, A. Maiti, and K. Stoner, Machine Learning: Science and Technology **2** (2021).

[21] M. Koch-Janusz and Z. Ringel, Nature Physics **14** (2018).

[22] L. Ziyin and M. Ueda, "Exact Phase Transitions in Deep Learning," (2022).

[23] S. P. Adam, S.-A. N. Alexandropoulos, P. M. Pardalos, and M. N. Vrahatis, in *Approximation and Optimization : Algorithms, Complexity and Applications*, Springer Optimization and Its Applications (Springer International Publishing, Cham, 2019).

[24] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, Heliyon **4**, e00938 (2018).

[25] P. Haley and D. Soloway, in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, Vol. 4 (1992) pp. 25–30 vol.4.

[26] A. Wilson and R. Adams, in *Proceedings of the 30th International Conference on Machine Learning* (PMLR, 2013) pp. 1067–1075.

[27] A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington, Jr., and S. Manzhos, The Journal of Chemical Physics **148**, 241702 (2018).

[28] C. E. Rasmussen, C. K. Williams, *et al.*, *Gaussian processes for machine learning*, Vol. 1 (Springer, 2006).

[29] M. R. Carbone, MRS Bulletin **47**, 968 (2022).

[30] J. Sous, C. Zhang, M. Berciu, D. Reichman, B. Svistunov, N. Prokof'ev, and A. Millis, arXiv:2210.14236 (2022).

[31] C. Zhang, J. Sous, D. R. Reichman, M. Berciu, A. J. Millis, N. V. Prokof'ev, and B. V. Svistunov, Phys. Rev. X **13**, 011010 (2023).

[32] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Rev. Mod. Phys. **94**, 015004 (2022).

[33] S. S. Gill, A. Kumar, H. Singh, M. Singh, K. Kaur, M. Usman, and R. Buyya, Software: Practice and Experience **52**, 66 (2022).

[34] J. Roffe, Contemporary Physics **60**, 226 (2019).

[35] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, "The Complexity of NISQ," (2022).

[36] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, Science **292**, 472 (2001).

[37] P. Rebentrost, M. Mohseni, and S. Lloyd, Phys. Rev. Lett. **113**, 130503 (2014).

[38] Y. Liu, S. Arunachalam, and K. Temme, Nature Physics **17**, 1013 (2021).

[39] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019).

[40] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019).

[41] Z. Krunic, F. F. Flöther, G. Seegan, N. D. Earnest-Noble, and O. Shehab, IEEE Transactions on Quantum Engineering **3** (2022).

[42] T. Hubregtsen, D. Wierichs, E. Gil-Fuster, P.-J. H. Derks, P. K. Faehrmann, and J. J. Meyer, Physical Review A **106**, 042431 (2022).

[43] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Nature Communications **11** (2020).

[44] S. K. Jeswal and S. Chakraverty, Archives of Computational Methods in Engineering **26**, 793 (2019).

[45] M. Schuld, F. Petruccione, M. Schuld, and F. Petruccione, Machine Learning with Quantum Computers , 217 (2021).

[46] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature Communications **9**, 4812 (2018).

[47] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes, "Exponential concentration and untrainability in quantum kernel methods," (2022).

[48] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, PRX Quantum **3**, 020365 (2022).

[49] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, Phys. Rev. Lett. **126**, 190501 (2021).

[50] L. Friedrich and J. Maziero, Phys. Rev. A **106**, 042433 (2022).

[51] A. Kulshrestha and I. Safro, in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)* (2022) pp. 197–203.

[52] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Quantum **6** (2022).

[53] R. Shaydulin and S. M. Wild, Phys. Rev. A **106**, 042407 (2022).

[54] Y. Suzuki, H. Kawaguchi, and N. Yamamoto, "Quantum Fisher kernel for mitigating the vanishing similarity issue," (2022).

[55] J. Kübler, S. Buchholz, and B. Schölkopf, Advances in Neural Information Processing Systems **34**, 12661 (2021).

[56] S. Barišić, J. Labbé, and J. Friedel, Phys. Rev. Lett. **25**, 919 (1970).

[57] D. J. J. Marchand, P. C. E. Stamp, and M. Berciu, Phys. Rev. B **95**, 035117 (2017).

[58] D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum, and G. Zoubin, in *Proceedings of the 30th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 28 (PMLR, 2013) pp. 1166–1174.

[59] R. A. Vargas-Hernández, J. Sous, M. Berciu, and R. V. Krems, Phys. Rev. Lett. **121**, 255702 (2018).

[60] M. R. Carbone, D. R. Reichman, and J. Sous, Phys. Rev. B **104**, 035106 (2021).

[61] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **473**, 20160751 (2017).

[62] H. Kadri, E. Duflos, P. Preux, S. Canu, and M. Davy, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 9 (PMLR, 2010) pp. 374–380.

[63] N. Aronszajn, Transactions of the American Mathematical Society **68**, 337 (1950).

[64] K. Asnaashari and R. V. Krems, Machine Learning: Science and Technology **3** (2021).

[65] D. K. Duvenaud, H. Nickisch, and C. Rasmussen, in *Advances in Neural Information Processing Systems*, Vol. 24, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger (Curran Associates, Inc., 2011).

[66] J. Dai and R. V. Krems, Mach. Learn.: Sci. Technol. **4**, 045027 (2023).

[67] S. Sun, G. Zhang, C. Wang, W. Zeng, J. Li, and R. Grosse, in *Proceedings of the 35th International Conference on Machine Learning* (PMLR, 2018) pp. 4828–4837.

[68] I. Roman, R. Santana, A. Mendiburu, and J. A. Lozano, Neurocomputing **462**, 426 (2021).

[69] A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 51, edited by A. Gretton and C. C. Robert (PMLR, 2016) pp. 370–378.

[70] "Statistical Physics, Part 1, 3rd Revised and Enlarged Edition - L.D. Landau; L.P. Pitaevskii; E. M. Lifshitz," .

[71] V. L. Ginzburg and L. D. Landau, Zh. Eksp. Teor. Fiz. **20**, 1064 (1950).

[72] J. C. Tolédano and P. Tolédano, *The Landau Theory of Phase Transitions: Application to Structural, Incommensurate, Magnetic and Liquid Crystal Systems*, World Scientific Lecture Notes in Physics, Vol. 3 (WORLD SCIENTIFIC, 1987).

[73] T. Mizokawa, A. Fujimori, T. Arima, Y. Tokura, N. Mōri, and J. Akimitsu, Phys. Rev. B **52**, 13865 (1995).

[74] E. Demler, W. Hanke, and S.-C. Zhang, Rev. Mod. Phys. **76**, 909 (2004).

[75] M. De Prato, A. Pelissetto, and E. Vicari, Phys. Rev. B **74**, 144507 (2006).

[76] P. Calabrese, A. Pelissetto, and E. Vicari, Phys. Rev. B **67**, 054505 (2003).

[77] I. P. Ivanov, Phys. Rev. E **79**, 021116 (2009).

[78] O. Can, T. Tummuru, R. P. Day, I. Elfimov, A. Damascelli, and M. Franz, Nature Physics **17**, 519 (2021).

[79] A. Bansil, H. Lin, and T. Das, Rev. Mod. Phys. **88**, 021004 (2016).

[80] Y. Ren, Z. Qiao, and Q. Niu, Reports on Progress in Physics **79**, 066501 (2016).

[81] J. Larson, E. Sjöqvist, and P. Öhberg, in *Conical Intersections in Physics: An Introduction to Synthetic Gauge Theories*, Lecture Notes in Physics (Springer International Publishing, Cham, 2020) pp. 55–91.

[82] T. Tummuru, E. Lantagne-Hurtubise, and M. Franz, Phys. Rev. B **106**, 014520 (2022).

[83] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **473**, 20160751 (2017).

[84] L. Brevault, M. Balesdent, and A. Hebbal, Aerospace Science and Technology **107**, 106339 (2020).

[85] A. E. Wiens, A. V. Copan, and H. F. Schaefer, Chemical Physics Letters **737**, 100022 (2019).

[86] C. Franchini, M. Reticcioli, M. Setvin, and U. Diebold, Nature Reviews Materials **6**, 560 (2021).

[87] F. Giustino, Rev. Mod. Phys. **89**, 015003 (2017).

[88] N. W. Ashcroft and N. D. Mermin, *Solid state physics* (Cengage Learning, 2022).

[89] X. Cai, Z.-X. Li, and H. Yao, Phys. Rev. Lett. **127**, 247203 (2021).

[90] B. J. Alder, K. J. Runge, and R. T. Scalettar, Phys. Rev. Lett. **79**, 3022 (1997).

[91] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

[92] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **106**, 162 (1957).

[93] J. Bardeen and D. Pines, Phys. Rev. **99**, 1140 (1955).

[94] H. Fröhlich, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **215**, 291 (1997).

[95] A. S. Alexandrov and A. B. Krebs, Soviet Physics Uspekhi **35**, 345 (1992).

[96] N. Mott, Physica C: Superconductivity **205**, 191 (1993).

[97] A. Alexandrov and J. Ranninger, Phys. Rev. B **24**, 1164 (1981).

[98] G. Wellein, H. Röder, and H. Fehske, Phys. Rev. B **53**, 9666 (1996).

[99] B. K. Chakraverty, J. Ranninger, and D. Feinberg, Phys. Rev. Lett. **81**, 433 (1998).

[100] A. S. Alexandrov, Europhysics Letters **56**, 92 (2001).

[101] B. Roy, J. D. Sau, and S. Das Sarma, Phys. Rev. B **89**, 165119 (2014).

[102] J. Sous, M. Chakraborty, R. V. Krems, and M. Berciu, Phys. Rev. Lett. **121**, 247001 (2018).

[103] M. R. Carbone, A. J. Millis, D. R. Reichman, and J. Sous, Phys. Rev. B **104**, L140307 (2021).

[104] C. Zhang, N. V. Prokof'ev, and B. V. Svistunov, Phys. Rev. B **104**, 035143 (2021).

[105] B. Cohen-Stead, O. Bradley, C. Miles, G. Batrouni, R. Scalettar, and K. Barros, Phys. Rev. E **105**, 065302 (2022).

[106] A. S. Mishchenko, N. Nagaosa, and N. Prokof'ev, Phys. Rev. Lett. **113**, 166402 (2014).

[107] T. Holstein, Annals of physics **8**, 325 (1959).

[108] T. Holstein, Annals of physics **8**, 343 (1959).

[109] B. Gerlach and H. Löwen, Phys. Rev. B **35**, 4297 (1987).

[110] A. J. Heeger, S. Kivelson, J. R. Schrieffer, and W. P. Su, Rev. Mod. Phys. **60**, 781 (1988).

[111] S. Barišić, J. Labbé, and J. Friedel, Phys. Rev. Lett. **25**, 919 (1970).

[112] in *Quantum Theory of the Electron Liquid* (Cambridge University Press, 2005) pp. 111–156.

[113] A. Peres, Phys. Rev. A **32**, 3266 (1985).

[114] J. K. Nauth and V. M. Stojanović, Phys. Rev. B **107**, 174306 (2023).

[115] M. Heyl, Reports on Progress in Physics **81** (2018).

[116] T. Gorin, T. Prosen, T. H. Seligman, and M. Žnidarič, Physics Reports **435** (2006).

[117] W. W. Ho and D. A. Abanin, Phys. Rev. B **95**, 094302 (2017).

[118] M. R. Carbone, S. Fomichev, A. J. Millis, M. Berciu, D. R. Reichman, and J. Sous, "The Generalized Green's function Cluster Expansion: A Python package for simulating polarons," (2022).

[119] M. Berciu, Phys. Rev. Lett. **97**, 036402 (2006).

[120] G. L. Goodvin, M. Berciu, and G. A. Sawatzky, Phys. Rev. B **74**, 245104 (2006).

[121] M. Berciu and G. L. Goodvin, Phys. Rev. B **76**, 165109 (2007).

[122] D. J. J. Marchand, G. De Filippis, V. Cataudella, M. Berciu, N. Nagaosa, N. V. Prokof'ev, A. S. Mishchenko, and P. C. E. Stamp, Phys. Rev. Lett. **105**, 266605 (2010).

[123] W. L. Loh, Ann. Stat. **24**, 2058 (1996).

[124] K. Hannewald, V. M. Stojanović, J. M. T. Schellekens, P. A. Bobbert, G. Kresse, and J. Hafner, Phys. Rev. B **69**, 075211 (2004).

[125] J. H. Fetherolf, D. Golež, and T. C. Berkelbach, Phys. Rev. X **10**, 021062 (2020).

[126] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, The Journal of Chemical Physics **138**, 114103 (2013).

[127] M. Grundner, T. Blatz, J. Sous, U. Schollwöck, and S. Paeckel, arXiv:2308.13427 (2023).

[128] M. N. Gastiasoro, J. Ruhman, and R. M. Fernandes, Ann. Phys. (N. Y.) **417**, 168107 (2020).

[129] J. Sous, M. Berciu, and R. V. Krems, Phys. Rev. A **96**, 063619 (2017).

[130] J. Sous, M. Chakraborty, C. Adolphs, R. Krems, and M. Berciu, Sci. Rep. **7**, 1169 (2017).

[131] E. Peters, J. Caldeira, A. Ho, S. Leichenauer, M. Mohseni, H. Neven, P. Spentzouris, D. Strain, and G. N. Perdue, Npj Quantum Inf. **7**, 1 (2021).

[132] T. Hubregtsen, D. Wierichs, E. Gil-Fuster, P.-J. H. S. Derks, P. K. Faehrmann, and J. J. Meyer, Phys. Rev. A **106**, 042431 (2022).

[133] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Nat. Rev. Phys. **3**, 625 (2021).

[134] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, arXiv:2405.00781 (2024).

[135] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Nat. Commun. **12**, 6961 (2021).

[136] A. V. Uvarov and J. D. Biamonte, J. Phys. A-Math. **54**, 245301 (2021).

[137] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, PRX Quantum **2**, 040316 (2021).

[138] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, PRX Quantum **3**, 010313 (2022).

[139] N. Innan, M. Khan, B. Panda, and M. Bennai, Quantum Inf. Process. **22**, 374 (2023).

[140] H.-K. Zhang, S. Liu, and S.-X. Zhang, Phys. Rev. Lett. **132**, 150603 (2024).

[141] "Theory for Equivariant Quantum Neural Networks," (2022).

[142] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, "Covariant quantum kernels for data with group structure," (2022).

[143] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Nature Communications **12** (2021).

[144] L. Schatzki, M. Larocca, Q. T. Nguyen, F. Sauvage, and M. Cerezo, "Theoretical Guarantees for Permutation-Equivariant Quantum Neural Networks," (2022).

[145] T. Hofmann, B. Schölkopf, and A. J. Smola, The Annals of Statistics **36**, 1171 (2008).

[146] Y. Xu and Q. Ye, *Generalized Mercer Kernels and Reproducing Kernel Banach Spaces*, Memoirs of the American Mathematical Society, Vol. 258 (American Mathematical Society, 2019).

[147] W. S. Noble, Nature Biotechnology **24** (2006).

[148] L. Deng, IEEE Signal Process. Mag. **29**, 141 (2012).

[149] A. Canatar, B. Bordelon, and C. Pehlevan, Nature Communications **12**, 2914 (2021).

[150] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Quantum **5**, 558 (2021).

[151] M. Cerezo and P. J. Coles, Quantum Science and Technology **6**, 035006 (2021).

[152] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Communications in Mathematical Physics **346**, 397 (2016).

[153] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80**, 012304 (2009).

[154] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nat. Commun. **12** (2021), 10.1038/s41467-021-21728-w.

[155] Z. Liu, L.-W. Yu, L.-M. Duan, and D.-L. Deng, Phys. Rev. Lett. **129**, 270501 (2022).

[156] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Quantum Science and Technology **7**, 045015 (2022).

[157] L. Schatzki, M. Larocca, Q. T. Nguyen, F. Sauvage, and M. Cerezo, "Theoretical Guarantees for Permutation-Equivariant Quantum Neural Networks," (2022).

[158] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, arXiv preprint arXiv:2105.03406 (2021).

[159] A. Holevo, Reports on Mathematical Physics **16**, 385 (1979).

[160] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, PRX Quantum **4**, 010328 (2023).

[161] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, "Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing," (2023), arXiv:2312.09121 [quant-ph, stat] .

[162] L. Leone, S. F. E. Oliviero, L. Cincio, and M. Cerezo, "On the practical usefulness of the Hardware Efficient Ansatz," (2022), arXiv:2211.01477 [quant-ph] .

[163] A. Basheer, Y. Feng, C. Ferrie, and S. Li, "Alternating Layered Variational Quantum Circuits Can Be Classically Optimized Efficiently Using Classical Shadows," (2022), 2208.11623 [quant-ph] .

[164] G. Schwarz, The Annals of Statistics **6**, 461 (1978).

[165] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, in *Proceedings of the 30th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 28 (PMLR, 2013) pp. 1058–1066.

[166] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Phys. Rev. Lett. **122**, 065301 (2019).

[167] T. Haug and M. Kim, arXiv preprint arXiv:2104.14543 (2021).

Check for updates

RESEARCH ARTICLE

Particle & Particle Systems Characterization
www.particle-journal.com

Gas-Sensing Properties of NO on Ce-Doped Zinc Oxide: A DFT Study

Neha Sharma, Umesh K. Dwivedi, Umesh T. Nakate, Mukhtiyar Singh, and Sandip Paul Choudhury**

One of the most prevalent pollutants that pollute the environment is nitrogen oxide (NO_x). NO and NO_2 gases, which are hazardous to both human health and the environment, are included in NO_x. The rare earth element Ce doped metal oxide semiconductor (MOS) ZnO is employed to reveal their NO gas sensing properties. Based on density functional theory (DFT) calculations, the optimized surface of ZnO (0001), Ce-doped ZnO (0001), adsorbate structure of NO, and adsorbate NO on the modified ZnO (0001) surface are obtained. The gas sensing properties are examined through adsorption energy, Bader charge analysis, charge density difference (CDD), charge transfer, band structure, total density of state (DOS), and partial density of states (PDOS). For the Ce-doped ZnO (0001) surface the NO adsorption energy is more negative than the bare ZnO. From the observation of Bader charge analysis, the charge transfer value increases from the substrate to adsorbate after doping with Ce, which indicates that the Ce-doped ZnO (0001) surface is more favorable for NO gas sensing. Favorable electronic properties and suitable adsorption energy of Ce-doped ZnO can be a potential gas sensor for NO molecule. The obtained DFT results are also compared with the existing experimental results.

1. Introduction

NO_x gases are common pollutant gases in the environment. These gases are formed during combustion processes in automotive engines, industrial factories, and power plants. NO_x includes NO and NO_2 gases which are dangerous for human health and the environment. NO gas emission provides health problems such as eye and throat irritation, and headache and also leads to environmental issues like acid rain, ozone depletion, and the greenhouse effect.[1,2] So, the detection of these gases at low ppm is important for environmental and health protection.

Chemical sensors play a vital role in detecting these harmful gases. Metal oxide semiconductors (MOSs),[3] such as ZnO,[4] SnO_2,[5] and WO_3,[6] are explored as the gas sensing material. Amongst all these materials, ZnO is more reliable due to its low cost and long-term stability. It is a responsive material for the detection of NO, NO_2, and methane gases. Gupta et al. reported that ZnO nanowire enhances the NO gas sensitivity due to higher oxygen vacancy in the nanowire structure.[7] Hieu et al. reported that the urchin-like structure of ZnO has a better response and selectivity toward NO gas at the optimum temperature of 200 °C also having a response value in the range of 3400–6000% per ppm.[8] They also reported that after annealing the sample at 400 °C temperature the response value increases by 20%. Madhusoodanan et al. reported that ZnO thin films prepared by spray pyrolysis technique show enhancement in the response value of NO_2 gas. They also reported that the response values are 3.32 (for 7 ppm NO_2 gas) and 1.4 (for 18 ppm H_2S gas).[9] Kusumam et al. have synthesized the ZnO nanosheets by hydrothermal method and found that more oxygen vacancies created on the surface reduce the response time (4 s) and recovery time (36 s) for NO_2 gas.[10] ZnO (MOS) is a type of chemical sensor, which is used as a sensing element in a chemical sensor. Recently, new kinds of sensing elements have been developed as a result of developments in materials science and fabrication technologies. Different chemical sensor types identify the same target, and certain chemical sensors use multiple sensing elements to identify multiple targets. Chen et al. discussed in their review paper different applications like health monitoring, multifunctional sensing, and system integration of metal oxide semiconductors and graphene-based metal oxide semiconductor materials.[3] Recent developments in chemical gas sensors have focused on enhancing sensitivity, selectivity, and reliability through innovations in materials and technology that are discussed below. Zinc oxide nanostructures have excellent optical, piezoelectric, and electrochemical properties as well as high surface area. Due to this it is used in pressure sensors, gas sensors, photoelectric sensors, and biosensors.[4]

N. Sharma, U. K. Dwivedi, S. P. Choudhury
Amity School of Applied Science
Amity University Rajasthan
Jaipur, Rajasthan 303002, India
E-mail: spchoudhury@jpr.amity.edu
U. T. Nakate
Department Polymer Nano Science and Technology
Jeonbuk National University
Jeollabuk-do, Jeonju-si 54986, Republic of Korea
M. Singh
Department of Applied Physics
Delhi Technological University
Rohini, Delhi 110042, India
E-mail: mukhtiyarsingh@dtu.ac.in

DOI: 10.1002/ppsc.202400126

 © 2024 Wiley-VCH GmbH

Improvement in selectivity and sensitivity of ZnO toward different gases can be efficiently enhanced by doping ZnO with other elements or modifying the structure of ZnO by introduction of bilayers of MOSs. Liu et al. reported that porous Al-doped ZnO film has a higher surface area and has better selectivity toward the NO gas in comparison to the CO_2 and O_2 gas.[11] Lin et al. reported that TiO_2 nanoparticle/ZnO thin film shows higher sensitivity in comparison to the single layer of TiO_2 nanoparticles and ZnO thin film.[12] Metal oxide semiconductors are conductometric gas sensors that are based on the change of the conductivity by adsorbing the gas molecule at the surface of the sensing material. Corlu et al. reported that Cu-doped ZnO thin films improved the NO gas sensing performance at the optimum temperature of 55 °C and also increased the selectivity toward the NO gas. They also showed that the zeolite-coated Cu-ZnO-based sensor reduced the operating temperature from 55 °C to room temperature.[13] Singh et al. reported that ZnO-$ZnGa_2O_4$ film shows higher selectivity and response toward NO gas compared to NO_2, SO_2, and CO_2 gas.[14] Moreover, the rare earth element doping in MOSs enhances the gas sensitivity and also reduces the optimum temperature because of higher distortion and large number of carriers.[15] Fidha et al. reported that Dy-doped ZnO shows a higher response toward NO_2 gas at the optimum temperature of 150 °C.[16] Chang et al. reported that Ce doped-ZnO sensing material improves the gas sensing performance and also reduces the operating temperature to 100 °C and detects (2 ppm) of NO_2 gas.[17] Sun et al. reported that Ce doping in ZnO increases the NO_2 gas sensing response and this sensor shows long-term stability.[18] Mokrushin et al. reported that praseodymium and europium doped ZnO show higher sensitivity toward 4–100 ppm concentration of NO_2 gas at the optimum temperature value of 125–150 °C. They also reported that average work function increases after doping of Pr and Eu.[19] Sarf et al. reported that Y, Ru, and Cs-doped ZnO enhance the NH_3 gas sensitivity and also improve the response value for 15 ppm NH_3 gas at room temperature.[20]

Currently, the Density Functional Theory (DFT) is applied for the detection of gases based on adsorption energy, bandgap, Bader charge transfer, density of states, Mullekin charge, and Molecular Dynamics (MD) simulations.[21] Rahman et al. performed a DFT study on the rare earth element doped ZnO and found that rare earth elements (REEs) doping in ZnO have negative formation energies which indicate that all samples are energetically favorable.[22] So, for this study, we use the Ce as a dopant atom because its doping in zinc oxide is energetically favorable.

DFT studies are performed to study the gas sensing behavior of H_2, NH_3, CO, and ethanol,[23] NO_2, NO, O, and N,[24] NO, and CO,[25] on the ZnO surface. There are DFT studies where the (0001) crystal face of ZnO is observed to display a good response toward the detection of these gases.[26,27] Motivated by these studies, we have investigated theoretically here the effect of Ce doping in the ZnO (0001) surface for NO gas sensing. In this DFT study, we investigated adsorption energy, charge transfer, Bader charge, band structure, TDOS, and PDOS for NO adsorption on bare and Ce-doped ZnO surfaces. Here, we have investigated that Ce-doping in the ZnO surface enhances charge transfer, adsorption energy, and bandgap.

2. Experimental Section

2.1. Computational Details

The calculations were performed using the Density Functional Theory (DFT) based Quantum Espresso (QE) package.[28] In these calculations, the electron-ion interaction was described by the projector-augmented wave (PAW) method and the exchange-correlation was treated within general gradient approximation (GGA) using Perdew–Burke–Ernzerhoff (PBE) functional. The band structure and other parameters in the current work were calculated using the energy cut-offs of 45 and 260 Ryd for the wavefunctions and charge densities. A 5 × 5 × 1 k-point mesh was used for all the calculations except DOS where a dense k-point grid (50 × 50 × 1) was used after optimizing. A dense k-mesh requires more computational time. Hence k-mesh of was only used for DOS calculation (50 × 50 × 1).[29] The NO adsorbate molecule is optimized in the 10 Å width of the vacuum layer. For NO molecule the optimized NO bond length of 1.166 Å is comparable to the reported results.[24] After doping of Ce in zinc oxide the bond length of NO becomes 1.267 Å. The site of the NO adsorbate molecule was taken on the upper layer of the ZnO (0001) surface at a distance of 3.495 Å (Zn-O_{ads}). After doping of Ce-atom on the upper layer surface of ZnO (0001), the nitrogen atom interacts more strongly than the oxygen atom which is shown in **Figure 1d**, and the distance from the surface to the adsorbate atom is 2.1045 Å.

2.2. Surface Models

The surface calculations were performed with the help of a slab model using the boundary conditions. For this, a 3 × 3 supercell with 2 surface layers (36 atoms) and a vacuum space of 10 Å was cleaved from a bulk structure of ZnO with a lattice constant of a = 3.22 Å and c = 5.20 Å.

To simulate the defect surface, the one zinc atom of the upper surface layer was removed by a Ce atom also called substitutional doping of Ce on the (0001) surface of ZnO. This model was optimized in a vacuum and the environment of adsorbate NO molecule.

The adsorption energy of the NO molecule to the surface was calculated by the following formula,[30]

$$\Delta E_{\text{adsorption}} = E_{\text{surface+NO}} - E_{\text{surface}} - E_{\text{NO}} \quad (1)$$

where, ΔE is the adsorption energy, $E_{\text{surface+NO}}$ is the total energy of the surface-NO model and E_{NO} is the total energy of the NO molecule. For the defect surface, the adsorption energy was calculated by the following formula

$$\Delta E_{\text{adsorption}} = E_{\text{ZnO/}Ce\text{+NO}} - E_{\text{ZnO/Ce}} - E_{\text{NO}} \quad (2)$$

where, $E_{\text{ZnO/Ce+NO}}$ is the total energy of the defect surface-NO model, $E_{\text{ZnO/Ce}}$ is the total energy of the defect surface and E_{NO} is the total energy of the NO molecule. From this adsorption energy, we calculate the stability of a dopant on the surface. The stable structures will have negative adsorption energy while unstable structures will have positive adsorption energy values.

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. Results and Discussion

In this work, we examine the adsorption of NO molecules on the surface of the ZnO (0001) polar facet by Density Functional Theory (DFT) using Quantum Espresso software. We study the NO adsorption on the cleaved surface and Ce-doped surface of ZnO because Ce doping enhances the sensitivity and reduces the operating temperature. For Ce-dopant we chose a top and bottom surface layer and found that the adsorption energies are quite similar for both the surfaces. So, for further calculation, we choose the position of the Ce dopant at the top surface layer. **Table 1** represents the comparison of our result of adsorption energy with previously reported results.

The geometries of pure ZnO and Ce-doped ZnO were examined in terms of structural properties and presented in **Figure 1**. The negative binding energy value of Ce doping in ZnO indicates

Table 1. Adsorption energy of ZnO and modified ZnO surface.

S.No.	Gas	Surface	Adsorption Energy	Refs.
1.	NO	ZnO (0002)	−0.49 eV	[25]
2.	NO	Pd@ZnO	−1.57 eV	[27]
3.	NO	ZnO $(2\bar{1}\bar{1}0)$	−0.99 eV	[31]
4.	NO	ZnO $(10\bar{1}0)$	−0.28 eV	[24]
5.	CO	ZnO $(10\bar{1}0)$	−46.27 kJ/mol	[32]
6.	CO_2	N-ZnO (101)	−1.61eV	[33]
7.	NO_2	ZnO-H (101)	−0.59 eV	[34]
8.	H_2S	h-ZnO	−0.54 eV	[35]
9.	CO	ZnO-Pt	−1.42 eV	[36]
10.	NO	ZnO (0001)	−0.42 eV	This work
11.	NO	Ce-doped ZnO (0001)	−0.96 eV	This work

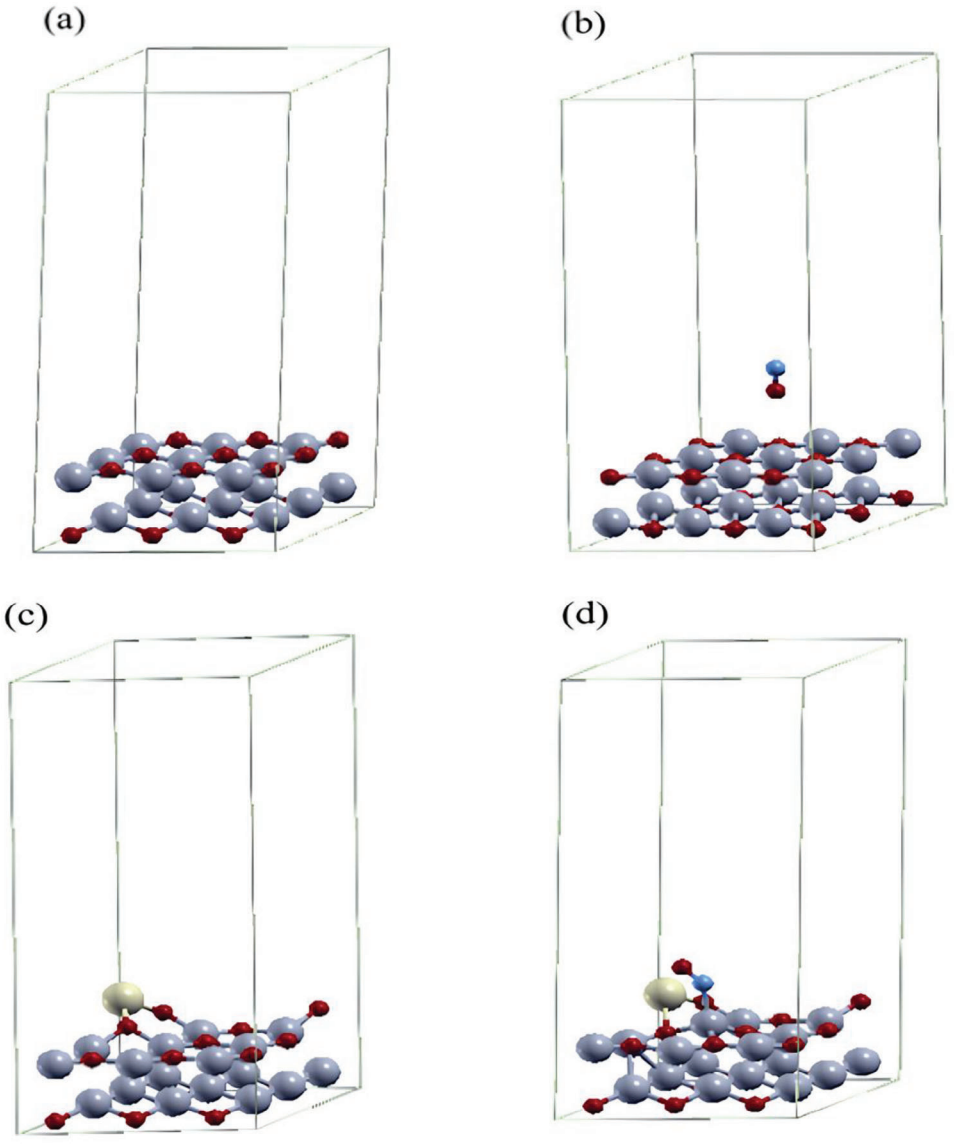


Figure 1. Geometry of optimized structures a) ZnO b) ZnO with NO molecule c) ZnO doped with Ce and d) ZnO doped Ce with NO molecule (grey color- Zn atom, red color- O atom, cream color- Ce atom, blue color- N atom).

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that Ce doping is favorable for the stable ZnO surface.[22] So, in the present work, we have investigated the effect of Ce-doping in the ZnO (0001) surface for NO gas sensing applications. The interaction of NO gas molecule with modified ZnO (0001) surface was estimated by adsorption energy values based on Equation (2). Figure 1b represents the optimized structure of the NO molecule on the surface of ZnO. The $E_{ads.}$ for pure ZnO-NO molecule is computed to be −0.42 eV. The NO gas adsorption on the Ce-doped ZnO surface is illustrated in Figure 1d. The $E_{ads.}$ value for NO gas on a doped surface is computed as −0.96 eV. This adsorption value is more negative than pure ZnO surface which is due to the doping of the Ce atom. Also, the charge transfer from the surface to the NO gas molecule is calculated. The charge transfer for pure and Ce-doped ZnO-NO surfaces are −0.02e and −0.74e, respectively. The details are discussed in section 3.3. For further study, we calculate the band structure from this calculation and we get the bandgap before adsorption and after adsorption of the NO molecule. The change in bandgap value indicates that the conductivity changes after gas adsorption. Using this property, we can calculate the sensitivity of the gas sensor by the following formula:

$$S = C\exp\left(\frac{E_{g_0} - E_{g_1}}{k_b T}\right) \tag{3}$$

where C is constant, E_{g_0}, E_{g_1}, K_b, and T represent the bandgap value before gas adsorption, bandgap after gas adsorption, Boltzmann constant, and temperature, respectively.[23]

3.1. NO Gas Adsorption on ZnO and Ce-doped ZnO Surface (Gas Sensing Mechanism)

Figure 1d shows that after the adsorption of NO on the Ce-doped ZnO surface the Zn atom makes the bond with two other Zn atoms which represent the strong bonding of NO molecule on the modified surface. When the NO molecule is adsorbed on the ZnO (0001) surface there is an interaction mechanism between the surface and adsorbed NO molecule. In the interaction process between the surface and NO molecule after adsorption, the NO molecule accepts the electrons from the surface which is shown in Bader analysis. This in turn changes the conductivity and resistance of the surface which can be used to measure the response value of NO at a particular temperature on the ZnO surface. The Ce doping in ZnO has a negative binding energy value. So, after adsorption of the NO molecule on the Ce-doped ZnO surface the NO molecule accepts more electrons in comparison to the pure ZnO surface bringing in more change in resistance and conductivity. Because of this, the Ce-doped ZnO surface behaves as a good sensing material for the NO molecule. After doping of Ce-atom the NO adsorption energy is more negative in comparison to the pure surface and the bond length decreases which indicates that the Ce-ZnO (0001) surface has more adsorption strength for the NO molecule.[35]

3.2. Charge Density Difference (CDD)

In charge density difference analysis, the system's charge density is subtracted from the isolated reference charge densities of the system's constituent components. It shows the charge redistribution caused by the interacting materials and shows the charges that have accumulated and been emptied. The formula for CDD is given by

$$\Delta\rho = \rho_{\mathrm{ZnO-NO}} - \rho_{\mathrm{ZnO}} - \rho_{\mathrm{NO}} \tag{4}$$

where $\Delta\rho$ represents the charge density difference, $\rho_{\mathrm{ZnO-NO}}$ is charge density of ZnO-NO surface, ρ_{ZnO} is the charge density of isolated ZnO surface and ρ_{NO} is the charge density of isolated NO molecule.[24]

Figure 2a,b represents the charge density difference (CDD) for ZnO-NO and Ce-doped ZnO-NO, respectively obtained by Vesta software. In Figure 2b charge accumulation is more around the NO in comparison to the Figure 2a which is consistent with the Bader charge analysis. In this Figure, the yellow color represents the charge accumulation and the cyan color represents charge depletion. The Ce-doped ZnO shows more sharing of electrons between the surface and NO molecule. This proves a strong interaction between NO and the Ce-ZnO surface. Figure 2 shows charge accumulation around the NO molecule which suggests that the NO behaves as a charge acceptor.

3.3. Bader Charge Analysis

Bader charge gives the results related to the charge transfer from the surface to the adsorbate gas. **Tables 2** and **3** show the Bader charge calculation for NO adsorption on the undoped ZnO surface and Ce-doped ZnO surface, respectively. The top layer surface atoms' charge changed dramatically after adsorption, indicating that the surface O and Zn atoms lose their charge to NO, which is reflected in the ZnO-NO system's computed Bader charges. The lower surface layers exhibit little adsorption-related change in their estimated charges, as would be predicted. Figure 1d shows strong interaction after doping of the Ce atom which is also consistent with Bader charge calculation. After, the doping of the Ce atom the top and lower surface atoms' charge changed dramatically after the adsorption of the NO molecule. The net charge transfer from the surface to the NO molecule is given by the following Equation (5)

$$Q_{\mathrm{total}} = Q_1 - Q_2 \tag{5}$$

where, Q_{total}, Q_1, and Q_2 represents the total charge transfer, charge on gas molecule before adsorption, and charge on gas molecule after adsorption of NO gas molecule respectively.[37]

After doping of the Ce atom into the ZnO surface the charge transfer from the surface to the NO molecule increases from −0.02e to −0.74e which indicates that the Ce-doped ZnO surface is more sensitive for NO gas sensing. The negative sign of Q_{total} indicates that the NO gas molecule withdraws the charge from the bare and Ce-doped ZnO (0001) surface, respectively. The charge transfer for the Ce-doped ZnO surface is observed at −0.74e indicating that the doped surface transferred more electrons to the NO molecule. These results are also consistent with the experimental findings which show after NO adsorption the resistance increases for Ce-ZnO nanoarrays.[38]

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

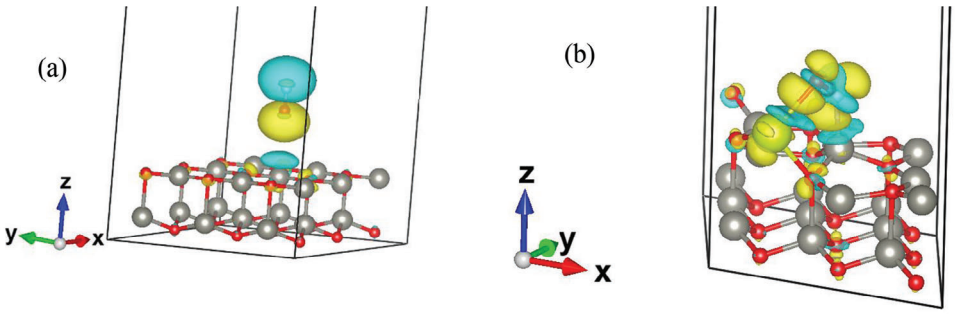


Figure 2. a) and b) represent CDD plots of ZnO-NO and Ce-doped ZnO-NO (grey color- Zn atom, red color- O atom, yellow color represents charge accumulation, cyan color represents charge depletion).

3.4. Band Structure and TDOS Analysis

3.4.1. Adsorption of NO on ZnO Surface

The ZnO—NO band structure shown in **Figure 3a** indicates that after the adsorption of NO, the bandgap increases, which is similar to the previously reported DFT and experimental results.[24] After the adsorption of NO gas, the bandgap increases up to 0.6 eV. Figure 3a shows that after the adsorption of NO gas, the band structure has the same property as the direct bandgap. To further confirm the electrical property, the DOS is calculated. NO adsorption gives the modified DOS shape. Figure 3b shows that levels are shifted to lower energy. This is consistent with the fact that NO acts as an electron acceptor.

3.4.2. Effect of Ce Doping on ZnO Surface

After doping with Ce, the fermi energy level shifts into the conduction band, which indicates the strong bonding of Ce on the surface of ZnO. **Figure 4a** indicates that after doping with Ce, the bandgap decreases. The bandgap value is 0.52 eV. The Ce-doped ZnO surface has a more compact structure, which indicates the strong bonding of the Ce atom with the other atoms. In the doping process, one Zn atom is replaced by a Ce atom, and the Ce ion behaves as a donor, so the energy level increases in the ZnO energy band. So, the carrier density increases in the conduction band, and from this result, the fermi energy level shifts upward.[18] The TDOS structure (Figure 4b) shows that the fermi energy level shifts into

Table 2. Bader charges of the NO, relaxed ZnO, and NO adsorbed ZnO surface. (O_1 represents oxygen atom of NO molecule).

Atom		Bader charges of Relaxed ZnO [0001]		Bader charge of Free NO	Bader Charges after NO adsorption on ZnO [0001]	
Zn, N, O_1	O	Zn	O	N, O_1	Zn, N, O_1	O
N				0.44	0.44	
O_1				−0.43	−0.45	
Zn1	O1	1.01	−1.14		1.24	−1.27
Zn2	O2	1.19	−1.06		1.12	−1.09
Zn3	O3	1.01	−1.15		1.24	−1.27
Zn4	O4	1.19	−1.06		1.12	−1.09
Zn5	O5	1.01	−1.14		1.24	−1.27
Zn6	O6	1.19	−1.06		1.12	−1.09
Zn7	O7	1.01	−1.15		1.24	−1.27
Zn8	O8	1.19	−1.06		1.12	−1.09
Zn9	O9	1.01	−1.15		1.26	−1.27
Zn10	O10	1.19	−1.06		1.12	−1.1
Zn11	O11	1.01	−1.15		1.24	−1.27
Zn12	O12	1.19	−1.06		1.12	−1.09
Zn13	O13	1.01	−1.14		1.25	−1.27
Zn14	O14	1.19	−1.06		1.12	−1.1
Zn15	O15	1.01	−1.15		1.24	−1.27
Zn16	O16	1.19	−1.06		1.12	−1.09
Zn17	O17	1.01	−1.14		1.24	−1.27
Zn18	O18	1.19	−1.06		1.12	−1.09

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3. Bader charges of the NO, relaxed Ce-doped ZnO, and NO adsorbed Ce-doped ZnO surface (Zn atom placed at position 1 is replaced by Ce atom which is indicated in bold).

Atom		Bader charges of relaxed Ce-doped ZnO		Bader charge of free NO	Bader Charges after NO adsorption on Ce-doped ZnO [0001]	
Zn, N, O_1, Ce	O	Zn, Ce	O	N, O_1	Zn, N, O_1, Ce	O
N				0.44	−0.14	
O_1				−0.43	−0.59	
Ce1	**O1**	**1.98**	**−1.25**		**2.06**	**−1.24**
Zn2	O2	0.98	−1.09		1.02	−1.07
Zn3	O3	1.1	−1.25		1.2	−1.23
Zn4	O4	1.16	−1.08		1.15	−1.07
Zn5	O5	1.08	−1.25		1.19	−1.24
Zn6	O6	0.98	−1.08		1.02	−1.08
Zn7	O7	1.08	−1.25		1.14	−1.26
Zn8	O8	0.98	−1.08		1.18	−1.07
Zn9	O9	1.11	−1.18		1.17	−1.27
Zn10	O10	1.12	−1.08		1.16	−1.11
Zn11	O11	1.19	−1.16		1.21	−1.22
Zn12	O12	1.12	−1.11		1.1	−1.09
Zn13	O13	1.11	−1.25		1.18	−1.23
Zn14	O14	1.16	−1.08		1.14	−1.07
Zn15	O15	1.26	−1.27		1.19	−1.22
Zn16	O16	1.16	−1.12		1.17	−1.09
Zn17	O17	1.11	−1.18		1.16	−1.22
Zn18	O18	1.12	−1.08		1.15	−1.07

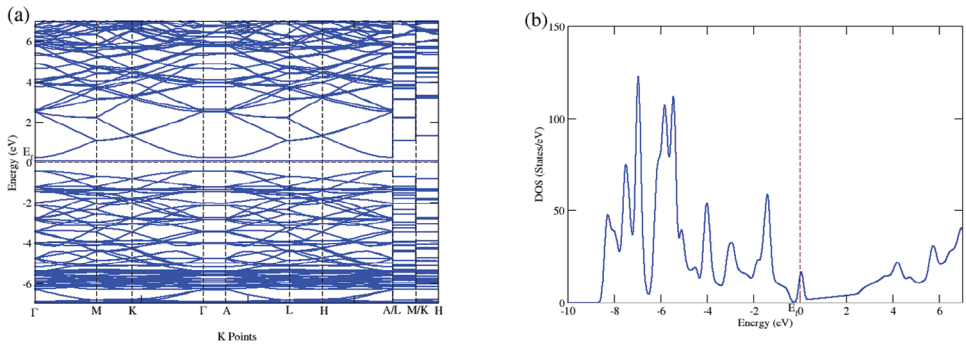


Figure 3. NO adsorption on ZnO surface a) Band structure and b) TDOS.

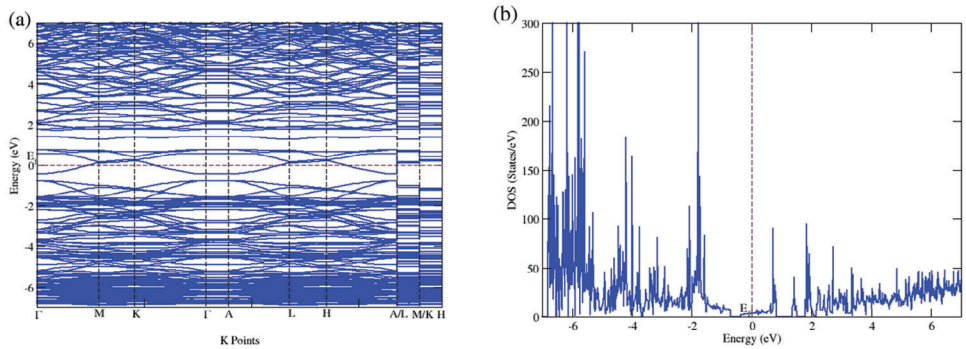


Figure 4. a) Band structure of Ce-doped ZnO b) Total density of states of Ce-doped ZnO.

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

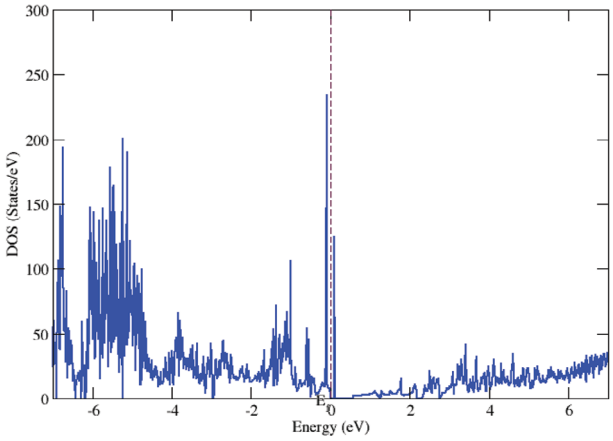


Figure 5. The total density of states for NO adsorption on Ce-doped ZnO surface.

the conduction band for the Ce-doped ZnO surface. The occupied states near the Fermi energy level are due to the Ce 4f orbitals.

3.4.3. NO Adsorption on Ce-doped ZnO Surface

After adsorption of NO gas on the Ce-doped ZnO surface the bandgap increases as shown in **Figure** 5. Also, an additional peak was observed in the region of the bandgap. The peak intensity at a higher value for the Ce-doped ZnO surface and after NO adsorption the peak intensity decreases showing the interaction of NO molecule with the Ce-doped ZnO surface.

4. PDOS (Partial Density of States) Analysis

To further confirm the gas-sensing properties on the bare and doped surface of ZnO we calculate the partial density of states. **Figure** 6a is the PDOS for the bare ZnO-NO surface, in this Figure, blue, red, green, and black lines represent the PDOS of ZnO-NO, Zn, O, and N respectively. Figure 6b is the PDOS for the Ce-doped ZnO surface, in this figure blue, red, green, and orange lines represent the PDOS of ZnO-Ce, Zn, O, and Ce respectively. Figure 6c is the PDOS for the NO molecule adsorbed on the Ce-doped surface, in this figure blue, red, green, black, and orange lines represent the PDOS of ZnO-Ce-NO, Zn, O, N, and Ce.

From the ZnO-NO PDOS, we conclude that there is a small no. of peaks for the nitrogen atom and their peak intensity is very low. But after doping of Ce on the ZnO surface the nitrogen atom shows more intense peaks in comparison to the ZnO-NO PDOS which represents that after doping of Ce, the nitrogen atom interacts more strongly with the surface (Ce-doped ZnO) which is consistent with the Figure 1d. From Figure 6a,b it is observed that the peak value is maximum at energy ~−1.8 eV but in Figure 6c at this energy value the peak value decreases which shows that NO molecule interacts strongly with the Ce-doped ZnO surface. The Figure 6b represents that the after doping of Ce in ZnO layer the gap between valance band and conduction band decreases that shows after doping of Ce conductance increases. The NO gas molecule interaction on the Ce-doped ZnO surface affects the gap between the valance band and conduction band this is due to the interaction. The fermi energy level shifts toward the valance band that represents the NO behaving as an acceptor molecule which is also discussed in the charge transfer section.

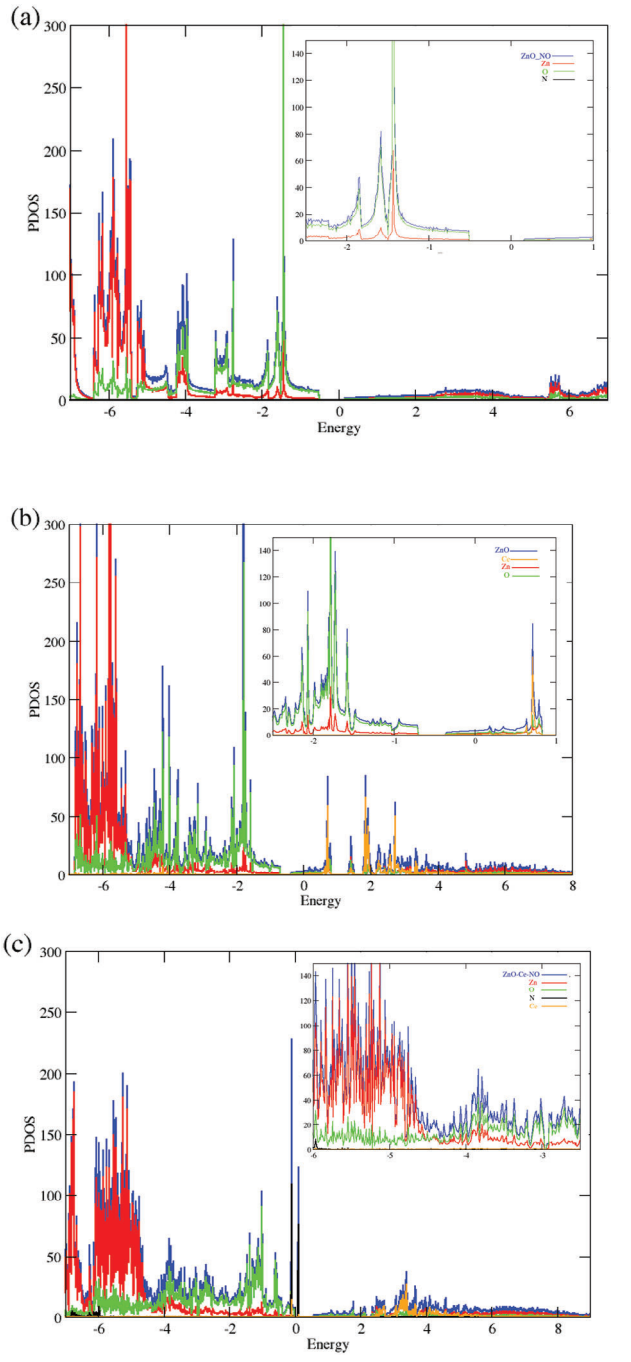


Figure 6. PDOS graph (in all fermi energy set to zero energy level energy in eV and PDOS in states/eV) a) NO adsorbed on ZnO b) Ce-doped ZnO surface c) NO adsorbed on Ce-doped ZnO surface, inset image represents the energy value that is discussed in section.

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

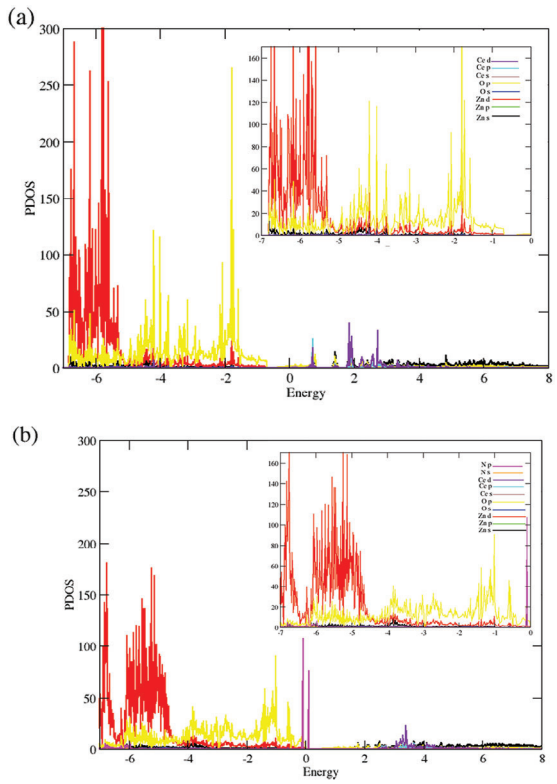


Figure 7. Orbital PDOS Analysis (in both figures fermi energy is set to zero energy level) a) Ce-doped ZnO surface and b) NO on the Ce-doped ZnO surface, In both figures' energy in eV and PDOS in states/eV. Inset images represent discussed energy values.

4.1. Orbital PDOS

For the confirmation of the interaction between orbitals, we also observed the partial density of states for orbitals. From **Figure** 7a which represents the PDOS of the Ce-doped ZnO surface, Zn d orbital energy peaks are in the energy range between −0.9 to −6 eV. After the adsorption of the NO molecule, the peaks are observed beyond −0.9 up to 0 eV. Also the peak intensity decreases. After adsorption of the NO molecule, the peaks of PDOS shifted to positive energy values which indicates that the NO molecule accepts the electrons from the surface.[39] The PDOS and TDOS of the Ce-doped ZnO surface indicate that the fermi energy level shifts into the conduction band as a result of Ce doping in pure ZnO. This is known as the Burstein-Moss effect.[40] After the adsorption of the NO molecule on the Ce-doped ZnO surface (Figure 7b) the fermi energy level moves into the valence band. From the charge transfer study, we observed NO molecules accept the electrons which increases the electron density and the fermi energy level will move to the valence band. In TDOS (NO adsorption on Ce-doped ZnO) the peaks observed near the fermi energy level are due to interaction between the N 2p orbital and Zn 3d orbital.

5. Conclusion

In this study zinc oxide (ZnO) and Ce-doped ZnO were used as sensing materials to calculate the NO gas sensing properties. For NO gas sensing properties, we employed Density Functional Theory (DFT) calculations by the Quantum Espresso Software. By analysing the adsorption energy, charge density difference (CDD), Bader charge, Band structure, DOS structure, and PDOS structure we found the following conclusion:

i. The adsorption energy of NO molecule on the Ce-doped ZnO surface increases from −0.42 to −0.92 eV which indicates that the Ce-doped ZnO surface is more favorable for NO gas sensing.
ii. The charge transfer value increases from −0.02 to −0.74e after doping of the Ce atom into the ZnO surface which indicates after doping NO accepts more electrons from the surface.
iii. The fermi energy level moves into the conduction band after doping of the Ce atom which indicates N-type behavior.
iv. The PDOS analysis confirms the Ce-doped ZnO surface is more favorable for NO gas sensing.

Acknowledgements

The authors acknowledge the DST PURSE grant no. (SR/PURSE/2021/77) awarded to Amity University Rajasthan, Jaipur, India by the Govt of India for the financial support and ACOAST, Amity University Rajasthan, Jaipur, India for computational facilities.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

bader charge, DFT, gas sensing, metal oxide semiconductor, REE

Received: June 5, 2024
Revised: July 6, 2024
Published online:

[1] C. Baratto, G. Sberveglieri, A. Onischuk, B. Caruso, S. Di Stasio, *Sens Actuators B Chem* **2004**, *100*, 261.
[2] J. Wang, D. Chen, W. Huang, N. Yang, Q. Yuan, Y. Yang, *Exploration* **2023**, 20210027.
[3] K.-Y. Chen, Y.-T. Xu, Y. Zhao, J.-K. Li, X.-P. Wang, L.-T. Qu, *Nano Mater. Sci.* **2023**, *5*, 247.
[4] L. Zhu, W. Zeng, *Sens. Actu. A Phys* **2017**, *267*, 242.
[5] F. Gyger, M. Hübner, C. Feldmann, N. Barsan, U. Weimar, *Chem. Mater.* **2010**, *22*, 4821.
[6] X. Wang, N. Miura, N. Yamazoe, *Sens. Actu. B Chem.* **2000**, *66*, 74.
[7] S. Gupta, A. Joshi, M. Kaur, *J. Chem. Sci.* **2010**, *122*, 57.
[8] H. N. Hieu, N. M. Vuong, H. Jung, D. M. Jang, D. Kim, H. Kim, S.-K. Hong, *J. Mater. Chem.* **2012**, *22*, 1127.
[9] K. Madhusoodanan, T. Vimalkumar, K. Vijayakumar, *Bull. Mater. Sci.* **2015**, *38*, 583.

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[10] T. A. Kusumam, V. Siril, K. Madhusoodanan, M. Prashantkumar, Y. Ravikiran, N. Renuka, *Sens. Actu. A Phys.* **2021**, *318*, 112389.
[11] C.-C. Liu, J.-H. Li, C.-C. Chang, Y.-C. Chao, H.-F. Meng, S.-F. Horng, C.-H. Hung, T.-C. Meng, *J. Phys. D: Appl. Phys.* **2009**, *42*, 155105.
[12] C.-Y. Lin, J.-G. Chen, W.-Y. Feng, C.-W. Lin, J.-W. Huang, J. J. Tunney, K.-C. Ho, *Sens. Actu. B Chem.* **2011**, *157*, 361.
[13] T. Çorlu, I. Karaduman, S. Galioglu, B. Akata, M. A. Yıldırım, A. Ateş, S. Acar, *Mater. Lett.* **2018**, *212*, 292.
[14] A. K. Singh, C.-C. Yen, C.-F. Wen, R.-H. Horng, D.-S. Wuu, *ACS Appl. Electron. Mater.* **2023**, *5*, 2574.
[15] N. Sharma, S. P. Choudhury, *Mater. Sci. Eng. B* **2024**, *307*, 117505.
[16] G. El Fidha, N. Bitri, S. Mahjoubi, F. Chaabouni, E. Llobet, J. Casanova-Chafer, *Sensors* **2022**, *22*, 5173.
[17] C.-J. Chang, C.-Y. Lin, J.-K. Chen, M.-H. Hsu, *Ceram. Int.* **2014**, *40*, 10867.
[18] K. Sun, G. Zhan, L. Zhang, Z. Wang, S. Lin, *Sens. Actu. B Chem.* **2023**, *379*, 133294.
[19] A. S. Mokrushin, I. A. Nagornov, T. L. Simonenko, N. P. Simonenko, P. Y. Gorobtsov, I. A. Arkhipushkin, E. P. Simonenko, V. G. Sevastyanov, N. T. Kuznetsov, *Appl. Surf. Sci.* **2022**, *589*, 152974.
[20] F. Sarf, I. Karaduman Er, E. Yakar, S. Acar, *J. Mater. Sci.: Mater. Electron.* **2020**, *31*, 10084.
[21] N. Sharma, S. P. Choudhury, *Mod. Phys. Lett. B* **2024**, *38*, 2342012.
[22] M. H. Rahman, M. Z. Rahaman, E. H. Chowdhury, M. Motalab, A. A. Hossain, M. Roknuzzaman, *Mol Syst Des Eng* **2022**, *7*, 1516.
[23] Q. Yuan, Y.-P. Zhao, L. Li, T. Wang, *J. Phys. Chem. C* **2009**, *113*, 6107.
[24] M. J. Spencer, I. Yarovsky, *J. Phys. Chem. C* **2010**, *114*, 10881.
[25] A. G. Saputro, M. K. Agusta, B. Yuliarto, H. K. Dipojono, F. Rusydi, R. Maezono, *Appl. Surf. Sci.* **2017**, *410*, 373.
[26] Y. V. Kaneti, Z. Zhang, J. Yue, Q. M. Zakaria, C. Chen, X. Jiang, A. Yu, *Phys. Chem. Chem. Phys.* **2014**, *16*, 11471.
[27] M. Liangruksa, T. Laomettachit, C. Siriwong, *Mater. Res. Express* **2021**, *8*, 045004.
[28] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, *J. Phys. Condens. Matter* **2009**, *21*, 395502.
[29] P. Kratzer, J. Neugebauer, *Front. Chem.* **2019**, *7*, 106.
[30] W. Gao, Y. Chen, B. Li, S.-P. Liu, X. Liu, Q. Jiang, *Nat. Commun.* **2020**, *11*, 1196.
[31] M. Breedon, M. Spencer, I. Yarovsky, *Surf. Sci.* **2009**, *603*, 3389.
[32] T. Yang, W. Jin, Y. Liu, H. Li, S. Yang, W. Chen, *Appl. Surf. Sci.* **2018**, *457*, 975.
[33] Y. Xia, A. Pan, Y.-Q. Su, S. Zhao, Z. Li, A. K. Davey, L. Zhao, R. Maboudian, C. Carraro, *Sens. Actu. B Chem.* **2022**, *357*, 131359.
[34] A. Tamvakos, K. Korir, D. Tamvakos, D. Calestani, G. Cicero, D. Pullini, *ACS Sens.* **2016**, *1*, 406.
[35] Y.-H. Zhang, L.-J. Yue, F.-L. Gong, F. Li, H.-L. Zhang, J.-L. Chen, *Vacuum* **2017**, *141*, 109.
[36] M. Shajahani, H. R. Dizaji, *Chem. Phys. Lett.* **2023**, *832*, 140884.
[37] Y. Chen, Y. Gui, X. Chen, *Results Phys.* **2021**, *28*, 104680.
[38] A. Umar, S. Akbar, R. Kumar, J. N. O. Amu-Darko, S. Hussain, A. A. Ibrahim, M. A. Alhamami, N. Almehbad, T. Almas, A. F. Seliem, *Chemosphere* **2024**, *349*, 140838.
[39] X. Gao, Y. Zhou, Y. Tan, Z. Cheng, Q. Tang, J. Jia, Z. Shen, *Energy Fuels* **2018**, *32*, 5331.
[40] J.-Q. Wen, J.-M. Zhang, Z.-G. Qiu, X. Yang, Z.-Q. Li, *Phys. B* **2018**, *534*, 44.

© 2024 Wiley-VCH GmbH

15214117, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppsc.202400126 by Delhi Technological University, Wiley Online Library on [13/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Generalized Switched-Capacitor-Based Hybrid Multilevel Inverter With Reduced Components Count and Inrush Current

Niraj Kishore, Kapil Shukla, *Senior Member, IEEE*, and Nitin Gupta, *Senior Member, IEEE*

***Abstract*— This article presents a novel topology for a three-phase switched-capacitor (SC) based hybrid multilevel inverter (HMLI) with boosted output voltage. The proposed topology employs generalised structure of SC-based inverters (SCBIs) to improve the output voltage levels utilising SC cells. The proposed structure additionally features self-charging and voltage-balancing capabilities of the SCs, without the need of auxiliary circuit/sensor. A modified PWM (MPWM) technique is utilized to modulate the system. The MPWM results in improved output voltage profile and reduction in voltage ripple across the SCs. The proposed structure is quantitatively compared with the state-of-the-art topologies to demonstrate its advantages in the terms of reduced components count, low maximum blocking voltage (MBV), low total standing voltage (TSV), lessened total harmonic distortion (THD), lowered cost function (CF) per level, and boosted output voltage. The performance of the proposed topology is verified in MATLAB/Simulink environment and a laboratory prototype is developed to confirm the feasibility to operate at steady-state and dynamic conditions.**

***Index Terms*— Level-shifted-carriers (LSCs), switched-capacitor (SC), switched-capacitor-based inverter (SCBI), two-level three-phase inverter (TTI).**

I. Introduction

MULTILEVEL inverters (MLIs) are the most advanced power electronics converters known for owing several benefits over conventional two-level inverters, such as low peak inverse voltage (PIV) as compared to the operating voltage, better harmonic profile, reduced dv/dt stress on the switches, modular structure, etc. [1], [2]. MLIs are well-suited for low-, medium-, and high-power applications that are commonly used in various industries, such as variable speed electric motor drives, uninterruptible power supply (UPS) systems, reactive power compensators, electric vehicles, etc. [3], [4], [5]. In renewable energy integration, particularly those utilising photo-voltaic panels, it is required to boost the input voltage for medium- and high-power application in addition to delivering higher quality power to the grid [6]. The conventional MLIs: 1) neutral-point clamped MLI (NPC-MLI); 2) flying-capacitor MLI (FC-MLI); and 3) cascaded H-bridge MLI (CHB-MLI), which have gained popularity in industry and academic research. However, NPC- & FC-MLI suffer from the problems of complex clamped circuits and unbalanced capacitors' voltage due to the series connection of power switches or capacitors to withstand DC-bus voltage. On the other side, the requirement of large number of power switches and DC-sources continues to be an area of concern for CHB-MLI. Furthermore, CHB-MLI designed with single DC-source requires line frequency transformer(s) for isolation [7], [8], which increases the cost and reduces the power density.

Due to the aforementioned issues, the switched-capacitor MLI (SC-MLI) topology is gaining popularity. SC-MLI topologies receive more attention due to their features like (i) high voltage gain (ii) self-balancing capacitor's voltage, and (iii) reduced DC sources. Many researchers have developed a significant number of SCMLIs and various types of boost inverters topologies, even though the large number of components count, high voltage stress across the switches, higher voltage-rated capacitors remain the challenges [9], [10], [11], [12], [13]. Numerous configurations for 13-level with high voltage gain have been described in [14], [15], [16], [17], [18], [19], [20], and [21]. The topology proposed in [15] uses more passive components and active devices. The power switches employed in [16], [17], and [18] are subjected to high voltage stress i.e. equal to the output voltage. Each topology experiences some limitations like requirement of several DC-sources in [16], two higher-rated capacitors in [17], and a large number of on-state switches in [18]. The SCMLI is subjected to high spikes in the source-current due to the parallel connection of SCs with the voltage-source during charging [17], [19], [21]. The high inrush current takes place due to the requirement of large energy-stored capacitor(s). Furthermore, SC-MLIs suffer disadvantages like voltage balancing issue of SCs at lower modulation indices (m), requirement of large-sized SCs for generalised structure, etc. Hence, SCMLIs are evaluated on the basis of parameters like total standing voltage (TSV) per unit, maximum blocking voltage (MBV) per unit, output voltage gain (V_{Gain}), cost function (CF) per level, size of SCs, etc.

In order to achieve cost-effective solution, a generalised structure of three-phase SC-based hybrid MLI (SC-HMLI) has been proposed in this article with level-shifted-carriers modified-PWM (LSCs-MPWM) technique [22]. The proposed structure consists of three power-modules, i.e. two SC-based inverter (SCBIs) and one two-level three-phase inverter (TTI). The SCBIs have generalised structure, which can be extended further on employing SC-cells in order to improve the output

Manuscript received 16 March 2024; revised 22 May 2024 and 14 July 2024; accepted 8 August 2024. This article was recommended by Associate Editor M. Wang. *(Corresponding author: Niraj Kishore.)*

The authors are with the Electrical Department, Malaviya National Institute of Technology, Jaipur, Rajasthan 302017, India (e-mail: 2020ree9543@mnit.ac.in).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSI.2024.3443188.

Digital Object Identifier 10.1109/TCSI.2024.3443188

1549-8328 © 2024 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 August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

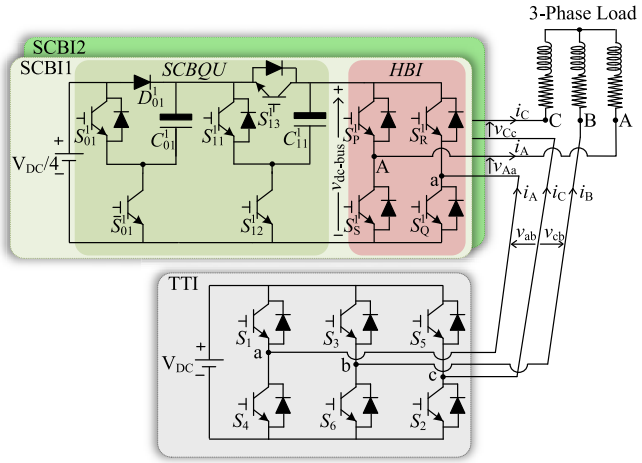


Fig. 1. Circuit diagram of proposed 17-level inverter.

voltage profile. The proposed inverter topology overcomes the disadvantages of the conventional MLIs as it inherits the following features:

1) The proposed topology requires least components per output voltage level.
2) Boosted output voltages are generated with low TSV & MBV.
3) The SCs' voltage remain balanced even at lower modulation indices.
4) The voltage ripple across the SCs is reduced by the factor of $\sqrt{3}/2$.
5) A significant proportion exceeding 50% of switches operate at line frequency (50 Hz) and the remaining switches operating at high frequency experience low voltage stress, hence leading to reduction in overall switching losses.
6) No additional circuit/sensor(s) is/are required for balancing the SCs' voltages.
7) The SCBIs can be extended further to increase the output voltage levels.

Furthermore, the proposed structure implements LSCs-MPWM technique, which improves the performance of the inverter. The redundancy states of the switches are utilised to reduce the voltage ripple across the SCs as well as balance the voltage.

II. Proposed Hybrid MLI Topology

A. Circuit Description

Hybrid-power-modules based proposed MLI is shown in Fig. 1 for three-phase 17-level (17L) output line voltage. The proposed structure is comprised of one TTI and two SCBIs. The SCBI consists one switched-capacitor-based quadruple unit (SCBQU) and one H-bridge inverter (HBI) for 17L inverter. Each quadruple unit consists of five switches, one diode, and two capacitors which generates $V_{DC}/4$, $V_{DC}/2$, $3V_{DC}/4$, and V_{DC} voltages across the $v_{dc\text{-}bus}$ using modified PWM techniques [23]. MPWM technique utilises three references and four LSCs signals to generate gate signals for three-phase 17L output line voltages. The TTI generates 3-level (peak-peak) line voltages, whereas SCBIs generates 9-level (peak-peak). The output voltage of the SCBIs (v_{Aa} & v_{Cc}) are cascaded with the output line voltages (v_{ab} & v_{cb}) of TTI, respectively to generate balanced three-phase output line voltages given by following equations:

TABLE I
Switching States of the Proposed 17-Level Inverter and Their Effects on SCs

States	S^1_{01}	S^1_{12}	S^1_{11}	S^1_{13}	S^1_P $\bar{S}^1_S$	S^1_R $\bar{S}^1_Q$	S_1 $\bar{S}_4$	S_3 $\bar{S}_6$	C^1_{01}	C^1_{11}	$v_{dc\text{-}bus}$	v_{AB}
V_{L8}	1	0	1	0	1	0	1	0	D↓	D↓	$+V_{DC}$	$+2V_{DC}$
V_{L7}	0	0	1	0	1	0	1	0	C↑	D↓	$+3V_{DC}/4$	$+7V_{DC}/4$
V_{L6}	1	1	0	0	1	0	1	0	D↓	C↑	$+V_{DC}/2$	$+3V_{DC}/2$
V_{L5}	0	0	0	1	1	0	1	0	C↑	I	$+V_{DC}/4$	$+5V_{DC}/4$
V_{L4}	1	0	1	0	1	0	0	0	D↓	D↓	$+V_{DC}$	$+V_{DC}$
V_{L3}	0	0	1	0	1	0	0	0	C↑	D↓	$+3V_{DC}/4$	$+3V_{DC}/4$
V_{L2}	1	1	0	0	1	0	0	0	D↓	C↑	$+V_{DC}/2$	$+V_{DC}/2$
V_{L1}	0	0	0	1	1	0	0	0	C↑	I	$+V_{DC}/4$	$+V_{DC}/4$
V_{L0}	1	1	0	0	0	0	0	0	D↓	C↑	$+V_{DC}/2$	0V
$-V_{L1}$	0	0	0	1	0	1	1	1	C↑	I	$+V_{DC}/4$	$-V_{DC}/4$
$-V_{L2}$	1	1	0	0	0	1	1	1	D↓	C↑	$+V_{DC}/2$	$-V_{DC}/2$
$-V_{L3}$	0	0	1	0	0	1	1	1	C↑	D↓	$+3V_{DC}/4$	$-3V_{DC}/4$
$-V_{L4}$	1	0	1	0	0	1	1	1	D↓	D↓	$+V_{DC}$	$-V_{DC}$
$-V_{L5}$	0	0	0	1	0	1	0	1	C↑	I	$+V_{DC}/4$	$-5V_{DC}/4$
$-V_{L6}$	1	1	0	0	0	1	0	1	D↓	C↑	$+V_{DC}/2$	$-3V_{DC}/2$
$-V_{L7}$	0	0	1	0	0	1	0	1	C↑	D↓	$+3V_{DC}/4$	$-7V_{DC}/4$
$-V_{L8}$	1	0	1	0	0	1	0	1	D↓	D↓	$+V_{DC}$	$-2V_{DC}$

D↓: Discharging of SCs, C↑: Charging of SCs, Idle: I

$$v_{AB} = v_{ab} + v_{Aa} \quad (1)$$

$$v_{BC} = v_{bc} + (-v_{Cc}) \quad (2)$$

Using (1) & (2), the third line voltage (v_{CA}) is given as:

$$v_{CA} = -(v_{AB} + v_{BC}) \quad (3)$$

B. Analysis of Operating Modes and Voltage Levels

The MPWM technique is used to generate the gate signals for SCBIs, whereas TTI operates in 180-degree conduction mode. Gate signals are generated in accordance with the output voltage levels as shown in Table I. The 'ON' and 'OFF' statuses of the switches are indicated as '1' and '0', respectively.

Figs. 2(a)-2(e) show the five different conduction states of the switches of the SCBI1 for positive half cycle of the output line voltage. SCs are marked with the green, red, and grey colour to show the charging, discharging, and idle states of the SCs, respectively. The charging path of SC is shown by dotted green line, whereas dotted red line shows the discharging path of SCs as well as load current. Dotted blue lines show the flow of negative current during reactive load.

In order to facilitate the analysis, all the components are assumed to be ideal. Moreover, C^1_{01} and C^1_{11} are kept large enough to keep their voltages constant at $V_{DC}/4$ and $V_{DC}/2$, respectively. Four different output voltages are generated across the dc-bus terminals by connecting SCs in series and parallel with DC-source as shown in Figs. 2(a)-2(d). The SCs with higher voltage level gets less time for charging as compared to lower voltage level. As a result, zero output

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

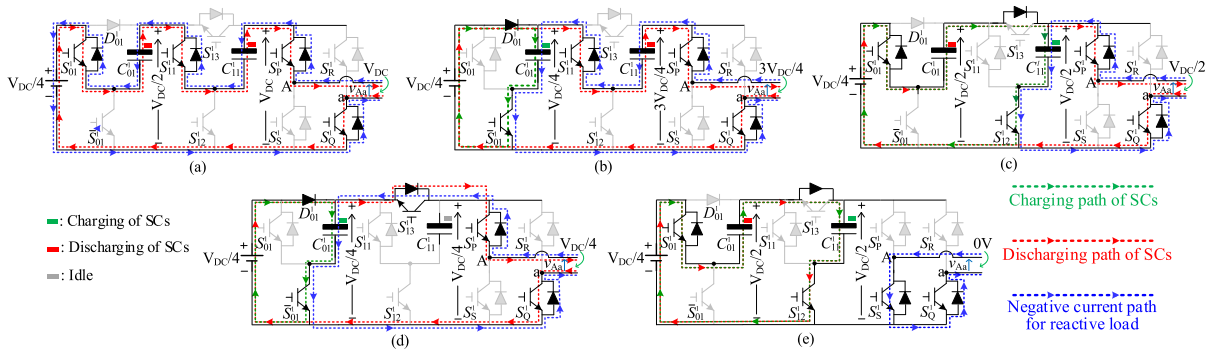


Fig. 2. Operating modes and active switches to generate positive output voltage levels for SCBI1. (a) $v_{\mathrm{Aa}} = +\mathrm{V_{DC}}$. (b) $v_{\mathrm{Aa}} = +3\mathrm{V_{DC}}/4$. (c) $v_{\mathrm{Aa}} = +\mathrm{V_{DC}}/2$. (d) $v_{\mathrm{Aa}} = +\mathrm{V_{DC}}/4$. (e) $v_{\mathrm{Aa}} = 0$.

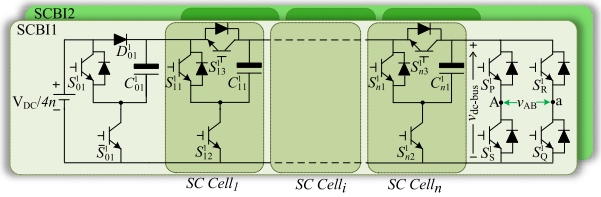


Fig. 3. Generalized structure of SCBIs.

TABLE II
SWITHING STATES OF SCBIS FOR $i = n$

'ON' states of switches	$v_{\mathrm{dc\text{-}bus}}$
S^k_{01},S^k_{11},S^k_{21},.....,S^k_{n1},S^k_P,S^k_Q	V_{DC}
S^k_{02},S^k_{11},S^k_{21},.....,S^k_{n1},S^k_P,S^k_Q	$(4n-1)V_{\mathrm{DC}}/4n$
S^k_{01},S^k_{12},S^k_{21},.....,S^k_{n1},S^k_P,S^k_Q	$(4n-2)V_{\mathrm{DC}}/4n$
S^k_{02},S^k_{13},S^k_{21},.....,S^k_{n1},S^k_P,S^k_Q	$(4n-3)V_{\mathrm{DC}}/4n$
.	.
.	.
S^k_{01},S^k_{11},S^k_{21},.....,S^k_{n2},S^k_P,S^k_Q	$V_{\mathrm{DC}}/2$
.	.
.	.
S^k_{01},S^k_{11},S^k_{22},.....,S^k_{n3},S^k_P,S^k_Q	$4V_{\mathrm{DC}}/4n$
S^k_{02},S^k_{11},S^k_{23},.....,S^k_{n3},S^k_P,S^k_Q	$3V_{\mathrm{DC}}/4n$
S^k_{01},S^k_{12},S^k_{23},.....,S^k_{n3},S^k_P,S^k_Q	$2V_{\mathrm{DC}}/4n$
S^k_{02},S^k_{13},S^k_{23},.....,S^k_{n3},S^k_P,S^k_Q	$V_{\mathrm{DC}}/4n$

voltage level states are utilised to charge the SC with higher most voltage level as shown in Fig. 2(e).

C. Extension of Proposed Topology

One of the major advantage of the proposed topology is the capability to increase the voltage levels by merely cascading the SC cells as shown in Fig. 3. Each cell is consist of three power switches (S^k_{i1}, S^k_{i2}, & S^k_{i3}) and one capacitor (C^k_{i1}), where i=$\{1, 2, 3, \ldots, n\}$ and k=$\{1, 2\}$ represent SC cells and SCBIs, respectively. Each SCBI generates $(8n+1)$-levels (peak-peak) of output voltage ($v_{\mathrm{Aa.}}$). Hence, the generalised structure is proposed to utilise the available DC voltage sources not necessarily available in the ratio of 1:1:4.

On cascading the output voltage of SCBIs with TTI, the proposed structure generates $(16n+1)$-levels (peak-peak) of line voltage (v_{AB}). The switching states for i=n is tabulated in Table II for the output voltage gain of $4n$ & $4n/(2n+1)$ for SCBIs and TTI, respectively. Despite the increment in the number of levels, the maximum blocking voltage (MBV) per unit and the total standing voltage (TSV) per unit remain the same. Hence, the proposed topology is more economical as far as other generalised similar structures are concerned for single-phase and three-phase extension.

III. ANALYSIS OF PROPOSED INVERTER WITH MPWM

A. LSCs-MPWM Strategy

As discussed earlier, the proposed structure can be extended for n-number of SC-cells. For n-number of SC-cells, three-phase sinusoidal signals and $4n$ level-shifted triangular signals are required to generate $16n+1$ levels of line voltage.

Three-phase sinusoidal reference signals ($v_{X,Ref.}$, $v_{Y,Ref.}$, & $v_{Z,Ref.}$) have been employed and logically compared to generate gate signals for TTI. Whereas, each SCBI requires a modified reference signal and $4n$ triangular carrier-signals. The modified reference signal ($|v_{X,Ref.}*|$) introduced in [22] is used to generate the gate signals for SCBI1 at different modulation index (m), which is defined as the ratio of peak value of the modified reference signal to the peak value of upper carrier signal. In the similar way, the switches of SCBI2 are gated using another modified reference signal ($|v_{Y,Ref.}*|$) and same level-shifted triangular carrier signals.

For SCBIs using single SC cell (n=1), four level-shifted triangular carrier signals (Cr_1−Cr_4) having same frequency and amplitude are compared with modified reference signal to get C_0, C_1, C_2, & C_3 as the comparators' output, which are logically operated to develop logic levels (L_0=$\bar{C}_0$, L_1=$C_0\bar{C}_1$, L_2=$C_1\bar{C}_2$, L_3=$C_2\bar{C}_3$, & L_4=C_4) in correspondence to the output voltage levels of SCBIs. Since, rectified reference signals are used to generate the gate for SCBIs, two complementary signals C_p & C_n are used for gating the switches of SCBIs in order to alternate the output voltage. Logic levels are further logically operated with due consideration of switching states of the Table I to generate the gate signals for SCBIs given by the following switching function:

$$v_{g,S^k_{01}} = (L_0 + L_2 + L_4) \times (C_p + C_n) \quad (4)$$

$$v_{g,S^k_{12}} = (L_0 + L_2) \times (C_p + C_n) \quad (5)$$

$$v_{g,S^k_{11}} = (L_3 + L_4) \times (C_p + C_n) \quad (6)$$

$$v_{g,S^k_{13}} = (L_1) \quad (7)$$

$$v_{g,S^k_P} = v_{g,S^k_Q} = v_{g,\bar{S}^k_R} = v_{g,\bar{S}^k_S} = (L_0) \times C_n + C_p \quad (8)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

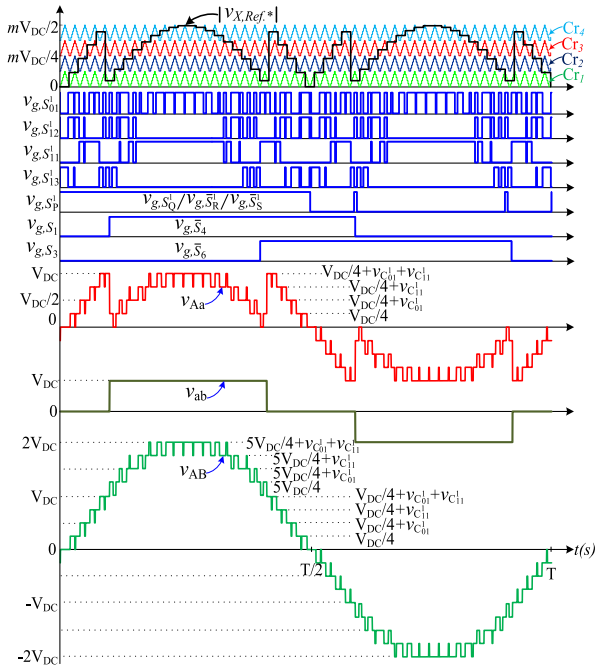


Fig. 4. Gate signals and 17-level line voltage (v_{AB}) generation by cascading the outputs of SCBI1 & TTI for single SC cell (n = 1).

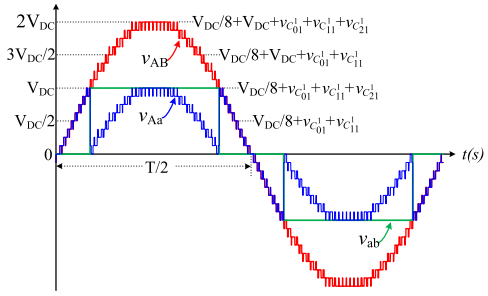


Fig. 5. 33-level line voltage (v_{AB}) generation by cascading the outputs of SCBI1 & TTI using two SC cells (n = 2).

where '+' and '×' denotes the boolean 'OR' and 'AND' function, respectively.

Fig. 4 shows the gate signals for the switches of TTI and SCBI1 for the formation of 17L output line voltage v_{AB} on cascading the output voltage (v_{Aa}) of SCBI1 with the output line voltage (v_{ab}) of TTI for $n = 1$. For two SC cells ($n = 2$), SCBI1 generates 17L output voltage (v_{Aa}), which goes on cascading with the output of TTI to develop 33-levels line voltage (v_{AB}) as shown in Fig. 5, keeping magnitude of output voltage same.

The simulation results of three-phase output voltage and current are shown in Fig. 6 for single SC cell (n = 1). The Fig. 6(a) shows the generation of three-phase 17L output line voltage, whereas the Fig. 6(b) shows 23L output phase voltage. The three-phase load current for RL-load is shown by Fig. 6(c). From the Fig. 6, it is evident that the proposed topology generates three-phase-balanced output voltage and current.

B. Sizing of SCs

The capacitance of the SCs depends on the longest discharging time (LDT) for C_{01}^k & C_{11}^k, which takes place for output line voltage levels of $2V_{DC}$ for the duration of

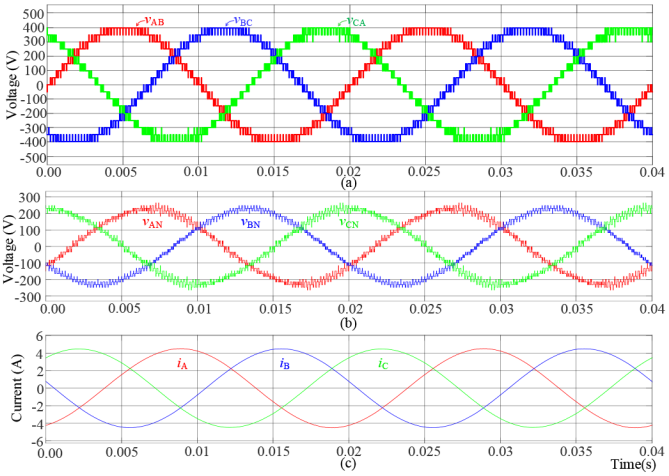


Fig. 6. (a) Generation of 17-level three-phase line voltage; (b) three-phase 23-level output phase voltage; (c) three-phase load-current for RL-load.

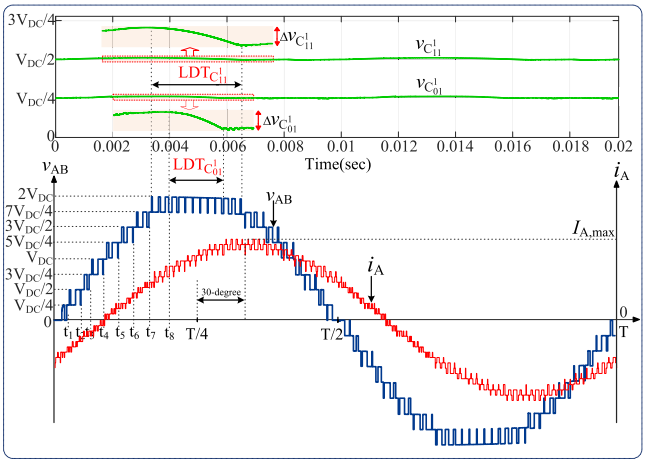


Fig. 7. SCs' voltage ($v_{C_{01}^1}$ & $v_{C_{11}^1}$), line voltage (v_{AB}), and line current (i_A) waveforms for resistive load.

$(T/2 - 2t_8)$ & $(T/2 - 2t_7)$, respectively as shown in Fig. 7. Therefore, the amount of charge discharged during LDTs for the SCs depends on the current flowing through the capacitors and the types of load. The capacitance of the SCs are calculated considering resistive load as it gives the maximum voltage ripple across the SC than any other load. For three-phase R-load system, the capacitor current (i.e load current during LDT) lags the line voltage by 30-degree. Hence, the following equations are used to calculate the capacitance [15].

$$\triangle Q_{C_{01}^1} = \int_{t_8}^{T/2-t_8} I_{A,max} sin(\omega t - \pi/6) d(t) \quad (9)$$

$$\triangle Q_{C_{11}^1} = \int_{t_7}^{T/2-t_7} I_{A,max} sin(\omega t - \pi/6) d(t) \quad (10)$$

As a result, the minimum capacitance (C_{min}) for permitted voltage ripple (V_{ripple}) are calculated as follows:

$$C_{01,min}^1 \geq \frac{2V_{DC}/\sqrt{3}}{\omega R_L \triangle V_{ripple}\%} \times 2(\cos \omega t_8)(\cos \pi/6) \quad (11)$$

$$C_{11,min}^1 \geq \frac{2V_{DC}/\sqrt{3}}{\omega R_L \triangle V_{ripple}\%} \times 2(\cos \omega t_7)(\cos \pi/6) \quad (12)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

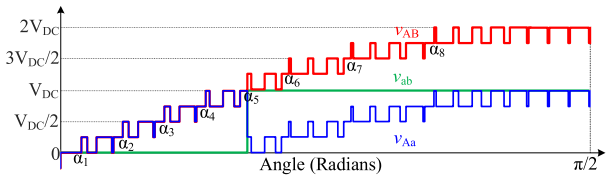


Fig. 8. Output voltage waveforms of SCBI1 (v_{Aa}), TTI (v_{ab}), and line v_{AB} using single SC cell (n = 1).

From the equations (11) & (12), the size of the SCs are reduced by the factor of $\sqrt{3}/2$ due to the phase difference of 30-degree between the line current (i.e. SCs' current during LDT) and the line voltage for R-load. If the size of the SCs are kept same, then the voltage ripple across the SCs is reduced by the same factor. Consequently, the reduced voltage ripple across the SCs reduces the inrush current.

C. Power Distribution Analysis

Power distribution among the power module is one of the important aspects of analysis and design of the proposed structure in the terms of 'm' [22]. For the analysis purpose, output voltage and currents are assumed to be sinusoidal. So

$$v_{\mathrm{AB}}(t) = mV_m \sin(\omega t) = m(2V_{\mathrm{DC}})\sin(\omega t) \tag{13}$$

$$i_{\mathrm{A}}(t) = I_m \sin\left(\omega t - \phi - \frac{\pi}{6}\right) \tag{14}$$

where ϕ is the load impedance angle. Active power rating of the inverter is given by:

$$P_n = \sqrt{3}\ V_L I_L \cos\phi = \sqrt{3}\ \frac{(2V_{\mathrm{DC}})}{\sqrt{2}} I_L \cos\phi \tag{15}$$

where V_L and I_L are the output line voltage and line current, respectively.

The output voltages of the power modules for quater-wave are shown in Fig 8. The per-unit active power shared by each module is discussed in [22]. Hence, the per-unit active power of the TTI in terms of modulation index (m) is given by:

$$P_{\mathrm{TTI(p.u.)}} = \frac{P_{\mathrm{TTI}}}{P_{n,Rated}} = \frac{\frac{4\sqrt{3}}{\pi\sqrt{2}}(mV_{\mathrm{DC}})I_A \cos(\alpha_5)\cos\phi}{\sqrt{6}\ V_{\mathrm{DC}} I_L}$$
$$= \frac{2}{\pi} m^2 \cos(\alpha_5) \times \cos\phi. \tag{16}$$

The per-unit active power of SCBI1 is given by:

$$P_{\mathrm{SCBI1(p.u.)}} = \frac{P_{\mathrm{SCBI1}}}{P_{n,Rated}} = \frac{2}{3\sqrt{3}\pi}\ m^2[\cos(\alpha_1) + \cos(\alpha_2) + \cos(\alpha_3) + \cos(\alpha_4) + \cos(\alpha_6) + \cos(\alpha_7) + \cos(\alpha_8) - 2\cos(\alpha_5)] \times \cos\left(\phi + \frac{\pi}{6}\right) \tag{17}$$

Using (16) and (17), the active power (p.u.) shared by TTI and SCBIs is shown in Figs. 9(a) & 9(b), respectively in the terms of 'm' and power factor. SCBIs share significantly less power as compared to TTI for higher value of 'm', as a result low-energy-stored SCs are required, which enhances the power density of the proposed structure.

Three types of losses are mainly observed in the power electronics converter, such as switching, conduction, and capacitor ripple losses. The efficiency of the proposed inverter

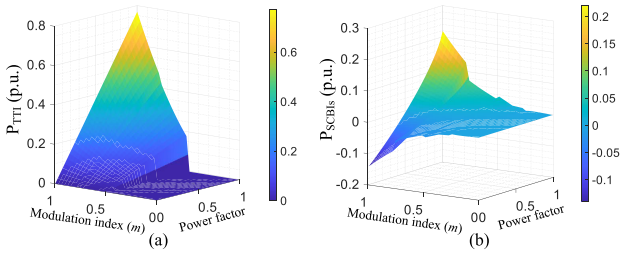


Fig. 9. Active power shared in the terms of modulation index (m) and power factor by: (a) TTI, and (b) SCBIs.

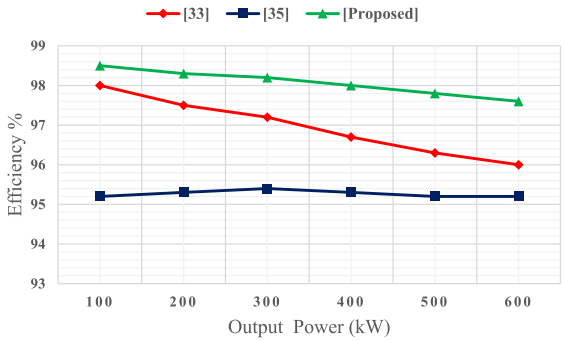


Fig. 10. Efficiency curves for different topology.

is evaluated on the basis of losses. For this, a thermal model of the proposed inverter is developed using the PLECS software. The efficiency curve for the proposed topology has been drawn and compared with the available data of the similar structure as shown in Fig. 10.

IV. Comparison With Other 17-Level Inverters

The recent and advanced existing switched-capacitors based MLI topologies with the ability to generate 17-level three-phase output voltages have been taken into the comparison with the proposed topology as shown in Table III. The proposed SC-based mutli-sourced topology has the advantage over single-sourced SCBIs topologies in the terms of power density. Single-sourced topologies require large number of SCs as well as large-sized capacitors, which hampers the power density for a particular kVA rating of an inverter as the rating of the DC-sources (i.e. kWh) remains the same.

A comparative study is carried out regarding the number of voltage sources (N_{DC}), switches (N_{S}), drivers (N_{Dr}), capacitors (N_{C}), and diodes (N_{D}). The MBV per unit ($\mathrm{MBV_{pu}}$), TSV per unit ($\mathrm{TSV_{pu}}$), voltage gain ($\mathrm{V_{Gain}}$), and cost function (CF) are also listed in the Table III. To make a fair comparison, the following relation of the CF introduced in [32] is utilised to evaluate the cost per level of the output voltage:

$$CF = \frac{N_{\mathrm{DC}} + N_{\mathrm{SW}} + N_{\mathrm{DR}} + N_{\mathrm{C}} + N_{\mathrm{D}} + TSV_{\mathrm{pu}}}{N_{\mathrm{L}}} \tag{18}$$

In order to conduct the comparative study between the proposed 17L three-phase structure and other state-of-the-art topologies having similar characteristics, the following factors have been taken into the account:

- The proposed structure stands first in the terms of least number of semiconductor devices (i.e. 24-IGBTs and 2-diodes for 17L output voltage) as well as requires least

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
COMPARISON AMONG THE PROPOSED HYBRID MLI AND OTHER ADVANCED THREE-PHASE MLI TOPOLOGIES

Ref.	N_L	N_{DC}	N_{SW}	N_{DR}	N_D	N_C	MBV_{pu}	TSV_{pu}	CF	$\eta\%$	V_{Gain}	*Remarks*
[15]	17	1	30	30	18	18	1.0	12.0	6.41	NA	8	High passive components counts with high voltage stress.
[24]	17	1	36	36	42	21	1.0	17.25	9.01	91.6% at 320 W	8	High active & passive components count with highvoltage on H-bridge switches.
[25]	17	1	117	117	21	21	0.13	14.64	17.15	91.7% at 900 W	8	Very high device count with low voltage stress on switches and capacitors.
[26]	17	1	51	48	0	15	0.5	16.14	7.71	96.1% at 1.2 kW	4	High switch count with medium voltage gain.
[27]	17	1	54	51	3	12	0.5	20.25	8.30	96.5% at 1 kW	4	High TSV and components count with medium voltage gain.
[28]	17	1	48	42	0	12	1.0	19.5	7.20	97.3% at 400 W	8	Switch count and TSV is high. No discussion on Inrush current.
[29]	17	2	54	54	6	18	1.0	21.75	9.16	91.0% at 260 W	4	High active & passive components counts with high voltage stress on switches and capacitors.
[30]	17	2	30	30	12	6	1.0	16.5	5.65	NA	2	High voltage stress on H-bridge switches with multi dc-source topology.
[31]	17	2	30	30	6	6	1.0	14.25	5.19	NA	2	High voltage stress on H-bridge switches with multi dc-source topology.
[32]	17	1	30	30	12	18	1.0	25.5	6.85	95.4% at 1 kW	8	Large number of capacitors with high voltage stress on the active components.
[33]	17	1	42	42	10	10	0.5	23.25	7.54	94.5% at 522 W	4	High active and passive components count with high TSV.
[34]	17	1	42	42	6	9	0.5	19.5	7.02	94.0% at 500 W	8	High active and passive components count with high TSV.
[35]	17	1	36	36	15	12	0.5	12.75	6.63	95.4% at 1 kW	8	MVB is low but active and passive components count is high.
Proposed	**17**	**3**	**24**	**24**	**2**	**4**	**0.5**	**9**	**3.88**	**97.8%** at 1.6 kW	**2**	Lesser number of diodes/capacitors/switches along with low MBV & TSV. Least CF.

$N_L/N_{DC}/N_{SW}/N_{DR}/N_D/N_C$: Number of voltage levels/dc-sources/switches/drivers/diodes/capacitors; $\eta\%$: Efficiency; V_{Gain}: Voltage gain

TABLE IV
COST COMPARISON WITH SIMILAR 3-ϕ 17-LEVEL INVERTERS [36]

Part	Part number	Voltage rating	Unit price($)	[26]	[27]	[31]	[35]	[Proposed] Con-, RB-IGBT
IGBTs	IRGP4072DPBF, FGW85N60RB	V_{DC}	3.94, 8.7	-	-	×6	×6	×4
	STGW30NC60VD, FGW85N60RB	$2V_{DC}$	4.57, 8.7	×36	×24	×6	×12	×6
	STGW30N90D	$3V_{DC}$	4.91	-	-	×6	-	-
	STGW30NC120HD	$4V_{DC}$	5.07	×15	×30	×12	×18	×14
Capacitors	LGY2A102MELZ35	V_{DC}	4.1	-	-	-	-	×2
	E36D101HLN222QA79U	V_{DC}	14.89	×6	×6	×3	×3	-
	E36D201HPN222MC54U	$2V_{DC}$	21.38	×9	×6	-	×6	×2
	E32D3B1HPN222MD79U	$3V_{DC}$	30.94	-	-	×3	-	-
	ERWE401LGN222MCD0U	$4V_{DC}$	37.57	-	-	-	×3	-
Diode	P600B-E3/54	V_{DC}	0.93	-	-	×3	×3	×2
	P600D-E3/54	$2V_{DC}$	0.93	-	×3	-	×6	-
	P600G-E3/54	$4V_{DC}$	0.93	-	-	×3	×6	-
Gate Driver	TLP250	-	1.89	×48	×51	×30	×36	×24
Total cost ($)	-	-	-	613.05	578.58	341.13	537.39	212.34, 231.38

number of conducting devices to generate output voltage levels. Consequently, the requirement of the gate-drivers is also least.

- The SCBIs play important roles in improving the output voltage profile, which utilise least number of low-energy-stored SCs due to the effective implementation of the MPWM technique and cascading of TTI with SCBIs. Hence, the voltage ripples across the SCs are reduced significantly according to (11) & (12).
- The CF is used to evaluate the cost per level of the output line voltage, which is found to be least for the proposed topology. Hence, the proposed topology is proved to be more economical than other topologies.
- MBV_{pu} is less than the peak of the output line voltage, which shows the low voltage stress across the switches. Moreover, the least value of TSV_{pu} makes the proposed topology even more economical.
- Cost comparison of the 17L proposed structure has been carried out in the Table IV considering the actual cost of the commercially available components. It is apparent for the proposed structure to be a cost-effective solution even if RB-IGBTs are used in place of conventional IGBTs

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
NUMBER OF DEVICE TRANSITIONS FOR CONSECUTIVE VOLTAGE LEVEL CHANGE

Voltage level	[33]	[34]	[35]	[Proposed]
V_{L8}	21	24	21	13
V_{L7}	20	21	22	12
V_{L6}	20	20	22	12
V_{L5}	21	24	23	12
V_{L4}	22	20	23	12
V_{L3}	21	20	22	12
V_{L2}	20	20	22	12
V_{L1}	21	20	23	12
V_{L0}	22	21	22	13

TABLE VI
NUMBER OF DEVICE TRANSITIONS FOR CONSECUTIVE VOLTAGE LEVEL CHANGE

Voltage level transition	[33]	[34]	[35]	[Proposed]
$V_{L7} \leftrightarrow V_{L8}$	23	12	15	6
$V_{L6} \leftrightarrow V_{L7}$	15	15	18	9
$V_{L5} \leftrightarrow V_{L6}$	15	11	11	7
$V_{L4} \leftrightarrow V_{L5}$	21	19	19	11
$V_{L3} \leftrightarrow V_{L4}$	21	13	23	6
$V_{L2} \leftrightarrow V_{L3}$	18	15	13	9
$V_{L1} \leftrightarrow V_{L2}$	9	11	13	7
$V_{L0} \leftrightarrow V_{L1}$	23	9	18	9

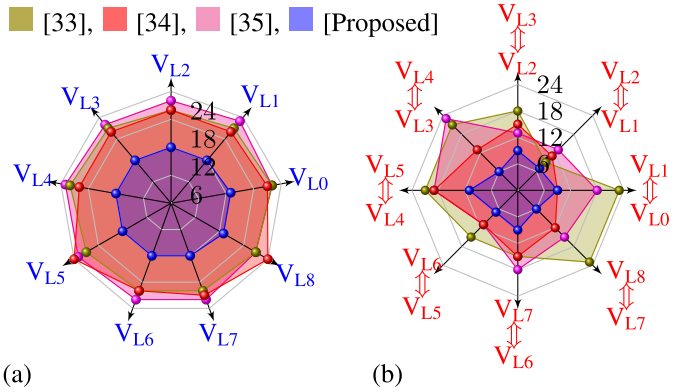


Fig. 11. (a) Number of conducting devices at each voltage level; (b) Number of device transitions for consecutive voltage level change.

(IGBT without anti-parallel diode) to block the reverse voltage.

- The number of conducting devices as well as device transitions at each level of output voltage of the proposed topology are compared with the similar 17L structure are tabulated in Tables V & VI, respectively as well as represented by radar charts. Fig. 11(a) exhibits that the proposed topology has least number of conducting devices for the generation of each output voltage levels, whereas Fig. 11(b) shows less number of device transitions for consecutive output voltage level change. The area enclosed by the radar chart in Figs. 11(a) & 11(b) represent the comparative conduction and switching losses, respectively. The proposed topology encloses smaller area than the similar 17-level state-of-the-art topologies. Hence, the proposed topology is subjected to low conduction and switching losses as depicted by the radar chart diagrams.

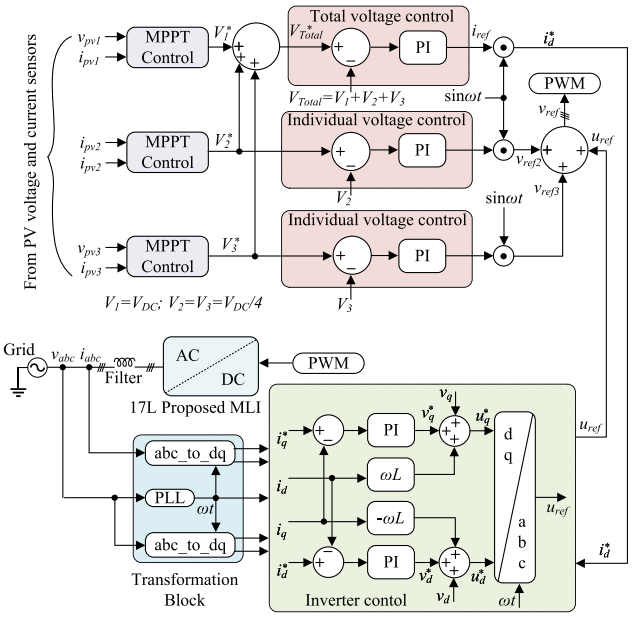


Fig. 12. Control block diagram of grid-tied PV system for Proposed MLI.

V. CONTROL STRATEGY

PV system performance varies with temperature and insolation levels throughout the day. The maximum power point tracking (MPPT) control mechanism is used to collect maximum power, ensuring that PV operates at MPP.

Closed-loop voltage control scheme ensures that the total DC-link voltage remains consistent with the reference set voltage, even when PV characteristics vary. The control loop estimates error by combining actual/measured DC-link voltages (V_{total}) with sum of the individual reference DC-link voltages (V^*_{total}) generated by the MPPT control algorithm. A proportional-integral (PI) controller is used to handle the voltage error to reduce it.

The proposed topology utilises three DC-sources, hence the individual DC-link voltage needs to be maintained constant. As a result, individual voltage control scheme is employed to balance each DC-link voltage. From the output of each individual voltage control module, the reference signal is generated for LSCs-MPWM approach.

MPPT estimates the power change in PV system and based on power change, it generates reference signals termed as V_1^*, V_2^*, and V_3^*. The generated reference signals are compared with the measured V_1, V_2, and V_3. Then the error signal is processed through the PI controller to produce reference current signal on the d-axis i_d^*. The three-phase current and voltage are converted into d- and q-axes current (i_d & i_q) and voltage (v_d & v_q) using three-phase PLL as shown in Fig. 12.

Then the i_d^* and i_d are compared and processed through the PI controller to generate v_d^*. Similarly, the v_q^* is produced on passing the compared value of i_q^* with i_q through the PI controller. On adding v_d^*, v_d, and the feed forward term $-\omega L i_q$ together produce u_d^*, wheres v_q^*, v_q, and feed forward term $\omega L i_d$ on adding together produce u_q^*. From u_d^* and u_q^*, three-phase reference signals for PWM technique (i.e. LSCs-MPWM) are produced by converting them into the dq-abc transformation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VII
PV PANELS SPECIFICATIONS

For SCBIs	
Maximum Power Pmax (Wp)	50.00
Maximum Power Voltage V@Pmax(V)	18.00
Maiximum Power Current I@Pmax(A)	2.78
Open Circuit Voltage Voc (V)	22.00
Short Circuit Current Isc (A)	3.28
3-panels are connected in series	
For TTI	
Maximum Power Pmax (Wp)	277.26
Maximum Power Voltage V@Pmax(V)	40.30
Maiximum Power Current I@Pmax(A)	6.88
Open Circuit Voltage Voc (V)	49.10
Short Circuit Current Isc (A)	7.48
5-panels are connected in series	

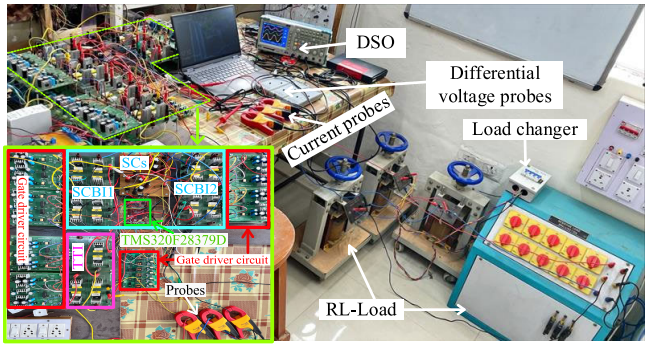


Fig. 13. Experimental setup.

TABLE VIII
SPECIFICATION & COMPONENTS OF THE 17L PROTOTYPE

Items	Specifications
DC Sources	200 V, 50 V, 50 V
Reference signal's frequency	50 Hz
Carriers' frequencies	5 kHz
Diodes (D_{01}^k)	P600B-E3/54
IGBTs (S_{01}^k & $\bar{S}_{01}^k$)	IRGP4072DPBF
IGBTs (S_{12}^k, S_{11}^k, & S_{13}^k)	STGW30NC60VD
IGBTs (S_P^k-S_S^k, & S_1-S_6)	STGW30NC120HD
Gate driver IC	TLP250
Capacitors (C_{01}^k & C_{11}^k)	50 V, 0.8 mF & 100 V, 1.3 mF
R-L Load	R = 50 ohms, L = 10 mH
Controller	TMS320F28379D

The PV array is selected and designed according to the specifications given by Table VII and the MPPT curves for TTI and SCBIs are shown in the next section.

VI. EXPERIMENTAL VERIFICATION

A laboratory prototype of the proposed structure for three-phase 17-level output line voltage has been developed as shown in Fig. 13. The components and their specification are enlisted in Table VIII. The LSCs-MPWM technique has been implemented on TI controller (TMS320F28379D) to gate the IGBTs as the switching devices. The capacitance of the SCs are calculated for 10% ripple from equations (11) & (12).

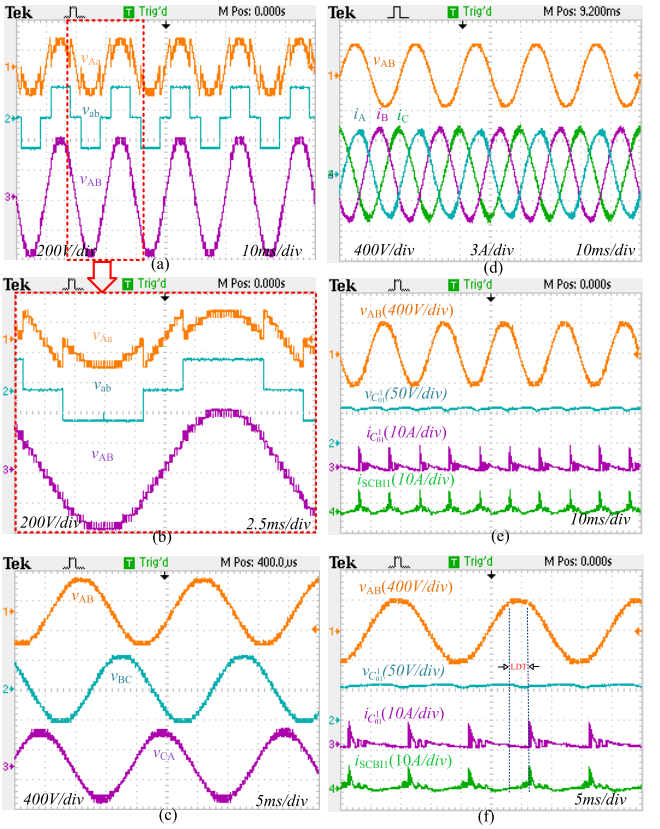


Fig. 14. Experimental steady-state results. (a)-(b) Generation of line voltage (v_{AB}) by cascading the outputs of TTI and SCBI1 and its zoomed-view; (c) Three-phase 17L balanced lines voltage; (d) Three-phase balanced lines current waveforms for R-load; (e)-(f) Line voltage, SCs (C_{01}^1) current/voltage, and current drawn by SCBI1 for R-load.

Capacitance of 800 μF and 1300 μF are taken for experimental purpose, which is near to the calculated value of 786.9 μF and 1296.1 μF, respectively for C_{01}^k & C_{11}^k.

For the parameters tabulated in Table VIII, the experimental results for the proposed topology at 'm' = 1 are presented to validate the performance of the proposal. The output line voltage of TTI (v_{ab}) is cascaded with the output of SCBI1 (v_{Aa}) to develop 17L line voltage (v_{AB}) as shown in Fig. 14(a) and its zoomed-view in Fig. 14(b). Similarly, the outputs of both the SCBIs are cascaded with the output of TTI, which generates three-phase voltage waveforms as shown in Fig. 14(c). The three-phase line currents, SC's current, voltage across the SC, and current drawn by SCBI for R-load are shown in Figs. 14(d)-(f).

Moreover, the SCBIs are cascaded with the line voltages of the TTI, which creates a phase difference of 30^0 (degree) between the peaks of capacitor's current and output voltage as shown in Fig. 14(e)-(f) for R-load. Hence, the ripple voltage across the capacitors are reduced by the factor of $\sqrt{3}/2$ according to (11) & (12). As a result, the spikes in the capacitors' currents are reduced significantly even in the case of R-load.

Figs. 15(a)-(g) show the dynamic results for balanced and unbalanced R/RL-loads. Fig. 15(a) has been captured during load transition from balanced R-load to balanced RL-load and its zoomed-view is shown in Fig. 15(b). The dynamic behaviour of the load current on transition from lagging to leading power factor is shown in Figs. 15(c)-(d). The proposed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

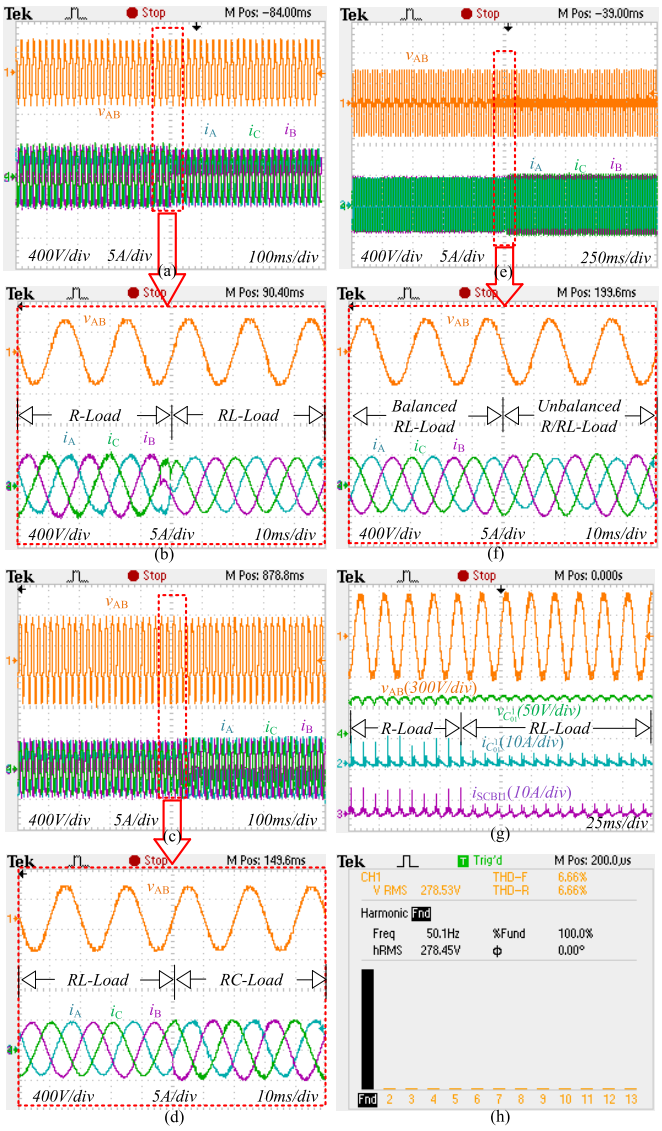


Fig. 15. Experimental dynamic results. (a)-(g) Step change in load from R to RL and vice-versa for balanced and unbalanced loads; (h) FFT analysis of output line voltage.

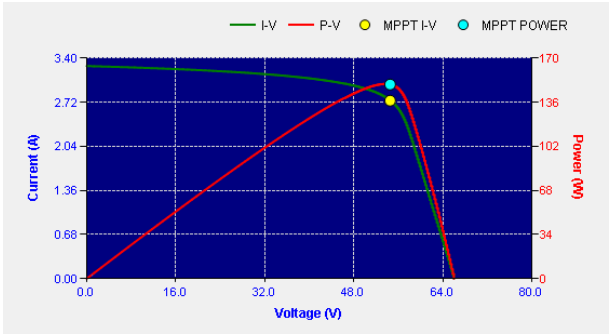


Fig. 16. MPP tracking curve for SCBIs (SCBI1 & SCBI2).

inverter is suitable for unbalanced loading conditions and it is verified through experimental results shown in Figs. 15(e)-(f), which show load transition from balanced RL-load to unbalanced RL-load. The effect of load transition from R-load to RL-load further reduces the voltage ripples and current spikes of the SCs which can be seen in Fig. 15(g). FFT analysis of the experimental line voltage is shown in Fig. 15(h). The MPPT

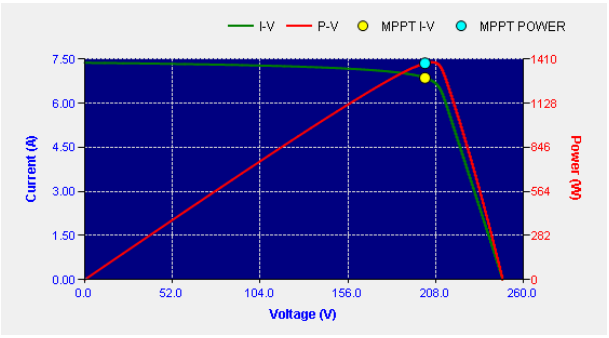


Fig. 17. MPP tracking curve for TTI.

curves for SCBIs and TTI have been shown by Fig. 16 and Fig. 17, respectively.

VII. Conclusion

A generalised cascading structure for three-phase HMLI topology with reduced components count and inrush current has been proposed in this article utilising SC cells. LSCs-MPWM technique is implemented to generate $16n+1$ levels of output voltage with lower harmonic distortion. Moreover, the ripple voltage across the capacitor is also reduced with a significant factor of $\sqrt{3}/2$ irrespective of load's nature due to the effective implementation and cascading of SCBIs with TTI. Low switching frequency of the TTI and Low voltage stress across the switches of SCBIs contribute to the reduction of switching losses. Dynamic performance and efficiency of the proposed inverter are verified experimentally. Detailed comparison of the proposed structure with the state-of-the-art topologies makes it a promising solution for the renewable energy sources like PV and fuel cells due to boosting ability, high quality output voltage, simple structure, etc.

References

[1] H. Abu-Rub, J. Holtz, J. Rodriguez, and G. Baoming, "Medium-voltage multilevel converters—State of the art, challenges, and requirements in industrial applications," *IEEE Trans. Ind. Electron.*, vol. 57, no. 8, pp. 2581–2596, Aug. 2010.

[2] R. Agrawal and S. Jain, "Comparison of reduced part count multilevel inverters (RPC-MLIs) for integration to the grid," *Int. J. Electr. Power Energy Syst.*, vol. 84, pp. 214–224, Jan. 2017.

[3] M. Sarebanzadeh, M. A. Hosseinzadeh, C. Garcia, E. Babaei, S. Islam, and J. Rodriguez, "Reduced switch multilevel inverter topologies for renewable energy sources," *IEEE Access*, vol. 9, pp. 120580–120595, 2021.

[4] M. A. Hosseinzadeh, M. Sarebanzadeh, M. Rivera, E. Babaei, and P. Wheeler, "A reduced single-phase switched-diode cascaded multilevel inverter," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 9, no. 3, pp. 3556–3569, Jun. 2021.

[5] E. Samadaei, M. Kaviani, and K. Bertilsson, "A 13-levels module (K-type) with two DC sources for multilevel inverters," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5186–5196, Jul. 2019.

[6] J. Zeng, W. Lin, D. Cen, and J. Liu, "Novel K-type multilevel inverter with reduced components and self-balance," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 8, no. 4, pp. 4343–4354, Dec. 2020.

[7] 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. Power Electron.*, vol. 36, no. 5, pp. 5204–5212, May 2021.

[8] S. Behara, N. Sandeep, and U. R. Yaragatti, "Design and implementation of transformer-based multilevel inverter topology with reduced components," *IEEE Trans. Ind. Appl.*, vol. 54, no. 5, pp. 4632–4639, Sep. 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[9] N. V. Kurdkandi et al., "A new seven-level transformer-less grid-tied inverter with leakage current limitation and voltage boosting feature," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 4, no. 1, pp. 228–241, Jan. 2023.

[10] X. Zhu, K. Ye, L. Jiang, K. Jin, W. Zhou, and B. Zhang, "Nonisolated single-phase quadratic switched-boost inverter with continuous input current and step-up inversion capability," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 4, no. 1, pp. 276–287, Jan. 2023.

[11] M.-K. Nguyen, V.-T. Tran, D.-T. Do, and C. Wang, "A single-stage boost-derived T-type inverter with self-balanced capacitor voltage," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 3, no. 3, pp. 715–724, Jul. 2022.

[12] X. Zhu, K. Ye, K. Jin, W. Zhou, and B. Zhang, "Nonisolated dual-output single-phase boost inverter with reduced active and passive components," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 4, no. 1, pp. 253–265, Jan. 2023.

[13] S. A. Khan et al., "Topology, modeling and control scheme for a new seven-level inverter with reduced DC-link voltage," *IEEE Trans. Energy Convers.*, vol. 36, no. 4, pp. 2734–2746, Dec. 2021.

[14] K. K. Gupta, A. Ranjan, P. Bhatnagar, L. K. Sahu, and S. Jain, "Multilevel inverter topologies with reduced device count: A review," *IEEE Trans. Power Electron.*, vol. 31, no. 1, pp. 135–151, Jan. 2016.

[15] M. Saeedian, M. E. Adabi, S. M. Hosseini, J. Adabi, and E. Pouresmaeil, "A novel step-up single source multilevel inverter: Topology, operating principle, and modulation," *IEEE Trans. Power Electron.*, vol. 34, no. 4, pp. 3269–3282, Apr. 2019.

[16] T. Roy, P. K. Sadhu, and A. Dasgupta, "Cross-switched multilevel inverter using novel switched capacitor converters," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8521–8532, Nov. 2019.

[17] Y. Ye, S. Chen, X. Wang, and K. E. Cheng, "Self-balanced 13-level inverter based on switched capacitor and hybrid PWM algorithm," *IEEE Trans. Ind. Electron.*, vol. 68, no. 6, pp. 4827–4837, Jun. 2021.

[18] K. P. Panda, P. R. Bana, and G. Panda, "A switched-capacitor self-balanced high-gain multilevel inverter employing a single DC source," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 12, pp. 3192–3196, Dec. 2020.

[19] N. Sandeep, "A 13-level switched-capacitor-based boosting inverter," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 3, pp. 998–1002, Mar. 2021.

[20] K.-M. Kim, J.-K. Han, and G.-W. Moon, "A high step-up switched-capacitor 13-level inverter with reduced number of switches," *IEEE Trans. Power Electron.*, vol. 36, no. 3, pp. 2505–2509, Mar. 2021.

[21] Y. Ye, G. Zhang, X. Wang, Y. Yi, and K. W. E. Cheng, "Self-balanced switched-capacitor thirteen-level inverters with reduced capacitors count," *IEEE Trans. Ind. Electron.*, vol. 69, no. 1, pp. 1070–1076, Jan. 2022.

[22] N. Kishore, K. Shukla, and N. Gupta, "A novel three-phase 13-level cascaded hybrid-module based multilevel inverter with level-shifted modified-PWM algorithm," *IEEE Trans. Ind. Appl.*, vol. 60, no. 2, pp. 3263–3272, Mar. 2023.

[23] K. Shukla, V. Malyala, and R. Maheshwari, "A novel carrier-based hybrid PWM technique for minimization of line current ripple in two parallel interleaved two-level VSIs," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 1908–1918, Mar. 2018.

[24] Y. Ye, K. W. E. Cheng, J. Liu, and K. Ding, "A step-up switched-capacitor multilevel inverter with self-voltage balancing," *IEEE Trans. Ind. Electron.*, vol. 61, no. 12, pp. 6672–6680, Dec. 2014.

[25] A. Taghvaie, J. Adabi, and M. Rezanejad, "A self-balanced step-up multilevel inverter based on switched-capacitor structure," *IEEE Trans. Power Electron.*, vol. 33, no. 1, pp. 199–209, Jan. 2018.

[26] W. Lin, J. Zeng, J. Liu, Z. Yan, and R. Hu, "Generalized symmetrical step-up multilevel inverter using crisscross capacitor units," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 7439–7450, Sep. 2020.

[27] P. Bhatnagar, A. K. Singh, K. K. Gupta, and Y. P. Siwakoti, "A switched-capacitors-based 13-level inverter," *IEEE Trans. Power Electron.*, vol. 37, no. 1, pp. 644–658, Jan. 2022.

[28] Y. Nakagawa and H. Koizumi, "A boost-type nine-level switched capacitor inverter," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 6522–6532, Jul. 2019.

[29] T. Roy and P. K. Sadhu, "A step-up multilevel inverter topology using novel switched capacitor converters with reduced components," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 236–247, Jan. 2021.

[30] E. Babaei and S. S. Gowgani, "Hybrid multilevel inverter using switched capacitor units," *IEEE Trans. Ind. Electron.*, vol. 61, no. 9, pp. 4614–4621, Sep. 2014.

[31] R. S. Alishah, S. H. Hosseini, E. Babaei, M. Sabahi, and G. B. Gharehpetian, "New high step-up multilevel converter topology with self-voltage balancing ability and its optimization analysis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 7060–7070, Sep. 2017.

[32] Y. Ye, S. Chen, T. Hua, M. Lin, and X. Wang, "Self-balanced multilevel inverter with hybrid double- and half-mode switched capacitor," *IEEE Trans. Ind. Electron.*, vol. 69, no. 6, pp. 5735–5744, Jun. 2022.

[33] J. Sathik, D. Almakhles, and M. F. Elmorshedy, "High boost seventeen level switched capacitor inverter topology with continuous input current," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 11, no. 3, pp. 2742–2754, Mar. 2023.

[34] A. K. Singh and R. K. Mandal, "A novel 17-level reduced component single DC switched-capacitor-based inverter with reduced input spike current," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 5, pp. 6045–6056, Oct. 2022.

[35] V. Anand et al., "Seventeen level switched capacitor inverters with the capability of high voltage gain and low inrush current," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 4, no. 4, pp. 1138–1150, Apr. 2023.

[36] *Digikey Home*. Accessed: Dec. 25, 2023. [Online]. Available: https://www.digikey.com/



Niraj Kishore was born in Muzaffarpur, India. He received the M.Tech. degree in power electronics and drives from the National Institute of Technology at Kurukshetra, Kurukshetra, India, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Malaviya National Institute of Technology at Jaipur, Jaipur, India.

From 2013 to 2020, he was an Assistant Professor with the School of Electronics and Electrical Engineering, Lovely Professional University, Jalandhar, India. His research interests include power electronics and drives, pulse width modulated converters, and multi-level inverters.



Kapil Shukla (Senior Member, IEEE) was born in Dungarpur, India. He received the M.Tech. degree in power systems from Delhi Technological University, New Delhi, India, in 2012, and the Ph.D. degree in electrical engineering from Indian Institute of Technology Delhi, New Delhi, in 2019.

He is currently an Assistant Professor with the Department of Electrical Engineering, Malaviya National Institute of Technology at Jaipur, India. His research interests include pulse width modulated converters, multi-level inverters, and power quality.



Nitin Gupta (Senior Member, IEEE) received the degree in electrical engineering from the Government Engineering College, Kota, in 2000, the Master of Engineering degree in power electronics from the College of Technology and Engineering (CTAE), Udaipur, in 2008, and the Ph.D. degree in electrical engineering from Indian Institute of Technology (IIT) Roorkee, Uttarakhand, in 2012, as a full-time research scholar under MHRD scheme.

He was a Post-Doctoral Researcher with the School of Engineering and Digital Sciences, Nazarbayev University, Astana, Kazakhstan, in 2023. He has been an Associate Professor with the Department of Electrical Engineering, MNIT, Jaipur, since 2023. His research interests include power electronics and drives, power quality, renewable energy grid integration, active power filter control, and power management in microgrids. He is an Academic Editor of *International Transactions on Electrical Energy Systems* (Hindawi). He is a reviewer of several international journals and international reputed conferences.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 27,2024 at 05:03:06 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Handwritten Amharic Word Recognition with Additive Attention Mechanism

Ruchika Malhotra[1] and Maru Tesfaye Addis[2,3]
[1]Department of Software Engineering, Delhi Technological University, Delhi 110042, India
[2]Department of Computer Science, Debre Tabor University, Amhara 251272, Ethiopia
[3]Department of Computer Science and Engineering, Delhi Technological University, Delhi 110042, India

Corresponding author: Maru Tesfaye Addis (marutcomp@gmail.com)

ABSTRACT Amharic stands as the second most widely spoken Semitic language globally, trailing only Arabic. As a result, it is a highly resourceful language for hardcopy document collections, and automatic recognition technology is required. Recognizing handwritten Amharic words is a difficult task due to factors, including variations in individual handwriting for the same words, connectivity of two words without spaces, similarities in the shapes of alphabet characters, and noise from scanned images. The Amharic alphabet has a vast character set that adapted most of its characters from the most popular script called Ethiopic script, formerly known as Ge'ez script, and added unique characters. Despite significant advancements in optical character recognition (OCR) research, Amharic script recognition has received less attention. This study employs a deep learning (DL) approach with additive attention to recurrent neural networks (RNN) to achieve precise recognition of handwritten Amharic words. Seven convolutional neural networks (CNN) and two RNN, using a connectionist temporal classification (CTC) strategy, make up this recognition model architecture, enabling efficient recognition through sequential feature extraction. The study addressed deep learning data insufficiency by using augmentation techniques to increase datasets. The study used an original dataset of 12,047 handwritten Amharic words and an augmented dataset of 22,00 images. The developed model achieved an average character error rate (CER) of 2.84% and an average word error rate (WER) of 9.75% for the testing dataset. These results are indeed promising, offering a glimpse into the potential of this attention-based approach for handwritten Amharic word recognition. This research represents a significant step toward bridging the gap in OCR technology for Amharic script and showcases the transformative capabilities of DL in pattern recognition.

INDEX TERMS Amharic handwriting, CNN, deep learning, handwritten word recognition, pattern recognition.

I. INTRODUCTION

Amharic stands as the second most widely spoken Semitic language globally, trailing only Arabic [1-3]. This rich language boasts a vast community of over 100 million speakers; spread across various nations, including Ethiopia, which is the origin country [4, 5]. Notably, Amharic holds the prestigious position of being the official language of Ethiopia [4, 6-8], underscoring its significance and influence.

One of the remarkable aspects of Amharic is its written form [9], which draws upon the indigenous symbols of the Ge'ez language [10], predominantly employed in Ethiopia and Eritrea [11, 12], giving rise to the unique and ancient script known as the Ethiopic script [13]. The Amharic alphabet includes a diverse array of characters, encompassing regular letters, digits, and punctuation marks. Here, the Amharic language adapted Arabic numerals rather than the Ge'ez digit [14]. In total, this alphabet comprises a staggering 300 distinct character symbols, as visually depicted in **FIGURE 1.** There are 267 alphabetic characters to create Amharic words, as well as 10 Arabic numerals to display numerical values. The Amharic writing system uses the remaining 23 characters as punctuation marks.

Currently, numerous federal and regional organizations in Ethiopia are actively seeking to reduce paper usage and transition to digital workflows [12, 15]. Institutions spanning hospitals, bank offices, primary school libraries, and locality offices find themselves inundated with the bulk of handwritten documents [13]. Consequently, there is a critical necessity to convert these handwritten resources into digital formats for streamlined management. The escalating demand

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

for effective OCR technology to meet this imperative is evident, as OCR serves as a pivotal component of any text recognition task [16].

OCR operates as a transformative system, capable of converting both printed and handwritten text from hardcopy documents into formats comprehensible by computers [15, 17, 18]. Its overarching goal is to convert these documents into electronic counterparts, making them easier to manage within computer systems. The implementation of an OCR system, which is the central problem and popular in pattern recognition [19, 20], offers manifold advantages to organizations. It bolsters document management operations, enhances data accessibility, and expedites the retrieval of information, collectively contributing to more efficient organizational processes.

Recognizing handwritten Amharic words is not an effortless task with OCR technology. This is due to factors such as individual writing style variation for the same words [9, 21], connectivity of two words without space, and small shape differences among the letters in the alphabet [22, 23]. Noise from scanned images and vast character sets from the alphabets contribute to this difficulty. So, identifying handwritten Amharic words in image documents presents a unique OCR challenge.

Despite research in pattern recognition for popular scripts, Amharic OCR has received less attention than other language OCR studies [2, 5, 24, 25]. Languages such as Latin, Chinese, and Arabic scripts [26, 27] were studied well with different approaches. Previous research [1, 2, 28, 29] on Amharic-related script OCR has mostly been based on printed texts. But the case of handwritten Amharic texts has not been studied much [24, 30] because there aren't many public research datasets [12, 30]. It is rare to find publicly available Amharic handwritten text datasets, which are typically required for deep learning algorithms [15, 26]. So, developing a DL model and preparing the required dataset is a must in the handwritten Amharic word recognition area to fill the gaps.

Conventional transcription and analysis methods often require manual involvement and can be prone to errors. Consequently, there is a demand for automated and effective methodologies capable of precisely recognizing Amharic words. Such techniques would be critical in preserving and comprehending handwritten documents, as well as addressing the challenges associated with traditional methods. Recent advances in artificial intelligence, specifically the use of DL techniques, have greatly improved pattern recognition [12]. They have proven to be more effective than traditional machine learning (ML) methods [31].

| | | |

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

conducting various experiments, including the selection of hyper parameters value, the model's performance is evaluated. The proposed method of handwriting Amharic word recognition appears to be effective based on these results. These preliminary findings signify a promising stride forward in the domain of OCR for handwritten Amharic script recognition.

These contributions collectively represent significant strides toward advancing the capabilities of OCR technology for the Amharic script, illustrating the promise of the proposed approach in recognizing handwritten Amharic words with enhanced precision and efficacy.

Throughout this paper, the following sections are divided into four parts: Section II presents an overview of relevant research related to handwritten Amharic word recognition. Section III describes the proposed approach, including the DL architecture and training procedures. In Section IV, we examine the experimental setup and present the evaluation results. Lastly, Section V provides a summary of this study.

II. LITERATURE REVIEW

Exploration in the realm of Amharic OCR remains an underdeveloped area, both within ML and DL techniques. This holds particularly true when addressing handwritten Amharic words. Within this context, this section explores the details of the Amharic alphabet and its writing system, exploring diverse methodologies employed for the recognition of handwritten Amharic words. Past research has predominantly leaned towards traditional ML methods, while contemporary studies showcase an increasing trend toward leveraging DL and ensemble approaches [12].

A. AMHARIC ALPHABET AND WRITING SYSTEM

Amharic follows the right direction in writing systems like Latin scripts and others. It has 34 base characters [24, 33], and the remaining characters are derived from those characters by adding small modifications to the base characters to create the vowel sounds [24, 25, 29], as shown in the sample in **TABLE 1**.

TABLE 1. Part of Amharic Alphabets [23].

	Base sound	Orders						
		1st (ä)	2nd (u)	3rd (ï)	4th (a)	5th (e)	6th (ə)	7th (o)
1	H	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
2	L	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
3	H	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
32	p’	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
33	F	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
34	P	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ

In addition to the characters found in **TABLE 1**, the Amharic alphabet includes special characters that generate unique “ua” sounds. Many of these unique characters are modified versions of the base alphabets, typically altered at the bottom. These special characters are displayed in **TABLE 2**.

TABLE 2. Amharic special characters

ሏ	ሟ	ሯ	ሷ	ሿ	ቧ
ቱ	ቿ	ኋ	ኟ	ኍ	ኗ
ዟ	ዧ	ጧ	ቈ	ቋ	ቍ

As shown in **FIGURE 2**, the modifications occur in the middle, bottom, and top. Some modifications also happened inside and outside of the base alphabet, which creates a complex structure. Character symbols in handwriting appear to be nearly identical to the base alphabet. This becomes difficult to identify due to the writer’s handwritten variation.

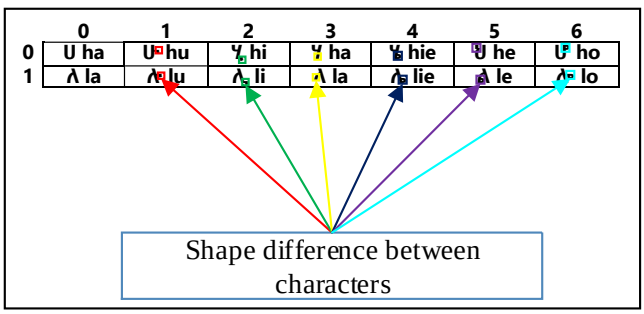


FIGURE 2. Amharic alphabets shape the difference between the base and derived characters [24].

B. HANDWRITTEN AMHARIC SCRIPT RECOGNITION

Recognizing handwritten Amharic words is a difficult task because of the intricate nature of the Amharic characters. Additionally, annotated datasets are not widely available, which makes it even more challenging. However, in recent times, some studies have been conducted to address the gaps that exist in handwritten Amharic script recognition.

In a study by Agegnehu et al. [34], deep learning was employed to identify Amharic punctuation marks and handwritten digits using a CNN architecture. In their study, they mainly focused on 20 handwritten Ge’ez digits and nine punctuation marks found in Amharic script. They collected 5800 handwritten images from 100 individuals. After preprocessing, they used 32-by-32-pixel binarized images for model development and evaluation. They developed four different models with different CNN architectures. Their models have two, three, four, and five convolutional layers. At the end of each model, there is one fully connected layer with 29 classes in addition to the input layer. The architecture is simple and precise according to the number of layers, which has the advantage of reducing model complexity. They utilized CNN’s automatic feature extraction, eliminating the need for handcrafted methods [2]. During the model configuration, they applied trial-and-error approaches. They

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

selected the model with three convolutional layers that generated the best result out of the four model configurations.

In his study, Tamir [6] investigated the application of CNN for the recognition of Amharic characters in handwriting. The architecture consists of two convolutional layers specifically designed for the classification of handwritten Amharic characters. Tamir compiled a dataset comprising 30,446 characters sourced from approximately 130 different writers. Among these, 27,413 and 3,033 characters were utilized for training and testing purposes, respectively. He implemented and assessed the suggested model design using the Keras and TensorFlow frameworks. Following the training phase, Tamir achieved approximately 99.52% training accuracy after the first epoch, representing the lowest accuracy among the epochs. Meanwhile, the testing accuracy reached approximately 99.71%.

Alemu and Fuchs [9] published a research paper outlining their approach to recognizing handwritten Amharic bank checks. The main technique they apply in their study is the Hidden Markova Random Field (HMRF). In their study, they took 7,240 handwritten character images to train the model written by different individuals to consider handwritten variation. They used 20 checks with 713 characters to test the model's performance. By applying contextual information, they achieved an accuracy rate of 99.44%, whereas without contextual information, they achieved an accuracy rate of 89.06%. They recommend using parallel processing to speed up HMRF computations.

Assabie and Begun [8] conducted a study focusing on the recognition of offline handwritten Amharic words. In their study, they used a direction tensor field rather than segmentation to extract features. 177 writers collected the data for the study from various documents written in an unrestricted environment. For training purposes, they extracted 10,932 distinct words from 307 pages with 300 dpi-resolution images.

The lack of openly available digit datasets limits research on Ge'ez numeral recognition. Before Agegnehu et al. [34], Nur et al. [30] proposed Ge'ez digit recognition by creating a recognition model with better accuracy and an experimental dataset consisting of 51,952-digit images written by 524 people. The researchers suggested exploring various deep-learning techniques for recognizing multiple digits. Their CNN model resulted in 96.21% accuracy.

In their study, "Amharic Handwritten Character Recognition Using Combined Features and Support Vector Machine," Reta et al. [35] integrated various feature extraction techniques to address the challenges and complexities associated with Amharic handwritten character recognition. The study utilized linear discriminant analysis (LDA) to reduce feature dimensions. A multiclass support sector machine (SVM) with error-correcting output codes (ECOC) served as the classifier. The algorithm was trained and tested on both the Amharic dataset and the Chars74K benchmark numeric dataset. Model validation was carried out using a 10-fold cross-validation technique. Because of this, the multiclass SVM classification method showed a strong algorithm that combined features well enough to correctly recognize handwritten characters.

The majority of researchers focused on the recognition of printed text in the case of the Amharic language. The availability of a public dataset for Latin text motivated researchers to focus on the area. However, there is no enough available dataset publicly for studying the Amharic words in this area. This discourages researchers from making efforts to recognize handwritten Amharic words. In this study, developing a recognition model with better accuracy and preparing the experimental dataset advance the state of the art in handwritten Amharic word recognition.

This study aims to identify handwritten Amharic words in images using an attention mechanism. To accomplish this, we incorporate additional DL techniques such as CNN for automatic feature extraction from the input image, RNN for temporal context modeling from the image, and CTC for efficient sequential classification.

III. METHODS

In this section, we explore handwritten Amharic word recognition by featuring an attention mechanism. We dedicate our study to developing a recognition model that can accurately identify handwritten words within scanned images. Here, we provide a comprehensive overview of the proposed model along with insights into the dataset used for experimentation.

A. THE DATASET PREPARATION

During this research, we realized that having a large dataset is crucial for developing and testing deep learning models [12]. We conducted a thorough search for a dataset that aligned with the specific needs of our deep learning task to overcome the challenge of insufficient data. We used the "HARD-I" dataset, which is accessible on the Google Drive repository (https://sites.google.com/view/hawdb-v1) in this study. The dataset comprises handwritten Amharic word images, prepared by Abdurahman et al. [7], and collected from 60 individuals of varying ages and educational backgrounds. This diversity helps the model effectively handle variations in handwriting.

This dataset contained 12,047 handwritten Amharic word images that featured 300 unique Amharic alphabets. We classified the dataset into three subsets: training, validation, and testing. To facilitate precise evaluation, a validation dataset was crafted, encompassing 10% of the training dataset.

When faced with the challenge of a limited dataset for DL tasks, data augmentation techniques were implemented to increase its size. Despite resource constraints, we successfully incorporated approximately 22,000 augmented images.

Utilizing image augmentation is instrumental in enabling DL models to effectively learn from limited labeled data, ultimately leading to enhanced recognition performance. This

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

technique has evolved into a crucial tool for overcoming data scarcity, maximizing DL algorithm potential, and finding applications across diverse domains [12]. The following sub-sections describe the steps and criteria for preparing synthetic datasets.

1) IMAGE AUGMENTATION STEPS

Generating the synthetic data follows the following algorithmic steps and the flowcharts of the steps are depicted in **FIGURE 3**. Image augmentation replicates real-world scenarios by applying various transformations to the original images, allowing the model to better recognize and handle different image variations [15]. These transformations increase the variability of the dataset, helping the model generalize better to new, unseen data. The augmentation steps applied in this study are rotation, width shifting, height shifting and zooming.

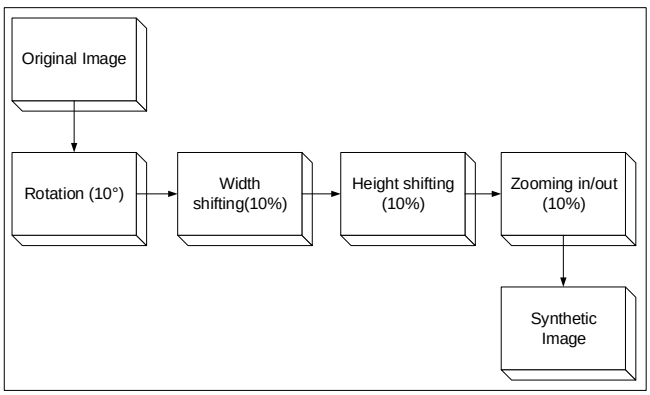
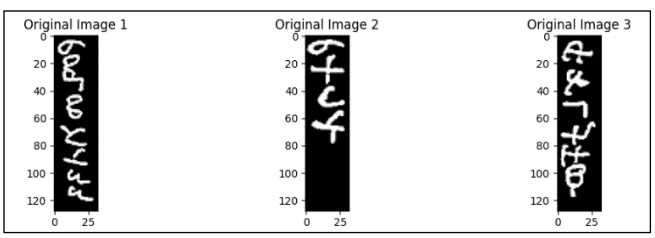


FIGURE 3. Synthetic dataset preparation steps

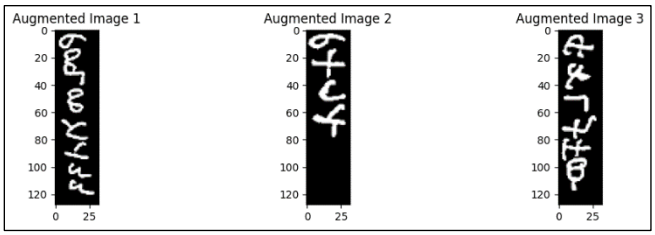
During augmentation, images randomly rotate up to 10 degrees. Additionally, we randomly adjust the width and height up to 10%. Also, randomly zooming in or out up to 10%.

2) SAMPLE AUGMENTED IMAGES

The synthetic generation of the dataset provides a more diverse and challenging set of images, enabling a comprehensive evaluation of the OCR model's capabilities in handling Amharic words. Sample handwritten Amharic word images are generated to reflect this diversity. **FIGURE 4** provides visual insights comparing the original dataset with the augmented dataset, highlighting improvements in our proposed model's performance.



(a)



(b)

FIGURE 4. Describes the presentation of sample images from the HARD-I database: (a) depicts the original image, while (b) features the augmented image.

B. THE PROPOSED MODEL

In this paper, we leverage the strengths of some DL algorithms: CNN, Attention Mechanism, RNN, and CTC. Each of these algorithms plays a crucial role in different stages of the recognition process [29]. CNN layers excel at capturing spatial hierarchies in data, making them ideal for extracting meaningful features from the raw pixel data of handwritten Amharic characters. This feature extraction transforms the input into an abstract representation that highlights relevant patterns. RNN layers, especially GRU cells, are adept at capturing sequential dependencies and temporal context. For handwritten Amharic word recognition, understanding the order of characters and their dependencies is vital. RNN layers model these dependencies, contextualizing characters within the entire word. By combining CNNs for spatial feature extraction and RNNs for temporal context modeling, our model leverages the strengths of both architectures. The CNN extracts spatial features, which are then processed by the RNN to understand the sequence. This approach ensures effective learning of both local and global features, crucial for accurate handwritten Amharic word recognition.

By combining their capabilities, we've crafted an inclusive and operative method for recognizing handwritten Amharic words. The selection and combination of these DL algorithms are key contributors to the accomplishment of this research, ensuring heightened accuracy and reliability in recognition.

The following sections provide a detailed breakdown of the specific roles each algorithm plays in our holistic approach.

1) INPUT LAYER

In this recognition model, the input layer is the initial part of the model that accepts the input image. It takes 32 by 128-pixel images from scanned images with handwritten Amharic words and makes them ready for the feature extraction part. We reshaped the input image to form 128 rows and 32 columns.

2) CNN LAYERS

Extracting deeper or lower features is important for the recognition of handwritten words. CNN has the capability to extract deeper features [36, 37]. So, this study aimed to extract deeper features from handwritten word images using CNN, which is known for its exceptional ability to identify patterns and significant information within images [11]. The scalability of CNN makes it a selective technique for different image

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

processing applications. Previous studies have successfully applied CNN to a variety of tasks [38], and CNN has demonstrated its power, particularly in document analysis [34]. Therefore, we employed these techniques to precisely identify the unique features of handwritten words.

To illustrate the effectiveness of CNN in feature extraction, let's explore the mathematical representation. Consider an input image of a handwritten word with pixel values denoted by $I_{(i,j)}$, where i and j are the spatial coordinates of the image. Let $O_{k,l}$ represent the output feature map at the spatial location (k,l), $W_{(m,n)}$ denote the learnable weights of the kernel at position (m,n). Each convolution applies a set of kernels to the input image, allowing for the computation of the output in "(1)". In equation "(2)", we compute the activation function element-wise after convolution to introduce non-linearity.

$$O_{k,l} = \sigma\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(I_{i+m,j+n}W_{m,n}\right)\right) \quad (1)$$

$$O_{k,l} = \mathrm{ReLU}(O_{k,l}) \quad (2)$$

Pooling, often called max-pooling, reduces the spatial dimensions of the feature maps, aiding in computational efficiency and mitigating overfitting [12]. We compute the pooled output in "(3)", which signifies the maximum value within the designated pooling region.

$$P = \max(I_{i:i+F,\,j:j+G}) \quad (3)$$

where P is the pooled output, and $\max(I_{i:i+F,\,j:j+G})$ represents the maximum value within the specified pooling region. The dimensions of the pooling filter are denoted by symbols F and G.

3) ADDITIVE ATTENTION LAYER

The attention mechanism, renowned in machine translation for its proficiency in unfixed order prediction [39], plays a central role in our recognition model. It is strategically designed to refine the model's focus on pertinent regions within handwritten words. Specifically, we leverage the additive attention mechanism, a subtype of soft attention mechanism, which dynamically assigns weights to different segments of the sequence during recognition.

In additive attention, the computation of attention weights involves a learned function that amalgamates information from both the source (input) and target (output) sequences. The term "additive" is aptly used because it encompasses a summation of weighted values.

This mechanism operates dynamically by adjusting the weights assigned to each element in the input sequence based on its relevance to the output. Notably, salient regions in the static representation can dynamically emerge to the forefront [40]. This dynamic adjustment supplements the model's attention to critical features and characters, proving particularly advantageous when addressing challenges like inter-class similarity and structural complexity [12].

Handwritten Amharic characters often exhibit subtle variations that can lead to ambiguity, especially among visually similar characters. The additive attention mechanism's ability to dynamically adjust attention weights helps mitigate confusion between characters with similar appearances. By emphasizing critical features during recognition, it improves the model's ability to distinguish between these visually similar characters.

The structural complexity of handwritten Amharic, characterized by diverse character forms and ligatures, poses significant challenges to recognition models. Additive attention enhances the model's ability to interpret and adapt to this complexity by allowing it to focus selectively on different parts of the sequence. This selective attention mechanism helps capture intricate details and variations within characters, improving overall recognition performance.

Our model describes the mathematical representation of the attention operation in "(4)".

$$c_i = \sum_{j=1}^{T}\alpha_{ij}h_j \quad (4)$$

Where, $i = 1, 2, ..., T$

Here, h_i is the hidden state at the time step i, T is the length of the sequence, α_{ij} represents the attention weights, and c_i corresponds to the context vectors.

The attention weights α_{ij} are computed using a scoring function that measures the relevance of the input at time step j to the output at time step i. In our approach, we adopt the dot product as the scoring function, as shown in "(5)".

$$\mathrm{score}_{ij} = h_i h_j \quad (5)$$

Subsequently, the scores are normalized using the SoftMax function to obtain attention weights, as outlined in "(6)".

$$\alpha_{ij} = \frac{e^{\mathrm{score}_{ij}}}{\sum_{k=1}^{T} e^{\mathrm{score}_{ik}}} \quad (6)$$

In summary, additive attention, as a soft attention mechanism, significantly enhances the interpretability and adaptability of the attention process in sequence-to-sequence tasks, particularly in the context of OCR.

4) RNN LAYERS

Recurrent neural networks which is segmentation-free networks generate impressive result [41]. In this study, we leverage BGRU networks which is less complex and more efficient computationally [42], to proficiently model intricate sequential dependencies inherent in handwritten words. The incorporation of BGRU layers substantially enhances the network's prowess in capturing contextual information efficiently [38], leading to a notable enhancement in

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

recognition accuracy [16]. Integrating BGRU layers into the recognition architecture empowers the model with a holistic understanding of the inherent structure and context of handwritten words. This strategic utilization contributes significantly to the overall effectiveness of our recognition model.

5) CTC LAYER

The utilization of the CTC technique proves to be indispensable for addressing sequential recognition tasks, a characteristic inherent in handwritten word recognition. This choice is grounded in the inherent sequential nature of handwritten word recognition tasks, where understanding the temporal order of characters is crucial.

In our specific case, employing CTC for recognizing handwritten Amharic words emerges as a sound technique. CTC provides the model with the ability to learn from sequences of varying lengths, eliminating the need for explicit alignment between input images and their corresponding labels. This inherent flexibility makes CTC a powerful tool for facilitating end-to-end training and decoding processes [12].

Central to the effectiveness of the CTC approach is the CTC loss function, which plays a pivotal role in training the model. This loss function, as represented in "(7)", quantifies the dissimilarity between the predicted sequence and the ground truth labels. Incorporating this loss function into our training regimen facilitates the model's learning of the intricate relationships within the sequential data, contributing to its ability to accurately recognize and decode handwritten Amharic words.

$$L_{CTC} = -\sum_{i,j} I(y_{i,j} \neq \text{blank}) \log(\hat{y}_{i,j}) \qquad (7)$$

Here, L_{CTC} denotes the CTC loss, $\hat{y}_{i,j}$ represents the predicted probability distribution for each time step, and $y_{i,j}$ corresponds to the ground truth label for the given time step.

The integration of the CTC layer, coupled with the use of its associated loss function, significantly enhances the model's ability to comprehend and recognize handwritten Amharic words in a sequential manner. This strategic combination aligns with the inherent characteristics of the task and contributes to the overall robustness and efficacy of our recognition model.

The probability $P(y|\ x)$ in "(8)" is calculated by summing over all valid alignments between the input sequence and the label sequence, denoted by A. This set of valid alignments considers the presence of a blank symbol as well as allowed repetitions and removals of labels during the alignment process.

$$P(y \mid x) = \sum_{a \in A} \prod_{t=1}^{T} \hat{y}_{t,a_t} \qquad (8)$$

This consideration ensures that the model accommodates potential variations in the alignment process, providing a robust mechanism for capturing diverse patterns within the handwritten Amharic words.

In essence, the integration of the CTC layer, coupled with a comprehensive understanding of the associated loss function and alignment considerations, contributes significantly to the adaptability and precision of our recognition model. This combination aligns seamlessly with the nuances of handwritten Amharic word recognition, ensuring a thorough and context-aware approach.

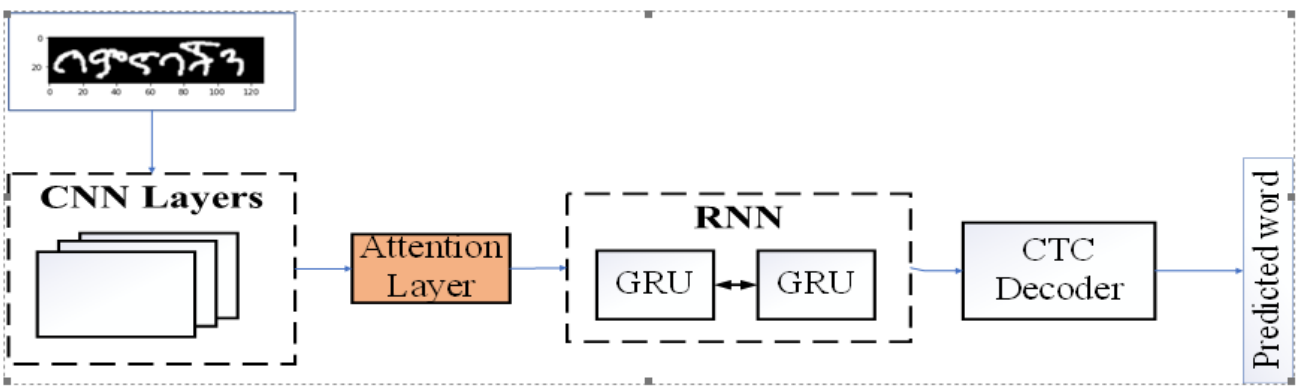


FIGURE 5. The proposed handwritten Amharic word recognition model architecture.

The overarching architecture for Amharic word recognition is depicted in **FIGURE 5**. The figure is customized from our previous study [12]. The figure illustrates the steps of the recognition process, commencing with the image data sourced from the HARD-I database. The image, measuring 32 by 128 pixels in our case, undergoes a sequence of transformations through seven convolutional layers, each equipped with a ReLU function. Additionally, two batch normalization layers with a momentum value of 0.8 each and four max-pooling layers contribute to the model's architecture, as outlined in **TABLE 3**. Following the CNN layers, the extracted features are reshaped into a 2D format and then directed to the attention

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

layers. Here, the reshape input is permuted with a permutation size of (2,1). Then we applied a fully connected layer (dense) with the Softmax activation function for each time step. Then the dimension is reduced by selecting the mean value. Then apply the repeat vector to the reduced dimension using the input dimensions. To get the probability of the region permutation, it is applied again with size (2,1). Finally, weighted average multiplication has been applied between the input and probability results. The attention layer hyper-parameter configuration and the number of parameters are depicted in **TABLE 4**. Subsequently, the sequence is processed by two BGRU layers, each comprising 128 hidden units, as specified in **TABLE 5**. A fully connected (dense) layer with a Softmax activation function further refines the features for classification. The number of units in the dense layer corresponds to the classes or categories in the recognition task, with a total of 301 classes, encompassing 300 alphabet symbols and a blank CTC space [12]. In this study, the model is compiled using the Adam optimizer and CTC loss.

TABLE 3. The proposed model CNN layers and hyper-parameter values configuration.

S. No	Layers	Configuration	Number of Parameters
1.	Input	Rows:128, Columns: 32, Channels:1, datatype:float32	0
2.	Conv_1	Feature:64, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	640
3.	MaxPooling_1	Pooling size: (2,2), strides:2	0
4.	Conv_2	Feature:128, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	73856
5.	MaxPooling_2	Pooling size: (2,2), strides:2	0
6.	Conv_3	Feature:256, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	295168
7.	Conv_4	Feature:256, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	590080
8.	MaxPooling_3	Pooling size: (1,2)	0
9.	Conv_5	Feature:512, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	1180160
10.	BN_1	Momentum: 0.8	248
11.	Conv_6	Feature:512, kernel:(3,3), padding: same, activation: ReLU, kernel_initializer: he_normal	2359808
12.	BN_2	Momentum: 0.8	248
13.	MaxPooling_4	Pooling size: (1,2)	0
14.	Conv_7	Feature:512, kernel:(2,2), activation: ReLU	1049088

TABLE 4. The attention layer hyper-parameter values configuration.

S. No	Layers	Configuration	Number of Parameters
1.	Reshape	Rows:31, Columns: 512	0
2.	permute	Size:(2,1), Rows:512, Columns:31	0
3.	dense	Units:timestep, activation: Softmax	992
4.	dim_reduction	Lambda:31	0
5.	repeat_vector	Rows:512, Columns:31	0
6.	attention_vector	Size:(2,1), Rows:31, Columns:512	0
7.	attention_multiply	Rows:31, Columns:512	0

TABLE 5. The proposed model RNN-CTC layers and hyper-parameter values configuration.

S. No	Layers	Configuration	
1.	Bidirectional_1	BGRU: 128, return sequence: True, droout:0.25	656384
2.	Bidirectional_2	BGRU: 128, return sequence: True, droout:0.25	394240
3.	last_dense	Units: no_class+1, activation: SoftMax	77357
4.	Output + CTC	max_len:11, input length: varies, Label length: varies	0

C. PERFORMANCE EVALUATION METRICS

In the field of deep learning, it's important to assess the performance of models before launching them in the market. This involves using different metrics to ensure that the model is functioning as expected. When it comes to recognition models, the WER is a commonly used metric for evaluating performance. The WER is calculated mathematically in "(9)", where the variables I_w, D_w, S_w, and GT represent the number of word insertions, deletions, substitutions, and the total number of words in the ground truth text, respectively.

$$\mathrm{WER} = \left(\frac{(I_w + D_w + S_w)}{\mathrm{GT}} \right) * 100 \tag{9}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

IV. EXPERIMENT SETUP, RESULTS, AND DISCUSSION

In this section, we elucidate the setup and execution of the experiments, present the outcomes, and deliberate on our findings. We did four tests to show how well our Amharic word recognition model works. We also discuss how we picked the best settings for our model and show the results we got with those settings. Finally, we take a close look at the error analysis in the results.

A. EXPERIMENT SETUP

In this study, we harnessed the computational power of a robust GPU to accelerate image processing tasks, leveraging the capabilities of Google Collaboratory—an efficient, cloud-based service enabling remote code execution. Python, a versatile scripting language, served as our programming foundation, while the Keras framework, backed by TensorFlow, facilitated the seamless development and training of our recognition model.

Our approach involved the creation of four distinct recognition models, each addressing specific considerations. In Case I, we examined the original image without incorporating attention mechanisms, aiming to establish a baseline performance. In Case II, we introduced attention mechanisms to the original images, enhancing the model's ability to focus on crucial features. Case III involved the exploration of augmented images without attention mechanisms, diversifying the training dataset for improved generalization. Finally, in Case IV, we combined the benefits of augmented images with attention mechanisms, seeking to achieve a comprehensive and high-performing recognition model. For training consistency across all cases, we opted for 25 epochs and a batch size of 128. This deliberate choice ensures a balance between model convergence and computational efficiency during the training process. The numerical selection of these parameters was driven by a careful consideration of the trade-off between training time and model accuracy.

B. RESULTS

In this section, we present the accomplishments of the four models developed for recognizing handwritten Amharic words. Each model underwent rigorous training, with both training and validation loss metrics being reported throughout the training phase. Additionally, the models were evaluated using the same testing dataset to ensure a fair and consistent comparison of their effectiveness. **FIGURE 6** shows the training and validation loss graphs for Case I, while **FIGURE 7** displays the same for Case II. **FIGURE 8** and **FIGURE 9** illustrate the training and validation losses for Cases III and IV, respectively. These figures provide valuable insights into each model's learning progress and performance throughout the training process.

By examining these graphs, we can assess how well each model configuration converges and generalizes to unseen data. The comparison of training and validation losses across different cases helps identify the most effective model for recognizing handwritten Amharic words.

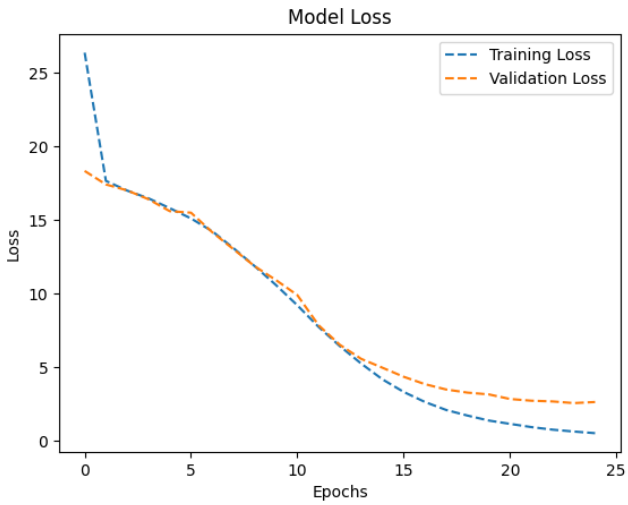


FIGURE 6. Loss of training and validation graph for case I.

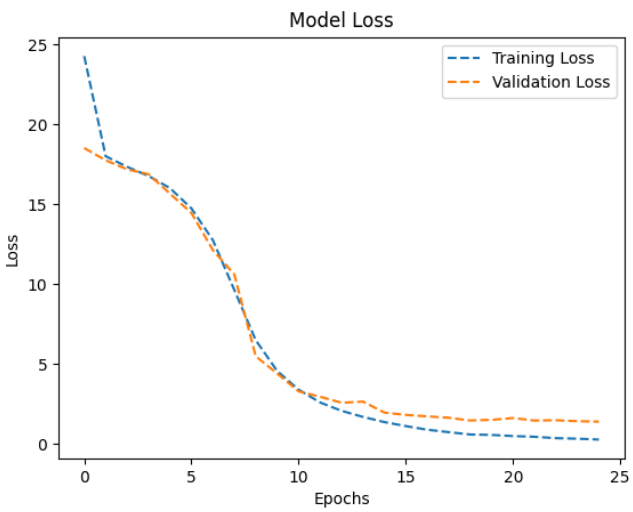


FIGURE 7. Loss of training and validation graph for case II.

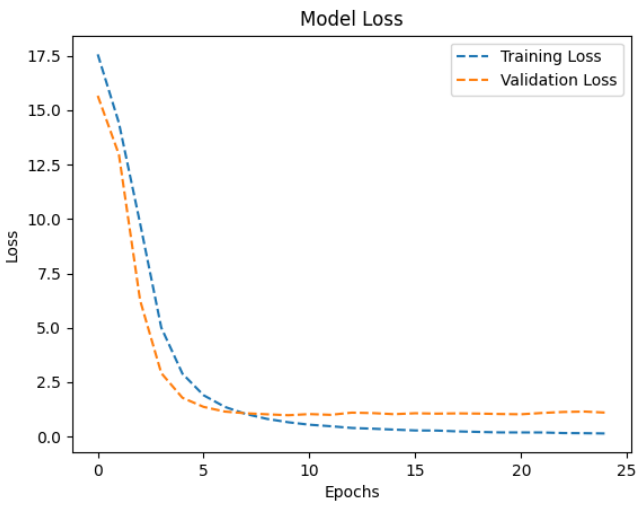


FIGURE 8. Loss of training and validation graph for case III.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

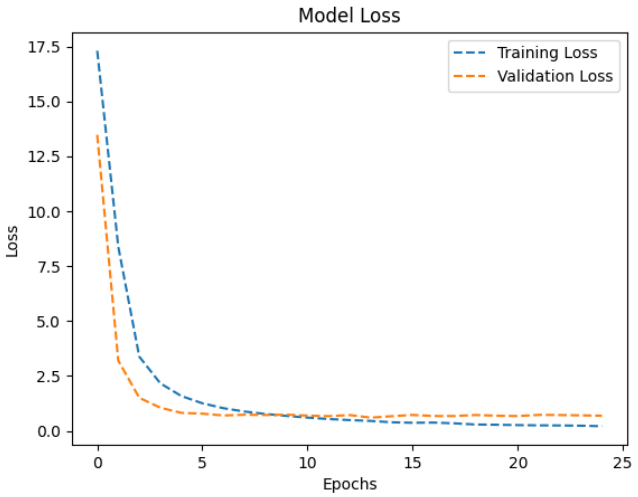


FIGURE 9. Loss of training and validation graph for case IV.

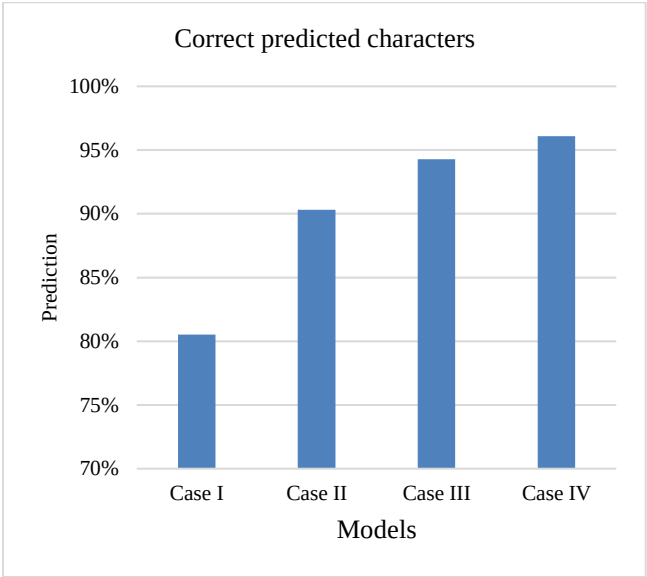


FIGURE 10. A graph of the correct characters' predictions for all cases.

TABLE 6. Testing results of the developed four models.

Cases	Augmentation	Attention	Average CER	Average WER	Correct characters predicted	Correct words predicted
Case I	no	no	11.44%	33.75%	80.51%	36.72%
Case II	no	yes	7.00%	21.42%	90.31%	43.56%
Case III	yes	no	4.75%	15.33%	94.29%	46.93%
Case IV	yes	yes	2.84%	9.75%	96.09%	50.02%

We have also provided the percentage of correctly predicted characters and words. **FIGURE 10** displays a comparison of the correct character prediction results for all four cases, while **FIGURE 11** shows the comparison of correct word prediction results for all four cases.

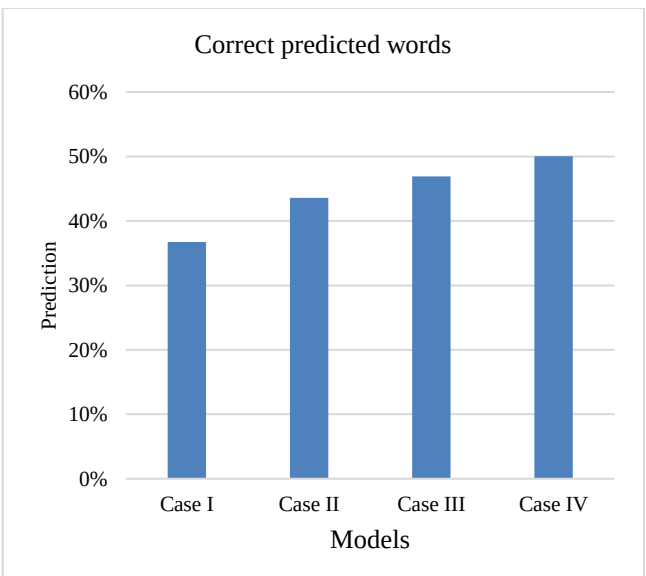


FIGURE 11. A graph of the correct word predictions for all cases.

C. DISCUSSION

Since our experimental dataset is small, we applied the augmented technique to increase the number of images from 12,047 to 34,047. The paper conducts a comparative analysis of four recognition models, taking into account the attention mechanism used for handwritten Amharic word recognition, through the implementation of four scenarios outlined in the experiment setup subsection.

As shown in **TABLE 6**, the results indicate that the models trained with augmented images perform better in terms of average CER and average WER on the testing dataset of 1200 images. The first model, Case I, achieved an average CER of 11.44% and an average WER of 33.75% for the testing dataset. The second model, Case II, achieved an average CER of 7.00% and an average WER of 21.42% for the testing dataset. The third model, Case III, achieved an average CER of 4.75% and an average WER of 15.33% for the testing dataset. The fourth model, Case IV, achieved an average CER of 2.84% and an average WER of 9.75% for the testing dataset.

The inclusion of an attention mechanism in the image training enabled the model to perform better. For further reference, it is possible to cross-check prediction results from **FIGURE 10** and **FIGURE 11**.

1) ERROR ANALYSIS

In this study, we tried to analyze errors that exist in the Amharic word recognition model. Two sample errors are

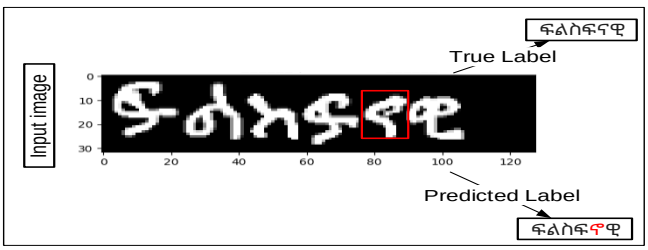
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

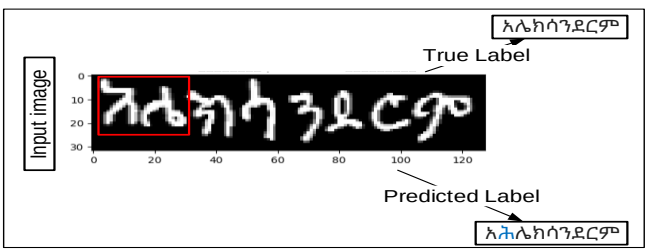
depicted in **FIGURE 12,** considering case IV. The first sample handwritten image, which is labeled as **(a),** has the word "ፍልስፍናዊ" and the second sample handwritten image, which is labeled as **(b)**, has the word "አሌክስአንደርንም". In this figure, characters incorrectly recognized are marked in a red rectangle.

From the first handwritten word image, the character "ና" is predicted as ኖ. In this case, the confusion happened due to the shape similarity of the characters when written by hand.

In the second handwritten word image, ሕ is predicted due to two አ and ሌ letters being connected to each other during handwriting. Each character is predicted correctly, but the character ሕ inserted a character during the prediction. The model is confused about character connectivity. When the space between characters is not sufficient, then the possibility of an error happening is high.



(a)



(b)

FIGURE 12. Sample error results generated for case IV.

2) METHOD COMPARISON

While comparing handwritten recognition systems across various scripts and experimental conditions can be challenging, we highlight two studies related to this paper to indicate progress in Amharic handwriting recognition.

TABLE 7. Effectiveness of the proposed model.

S.No	Study	Technique	Test Data Samples	Results (CER) %
1.	[12]	CNN, RNN, CTC	6,150	17.95
2.	[43]	Gated CNN , Self-Attention encoder	58,000	8.22
3.	Proposed	CNN, Attention, RNN, CTC	1,200	2.84

Malhotra and Addis [12] employed CNN, RNN, and CTC techniques, achieving a Character Error Rate (CER) of 17.95% on 6,150 testing samples. Tadesse et al. [43] proposed a "Gated Convolution and Stacked Self-Attention Encoder–Decoder-Based Model for Offline Handwritten Ethiopic Text Recognition," resulting in a CER of 8.22%.

Our proposed method, utilizing CNN, Attention, RNN, and CTC, achieved a CER of 2.84%, outperforming the previous two studies. As shown in Error! Reference source not found., our model demonstrates superior effectiveness in recognizing Amharic characters.

3) FUTURE WORKS

One direction for future research involves incorporating a multilingual word recognition model into our study. Expanding the model to support multiple languages will make it more versatile and applicable in diverse linguistic contexts. This enhancement would require the collection and processing of multilingual handwriting datasets and the development of advanced techniques to handle the intricacies of different scripts and writing styles.

Currently, our study relies on synthetic datasets for training and evaluation. Future work will focus on utilizing real-life data sources to better capture the variability and complexity of handwritten documents encountered in everyday situations. This transition to authentic datasets will improve the model's generalizability and reliability when applied to practical applications.

Handwritten documents in real-world scenarios often contain various forms of noise, degradation, and variability, such as smudges, blurs, ink bleed, and inconsistent writing styles. Enhancing the model's robustness to these factors is crucial for reliable performance. Future research will aim to develop and integrate techniques that make the model more resilient to these common issues. Potential approaches include advanced preprocessing algorithms, noise reduction techniques, and augmentation strategies to simulate realistic conditions during training.

By pursuing these future directions, we aim to enhance the capabilities and applicability of our handwriting recognition model, making it more robust, versatile, and user-friendly in real-world scenarios.

V. CONCLUSION

In conclusion, this research on handwritten Amharic word recognition with an additive attention mechanism has successfully addressed the challenges associated with recognizing handwritten Amharic words. Leveraging a deep learning approach, specifically, the integration of additive attention with RNN, our proposed recognition model, consisting of seven CNNs and two RNNs using the CTC strategy, has demonstrated promising results.

The study not only identified and navigated the complexities of variations in individual handwriting, connectivity of words without spaces, similarities in alphabet characters, and noise from scanned images, but also tackled the issue of data insufficiency through effective augmentation techniques. The utilization of an original dataset of 10,784 handwritten

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

Amharic words, augmented with 24,000 images, contributed to a more robust and diverse training set.

The positive outcomes achieved in the recognition of handwritten Amharic words demonstrate the additive attention mechanism's effectiveness in improving precision. This research represents a significant stride in bridging the existing gap in OCR technology for the Amharic script, emphasizing the transformative capabilities of deep learning in pattern recognition.

Moving forward, the insights gained from this study can pave the way for further advancements in the field of handwritten language recognition, not only for Amharic but also for other languages facing similar challenges. The success of our approach highlights the potential for broader applications of additive attention mechanisms to improve the accuracy and efficiency of handwritten word recognition systems.

ACKNOWLEDGMENT

We extend our heartfelt gratitude to the esteemed faculty members, dedicated staff, and fellow students at Delhi Technological University. Their unwavering guidance and support have been invaluable throughout the completion of this manuscript, playing a pivotal role in our research project's success. Their continuous assistance and expertise have not only enriched our academic journey, but they have also significantly contributed to our growth as researchers.

We also acknowledge the indispensable role that "ChatGPT (GPT-3.5)" and "Wordtune" played in the development of this study. Their assistance has proven to be extremely beneficial, enhancing the overall quality and depth of our research. We used these tools for language refinement, grammar checking, and enhancing the clarity of our writing.

REFERENCES

[1] M. Meshesha and C. V. Jawahar, "Recognition of printed Amharic documents," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, Seoul, Korea (South), 2005, vol. 2, pp. 784-788 IEEE.

[2] B. H. Belay, T. A. Habtegebrial, and D. Stricker, "Amharic Character Image Recognition," in *18th IEEE International Conference on Communication Technology*, Chongqing, China, 2018, pp. 1179-1182: IEEE.

[3] M. Meshesha, "Recognition and Retrieval from Document Image Collections," Ph.D, Computer Science and Engineering, International Institute of Information Technology Hyderabad, Hyderabad, India, 2008.

[4] Y. Assabie and J. Bigun, "Offline handwritten Amharic word recognition," (in English), *Pattern Recognition Letters*, vol. 32, no. 8, pp. 1089-1099, Jun 1 2011.

[5] E. Y. Obsie, H. QU, and Q. Huang, "Amharic Character Recognition Based on Features Extracted by CNN and Auto-Encoder Models," in *The 13th International Conference on Computer Modeling and Simulation*, Melbourne VIC, Australia, 2021, vol. 21, pp. 58-66: ACM.

[6] K. Tamir, "Handwritten Amharic characters Recognition Using CNN," in *2019 IEEE AFRICON*, Accra, Ghana, 2019, pp. 1-4: IEEE.

[7] F. Abdurahman, E. Sisay, and K. A. Fante, "AHWR-Net: offline handwritten amharic word recognition using convolutional recurrent neural network," *SN Applied Sciences*, vol. (2021) 3:760, pp. 1-11, 2021.

[8] Y. Assabie and J. Bigun, "HMM-Based Handwritten Amharic Word Recognition with Feature Concatenation," in *2009 10th International Conference on Document Analysis and Recognition*, 2009, pp. 961-965.

[9] W. Alemu and S. Fuchs, "Handwritten Amharic Bank Check Recognition Using Hidden Markov Random Field," in *2003 Conference on Computer Vision and Pattern Recognition Workshop*, 2003, vol. 3, pp. 28-28.

[10] D. M. Kassa and H. Hagras, "An Adaptive Segmentation Technique For the Ancient Ethiopian Ge'ez Language Digital Manuscripts," in *2018 10th Computer Science and Electronic Engineering (CEEC)*, 2018, pp. 83-88.

[11] H. T. Weldegebriel, H. Liu, A. U. Haq, E. Bugingo, and D. Zhang, "A New Hybrid Convolutional Neural Network and eXtreme Gradient Boosting Classifier for Recognizing Handwritten Ethiopian Characters," *IEEE Access*, vol. 8, pp. 17804-17818, 2020.

[12] R. Malhotra and M. T. Addis, "End-to-End Historical Handwritten Ethiopic Text Recognition Using Deep Learning," *IEEE Access*, vol. 11, pp. 99535-99545, 2023.

[13] M. Meshesha and C. V. Jawahar, "Optical Character Recognition of Amharic Documents," *African Journal of Information and Communication Technology*, vol. 3, pp. 53-66, 2007.

[14] E. G. Beyene, "Handwritten and Machine printed OCR for Geez Numbers Using Artificial Neural Network," in *33rd NeurIPS 2019 Workshop on Machine Learning for the Developing World (NIPS 2019)*, Vancouver, CANADA, 2019, vol. 1, pp. 1-5: arXiv.

[15] P. Goel and A. Garantra, "Handwritten Gujarati Numerals Classification Based on Deep Convolution Neural Networks Using Transfer Learning Scenarios," *IEEE Access*, vol. 11, pp. 20202 - 20215, 2023.

[16] Y. S. Chernyshova, A. V. Sheshkus, and V. V. Arlazarov, "Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images," *IEEE Access*, vol. 8, pp. 32587-32600, 2020.

[17] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access*, vol. 8, pp. 142642-142668, 2020.

[18] T. Nasir, M. K. Malik, and K. Shahzad, "MMU-OCR-21: Towards End-to-End Urdu Text Recognition Using Deep Learning," *IEEE Access*, vol. 9, pp. 124945-124962, 2021.

[19] L. Xu, Y. X. Wang, X. X. Li, and M. Pan, "Recognition of Handwritten Chinese Characters Based on Concept Learning," (in English), *Ieee Access*, vol. 7, pp. 102039-102053, 2019.

[20] B. H. Belay, "Deep Learning for Amharic Text-Image Recognition: Algorithm, Dataset and Application," PhD, Department of Computer Science, Technische Universit¨at Kaiserslautern, Germany, D 386, 2021.

[21] A. B. M. A. RAHMAN *et al.*, "Two Decades of Bengali Handwritten Digit Recognition: A Survey," *IEEE Access*, vol. 10, pp. 92597-92632, 2022.

[22] B. Belay, T. Habtegebrial, M. Meshesha, M. Liwicki, G. Belay, and D. Stricker, "Amharic OCR: An End-to-End Learning," (in English), *Applied Sciences-Basel*, Open Access vol. 10, no. 1117, pp. 1-13, Feb 2020.

[23] Y. Assabie and J. Bigun, "A Hybrid System for Robust Recognition of Ethiopic Script," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, Curitiba, Brazil, 2007, vol. 1, pp. 1-5: IEEE.

[24] B. Belay, T. Habtegebrial, M. Liwicki, G. Belay, and D. Stricker, "Factored Convolutional Neural Network for Amharic Character Image Recognition," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 2906-2910.

[25] Y. Assabie and J. Bigun, "Lexicon-based offline recognition of Amharic words in unconstrained handwritten text," in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 1-4: IEEE.

[26] H. T. Weldegebriel, J. Chen, and D. Zhang, "Deep learning for Ethiopian Ge'ez script optical character recognision," in *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, Xiamen, China, 2018, pp. 540-545: IEEE.

[27] A. Worku, "The application of OCR techniques to the Amharic scrip," Master Information Science, Addis Ababa University Addis Ababa, Ethiopia, 1997.

[28] R. Malhotra and M. T. Addis, "Ethiopic Base Characters Image Recognition using LSTM," in *2021 2nd International Conference on*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3444897

Computational Methods in Science & Technology (ICCMST), Mohali, India, 2021, pp. 94-98: IEEE.
[29] D. Addis, C. D. Liu, and V. D. Ta, "Printed Ethiopic Script Recognition by Using LSTM Networks," in *2018 International Conference on System Science and Engineering (ICSSE)*, New Taipei, Taiwan, 2018, pp. 1-6: IEEE.
[30] M. A. Nur, M. Abebe, and R. S. Rajendran, "Handwritten Geez Digit Recognition Using Deep Learning," *Applied Computational Intelligence and Soft Computing*, Open access vol. 2022, pp. 1-12, 2022.
[31] S. Aly and A. Mohamed, "Unknown-Length Handwritten Numeral String Recognition Using Cascade of PCA-SVMNet Classifiers," *IEEE Access*, vol. 7, pp. 52024 - 52034, 2019.
[32] B. Gatos, K. Ntzios, I. Pratikakis, S. Petridis, T. Konidaris, and S. J. Perantonis, "An efficient segmentation-free approach to assist old Greek handwritten manuscript OCR," *Pattern Analysis and Applications*, vol. 8, no. 4, pp. 305-320, 2006/02/01 2006.
[33] Y. Assabie and J. Bigun, "Ethiopic Character Recognition Using Direction Field Tensor," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, vol. 3, pp. 284-287.
[34] M. Agegnehu, G. Tigistu, and M. Samuel, "Offline Handwritten Amharic Digit and Punctuation Mark Script Recognition using Deep learning," in *2nd Deep Learning Indaba-X Ethiopia Conference 2021*, Adama, Ethiopia, 2022, pp. 38-47.
[35] B. Y. Reta, D. Rana, and G. V. Bhalerao, "Amharic Handwritten Character Recognition Using Combined Features and Support Vector Machine," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2018, pp. 265-270.
[36] S. B. Ahmed, I. A. Hameed, S. Naz, M. I. Razzak, and R. Yusof, "Evaluation of Handwritten Urdu Text by Integration of MNIST Dataset Learning Experience," *IEEE Access*, vol. 7, pp. 153566-153578, 2019.
[37] M. I. Sharif *et al.*, "Deep Learning and Kurtosis-Controlled, Entropy-Based Framework for Human Gait Recognition Using Video Sequences," *Electronics*, vol. 11, no. 334, pp. 1-23, 2022.
[38] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 11, pp. 2298-2304, Nov 2017.
[39] Y. Zhu, Z. Xie, L. Jin, X. Chen, Y. Huang, and M. Zhang, "SCUT-EPT: New Dataset and Benchmark for Offline Chinese Text Recognition in Examination Paper," *IEEE Access*, vol. 7, pp. 370-382, 2019.
[40] J. Zhang, J. Du, and L. Dai, "Track, Attend, and Parse (TAP): An End-to-End Framework for Online Handwritten Mathematical Expression Recognition," *IEEE Transactions on Multimedia*, vol. 21, no. 1, pp. 221-233, 2019.
[41] B. H. Belay, T. Habtegebirial, M. Liwicki, G. Belay, and D. Stricker, "Amharic Text Image Recognition: Database, Algorithm, and Analysis," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 1268-1273.
[42] Y. Jiang, H. Dong, and A. E. Saddik, "Baidu Meizu Deep Learning Competition: Arithmetic Operation Recognition Using End-to-End Learning OCR Technologies," *IEEE Access*, vol. 6, pp. 60128-60136, 2018.
[43] D. A. Tadesse, C.-M. Liu, and V.-D. Ta, "Gated Convolution and Stacked Self-Attention Encoder–Decoder-Based Model for Offline Handwritten Ethiopic Text Recognition," *Information* vol. 14, no. 654, pp. 1-15, 2023.



RUCHIKA MALHOTRA received the master's and Ph.D. degrees in software engineering from the University School of Information Technology, Guru Gobind Singh Indraprastha University, Delhi, India. She is currently a Professor with the Department of Software Engineering, Delhi Technological University, Delhi. She has published more than 200 research papers in international journals and conferences. Her research interests include software testing, improving software quality, statistical and adaptive prediction models, software metrics, neural nets modeling, and the definition and validation of software metrics.
.



MARU TESFAYE ADDIS received the bachelor's degree in computer science and information technology from Arba Minch University, Ethiopia, in 2010, and the master's degree in computer science from Bahir Dar University, Ethiopia, in 2017. He is currently pursuing the Ph.D. degree in computer science and engineering with Delhi Technological University, India. He is also a Lecturer with the Department of Computer Science, Debre Tabor University. His research interests include pattern recognition, deep learning, artificial intelligence, and image processing. He is also dedicated to advancing knowledge and contributing to the development of innovative solutions in these areas.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)
Karnataka, India. Jun 28-29, 2024

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) | 979-8-3503-8269-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITEICS61368.2024.10625256

Impact of Device Parametrics on a 10nm Gate Length SOI n-FinFET

Anurag Somayajula
Department of Applied Physics
Delhi Technological University
Delhi, India
sanurag_ep20a6_03@dtu.ac.in

Tanish Aggarwal
Department of Applied Physics
Delhi Technological University
Delhi, India
tanishaggarwal_ep20a6_16@dtu.ac.in

Vishnu Sajith
Department of Applied Physics
Delhi Technological University
Delhi, India
vishnusajith_ep20b1_49@dtu.ac.in

Rishu Chaujar
Department of Applied Physics
Delhi Technological University
Delhi, India
chaujar.rishu@dtu.ac.in

***Abstract*—This research paper presents a comprehensive analysis of the behaviour of a Tri-gate Silicon-On-Insulator (SOI) n-FinFET with a gate length of 10nm. Innovative gate stacking featuring SiO_2, HfO_2 oxide layer, and a ferroelectric film along with a TiO_2 spacer are two features of the device that improve performance in low-power applications. Utilizing TCAD models, the research delves into crucial subthreshold variables, including threshold voltage and subthreshold swing and other output parameters like the drain induced barrier lowering (DIBL) which are analysed using Silvaco simulations. Systematic investigation of the effects of the incorporating TiO_2 spacer and gate stacking found the subthreshold swing value of the optimised FinFET design to be 136 mV/dec and the DIBL value for the optimised design is found to be 63.10 mV/V. Changes in the I-on and I-off values for the optimised design are also observed. The effects of the variation of the gate metal work function (Φ_m) are also studied. Our research helps to better comprehend unique design configurations for sub-10nm FinFETs, to continue advancing semiconductor technology and create more effective integrated circuits.**

***Index Terms*—SOI n-FinFET, Subthreshold Swing, DIBL, Gate Stacking, TiO_2 Spacers.**

I. INTRODUCTION

Advances in MOSFET technology have been driven by the pursuit of semiconductor downsizing, with a primary focus on decreasing channel length to improve performance and power efficiency. Nevertheless, substantial short-channel effects (SCEs), such as an increase in off-state current and standby power, have emerged because of aggressive scaling. Novel design approaches have been investigated to overcome these difficulties, like adding a second gate and using insulating substrates to improve gate control. This work focuses on a 10nm gate length Tri-gate Silicon-On-Insulator (SOI) n-FinFET that has a ferroelectric film, an oxide layer HfO2 and SiO2, and a TiO2 spacer in a novel gate stacking. The goal of the study is to examine this advanced FinFET configuration's subthreshold behavior and offer insights into how well it performs in low-power applications.

Moreover, adjusting the channel length adversely affects the gate's ability to regulate channel behavior. To enhance gate control, one approach involves introducing a secondary gate to complement the existing one, or situating the device on an insulator, enabling the body to serve as a bottom gate. Both methods have shown enhanced performance in downsized conventional MOSFET devices [1].

A Silicon-on-Insulator FinFET offers numerous benefits:

- It functions effectively with a reduced supply voltage, and lesser V_{th} (threshold voltage) makes it susceptible to variations in gate length, rendering it suitable for low standby power (LSTP) applications [2].
- It delivers increased I-on (on-current) and decreased I-off (off-current).
- Its subthreshold swing is diminished.
- The occurrence of short-channel effects (SCEs) and leakage levels are minimized compared to those observed in bulk FinFETs [3].

The integration of SOI structure alongside scaling the MOSFET device channel has not only enhanced the device's speed but also its on-state current. SOI mitigates capacitance at the source and drain junctions by eliminating depletion regions extending into the substrate. Moreover, the SOI substrate offers dielectric isolation for the devices, thereby minimizing the influence of parasitic effects commonly observed in bulk devices.

The introduction of FinFETs stemmed from reduced process complexity and their compatibility with the CMOS fabrication process flow. The emergence of 3D FET devices like Tri-gate/FinFET architectures has been explored to prolong Moore's Law and to reduce power consumption when compared to planar MOSFETs. FinFETs feature a distinctive geometry that enables better management of short-channel effects [4][5].

The Tri-gate FinFET configuration enhances gate electrostatic control over the channel, resulting in improved performance. The combination of SOI structure with MOSFET channel scaling has been demonstrated to increase device speed and on-state current. Specifically, the study explores the subthreshold behavior analysis of the Tri-gate SOI n-FinFET with a gate length of 10 nm, building upon this

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

background. A major focus is on the addition of a ferroelectric film, SiO2, HfO2 oxide layer, and a TiO2 spacer to the advanced gate stacking materials. We seek to investigate crucial subthreshold characteristics, current values, threshold voltage, and subthreshold swing, using TCAD simulations. It is anticipated that the research's findings will provide important new perspectives on how this FinFET arrangement might be used in low-power applications to achieve energy efficiency. An overview of the results and their implications for the ongoing advancement of semiconductor technologies are presented in the paper's further sections.

II. Simulation and Structure of SOI n-FinFET Device Model

The proposed device architecture of the FinFET underwent simulation in Silvaco using the BQP model, an extension of the Wigner equation. Developed in partnership with the University of Pisa and integrated into ATLAS, this model accurately considers C-V curves and the effects of quantum confinement on electron and hole concentrations [6]. Compared to the Density Gradient approach, the BQP model offers improved convergence properties, which is a significant advantage. It allows for calibration against results obtained from the Schrödinger-Poisson equation with minimal current flow. The simulation is grounded on the Bohm interpretation, which is non-local and non-relativistic, yet causal. This investigation delves into the simulation outcomes, providing valuable insights into the performance characteristics of the FinFET structure.

TABLE I. DEVICE DESIGN PARAMETERS

Device Parameters	Values (Units)
Gate Length	10 (nm)
Fin Height (HFin)	15 (nm)
Fin Width (WFin)	10 (nm)
Buried Oxide Thickness (BOX)	30 (nm)
Gate Oxide Thickness (GOX)	1 (nm) - SiO2 1 (nm) - HfO2
Ferroelectric Film	1 (nm)
Spacer Length	2 (nm)
Acceptor Concentration (NA)	1e16
Donor Concentration (ND)	1e20
Band Gap	1.1245 (eV)
Electron Affinity	4.05 (eV)
Relative Permittivity	11.9

A smaller W_{Fin} is necessary to improve gate control over the channel, consequently leading to a decreased subthreshold swing [7].

The gate geometry is pivotal in governing the performance of FinFETs. The dimension referred to as "Fin width" (W_{Fin}), illustrated in Fig. 1, influences the device's performance. Additionally, the height of the fin (H_{Fin}) restricts the channel width of a single fin device. A multigate device's channel width is specified as follows:

$$W \sim 2H_{Fin} + W_{Fin} \quad (1)$$

The structure shown in Fig. 1 is a 3D rendition of the proposed FinFET design based on the parameters defined in Table 1, a proposed novel design of FinFET architecture, below which follows a detailed elucidation of the key components, each meticulously chosen to augment the performance and capabilities of the transistor.

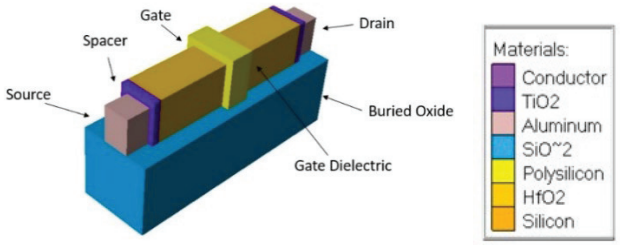


Fig. 1. A three-dimensional representation depicting the structure of an SOI n-FinFET with a gate length of 10 nm.

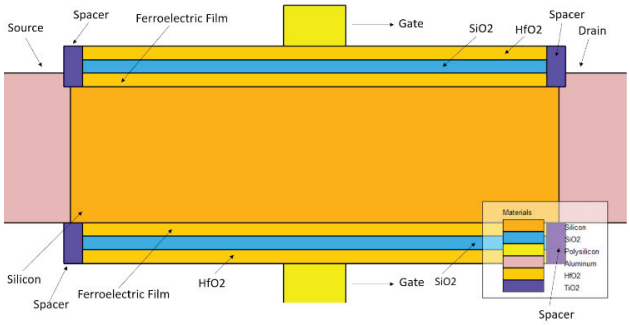


Fig. 2. A two-dimensional visualization portraying the structure of an SOI n-FinFET with a gate length of 10 nm.

An essential component of the design is the Buried Oxide (BOX) layer, which is shown in blue. This demonstrates our commitment to Silicon-On-Insulator (SOI) technology, which is essential for increased insulation, reduced parasitic effects, and enhanced device performance in SOI FinFETs. A silicon substrate acts as the fundamental layer and conduction conduit, positioned above the BOX layer. The electrical properties of the FinFET are largely determined by this elemental layer. We incorporate a new element surrounding the silicon substrate: a thin coating of ferroelectric material, called the spacer [8]. It is anticipated that as scaling occurs, the overall parasitic capacitance will rise and account for a significant portion of the total capacitance [9]. Power dissipation, propagation delay, and capacitive load in circuits are all increased by increased parasitic capacitance [10]. To overcome this a layer of Tio_2 as the spacer material is used to lessen the parasitic capacitance. With this, the transistor gains dynamism and may find use in non-volatile memory and enhanced switching capabilities.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

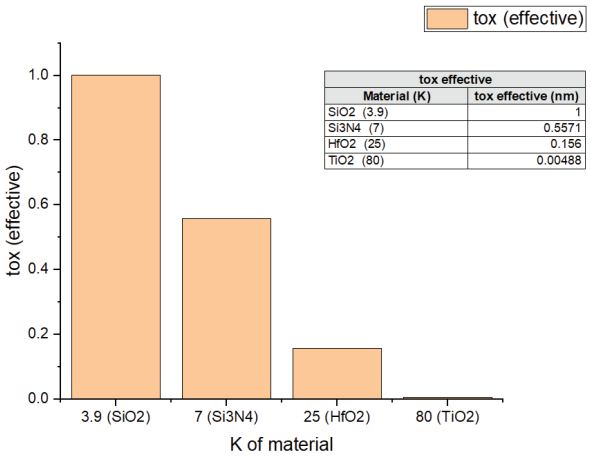


Fig. 3. Change in effective TOX with change in stack material.

Arranged above the ferroelectric film are stacked oxide layers, which are crucial for device functionality [11]. However, the performance of the device deteriorates as the gate oxide thickness (GOX) decreases and approaches 2 nm. In the proposed structure, the GOX value is 1 nm, resulting in an escalation of gate direct tunneling current. Consequently, the off-state leakage current increases exponentially [12]. To solve this issue, high-κ dielectric material is employed. Nonetheless, the direct deposition of high-dielectric materials on silicon substrate continues to be a difficult field. Another option or an alternative approach is to integrate Gate Stack (GS) architecture, which preserves a consistent effective oxide thickness (TOX) by introducing a thin interfacial SiO_2 layer to serve as a protective barrier between the Si substrate and high-κ dielectrics.This stack structure, which consists of a blend of SiO_2 and HfO_2 oxides, is specifically designed to maximize gate dielectric characteristics [13][14]. It tackles issues associated with gate leakage and guarantees accurate channel management. The gate structure is located on top of the stacked oxide layers [15]. This part acts as the main control element and has a big impact on the transistor's current flow. The material properties of the gate are designed as to provide efficient channel modulation and improved electrostatic control. The ensemble is completed with a spacer placed strategically around the entire construction. This spacer contributes to the overall stability and precision of the device, playing a pivotal role in defining physical dimensions and ensuring uniformity in electrical characteristics [16][17].

III. RESULTS AND DISCUSSIONS

The construction and properties of the FinFET device were built, examined, and simulated in three dimensions using the software package Silvaco-Atlas, as discussed in section 2 of this paper. This section will focus on the results obtained from the simulation run.

A. Transfer and Output Characteristics

In FinFETs, the analysis of transfer and output characteristics refers to the examination of how the transistor responds to changes in the Gate source voltage (V_{gs}) and Drain source voltage (V_{ds}), respectively. For the analysis of this FinFET design, the values of Drain Current (I_d) under the changing values of V_{gs} and V_{ds} are noted.

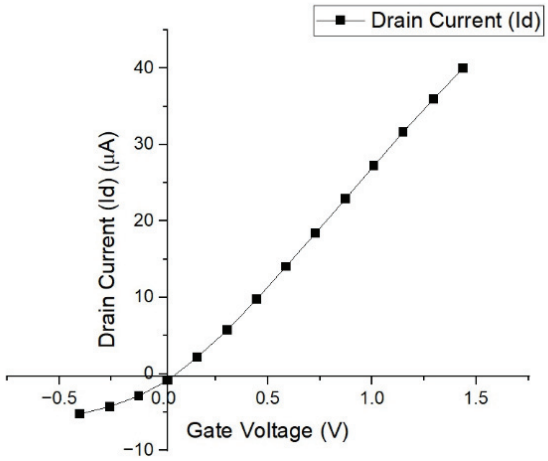


Fig. 4. Characteristics of Ids - Vgs for the FinFET at Vds = 0.1V, depicted on a linear scale.

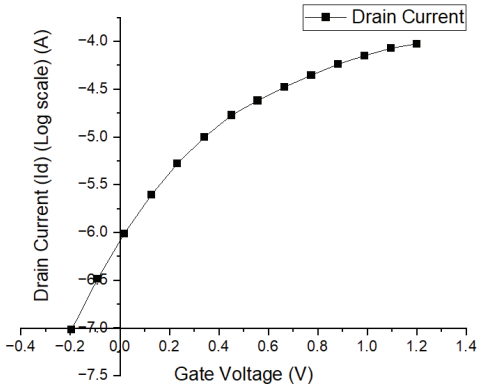


Fig. 5. Characteristics of Ids - Vgs for the FinFET at Vds = 0.1V, illustrated on a logarithmic scale.

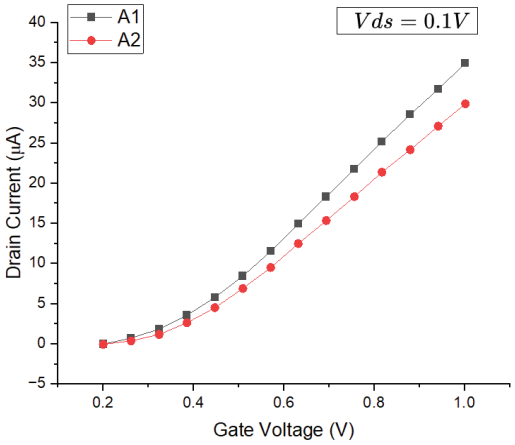


Fig. 6. Comparison of the Ids - Vgs characteristics for the FinFET designs A1 and A2 at Vds = 0.1V.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

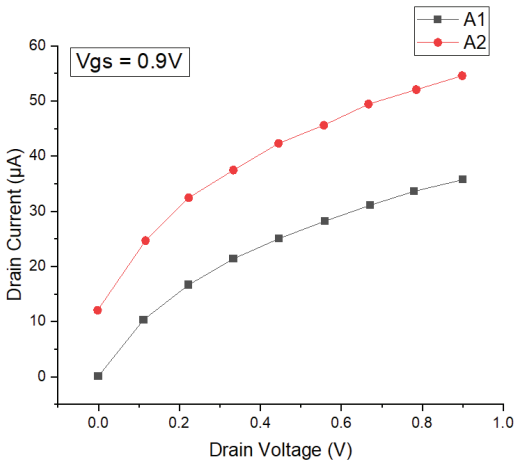


Fig. 7. Comparison of the Ids - Vds characteristics for the FinFET designs A1 and A2 at Vgs = 0.9V.

In Fig. 4 and 5 the I_d-V_{gs} transfer characteristic of the SOI n-FinFET are shown in the linear and log scale respectively. Fig. 6 and 7 shows us the transfer characteristics, I_d-V_{gs} and I_d-V_{ds} respectively of the SOI n-FinFET designs A1 and A2. The optimized structure incorporating a spacer and a gate stack is labeled as A1 and the FinFET design without any structural modifications but dimensionally alike to the optimized structure is termed as A2, this classification is followed in further sections of the paper. It can be observed from Fig.6 and 7 that on comparing the transfer characteristic curve of the FinFET structures with and without the design enhancements, the I_d-V_{gs} curve for the improved structure is steeper than the basic structure.

A steeper I_d-V_{gs} curve indicates a more abrupt increase in drain current (I_d) with respect to gate-source voltage (V_{gs}) and an enhanced transistor sensitivity to gate voltage changes, which can be advantageous for optimizing power efficiency and achieving better performance in low-power applications.

In Fig. 4, the calculated Threshold Voltage (V_{th}) value of the SOI n-FinFET design, incorporating design optimizations, is 0.24V at V_{ds} = 0.1V. In FinFETs, the threshold voltage (V_{th}) denotes the gate-source voltage at which the transistor begins conducting current. This parameter is crucial as it defines the device's switching behavior: below V_{th}, the transistor remains "off," permitting minimal current flow between the source and drain; surpassing V_{th} transitions it to the "on" state, enabling significant current flow. The threshold voltage for a Multigate Field Effect Transistor (MuGFET) device structure is formulated as:

$$V_{Th} = \phi_{ms} + 2\phi_f + \frac{Q_d}{C_{ox}} + \frac{Q_{ss}}{C_{ox}} + V_{in} \quad (2)$$

In this context, Φ_{ms} symbolizes the work function of the metal-semiconductor interface. Φ_f indicates the Fermi potential, while Q_{ss} represents the charge stored in the gate dielectric. Moreover, Cox denotes the gate capacitance, and Q_d refers to the depletion charge within the channel.

The threshold voltage (V_{th}) value obtained for the optimized SOI n-FinFET design is utilized to simulate its subthreshold behavior in Silvaco ATLAS. An essential parameter for analyzing the structure's subthreshold behavior is the Subthreshold Swing (SS). Subthreshold swing (SS) of an n-FinFET, representing the rate of change of the gate-source voltage (V_{gs}) concerning the drain current (I_d) in the subthreshold region, plays a crucial role. This region is where the transistor operates at low currents. SS serves as a vital metric for evaluating the efficiency of the transistor in low-power applications [18].

Furthermore, the subthreshold swing (SS) can be calculated as per the formula:

$$SS\ (mV/dec) = \frac{dV_{GS}}{d(\log_{10} I_{DS})} \quad (3)$$

Upon calculation the subthreshold swing (SS) value obtained from the slope mentioned in Fig.8 for the optimized FinFET design (Structure A1) comes out to be 136mV/dec i.e., a 136mV A tenfold alteration in drain current (I_d) occurs due to a variation in gate voltage (V_{gs}). For reference the SS value of structure A2 is 161mv/dec, this shows that there is a change of 15.52% in the observed SS values for structures A1 and A2.

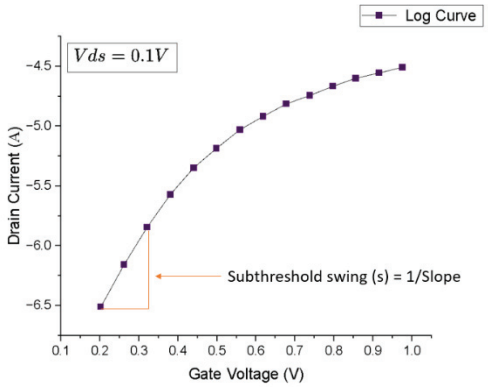


Fig. 8. Characteristic Id - Vgs curve in logarithmic scale, emphasizing the subthreshold slope of the FinFET designs at Vds = 0.1V.

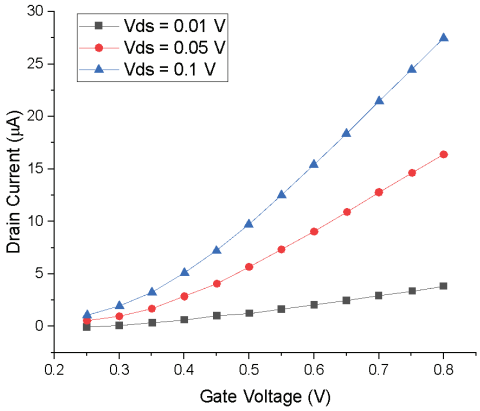


Fig. 9. Current (Ids) - Gate-Source Voltage (Vgs) characteristics for an SOI n-FinFET at various drain voltages, represented in linear scale.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

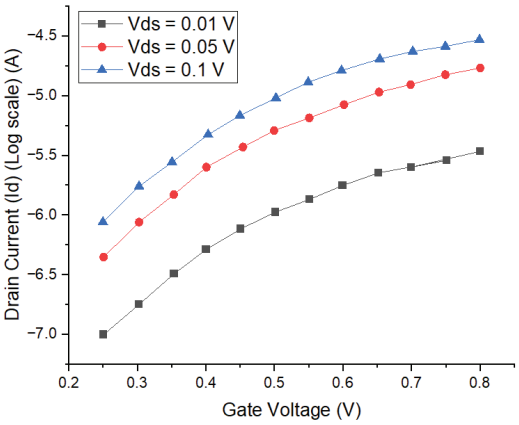


Fig. 10. Current (Ids) - Gate-Source Voltage (Vgs) characteristics for an SOI n-FinFET at various drain voltages, depicted in logarithmic scale.

At a constant V_{gs} value, the mean velocity of electrons moving from the source to the drain escalates. with an elevated drain voltage. As illustrated in Fig. 9, the conductance escalates when the gate voltage exceeds the threshold voltage. Consequently, if the gate voltage exceeds the threshold value and the drain voltage is elevated, a higher current will flow. The results obtained from Fig. 9 and 10 indicate that the device exhibits a satisfactory saturation region under increased gate bias.

The I-on and I-off values were calculated for the FinFET structure without any design optimizations and the FinFET structure with added design optimizations like the spacer and gate stack. The values of the I-on and I-off are given below in the table 2:

TABLE II. COMPARISON OF I ON AND I OFF VALUES

VALUES	STRUCTURE A1 (OPTIMIZED DESIGN)	STRUCTURE A2
I ON	35.69µA	31.18µA
I OFF	18.775nA	34.439nA
Switching Ratio	1903	905

The decrease in the I-off value observed in the optimized design is a beneficial outcome, as it helps mitigate static power dissipation of the device, particularly during standby mode. Table 2 also makes it evident that the device's switching ratio has been enhanced, which improves speed, dependability, and power efficiency. It improves noise immunity, quickens switching, and decreases leakage current. It also increases the voltage margin and facilitates low-power design techniques for circuits that are energy-efficient without sacrificing functionality.

Equation 4 is used to determine the drain induced barrier lowering (DIBL).

$$\text{DIBL (mV/V)} = \frac{\triangle V_{Th}}{\triangle V_{DS}} \quad (4)$$

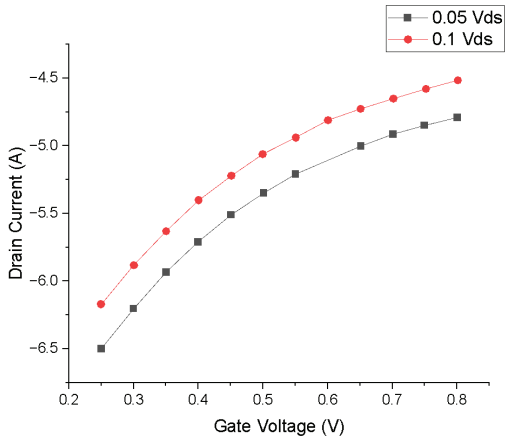


Fig. 11. DIBL in SOI n-FinFET Structure A1

The change in threshold voltage resulting from an increase in drain voltage from 0.01 V to 0.05 V is referred to as Drain Induced Barrier Lowering (DIBL) [19]. In our case, illustrated in Fig. 11, DIBL measures 63.10mV/V for structure A1 (the SOI n-FinFET design, incorporating gate stacking and spacers). The calculated DIBL value for structure A2 is 67mV/V.

B. Impact of Gate Metal Work Function Variation

Variations in the gate metal work function have a substantial impact on FinFET performance. It regulates the subthreshold swing, leakage current, threshold voltage, and general operation of the device.

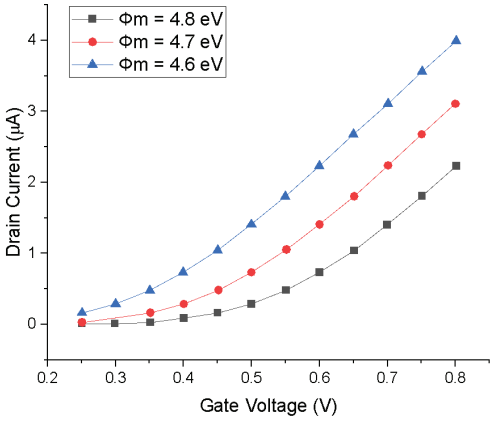


Fig. 12. Characteristics of Ids - Vgs for the SOI n-FinFET represented in a linear scale.

By optimizing the work function, device features can be adjusted to better suit application needs and increase energy efficiency. Analyzing the effect of the gate metal work function (Φm) on the device's performance and current gives us valuable insights [20]. The transfer characteristics of the optimized device (structure A1) are depicted in Figures 12 and 13, demonstrating the effect of increasing Φm from 4.6eV to 4.8eV.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

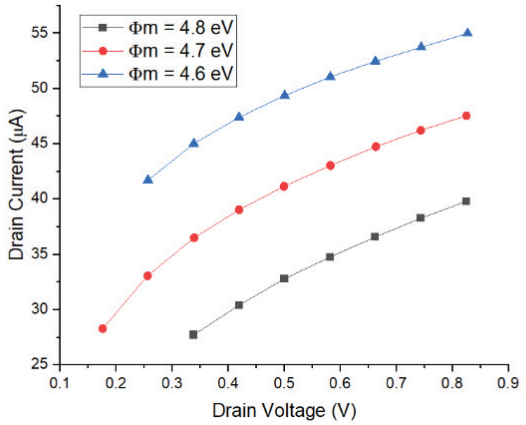


Fig. 13. Characteristics of Ids - Vds for the SOI n-FinFET displayed in a linear scale.

As illustrated in Figs. 12 and 13, we can see that when the gate work function grows, the drain current characteristic slope decreases.

IV. CONCLUSION

Several significant conclusions can be made from the analysis and simulation findings of the SOI n-FinFET model utilizing Silvaco-Atlas. The improved design of the FinFET seeks to maximize efficiency. It has a unique component arrangement that includes the ferroelectric film, stacked oxide layers, and buried oxide (BOX) layer. The improved FinFET structure (structure A1) shows steeper Id-Vgs curves, indicating enhanced transistor sensitivity to gate voltage changes, which is important for low-power applications, according to assessments of transfer and output characteristic data. In addition, the A1 FinFET structure or the optimized design leads to the substantial decrement in the subthreshold swing by a factor of 15.52% and drain induced barrier lowering (DIBL) by a factor of 5.82% along with the decrement in the off-state current by a factor of 45.48% and an increment in the on-state current by the factor of 14.46%. The decreased DIBL value and enhanced on/off current ratios enhances speed, dependability, and power efficiency.

The key features of the optimized design which make them apt for a variety of applications include:

- Energy efficiency – As the subthreshold swing and drain-induced barrier lowering decrease, power consumption diminishes significantly, thereby prolonging the device's operation. The modified structure of A1 offers lowering of off state current which aligns itself to the energy efficiency characteristic of the device.
- Higher integration density – As the channel length decreases, it becomes easier to maintain higher transistor density on the chip. Subthreshold swing and DIBL are short channel effects and arise due to high density packing of transistors but with the reduction in their values the effect of short channel effects is reduced to an appreciable extent.
- Improved performance – with the mentioned improvements the computational throughput of the device increases with high switching speed (I_{on}/I_{off}) which allows for better pattern recognition and faster learning of the device structure and allows for faster decision making.

These improvements find their applications in the development of ultra-low powered devices, moreover the added advantage of miniaturization of the transistor has allowed it to be used in the medical sector for the creation of neural simulators. It can be an appreciable contributor in the field of neuromorphic computing, which is the process of mimicking the human brain for performing complex cognitive tasks.

Furthermore, changes in the gate metal work function have a considerable impact on the device's performance and its transfer characteristics of different Φm values are plotted. All things considered, these results demonstrate the promise of the suggested FinFET design and offer insightful information about how to best optimize its performance characteristics for a range of applications.

REFERENCES

[1] A. Es-Sakhi and M. Chowdhury, "Analysis of device capacitance and subthreshold behavior of Tri-gate SOI FinFET," Microelectronics Journal, vol. 62, pp. 30–37, Apr. 2017

[2] K. Bailey and K. S. Gurumurthy, "3D Device Modeling and Assessment of Triple Gate SOI FinFET for LSTP Applications," International Journal of Computer Theory and Engineering, pp. 831–834, 2012

[3] A. Gill, Charu Madhu, and P. Kaur, "Investigation of short channel effects in Bulk MOSFET and SOI FinFET at 20nm node technology," Dec. 2015

[4] M. Saremi, A. Afzali-Kusha, and S. Mohammadi, "Ground plane fin-shaped field effect transistor (GP-FinFET): A FinFET for low leakage power circuits," Microelectronic Engineering, vol. 95, pp. 74–82, Jul. 2012

[5] J. Lacord, G. Ghibaudo, and F. Boeuf, "Comprehensive and Accurate Parasitic Capacitance Models for Two- and Three-Dimensional CMOS Device Structures," IEEE Transactions on Electron Devices, vol. 59, no. 5, pp. 1332–1344, May 2012.

[6] S. Tripathi, R. Mishra, V. Narendra, and R. Mishra, "Optimization of Pie-gate Bulk FinFET Structure," International Journal of Computer Applications, vol. 59, no. 2, pp. 975–8887, 2012, Accessed: Mar. 27, 2024.

[7] E. Shang, Y. Ding, W. Chen, S. Hu, and S. Chen, "The Effect of Fin Structure in 5 nm FinFET Technology," Journal of Microelectronic Manufacturing, vol. 2, no. 4, pp. 1–8, 2019

[8] Z. Zhang et al., "FinFET with Improved Subthreshold Swing and Drain Current Using 3-nm Ferroelectric Hf0.5Zr0.5O2," IEEE Electron Device Letters, vol. 40, no. 3, pp. 367–370, Mar. 2019

[9] M. Saremi, B. Ebrahimi, and A. Afzali-Kusha, "Ground plane SOI MOSFET based SRAM with consideration of process variation," Dec. 2010

[10] B. Kumar and Rishu Chaujar, "Numerical simulation of analog metrics and parasitic capacitances of GaAs GS-GAA FinFET for ULSI switching applications," The European Physical Journal Plus, vol. 137, no. 1, Jan. 2022

[11] B. Kumar and R. Chaujar, "Analog and RF performance evaluation of junctionless accumulation mode (JAM) gate stack gate all around (GS-GAA) FinFET," Silicon, vol. 13, pp. 919–927, 2021

[12] Y. Omura, H. Konishi, and K. Yoshimoto, "Impact of Fin Aspect Ratio on Short-Channel Control and Drivability of Multiple-Gate SOI MOSFET's," JOURNAL OF SEMICONDUCTOR TECHNOLOGY

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

AND SCIENCE, vol. 8, no. 4, pp. 302–310, Dec. 2008, Accessed: Mar. 27, 2024

[13] A. B. Sachid, M.-C. Chen, and C. Hu, "FinFET with High- kappa Spacers for Improved Drive Current," IEEE Electron Device Letters, vol. 37, no. 7, pp. 835–838, Jul. 2016

[14] A. B. Sachid, M.-C. Chen, and C. Hu, "Bulk FinFET with Low- kappa Spacers for Continued Scaling," IEEE Transactions on Electron Devices, vol. 64, no. 4, pp. 1861–1864, Apr. 2017.

[15] Nazanin Baghban Bousari, M. K. Anvarifard, and Saeed Haji-Nasiri, "Improving the electrical characteristics of nanoscale triple-gate junctionless FinFET using gate oxide engineering," AEU - International Journal of Electronics and Communications, vol. 108, pp. 226–234, Aug. 2019

[16] B. Kumar, A. Kumar, and Rishu Chaujar, "The Effect of Gate Stack and High-κ Spacer on Device Performance of a Junctionless GAA FinFET," 2020 IEEE VLSI DEVICE CIRCUIT AND SYSTEM (VLSI DCS), Jul. 2020

[17] A. Chhabra, A. Kumar, and Rishu Chaujar, "GaAs Junctionless FinFET Using Si3N4 Spacer for High Performance Analog Application," 2018 International Conference on Advances in Communication and Computing Technology (ICACCT), Feb. 2018

[18] Azzedin Es-Sakhi and M. H. Chowdhury, "Analytical model to estimate the subthreshold swing of SOI FinFET," Dec. 2013

[19] Q. Berlingard et al., "Study of threshold voltage extraction from room temperature down to 4.2 K on 28 nm FD-SOI CMOS technology," Solid-State Electronics, vol. 194, pp. 108325–108325, Aug. 2022

[20] Nour, Baghdad Hadri, A. Caddemi, G. Crupi, and S. Patanè, "3-D Simulation of Nanoscale SOI n-FinFET at a Gate Length of 8 nm Using ATLAS SILVACO," Transactions on Electrical and Electronic Materials, vol. 16, no. 3, pp. 156–161, Jun. 2015

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 28,2024 at 05:08:19 UTC from IEEE Xplore. Restrictions apply.

Journal of Thermal Analysis and Calorimetry
https://doi.org/10.1007/s10973-024-13558-x

Improving single-slope passive solar still efficiency through integration of phase change materials and Al_2O_3 nanoparticles

Yunis Khan[1,2] · Zafar Said[3,4] · Roshan Raman[2,5] · Parminder Singh[6] · Mohammad Mehdi Rashidi[7] · Hakan Caliskan[8] · Aman Garg[5,9]

Received: 31 July 2023 / Accepted: 3 August 2024
© Akadémiai Kiadó, Budapest, Hungary 2024

Abstract

This research explores the enhancement of single-slope solar still performance using phase change material (PCM), specifically paraffin, incorporating Al_2O_3 nanoparticles. The application of paraffin, a PCM, improves energy storage density and maintains a consistent temperature during the phase transition. Adding Al_2O_3 nanoparticles to the PCM improves its thermal properties, increasing production rates. Three scenarios were tested for comparison: (1) a standalone solar still, (2) a solar still with PCM, and (3) a solar still with PCM containing Al_2O_3 nanoparticles. The productivity yields for these systems were 0.837 kg, 0.924 kg, and 1.145 kg, respectively. The results indicate a significant improvement in the solar still's performance upon adding PCM and Al_2O_3 nanoparticles, yielding a 10.38% increase in daily output and a 36.77% increase in daily distillate compared to the standalone solar still. Optimizing the temperature difference between the water and the glass surface through ideal water spraying conditions also bolstered the distillate production rate. The outcomes from this research suggest that solar distillation plants, which provide an efficient source of clean drinking water, can significantly improve performance and productivity by leveraging the benefits of innovative materials such as nanoparticles.

Keywords Experimental analysis · Single-slope passive solar still · Energy-absorbing material · Nanoparticle

List of symbols

PCM	Phase change material
I	Intensity of solar radiation/W m^{-2}
P	Productivity/m^{-3}
A	Area/m^2

Suffix

n	Nanoparticle

Greek letters

η	Efficiency
ϕ	Mass fraction of the nanoparticle

✉ Roshan Raman
roshanraman@ncuindia.edu

[1] Department of Mechanical Engineering, Indian Institute of Technology (ISM), Dhanbad, Jharkhand 826004, India

[2] Department of Mechanical Engineering, Delhi Technological University, Bawana Road Delhi, Delhi 110042, India

[3] Department of Sustainable and Renewable Energy Engineering, University of Sharjah, P.O.Box 27272, Sharjah, United Arab Emirates

[4] Department of Industrial and Mechanical Engineering, Lebanese American University (LAU), Byblos, Lebanon

[5] Department of Multidisciplinary Engineering, The NorthCap University, Gurugram, Haryana 122017, India

[6] Chemical Engineering Department, Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India

[7] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610056, Sichuan, China

[8] Department of Mechanical Engineering, Faculty of Engineering and Natural Sciences, Usak University, 64200 Usak, Türkiye

[9] State Key Laboratory of Intelligent Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Published online: 22 August 2024

Introduction

The exponential growth of the population and the escalating energy demand have led to the depletion of fossil fuel reserves, environmental contamination, and increased expenses associated with production and consumption [1]. In light of these challenges, solar energy has emerged as a promising alternative for various applications, including wastewater treatment [2]. However, the intermittent solar irradiation poses a substantial obstacle to the energy management of this technology. Moreover, in areas devoid of contemporary technological advancements, the process of distillation continues to be employed for the purpose of generating drinkable water [3]. This process can benefit from the use of phase change materials (PCMs), which are substances capable of storing and releasing heat during phase changes, such as the transition from solid to liquid. The incorporation of PCMs in distillation systems holds the potential for improving energy efficiency and water purification processes [4]. Bo et al. [5] proposed a new solar-driven PCMs that integrated an interfacial evaporation system to achieve high energy-efficiency desalination. Moreover, this technique proved to be very efficient in utilizing intermittent solar energy. Additionally, PCMs, such as paraffin, salt hydrates, and fatty acids, are frequently employed in various technical applications. In a different study, Liu et al. [6] aimed to improve the performance of PCMs by integrating halloysite nanotubes (HNTs) into organic PCM microcapsules. As a result, the inclusion of HNTs in the PCM microcapsules led to enhanced thermal storage, temperature regulation, flame retardancy, and thermal conductivity of the composite. Another experimental study conducted by Yousef and Hassan et al. [7] focused on solar distillation systems that utilized PCM and wool fibers to increase daytime freshwater output. Through their research, they discovered that PCM significantly improved the exergetic and energetic performance of the distillation system. In the field of desalination processes, nanomaterials have become an increasingly popular choice due to their ability to enhance solar thermal performance [8]. In fact, solar stills, which play a critical role in solar energy systems and offer eco-friendly solutions for hard water or brine treatment, have leveraged the use of nanomaterials to boost their efficiency [9]. Various researchers have employed nanofluids to optimize the performance of solar stills [10]. For instance, Kabeel et al. [11] developed an experimental model of solar still that effectively incorporated an oil heat exchanger and PCMs to substantially increase the production of freshwater. In addition, researchers such as Chaichan et al. [12] have discovered that the addition of nanoparticles to a solar distillation system enhances heat transfer and boosts distillation yield. It has been demonstrated that nanofluids, which consist of nanoparticles dispersed within a base fluid, have the potential to improve the performance and thermophysical properties of heat exchanger fluids [10]. Overall, it is evident that nanomaterials and nanofluids offer optimistic results in solar desalination and the treatment of hard water or brine [10]. Additionally, it has been observed that the incorporation of nanofluids increases the rate of evaporation of nanoparticle/fluid-based solar stills, particularly in regions with limited energy supplies; however, difficulties like long-term stability and life cycle assessment still need to be resolved [13]. The productivity of solar stills depends on efficiency parameters including thermal conductivity, evaporation, and condensation rates, which can be improved by incorporating nanofluids into solar stills [9]. Various design and operational aspects, such as still type, inclination angle, and energy storage, must be taken into account to achieve optimal production [14]. The efficiency and productivity of solar stills can be significantly increased, making them more efficient at addressing the freshwater shortage if correct optimal design considerations are made with the usage of nanofluids [8]. This is corroborated by studies that demonstrate water outputs from 4 kg m^{-2} to 15 kg m^{-2} per day may be produced by active solar stills, which function by utilizing solar energy [15]. Additionally, the use of nanofluids in these distillation systems has shown a notable improvement in both thermal conductivity and absorptivity, resulting in a general increase in system efficiency [15]. Overall, the use of nanofluids in solar stills is a promising approach to boosting productivity and efficiency, especially in remote places with a lack of fresh water. Nevertheless, solar stills continue to encounter constraints in terms of their efficiency and the deposition of salt [16]. Numerous studies have been undertaken with the aim of enhancing the efficiency, productivity, and cost-effectiveness of solar stills. A comparative study was conducted in Chennai, India, to assess the energy efficiency of single-slope solar stills utilizing cotton cloth energy storage [17]. The results indicated that the solar still equipped with a 6-mm cotton cloth exhibited the highest energy efficiency, measuring 23.8%. Furthermore, this configuration demonstrated the highest exergy efficiency, reaching 2.6%. Consequently, the implementation of the 6 mm cotton cloth solar still resulted in a notable 24.1% increase in productivity [18]. The investigation determined that a solar still of moderate size, equipped with reflectors and a glass cover for cooling, exhibited superior performance in terms of freshwater productivity and energy efficiency [19]. Pathak et al. investigated a hybrid system for the distillation and production of hot water using a heat pipe-equipped vacuum tube collector and a solar still. The system increased the productivity of distilled water by 152.9% at 40% depth and 162% at

60% depth for variable flow rates. The energy and exergy efficiencies increased by 40–22% and 0.7% to 4.1%, respectively [20]. A comparative study was conducted to evaluate three different configurations of solar stills. The results indicated that the solar stills equipped with a packed layer of glass balls as a thermal storage medium exhibited the highest efficiency, reaching 14.96% [21]. Furthermore, these solar stills demonstrated the potential for integration with parabolic trough collectors, which enhances their viability for deployment in water-scarce regions [21]. Furthermore, an examination was conducted to assess the influence of nanoparticles on the thermal efficiency of single-basin dual-slope solar stills [22]. The findings of the study indicated that the utilization of copper oxide nanoparticles yielded the highest degree of enhancement in the thermal efficiency of solar stills [22]. Subsequent investigations have delved into the utilization of phase change materials (PCMs), wherein passive stills have exhibited enhancements of up to 120%, while active stills have achieved a remarkable improvement of 700% [4]. Moreover, in another investigation researchers suggested that the incorporation of γ-Al_2O_3 nanoparticles in a solar still can improve the thermal characteristics of saline water and enhance the distillate yield. Specifically, the addition of 0.3 mass% of γ-Al_2O_3 nanoparticles resulted in a substantial increase in distillate yield [23]. Additionally, a separate study examined the utilization of novel substances, such as phase change material and nanoparticles, in combination with water flow across a glass cover, intending to improve the daily production rate of distilled water in a solar thermal application involving single-slope single-basin solar still [24]. A study conducted in Algeria revealed that the incorporation of CuO nanoparticles into solar stills resulted in enhanced productivity during the winter season [25]. The conventional stills achieved productivity of 3.5 kg m^{-2} in summer and 2.2 kg m^{-2} in winter. However, the modified distillery achieved a significantly improved outcome, approximately 1.4 times better, during the winter season [25]. Additionally, this modification led to a notable increase in exergy efficiency, with a 52.5% improvement observed during the winter season [25]. Furthermore, a comparative study was conducted utilizing CuO, Al_2O_3, Ag, Fe_2O_3, and SiC-water nanofluids in passive single-slope solar stills. Al_2O_3-water nanofluid generated 14.22% more thermal energy per day than a simple solar still without nanofluid, followed by CuO (10.82%), Ag (8.11%), Fe_2O_3 (7.63%), and SiC (7.61%). The outcomes were compared to experimental and theoretical research [26]. Finally, a research study was conducted to examine the influence of the number of baffles on the efficiency of solar stills [27]. The findings of the study indicate that there was a positive correlation between the number of baffles and the increase in water temperature, with a 26.37% rise observed [27]. Additionally, freshwater productivity experienced a boost ranging from 9.7% to 10.925%. These results suggest that increasing the number of baffles can be considered as an effective approach to enhance the efficiency of solar stills [27]. In general, the aforementioned studies provide evidence that the efficiency, productivity, and cost-effectiveness of solar stills can be improved by integrating new materials and inventive designs. This enhancement is particularly valuable in tackling the issues of water scarcity and energy demand [28]. In summary, solar energy has demonstrated its efficacy and cost-effectiveness as a viable alternative to conventional energy sources. Significant advancements have been achieved in enhancing the efficacy of water purification and desalination procedures through the utilization of solar stills and the integration of nanofluids [29]. However, additional research and development efforts are required to address current limitations and enhance the efficiency of solar stills. By effectively tackling these challenges, solar energy has the potential to further contribute to sustainable solutions for the pressing issues of water scarcity and environmental concerns [30]. Despite the extensive body of research on the utilization of PCMs and nanomaterials to enhance solar distillation processes, there is a notable gap in the existing literature regarding the improvement of performance in single-slope solar stills. Previous research has predominantly concentrated on examining either PCMs or nanomaterials with the aim of augmenting the efficiency of solar stills. However, there has been limited research conducted on the utilization of both combined composite materials in a singular inclined solar energy system. The main aim of this study is to enhance the efficiency of water distillation methods, specifically targeting the pressing need for improved access to potable water in remote regions that heavily rely on solar distillation techniques. In order to accomplish this objective, PCM, specifically paraffin, is utilized along with Al_2O_3 nanoparticles to improve the performance of the single-slope solar still. Furthermore, a comparative analysis is also conducted on standalone solar stills and solar stills integrated with PCM. This distinctive approach effectively assesses both thermal efficiency and distillation productivity, thereby offering valuable insights for potential advancements in this domain.

Experimental setup

The solar still's base was constructed using a galvanized iron (GI) painted sheet, whereas the basin was painted black to increase its ability to absorb solar radiation. The apparatus contained water, and as solar radiation permeated the glass enclosure, the water's temperature gradually



Fig. 1 Installed solar still on the roof of DTU

rose until it reached a predetermined threshold, prompting evaporation to occur within the distillation apparatus. In order to ensure optimal performance, it is imperative that the basin liner possesses impermeability to leak and exhibits a substantial absorption capacity. Figure 1 illustrates the experimental arrangement, wherein the base was thermally insulated using thermocol material and positioned on a stand with a height of 0.75 m. A layer of black paint was put on the basin's surface to improve the absorption of solar energy. The basin possessed a surface area measuring 0.185 square meters and was equipped with insulation in the form of a 17-mm thick layer of thermocol material applied to its base. The glass cover was positioned at an inclination of 30 degrees relative to the horizontal plane and securely sealed using silicon sealant and tape. A V-shaped channel comprising three sides was fabricated, accompanied by a 4-cm drainage channel featuring a nozzle with a diameter of 6 mm for water discharge. The liquid measurement cylindrical jar, with an accuracy of ± 0.1 mL, was connected to the outlet through a pipe to collect the distilled water. A comprehensive examination was conducted on the entire setup to ascertain the absence of leaks and optimize heat dissipation. In order to address concerns regarding fragility, a 4-mm glass cover with an average transmissivity of 0.087 was employed as a substitute for high transmissivity glass. The implementation of adequate insulation played a crucial role in ensuring the optimal functioning of the system. Moreover, thermocol was employed on the lower surface to mitigate heat loss, while the side wall was constructed using a GI sheet that had been painted. To further reduce heat loss and enhance the evaporation process, a sealant was employed. Furthermore, to gather the distilled water, a V-shaped channel was affixed to the GI sheet. The installation of a pyranometer set up by the National Institute of Wind Energy was undertaken to quantify wind energy and solar radiations. The data utilized in this experiment were acquired from a website specifically dedicated to global radiation incidents [31]. Using a digital thermosensor manufactured by ApTechDeals, the temperatures of various components of the solar still, including the basin liner, inner glass, and water vapor, were measured. The instrument's sensitivity was taken into consideration for accurate temperature readings. The thermal sensor exhibited a temperature range spanning from − 50 ℃ to 110 ℃, with an accuracy level of ± 1 °C. Moreover, two sensors were used to measure the properties of the inner and outer glass, as well as four sensors for the sidewalls, one for the basin liner, one for water temperature, and one for water vapor. Finally, the outlet water was gathered utilizing a liquid measurement cylindrical container with a precision of ± 0.1 mL. Table 1 shows the geometric characteristics of the setup.

Materials selection

The basin of the solar still employed commercial-grade paraffin (CAS NO.: 8002–74 − 2) as the PCM. To separate the paraffin from the water, a 0.5-mm-thick aluminum sheet was utilized. Table 2 shows the thermophysical properties of paraffin wax.

Besides this aluminum oxide nanoparticle (Alpha), CAS: 1334–28 − 1 was used as nanomaterials in this experiment.

Table 1 Geometric properties of the setup

Parameters	Values
Basin area/m^2	0.185
Glass thickness/mm	4
Number of glasses	1
Slope of glass/ °	30
Effective absorption area/m^2	0.165
Condensation aera/m^2	0.207

Table 2 Thermophysical properties of paraffin wax [17]

Parameters	Values
Melting temperature/°C	58–60
Latent heat of fusion/kJ kg^{-1}	226
Thermal conductivity/W m^{-1} K^{-1}	0.24
Density solid/liquid/Kg m^{-3}	818
Specific heat solid/liquid/kJ kg^{-1} K^{-1}	2.14
Application	Solar still

Experimental procedure

In June 2022, the experimental setup was installed on the rooftop of DTU Delhi, India, to obtain precise and reliable results. The experiment was carried out under clear weather conditions, with the solar still being oriented north–south to measure its entire output accurately. Thermosensors took hourly temperature measurements for the basin liner, cover glass, and water vapor. The estimation of distillate yields on an hourly and daily basis was conducted by evaluating liquid containers. The analysis was conducted between 9 AM and 8 PM. The experiment was conducted in three distinct stages, with each stage focusing on examining different cases. The solar still underwent testing from June 10 to June 13 in the initial instance. The solar still with PCM was tested from June 21 to June 25 in the second case. Finally, the third case study entailed the utilization of a solar still incorporating PCM-Al_2O_3 from June 26 to June 29. Figure 2 displays the hourly readings that were obtained. Water was sprayed after 5 PM. The aluminum oxide (Al_2O_3) was uniformly distributed onto the PCM with a mass percentage of 2% aluminum oxide. The experimental procedure initially subjected the wax to heat until it reached its melting point. Subsequently, an approximate quantity of 5 g of aluminum particles was evenly dispersed within the container, which was then subjected to agitation for 60 min. Following the cessation of the heat supply, an approximate quantity of 5 g of aluminum oxide was dispersed onto the molten wax, resulting in its solidification. Subsequently, an additional quantity of 10 g was evenly distributed onto the upper surface of the solidified wax. The required amount of nanoparticles can be calculated below the equation:

$$\varnothing = \frac{m_{\text{n-Al2O3}}}{m_{\text{PCM}} + m_{\text{n-Al2O3}}} \tag{1}$$

where $\varnothing$ represents the mass fraction of the nanoparticle, m_{PCM} denotes the mass of the PCM, and $m_{\text{n_}}Al_2O_3$ indicates the mass of the nanoparticle that has been introduced to the PCM. The consistent and enduring diffusion of nanoparticles into paraffin wax poses significant challenges in preserving the thermophysical characteristics of the mixture. Consequently, even a small dispersion of nanoparticles can significantly impact the properties of the mixture. This approach entails preserving consistency within solutions containing wax and nanoparticles.

The hourly efficiency calculated by the given formula

$$\eta_h = \frac{p \bullet h_{\text{fs}}}{I \bullet A} \tag{2}$$

where, A is the area of solar still (m^2), I is the intensity of solar radiation (W m^{-2}), P is the productivity (m^{-3}), and h_{fs} is the latent heat of the distiller water.

Result and discussion

The experimental observation was carried out in Delhi (Latitude: 28.7°N, 77.10°E) from June 10, 2022, to June 13, 2022, without PCM. In the second case, an experiment was conducted with a layer of PCM on the basin bed from June 21 to June 25, 2022. In the third case, another experiment was conducted using PCM–Al_2O_3. An appropriate instrument was used to measure all the temperatures. Figure 2 displays the global solar radiation variation for the solar still alone. The values for June 10 and June 11 range from 700 to 900 W m^{-2} during the 9-AM to 2-PM interval. Figure 3

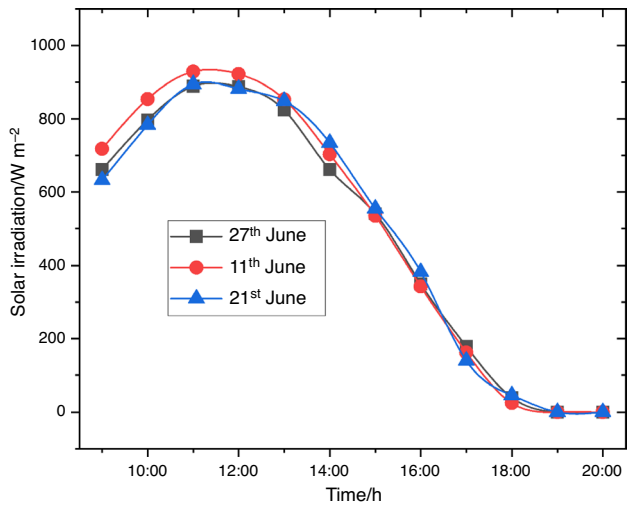


Fig. 2 Solar radiation for a different day (solar still alone)

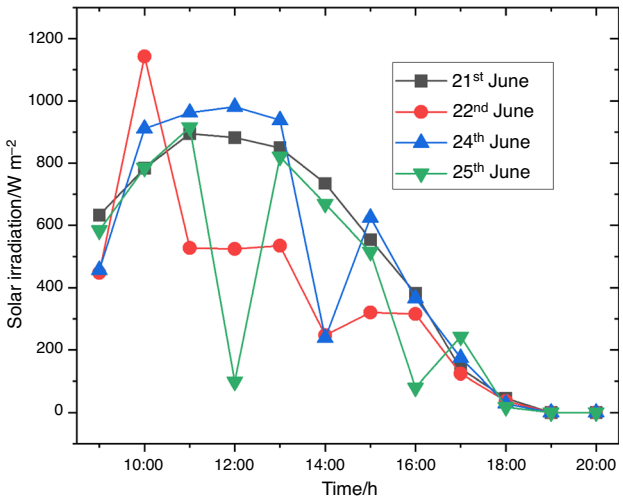


Fig. 3 Solar radiation for a different day (solar still with PCM)

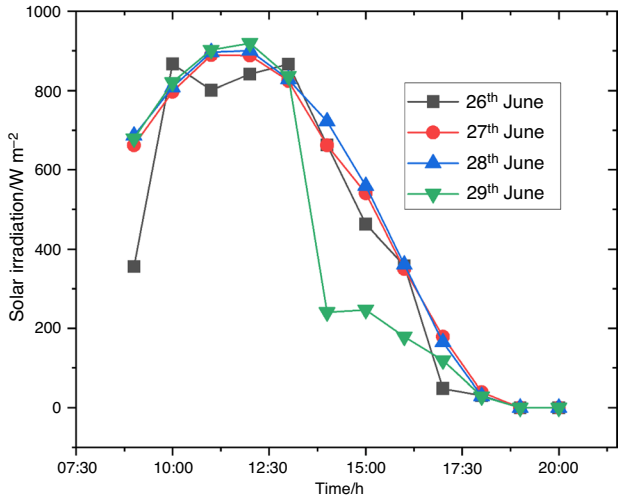


Fig. 4 Variation of solar radiation for a particular date (solar-PCM–Al_2O_3)

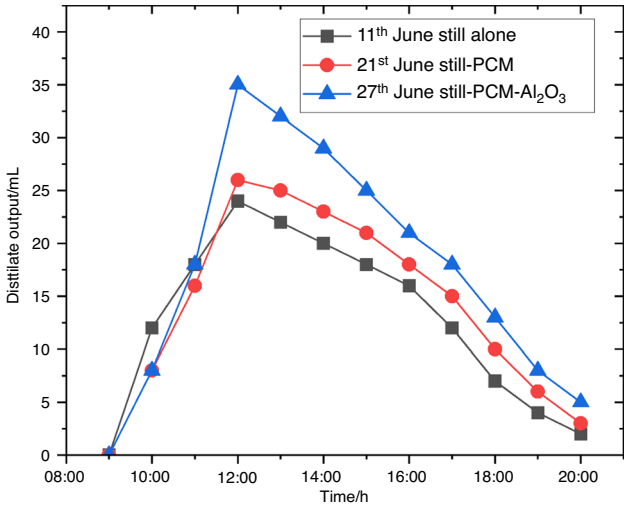


Fig. 6 Variation of distilled water output for different case

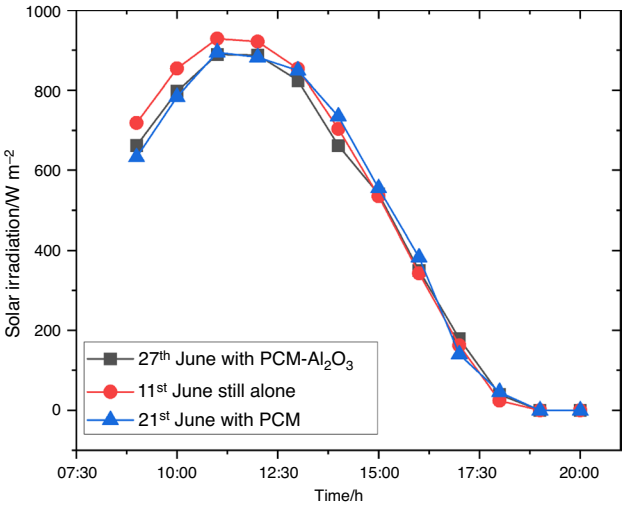


Fig. 5 Solar radiation was observed for solar sill alone, with PCM, with PCM–Al_2O_3

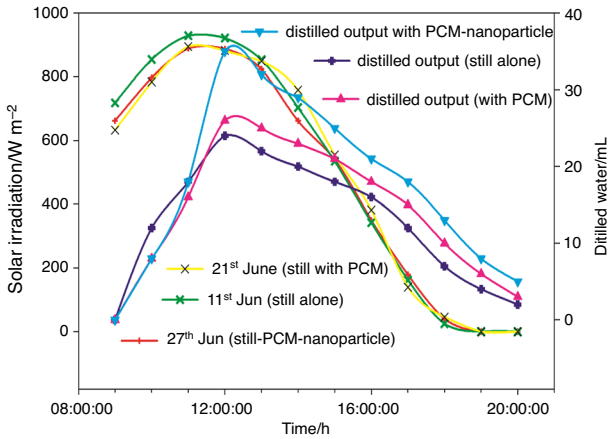


Fig. 7 Variation of distilled water output with respect to solar radiation

shows the global solar radiation variation for the solar still with PCM. The values for June 21 and June 24 range from 550 to 750 W m^{-2} in the 9-AM to 2-PM interval. Figure 4 provides an overview of the global solar radiation variation for the solar still with PCM–Al_2O_3. The values for June 27 and June 29 range from 650 to 900 W m^{-2} within the 9-AM to 2-PM interval.

Figure 5 displays the estimated solar irradiation observed for a solar still without PCM on June 11, a solar still with PCM on June 21, and a solar still with PCM–Al_2O_3 on June 27. The solar radiation during this period was determined to be approximately the same. Figure 5 shows the radiation variation in different scenarios: which is still alone, one with PCM, and one with PCM–Al_2O_3. The figure demonstrates that the solar radiation is nearly identical for the specific date used in this analysis. This radiation data was utilized to compare various parameters of the solar still.

Figure 6 displays the distilled output data for three different configurations: solar still alone, solar still with PCM, and solar still-PCM–Al_2O_3. The findings suggest that productivity is enhanced during off-peak periods, specifically in the context of solar still- PCM–Al_2O_3. The results demonstrate an overall increase in productivity during both on and off-peak times. Figure 7 illustrates the fluctuations in distilled water in response to solar radiation across various scenarios, namely the use of a standalone still, a still incorporating PCM, and a still incorporating PCM–Al_2O_3. The graph indicates that the distilled water output of the still with PCM–Al_2O_3 is greater than that of both the still

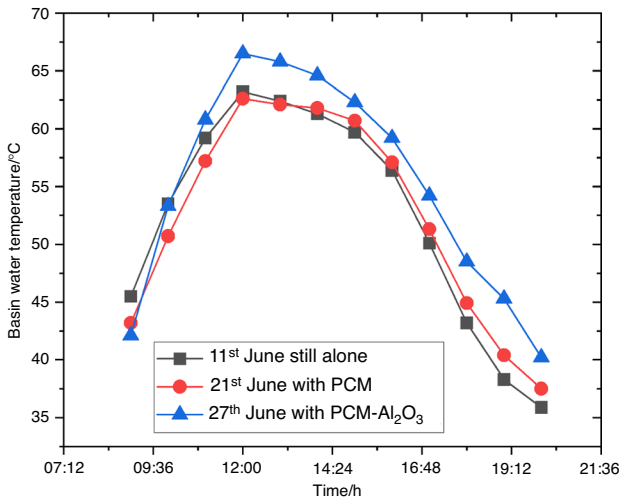


Fig. 8 Variation of basin water temperature with time

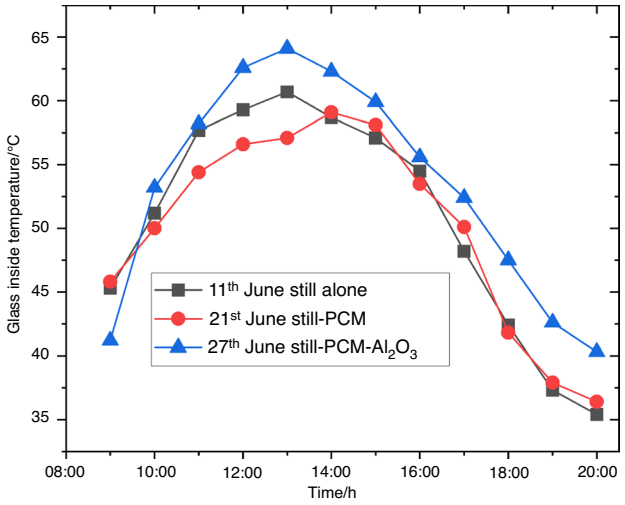


Fig. 10 Variation of glass inside temperature with time for three conditions

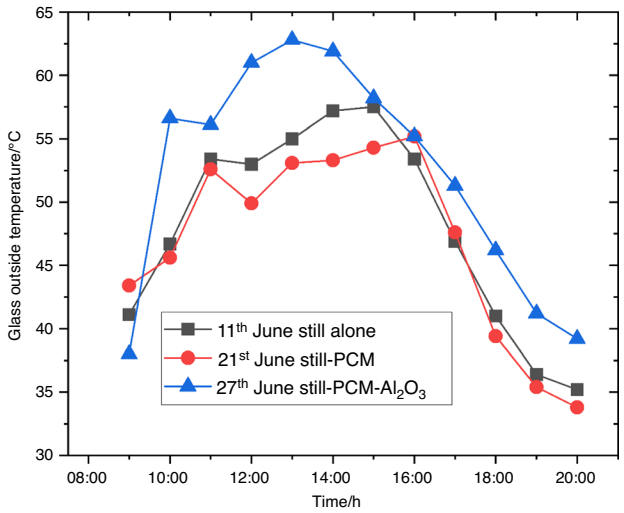


Fig. 9 Variation of glass outside temperature with respect to time for all three cases

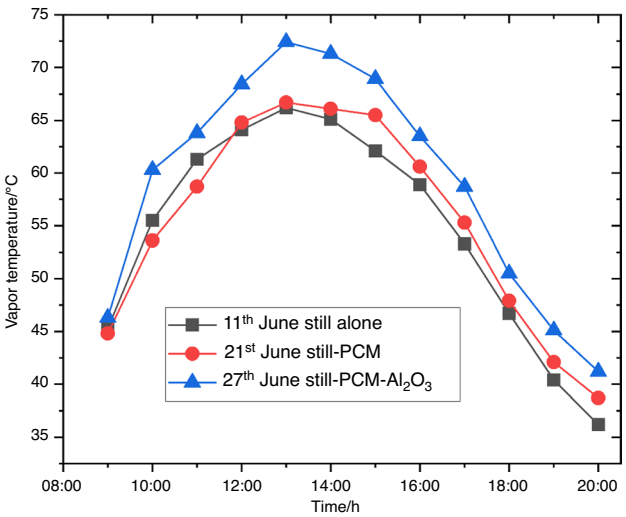


Fig. 11 Variation of water vapor of still for different cases with time

alone and the still-PCM. Figure 8 illustrates the fluctuations in basin water temperature for three different scenarios: solar still alone, solar still with PCM, and solar still with PCM–Al_2O_3. The graph illustrates that the rate of increase in basin water temperature is significantly higher in solar stills equipped with PCM and solar stills incorporating both PCM and Al_2O_3, compared to solar stills operating without these enhancements.

The comparison between the hourly solar still performance with the incorporation of PCM and PCM–Al_2O_3, and the performance of the solar still alone, is shown in Fig. 9. The solar still's yield was measured at 837 mL m^{-2} day^{-1}, 924 mL m^{-2} day^{-1}, and 1145 mL m^{-2} day^{-1} for the solar still alone, solar still with PCM, and solar still with PCM and Al_2O_3, respectively. The increase in yield was approximately 10.38% and 36.77%, respectively. The solar still yielded the highest quantity of distilled water at noon due to the prevailing hot weather conditions. The nocturnal condensed product of the distillation process resulted in an augmentation of solar energy production and an enhancement of the overall distillation output. At 5 PM, water was sprayed onto the glass in order to sustain the temperature differential and optimize the distillation process. Figure 9 illustrates the fluctuations in the external temperature of the glass for three scenarios: the solar still alone, the solar still with PCM, and the solar still with PCM and Al_2O_3. It is evident from the graph that the outside glass temperature of the

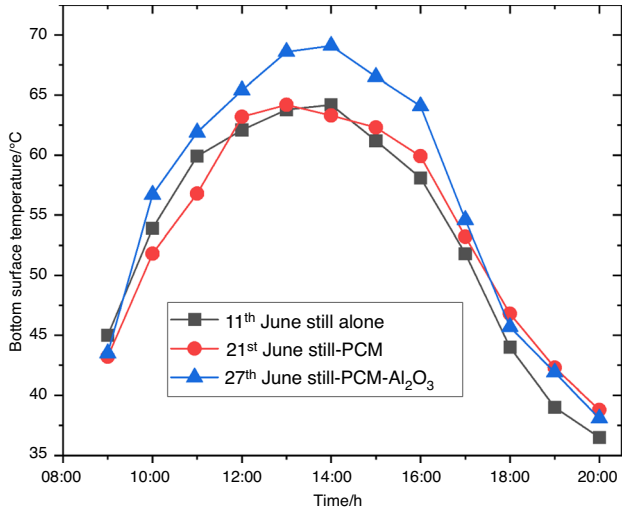


Fig. 12 Variation of still bottom surface temperature with hourly period

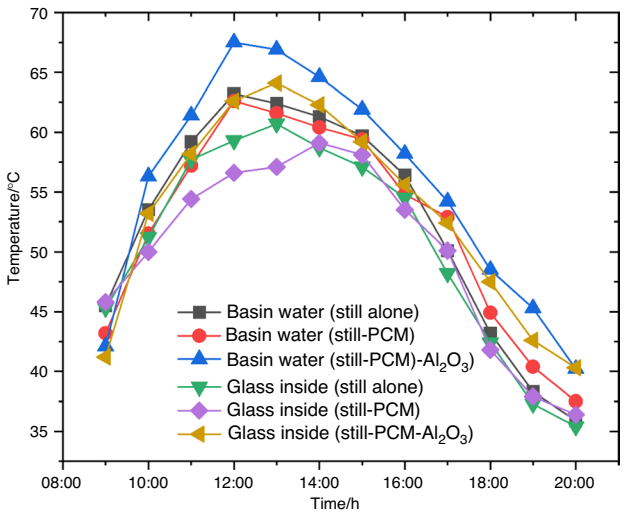


Fig. 13 Variation of the inside of glass and water basin temperature with time for three cases

solar still with Al_2O_3 is higher compared to both the solar still alone and the solar still with PCM.

Figure 10 shows the variation of the inside temperature of the glass concerning the hourly time for different cases: solar still alone, still-PCM, and still- PCM–Al_2O_3. The results indicate that the temperature inside the glass for the solar still-PCM and still- PCM–Al_2O_3 cases was slightly higher than the other two cases. Figure 11 displays the hourly measurement of water vapor for various cases. The graph indicates that the variation of water vapor ranges from approximately 30 ℃ to 36 ℃ across different cases.

Figure 12 depicts the hourly measurements of the temperature variation on the bottom surface of the still basin for three different configurations: solar still alone, solar still with PCM, and solar still with PCM and Al_2O_3. The graph illustrates the range of temperature variations observed on the lower surface, which spans from 45 ℃ to 68 ℃. It was noted after careful observation that the basin's temperature experienced a slight decrease. This can be attributed to the temperature reaching the melting point of the PCM, which is within the range of 58 ℃ to 60 ℃. Between the hours of 12:00 PM and 3:00 PM, the behavior of paraffin wax was characterized by undercharging conditions. Upon completion of the charging process, the glass is covered using a protective cover, and the PCM discharge is underway. Subsequently, the stored energy is transferred to the activated solar still during nighttime. A marginal decrease in yield was observed during the day when comparing the use of still and non-still PCM that is stored below the basin.

Figure 13 illustrates the relationship between the water temperature in the basin and the temperature of the glass placed inside. This suggests that it is important to maintain a distinction to achieve high water distillation productivity for three different systems: solar still alone, solar still with PCM and solar still with PCM and Al_2O_3. The process of distillation conducted by the simple apparatus yields a greater volume of purified water during the time period spanning from 8 to 11 AM. However, the production of other distillates is expected to increase after this hour. This increase in production results in a portion of the heat being transferred to the wax, causing the temperature of the wax to rise gradually. Consequently, the temperature of the water decreases due to the high capacity of the wax to absorb heat. According to the findings, the temperature of the bottom surface increased to a greater extent compared to the temperatures of the PCM and nano-PCM during the charging phase. This observation aligns with the well-established understanding that metals have a higher heat absorption rate than other materials. During the period of discharge, there was a rapid decrease in temperature observed on the bottom surface in comparison to the paraffin wax. Although the majority of the temperature data obtained from the plates and subsequently transferred to the wax is typically stored as sensible heat and latent heat. Incorporating a solar still involves the utilization of Al_2O_3–PCM to achieve a slightly higher temperature than the temperature observed when using PCM alone. The incorporation of Al_2O_3 in the PCM resulted in an enhancement of thermal conductance. This improvement allows the PCM to efficiently accumulate more heat energy within a shorter duration, making it suitable for integration into solar stills. Therefore, using Al_2O_3–PCM in solar stills results in a higher distillation efficiency compared to solar stills without PCM. Additionally, PCM's presence impacts the temperature of the water inside the solar still.

Comparison with the previous study

The findings of this study build upon the work of Ramzy et al. [32], who investigated the performance of solar stills using various absorbing materials, including steel wool pads. Ramzy et al. reported a thermal efficiency of 32.74% and a yield of 4.384 l m^{-2} using steel wool pads as the absorbing material. In contrast, the current research improved solar still performance by integrating PCM with Al_2O_3 nanoparticles. This method significantly increased thermal efficiency to 36.77% and a maximum water yield of 1.245 kg using the PCM–Al_2O_3 composite. Therefore, the findings of this study suggest that incorporating nanoparticles into PCM can potentially enhance the efficiency of solar distillation systems.

Conclusions

This study investigates solar distillation systems, focusing on the integration of PCM and PCM–Al_2O_3. The results show that adding PCM and Al_2O_3 nanoparticles significantly improved solar still performance. Adding Al_2O_3 significantly enhanced the thermal performance of paraffin wax, boosting distilled water yield by 10.38%. The daily distillate yield increased by 36.77% when Al_2O_3 nanoparticles were included in the PCM design. The PCM–Al_2O_3 composite also showed lower charging and discharging periods, indicating improved thermal storage. Maintaining an ideal water spray rate on the glass surface also increased the output rate, allowing for a more favorable temperature difference between the water and glass surface, boosting the condensation rate. Future studies will explore potential yield increases by swapping Al_2O_3 for other materials like titanium oxide and copper oxide. The urgent requirement to enhance the efficiency of solar distillation facilities, which hold the potential to provide uncontaminated potable water, underscores the significance of ongoing investigations into novel materials for solar-driven water purification systems.

Data availability No data associated with this study.

References

1. Akram N, Sadri R, Kazi SN, Nashrul M, Zubir M, Ridha R, Ahmed W, Soudagar MEM, Arzpeyma M. A comprehensive review on nanofluid operated solar flat plate collectors. J Therm Anal Calorim. 2020;139:1309–43.
2. Pathak AK, Tyagi VV, Anand S, Pandey AK, Kothari R. Advancement in solar still integration with phase change materials-based TES systems and nanofluid for water and wastewater treatment applications. J Therm Anal Calorim. 2022;147:9181–227.
3. Khan Y, Raman R, Rashidi MM, Caliskan H, Chauhan MK, Chauhan AK. Thermodynamic analysis and experimental investigation of the water-spray cooling of photovoltaic solar panels. J Therm Anal Calorim. 2023;148:5591–602.
4. Omara AAM, Abuelnuor AAA, Mohammed HA, Khiadani M. Phase change materials (PCMs) for improving solar still productivity: a review. J Therm Anal Calorim. 2020;139:1585–617.
5. Gong B, Yang H, Wu S, Tian Y, Yan J, Cen K, Bo Z, Ostrikov KK. Phase change material enhanced sustained and energy-efficient solar-thermal water desalination. Appl Energy. 2021;301:117463.
6. Liu Y, Kang M, Lin W, Liang C, Qu W, Wang Y, Guan Y, Cheng J. Enhanced thermal performance of phase change material microcapsules using halloysite nanotubes. Appl Therm Eng. 2023;231:120965.
7. Khan Y, Raman RR, Rashidi MM, Said Z, Caliskan H, Hoang AT. Thermodynamic and exergoenvironmental assessments of solar-assisted combined power cycle using eco-friendly fluids. J Therm Anal Calorim. 2024;149:1125–39.
8. Vaithilingam S, Gopal ST, Srinivasan SK, Manokar AM, Sathyamurthy R, Esakkimuthu GS, Kumar R, Sharifpur M. An extensive review on thermodynamic aspect based solar desalination techniques. J Therm Anal Calorim. 2021;145:1103–19.
9. Seyednezhad M, Sheikholeslami M, Ali JA, Shafee A, Nguyen TK. Nanoparticles for water desalination in solar heat exchanger. J Therm Anal Calorim. 2020;139:1619–36.
10. Okonkwo EC, Wole-Osho I, Almanassra IW, Abdullatif YM, Al-Ansari T. An updated review of nanofluids in various heat transfer devices. J Therm Anal Calorim. 2021;145:2817–72.
11. Kabeel AE, Abdelgaied M. Observational study of modified solar still coupled with oil serpentine loop from cylindrical parabolic concentrator and phase changing material under basin. Sol Energy. 2017;144:71–8.
12. Chaichan MT, Kazem HA. Single slope solar distillator productivity improvement using phase change material and Al2O3 nanoparticle. Sol Energy. 2018;164:370–81.
13. Rashidi S, Karimi N, Mahian O. A concise review on the role of nanoparticles upon the productivity of solar desalination systems. J Therm Anal Calorim. 2019;135:1351145–59.
14. Jathar LD, Ganesan S, Shahapurkar K, Soudagar MEM, Mujtaba MA, Anqi AE, Farooq M, Khidmatgar A, Goodarzi M, Safaei MR. Effect of various factors and diverse approaches to enhance the performance of solar stills: a comprehensive review. J Therm Anal Calorim. 2022;147:4491–522.
15. Dharamveer, Samsher, Singh DB, Singh A, Kumar N. (2019) Solar Distiller Unit Loaded with Nanofluid—A Short Review. Lecture Notes in Mechanical Engineering. https://doi.org/10.1007/978-981-13-6577-5_24
16. Elgendi MA. Review of the Applications of Nanomaterials to Augment Solar Still Productivity. Water Resources Management and Sustainability: Water Science and Technology Library; 2023. https://doi.org/10.1007/978-3-031-24506-0_30.
17. Sathish D, Jegadheeswaran S. Evolution and novel accomplishments of solar pond, desalination and pond coupled to desalination systems: a review. J Therm Anal Calorim. 2021;146:1923–69.
18. Sakthivel TG, Arjunan TV. Thermodynamic performance comparison of single slope solar stills with and without cotton cloth energy storage medium. J Therm Anal Calorim. 2021;137:351–60.
19. Sharshir SW, Salman M, El-Behery SM, Halim MA, Abdelaziz GB. Enhancement of solar still performance via wet wick, different aspect ratios, cover cooling, and reflectors. Int J Energy Environ Eng. 2021;12:517–30.
20. Pathak AK, Chopra K, Tyagi VV, Anand S, Kothari A, Sari A, Pandey AK. Solar heat pipe ETC integrated with solar still system for water treatment and hot water production: novel hybrid experimental approach. J Therm Anal Calorim. 2023;148:8969–89.

21. Madiouli J, Saleel CA, Lashin A, Badruddin IA, Kessentini A. An experimental analysis of single slope solar still integrated with parabolic trough collector and packed layer of glass balls. J Therm Anal Calorim. 2021;146:2655–65.
22. Modi KV, Jani HK, Gamit ID. Impact of orientation and water depth on productivity of single-basin dual-slope solar still with Al2O3 and CuO nanoparticles. J Therm Anal Calorim. 2021;143:899–913.
23. Faridani HZ, Ameri A. Performance enhancement of a basin solar still using γ-Al2O3 nanoparticles and a mixer: an experimental approach. J Therm Anal Calorim. 2022;147:1919–31.
24. Suresh C, Shanmugan S. Effect of water flow in a solar still using novel materials. J Therm Anal Calorim. 2019. https://doi.org/10.1007/s10973-019-08449-5.
25. Attia MEH, Karthick A, Manokar AM, Driss Z, Kabeel AE, Sathyamurthy R, Sharifpur M. Sustainable potable water production from conventional solar still during the winter season at algerian dry areas: energy and exergy analysis. J Therm Anal Calorim. 2021;145:1215–25.
26. Dhindsa GS, Kumar V, Mittal MK, Sokhal GS, Khalilpoor N, Sharifpur M, Issakhov A, Tonk R. Performance comparison of single-slope solar still loaded with various nanofluids. Energy Sci Eng. 2022;10:4318–31.
27. Al-Adel Z, Bouabidi A, Chrigui M. 3D CFD simulation and experimental validation of the baffle number effect on the solar still performance. J Therm Anal Calorim. 2023;148:2171–88.
28. Mohammed AH, Shmroukh AN, Ghazaly NM, Kabeel AE. Active solar still with solar concentrating systems. Review J Therm Anal Calorim. 2023;148:8777–92.
29. Kumar TRS, Jegadheeswaran S, Chandramohan P. Performance investigation on fin type solar still with paraffin wax as energy storage media. J Therm Anal Calorim. 2019;136:101–12.
30. Shoeibi S, Rahbar N, Esfahlani A, Kargarsharifabad H. Energy matrices, economic and environmental analysis of thermoelectric solar desalination using cooling fan. J Therm Anal Calorim. 2022;147:9645–60.
31. https://nsrdb.nrel.gov/ accessed on 12/07/2022.
32. Ramzy K, Abdelgaleel M, Kabeel AE, Mosalam H. Performance of a single slope solar still using different porousabsorbing materials: an experimental approach. Environ Sci Pollut Res. 2023;30:72398–414.

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.

Macromolecular Research
https://doi.org/10.1007/s13233-024-00300-7

Online ISSN 2092-7673
Print ISSN 1598-5032

ARTICLE

Injectable mesquite gum and carboxymethyl chitosan hydrogel using schiff base crosslinks: a versatile platform for drug delivery in wound care

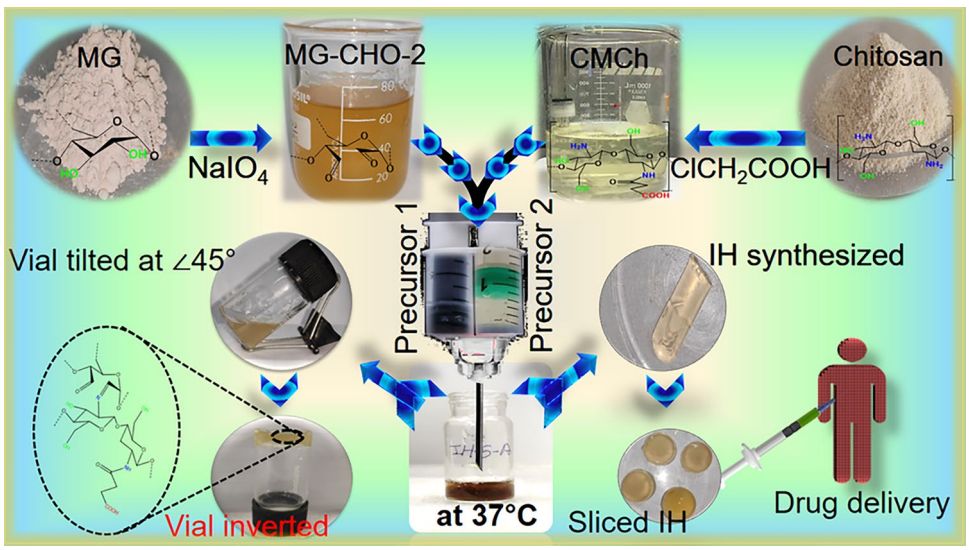
Jitendra Kumar[1] · **Roli Purwar**[1]

Received: 22 May 2024 / Revised: 10 July 2024 / Accepted: 15 July 2024
© The Author(s), under exclusive licence to The Polymer Society of Korea 2024

Abstract

The novel injectable hydrogel (IH) is developed at 37 ℃ using the dynamic imine bond between mesquite gum with multi-aldehyde groups (MG-CHO) and carboxymethyl chitosan (CMCh). The investigation consists of determining the ideal concentration of an oxidizing agent to maximize the amount of aldehydes in mesquite gum. Then, the oxidized mesquite gum with the optimized aldehyde content (47.6%) determines the minimum gelation time (7 to 2 min.). Structural characterization is conducted through Fourier transform infrared spectroscopy (FT-IR) and proton nuclear magnetic resonance spectroscopy (^{1}H-NMR). The scanning electron microscopy (pore size = 14 to 34 µm) and rheometer examine surface morphology and rheological properties. The swelling ratio in phosphate buffer saline (PBS) at varying pH levels (5.5, 7.4, and 8.5) is measured for both dry and gel forms, revealing a decrease in the swelling ratio with increasing pH (5.5 to 7.4), followed by an increase at pH 8.5. Ciprofloxacin HCl is employed as a model drug for release experiments, and drug release behavior is compared in PBS at pH 5.5, 7.4, and 8.5, using the Korsemeyer–Peppas model to determine the release mechanism. Biocompatibility of injectable hydrogels is assessed regarding in vitro cytotoxicity using L-929 fibroblast cell lines and hemolysis assay. Additionally, the antibacterial study is analyzed using gram-positive and gram-negative bacteria. Furthermore, the hydrolytic biodegradability of IHs in phosphate buffer saline at pH 7.4 is evaluated.

Graphical abstract



Synthesis of self-healing, biocompatible, and biodegradable injectable hydrogel using mesquite gum and carboxymethyl chitosan crosslinked through Schiff base mechanism utilized for drug delivery application

Extended author information available on the last page of the article

Published online: 09 August 2024

Keywords Carboxymethyl Chitosan · Mesquite gum · Injectable Hydrogel · Self-healing · Schiff base · Drug delivery

1 Introduction

In recent times, the field of drug delivery has made significant progress due to the search for more patient-friendly, safe, and effective treatment approaches. Hydrogels are a cutting-edge material that is making a big splash in this area [1]. They can precisely control the release kinetics of a wide range of therapeutic agents, including proteins, peptides, and small molecules [2]. Specifically, injectable hydrogels (IHs) provide a less intrusive method of localized drug administration, with benefits such as decreased systemic adverse effects and increased patient compliance [3]. Different ways can be utilized to synthesize injectable hydrogels, which can be classified as physical or chemical. Physical hydrogels form spontaneously due to modest secondary forces, responding to temperature, pH, or ionic concentration changes. Chemical crosslinked methods include enzymatic, Schiff base cross-linking, Michael addition, click chemistry, and photo-crosslinking [4]. A primary amine and an aldehyde react to form Schiff base crosslinking, which has numerous benefits for injectable hydrogel design: mild reaction conditions, controlled crosslinking kinetics, and the possibility of customizable mechanical and drug release features. The resultant Schiff base-based injectable hydrogels are a desirable alternative for drug delivery because they show strong stability and the capacity to change dynamically in response to external stimuli [5]. Natural polymers—particularly polysaccharides, have become more prevalent in hydrogel formulation because of their biocompatibility, biodegradability, and innate capacity to interact with biological systems. Examples of these include guar gum [6], chitosan [7], and gum arabic [8].

In recent decades, chitosan, a polysaccharide, has drawn more and more attention. It is produced by deacetylation of chitin from marine trash, the second most prevalent material after cellulose [9]. Because of the amino groups on its chains, chitosan is much more reactive than the original polysaccharide, making it a desirable biopolymer [10]. It is soluble in an acidic medium, which is a severe concern for use in the medicinal field [11]. Therefore, carboxymethyl chitosan is a significant chitosan derivative with high solubility in physiological fluid and keeps its characteristics [12] [13] [14].

Mesquite gum (MG), also called "vilayati babool," derived from the mesquite tree (*Prosopis juliflora*), primarily consists of polysaccharides, including galactomannans. These galactomannans are composed of a mannose backbone with galactose side units. The ratio of mannose to galactose and the branching patterns in the structure can vary among different species and sources of mesquite gum [15]. Mesquite gum's biocompatibility makes it a potentially safe choice for biomedical applications [16]. Ana M. Pinilla-Torres et al. published a research article based on mesquite gum for its potential in the formulation of microspheres or nanoparticles to encapsulate drugs for controlled release and targeted delivery [17]. Cristobal Rodriguez et al. synthesized nanofibers based on mesquite gum combined with pollulan and chitosan and utilized them in wound healing applications [18]. Rogelio Rodriguez-Rodriguez et al. published a research work in which they synthesized polyelectrolyte complexes based on mesquite gum and chitosan, which can be considered for future studies in foods and pharmaceutical applications [19].

Various advantages of the natural gum-based injectable hydrogel are present, such as biocompatibility, biodegradability, non-toxicity, high water absorption capacity, tunable mechanical properties, ease of injection, potential for drug delivery, and promotes cell adhesion and proliferation. Besides this, some limitations are also inherited, such as low mechanical properties, rapid degradation upon contact with body fluids, and initial burst release, which often causes harmful side effects and ineffective treatment of diseases [20]. However, the present study overcome these limitations.

Our hypothesis centers on the potential of multi-aldehyde mesquite gum to engage in dynamic Schiff base chemistry with carboxymethyl chitosan, resulting in an injectable hydrogel formation at 37 °C. The IH is anticipated to possess favorable syringeability, injectability, and self-healing capabilities and be independent of additional cross-linkers. We examined rheological properties, surface morphology, swelling behavior, and drug release kinetics comprehensively. Additionally, we explored biocompatibility assays regarding in vitro cytotoxicity, hemolysis assays, and hydrolytic degradation to ascertain the safety profile.

A literature review revealed no mesquite gum-based injectable hydrogel synthesized utilizing Schiff base chemistry. Thus, the author hypothesized that multi-aldehyde mesquite gum could crosslink with carboxymethyl chitosan using dynamic Schiff base chemistry, resulting in an injectable hydrogel with excellent syringeability, injectability, and self-healing ability at 37 °C without additional crosslinker. Adding crosslinkers adds a step to the synthesis process, but this research article does not require them. Thus, this study is cheaper and safer. Additionally, researchers studied the swelling and drug release of injectable hydrogels using a dry form. This research article compares injectable hydrogels'

dry and gel forms, examining swelling, drug release, and hydrolytic degradation. The novelty of this research article is the usage of mesquite gum biomaterial, lack of hazardous crosslinker, and comparison of dry form with gel form injectable hydrogels.

2 Materials and methods

2.1 Materials

Chitosan was bought from Otto Chemical Pvt. Ltd. (C 1906), which has a deacetylation degree of 85%, $M_v = 2,18,112$ Da (supplementary data 1.2), pH 7–9, and viscosity (1% w/v, 200 cps), while mesquite gum (product code-231545, $M_v = 2$, 71,034 Da, supplementary data 1.1) was obtained from Indianjadibooti Pvt. Ltd. Tamilnadu, India. Sodium hydroxide (030167), sodium periodate (645645) ($NaIO_4$), and hydroxylamine hydrochloride (565325) were purchased from CDH chemicals. Ethylene glycol (71,883), isopropyl alcohol (67–63-0), and monochloroacetic acid (79–11-8) were obtained from Thomas Baker Chemicals Pvt. Ltd. Ethanol (57,101) was purchased from SRL Chemicals, India. The antibiotic ciprofloxacin HCl was given as a gift from M/S Unicure India Limited. Double-distilled water (MQ) was employed for all reactions. Chemicals and solvents were employed as received.

2.2 Oxidation of mesquite gum

The oxidation of mesquite gum and determination of the aldehyde percentage of oxidized mesquite gum were analyzed according to the protocol described by Mohamad Sadegh Aghajanzadeh et al. methodology [21]. A detailed description of the procedure is given in supplementary data 1.3.

2.3 Derivatization of chitosan

Carboxymethylation of chitosan was carried out according to the procedure reported by Feng and coworkers [12]. The degree of carboxymethylation was calculated using back titration (supplementary data 1.4) [22].

2.4 Preparation of injectable hydrogel

For a whole day at 700 RPM and 27 °C, the known amount of oxidized MG gum was dissolved in MQ water (precursor 1). CMCh (5% w/v) was dissolved using a similar process in phosphate buffer saline at pH 7.4, 700 RPM, and 27 °C (precursor (2). Next, at physiological temperature (37 °C), equal volumes of both precursors were concurrently extruded into two distinct syringes of 21 gauzes, each of which carried 5 mL of precursors. Gel formation began immediately. A total of 6 IH samples were synthesized and abbreviated as IH-MG-a, where "a" denotes the concentration of oxidized MG using 2% w/v of $NaIO_4$ (Table S1).

Self-healing behavior, syringeability, injectability evaluation, and network parameters of the IHs are described in supplementary data 1.5 and 1.6.

2.5 Gelation time

It was determined by two approaches.

Method 1: Tube inversion test: This qualitative assessment involves simultaneously extruding precursors 1 and 2 in a glass vial with a 10 mL capacity at 37 °C in a 1:1 ratio, although at different concentrations. To assess the physical state of the IH at 37 °C, the vial was tilted at ∠ 45 and then completely inverted. The gel flow was visually inspected. IH was unable to flow gravimetrically, which helped to support the gelation. The exact gelation time of the synthesized IHs was ascertained by performing tube inversion tests five times (Fig. S1).

Method 2: Using a rheometer—Precursors 1 and 2 were concurrently extruded at the rheometer platform at 37 °C in a 1:1 ratio and allowed to get crosslinked on the platform itself. This was carried out as part of a time sweep research.

2.6 Gel content

The approach of Shaohua et al. [23] was applied to the gel content. To put it briefly, IH was submerged in 100 mL of MQ water at 37 °C for 48 h (A_1). After extracting the water-insoluble injectable hydrogel, it was dried in a hot air oven at 50 °C until it reached a constant weight (A_2). The percentage of gel content (G %) was calculated using Eq. 1. The trials were conducted in triplicate, and the mean value of the three trials was reported.

$$G\,\% = \frac{A_2}{A_1} \times 100 \tag{1}$$

2.7 Porosity

With some changes, the liquid displacement method was employed to calculate the injectable hydrogel's porosity [21]. In summary, a vial containing ethanol was filled with a known quantity of oven-dried IHs (W_2) having volume "*v*" and left for 24 h. Parafilm was used to completely enclose the vial to prevent the ethanol from evaporating and spilling. After a day, the samples were taken from the vial and weighed (W_1). Equation 2 was utilized to ascertain the injectable hydrogel's porosity (*P* %).

$$P\% = \frac{W_1 \times v}{W_2} \times 100 \quad (2)$$

2.8 Rheological properties of synthesized injectable hydrogel

At 37 °C, a parallel plate-equipped MCR-101 rheometer (Anton Paar, Japan) was used to examine various rheological characteristics. A 1 mm space was maintained between the parallel plate geometry and the plate throughout each session. Before testing, the hydrogels were incubated for one hour at 37 °C. A time sweep study in isothermal mode was used to investigate time-dependent behavior. A flow sweep experiment was conducted within the 0.1–100 s^{-1} range to assess the efficacy of shear-thinning and viscosity adjustments. An amplitude sweep test was used to ascertain the hydrogel's linear viscoelastic range (LVE). Next, to verify the hydrogel's capacity for self-healing, they were exposed to a range of strain amplitudes (high 200% and low 1%). After that, it was observed that the storage modulus values changed. Furthermore, these hydrogels underwent frequency sweep tests covering the 0.1–100 rad s^{-1} frequency range at 1% amplitude.

2.9 Swelling ratio

Two methods were used to measure the swelling ratio from IH. The swelling ratio was measured in PBS at 37 °C and pH 5.5, 7.4, and 8.5.

Method 1: To achieve total crosslinking, synthesized IHs were first incubated for 1 h at 37 °C. It was then stored in a hot air oven at 50 °C to dry out entirely. To observe the swelling behavior, button-form IHs were placed in 100 mL of PBS at pH 5.5, pH 7.4, and pH 8.5 at 37 °C. IHs were taken out of their beakers at fixed intervals, the surface water was wiped off using filter paper, and the gels were weighed.

Method 2: Freshly synthesized IHs gel form is immediately added to 100 mL of PBS solution at pH 5.5, pH 7.4, and pH 8.5 at 37 °C.

The swelling ratio (*S* %) was determined by the amount of water absorbed by the IHs using Eq. 3.

$$S\% = \left[\frac{Ws - Wd}{Wd}\right] \times 100 \quad (3)$$

where,

Wd—Dried weight of the injectable hydrogel.

Ws – Swollen weight of the injectable hydrogel.

The network parameters of IHs were calculated and described in supplementary data 1.7.

2.10 Drug loading and release kinetics

Ciprofloxacin HCl, an antibiotic, is used as a model drug. The calibration curve for the drug was made using a UV–visible spectrophotometer. To make precursor 3, MG-CHO-2 solutions were combined with a 3 mL aqueous solution of ciprofloxacin HCl (6 mg mL^{-1}). To prepare the drug-loaded IHs, precursors 2 and 3 were mixed in equal parts. There are two different methods for measuring the amount of drug released by IHs.

Method 1: The synthesized IHs were first incubated for 1 h at 37 °C, followed by drying in a hot air oven at 50 °C. Button form drug-loaded IHs were kept in 100 mL PBS solutions at pH 5.5, pH 7.4, and pH 8.5 at 37 °C to release the drug.

Method 2: The freshly synthesized gel form of IHs immersed in 100 mL PBS solution with pH values of 5.5, 7.4, and 8.5. To maintain the same overall volume of the release medium for every drug release test, aliquots from each medium were removed every two hours and concurrently replaced with fresh media. In this case, the drug molecules were thought to be uniformly distributed throughout the IHs. The observations were made in triplicate, and the mean value of the three observations was used for the analysis. The amount of drug released was determined using UV–visible spectroscopy at an absorbance of 275 nm (PBS 5.5) and 271 nm (PBS 7.4 and PBS 8.5).

The experimental data was fitted with the Korsmeyer-Peppas model (K-P) model. The K-P model is based on the power law equation and is represented as

$$\frac{M_t}{M_\infty} = kt^n$$

The cumulative drug concentrations released at time "*t*" and time infinity are denoted by M_t and M_∞, respectively. The release exponent "*n*" describes the release mechanism, and the release rate constant "*k*" depends on the matrix's geometry and structure.

2.11 Antibacterial assay

The antibacterial activity was evaluated using the zone inhibition method, also known as the Kirby-Bauer method. Both *S. aureus* (Gram-positive) and *E. coli* (Gram-negative bacteria) were selected as representative models. The Mueller–Hinton agar plates were inoculated by spreading with 100 µl of each bacterium culture (approx cell density 1.5×10^8 CFU mL^{-1}) and followed by placing the discs containing 10 µl of different concentrations (0 to 100 mg mL^{-1}). A 10% sample was taken and serially diluted to achieve the required load on the disc. One disc (5 mm) on each plate was

loaded with solvent alone, serving as the vehicle control, while a Ciprofloxacin disc (10 µg) was used as the positive control. The plates were then incubated at 37 °C for 24 h. The clear zones around the discs were measured and recorded to assess the antibacterial activity. Three replicates were conducted for each compound against each bacterium.

2.12 Hemolysis assay

Human blood was centrifugated at 1000 g for 15 min. at 4 °C to separate plasma and buffy coat, which were then removed. The erythrocytes were washed three times with PBS and prepared as a 1% suspension in PBS. Subsequently, 500 µl of the 1% erythrocyte suspension was dispensed into a 96-well plate and incubated for 30 min. at 37 °C in the presence or absence of 100 µl test samples across a concentration range of 1–1000 µg mL^{-1}. After incubation, 400 µl of PBS was added to the samples. For the positive control representing 100% hemolysis, 400 µl of RBC lysis buffer was added to the control sample. Following centrifugation, 100 µl of the supernatant solution was transferred to a fresh 96-well plate, and the absorbance was measured at 540 nm using a microplate reader (iMark, BioRad).

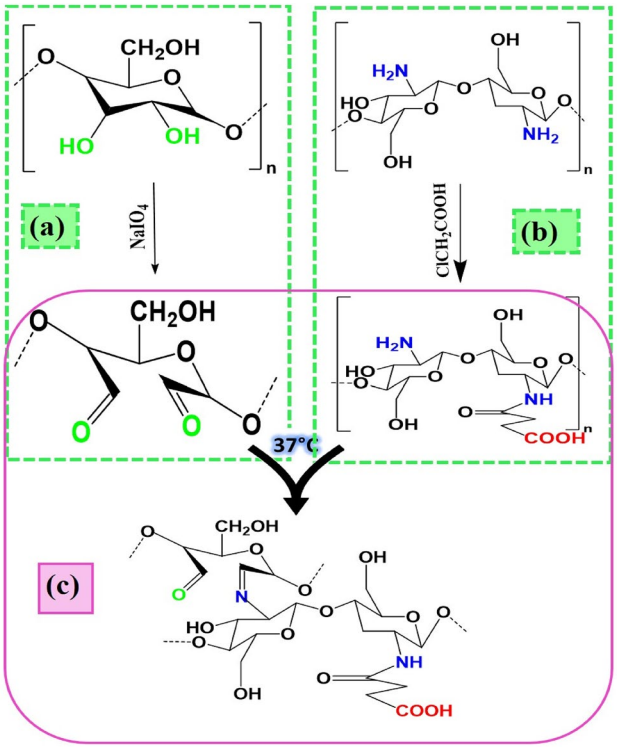
2.13 In vitro cytotoxicity assay

Using the NCCS Pune cell line L-929, the cytotoxicity of the supplied samples was assessed using the MTT Assay. The cells (5000–8000 per well) were grown in DMEM media with 10% FBS and 1% antibiotic solution for 24 h at 37 °C with 5% CO_2 in 96-well plates. Cells were treated with formulations ranging from 1 to 1000 µg mL^{-1} the following day. MTT Solution was added to the cell culture and incubated for an additional two hours at a final concentration of 250 µg mL^{-1}, following a 24-h incubation period. After this, the culture supernatant was extracted, and the cell layer matrix was dissolved in 100 µl dimethyl sulfoxide (DMSO) and measured at 540 nm and 660 nm using an Elisa plate reader (iMark, Biorad, USA) computed by using the software Graph Pad Prism -6. Images were captured under an inverted microscope (Olympus ek2) using a Camera (AmScope digital camera 10 MP Aptima CMOS).

2.14 In vitro degradation in PBS at pH 7.4

Two techniques were used to carry out the in vitro degradation of IHs.

Method 1. The methodology of Rui Chen et al. was applied to ascertain in vitro degradation [24]. A known amount of dried IHs (0.2 g) was weighed (X_1) after they attained equilibrium swelling after being submerged in PBS at pH 7.4 and 37 °C. After removing the samples from the solution and cleaning the surface, they were weighed again



Scheme 1 (a) Oxidation of mesquite gum (MG-CHO) (b) Derivatization of chitosan (CMCh) (c) Fabrication of injectable hydrogel (IH)

(X_2) at predetermined intervals. Equation (4) was utilized to calculate the percentage degradation (IVD %) based on weight loss. The three trials were conducted, and the mean value of the three trials was used for the analysis.

Method 2. The freshly prepared gel form of IHs was cut into slices and put in PBS at pH 7.4 at 37 °C. Then, weight loss was checked using Eq. (4).

$$IVD\% = \frac{X_1 - X_2}{X_1} \times 100 \tag{4}$$

3 Result and Discussion

The present work reports the synthesis of multi-aldehyde groups on mesquite gum by employing $NaIO_4$ as an oxidant (Scheme 1a). To make chitosan more soluble in water in neutral environments, carboxymethyl groups were added to the molecule to form CMCh (Scheme 1b). The aldehyde groups of MG-CHO can covalently crosslink with the NH_2 of CMCh through Schiff base crosslinking at 37 °C without the requirement for any hazardous or chemical crosslinking agents (Scheme 1c). The study investigated the synthesized injectable hydrogel's rheological behavior, swelling ability, self-healing capability, drug release assay at varying pH

levels, biocompatibility, and antibacterial properties. The actual application of injectable hydrogel is in a gel state. Therefore, this work presents comparison analyses between dry and gel forms of IHs, examining their swelling, drug release, and hydrolytic degradation behavior.

3.1 Effect of concentration of $NaIO_4$ on aldehyde content

The impact of MG derivatization on aldehyde content than on gelation time is displayed in Table 1. Gelation time drops from 30 to 7 min. An increase in $NaIO_4$ concentration (0.5 to 2% w/v) nearly stays constant (8 min.), with an additional rise in $NaIO_4$ concentration (3% w/v). This behavior was caused mainly by an increase in the oxidizing agent's concentration, which increases the conversion of hydroxyl groups to aldehyde groups. After analysis, the percentage of aldehyde was discovered to have increased from 33.5 to 50.2%. This behavior was explained by determining the ratio of aldehyde % to the amine group on CMCh.

The results show that adding 2% (w/v) of $NaIO_4$ results in a value of 47.6%, whereas adding 3% (w/v) of $NaIO_4$ results in an almost constant value of 50.2%. To accomplish a successful Schiff base reaction, the ratio of the aldehyde group to the amino group must be adjusted for different polysaccharide hydrogels. An incorrect ratio would lead to longer gelation times. Hyaluronic acid that had been derivatized using $NaIO_4$ and chitosan were combined to create an injectable hydrogel by Nair et al. After adding the necessary quantity of $NaIO_4$, successful oxidation degrees of 10, 15, and 25% were obtained with 8-s gelation time [25]. Based on this investigation, the optimal $NaIO_4$ concentration for subsequent research was determined to be 2% w/v.

3.2 Effect of concentration of oxidized mesquite gum on gelation time

The impact of oxidized MG content on gelation time is displayed in Table 2. The $NaIO_4$ concentration was maintained at 2% w/v (MG-CHO-2). The table unequivocally demonstrates that a reduction in gelation time (7 min. to 2 min.) was noted with a rise in MG-CHO-2 concentration (1% w/v to 5% w/v). This is mainly because derivatized chitosan's amine groups were more likely to react with an overall number of aldehyde functional groups. At the same time, adding 7% w/v of MG-CHO-2 increases the gelation time (3 min.) because there are now more aldehyde groups in the MG-CHO-2 than amine groups in the CMCh. As a result, too many aldehyde groups could prevent the formation of gels. Moreover, the basic unitary approach was used to explain this tendency. A comprehensive analysis of the computation was conducted, and the magnitude value of the ratio increased steadily (0.7 to 5) with the increased MG-CHO-2 (1% w/v to 7% w/v). According to a study by Ashiq Hussain Pandit et al., they created injectable hydrogel using oxidized gum arabic and 0.5 g of $NaIO_4$. They were able to reach an oxidation degree of about 44% by varying the amount of derivatized gum arabic (20% to 40%), which resulted in a reduction in gelation time (42 s to 29 s) [26].

3.3 Physical properties of injectable hydrogel

Table 3 lists the physical characteristics, including gel content, porosity, syringeability, and network parameters. It is discovered that IH-MG-1, IH-MG-3, and IH-MG-5 have gel contents of 62%, 73%, 81%, and porosities 75.42%, 61.24%, and 49.23%, respectively. The increased gel content and lower porosity of IHs are attributed to well-formed imine

Table 1 Effect of concentration of $NaIO_4$ on gelation time of injectable hydrogel

Sample code	MG % (w/v)	$NaIO_4$ % (w/v)	Aldehyde % (A)	Amine group % present in CMCh (B)	Ratio = $\frac{A}{B}$	Gelation time (min.)
IH-$MG_{0.5}$–1	1	0.5	33.5	67	0.5	30
IH-MG_1-1	1	1	40.8	67	0.6	15
IH-MG-1	1	2	47.6	67	0.7	7
IH-MG_3-1	1	3	50.2	67	0.7	8

Table 2 Effect of concentration of oxidized MG (MOG-CHO-2) on gelation time

Sample code	Conc. of MG-CHO-2 (g) (B)	Aldehyde % of MG-CHO-2 (A)	Amine group % present in CMCh (C)	Ratio = $\frac{A \times B}{C}$	Gelation time (min.)
IH-MG-1	1	47.6	67	0.7	7
IH-MG-3	3	47.6	67	2.13	5
IH-MG-5	5	47.6	67	3.55	2
IH-MG-7	7	47.6	67	5	3

Table 3 Physical properties of injectable hydrogel

Property	Value		
	IH-MG-1	IH-MG-3	IH-MG-5
Gel content	62 ± 1%	73 ± 1%	81 ± 2%
Porosity	75.42 ± 1.5%	61.24 ± 1.2%	49.23 ± 1.4%
Syringeability	99.21 ± 0.5%	98.76 ± 04%	98.11 ± 0.6%
Average molecular weight between crosslinks	43.2×10^{-2} g mol^{-1}	51.3×10^{-2} g mol^{-1}	67.4×10^{-2} g mol^{-1}
Crosslink density	1.1 mol cm^{-3}	1.5 mol cm^{-3}	1.9 mol cm^{-3}
Mesh size	7.2×10^{-2} nm	5.1×10^{-2} nm	4.3×10^{-2} nm

bond crosslinks. Additionally, with increasing porosity, the average molecular weights between crosslinks increase, the mesh size increases, and the crosslink density decreases. This is because a more compact structure developed when the concentration of MG-CHO-2 was raised (1% to 5% w/v). Determining the porosity of the IH also aids in understanding the spaces where liquids get trapped or stuck. In a study by Mohamad Sadegh Aghajanzadeh et al., % porosity determination of 50% was observed using oxidized xanthan gum and gelatin via Schiff base crosslinking [21].

Self-healing behavior was checked by placing two half-sliced, cylindrical injectable hydrogels of distinct colors together; the self-healing property was ascertained qualitatively. The self-healing IH in both hues was put into a 20 mL syringe and extruded using a 21 G needle. After 30 min., a picture of "DTU" was produced. The image displays the blending of yellow with the IH's existing color (dirty white). The continuous strands emerge from a needle to provide a "DTU" image, confirming the IHs' self-healing properties by showing the reversibility of the Schiff-base bond, which is further utilized to fix the cut surfaces of wounds (Fig. 1).

Syringeability and injectability evaluation were described in supplementary data 3.1 (Fig. S2).

3.4 Structural properties

FT-IR and proton NMR characterize the injectable hydrogel's structural characteristics. The FT-IR spectra of the MG, MG-CHO-2, chitosan, CMCh, and IH-MG-5 are displayed in Fig. 2a to 2c. The FT-IR spectra of pure MG exhibit absorptions attributed to O–H and C–H at 3290 and 2922 cm^{-1}, respectively. A peak on 1600 cm^{-1} is designated to amide I, indicative of the protein content within the MG. The vibration modes of C–O and C–O–H groups of carbohydrates (glucose, mannose, and galactose) are evidenced by peaks around 1150 cm^{-1} and 900 cm^{-1}. Furthermore, the occurrence of pyranose glycosidic acetal groups is suggested by the peaks at 834 and 776 cm^{-1}[27]. A significant peak is observed at 1727 cm – 1, which confirms the presence of the aldehyde group [28] (Fig. 2a), which qualitatively supports observation obtained from the oxidation of MG. The FT-IR spectra of pure chitosan show that the bending vibrational peak of the amino group is located at 1570 cm^{-1}. The C–O–C asymmetric structure of the chitosan saccharide is located at 1250 cm^{-1}, and the C-O stretching vibration is located at 1024 cm^{-1}. The presence of two additional peaks in the FT-IR spectrum

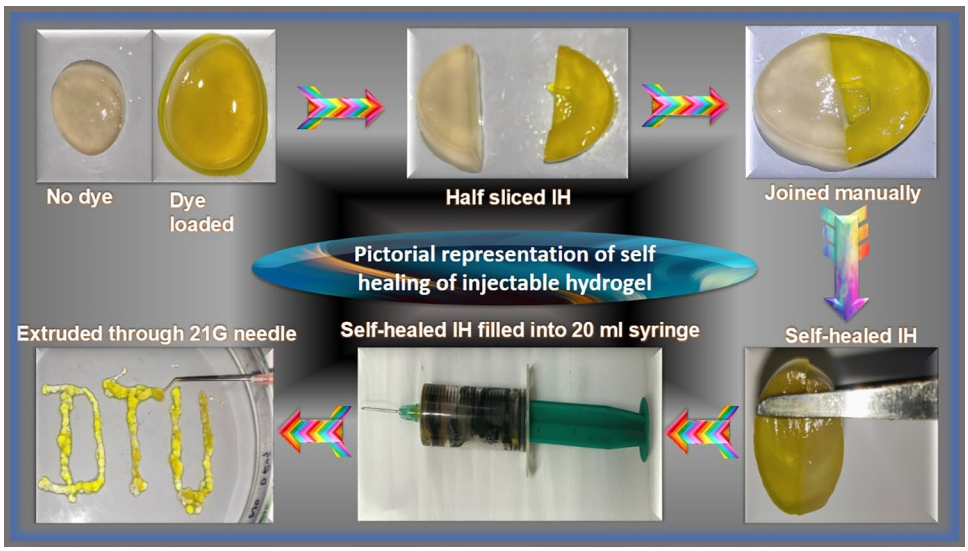


Fig. 1 Pictorial representation of self-healing behavior of injectable hydrogel

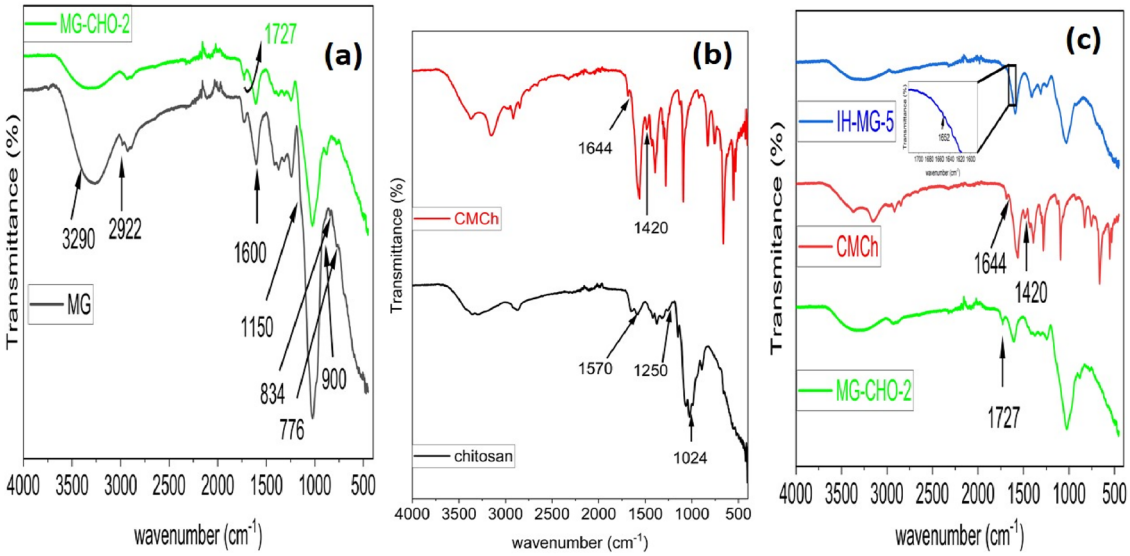


Fig. 2 FT-IR spectra of **a** MG, MG-CHO-2 **b** chitosan, CMCh and **c** MG-CHO-2, CMCh, IH-MG-5

of CMCh at 1644 and 1420 cm^{-1} indicates symmetric and asymmetric stretching vibrational modes of the -COO group, respectively. These peaks show that chitosan derivatization was successful and that carboxymethyl groups are present on CMCh [28] (Fig. 2b). A peak at 1642 cm^{-1} in the FT-IR spectrum of the injectable hydrogel (IH-MG-5) indicates the formation of an imine structure as a result of a dynamic Schiff base crosslinking reaction between the amine group of CMCh and the aldehyde groups of MG-CHO-2. Removing the aldehyde peak in the IH-MG-5 spectrum further illustrates the evolution of the Schiff base linkage [29] (Fig. 2c).

As shown by FT-IR, structural properties are also supported by the proton NMR spectroscopy technique. The ^{1}H-NMR spectra of the chitosan show a peak at 1.24 ppm due to the presence of protons in the acetyl group. The anomeric protons of glucosamine show a peak at 2.96 ppm, while the non-anomeric protons of the chitosan structural backbone are responsible for the peaks ranging from 3.68 to 4.84 ppm. Additionally, a new peak was observed at 3.87 ppm due to the presence of the $–CH_3COO-$ group (Fig. S3) [30]. D_2O was used as a solvent to oxidize MG before ^{1}H-NMR analysis. The presence of methyl groups is the cause of the peak at 1.3 ppm. The spectra showed congested signals in the 2.16, 3.65–4.30, and 4.9 ppm spectrum bands, which corresponds to the distinctive characteristics of the anomeric proton in polysaccharides due to the existence of sugar residues. The primary peak, identified by D_2O, was observed at 4.7 ppm (Fig. 3a). A new distinctive peak of the aldehydic proton is observed at 9.31 ppm, which confirms the oxidation of MG, which supports the data obtained from the FT-IR graph (Fig. 3b) [31].

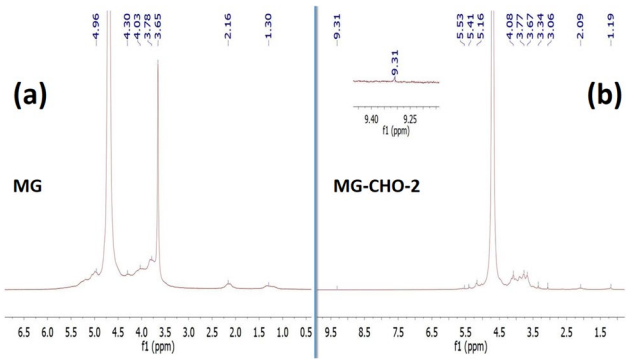


Fig. 3 ^{1}H-NMR spectra of **a** MG **b** MG-CHO-2

3.5 Rheological properties

A rheometer operating at a constant strain of 1% and a set frequency of 1 rad s^{-1} was used to conduct the time sweep research. Precursors 1 and 2 were concurrently extruded in a 1:1 ratio at 37 °C on the rheometer platform, and they were then left to crosslink on the platform itself. In isothermal mode (37 °C), the variations in storage modulus (G') and loss modulus (G") over time were measured. After the initial few seconds, the G' value increased because the viscosity of the liquid increased, and the Schiff base reaction proceeded slowly. The G' value peaked at 2086.2 Pa for IH-MG-5 after 300 s, surpassing that of IH-MG-3 at 760.02 Pa and IH-MG-1 at 257.35 Pa. This is because the concentration of MG-CHO-2 increased (1% w/v to 5% w/v) along with the concentration of the Schiff base bond, which led to compact polymer networks and more

excellent mechanical characteristics. This interaction between the NH_2 groups of CMCh and the aldehyde group of MG-CHO-2 caused the G' to grow dramatically quickly. Additionally, the G'' values for IH-MG-1, IH-MG-3, and IH-MG-5 are 6.12, 23, and 272 Pa, respectively. Figure 4a indicates no discernible variations in the G'' values of the IHs with varying weight ratios of MG-CHO-2. Yongping Liang et al. synthesized an injectable hydrogel using Schiff base chemistry between chitosan-grafted-dihydrocaffeic acid and oxidized pollulan; the time sweep study shows that G' surpassed the G'' with time [32].

To assess the shear thinning behavior, IHs were placed between the parallel plates of the rheometer and allowed to squeeze at increasingly more excellent shear rates during a continuous flow sweep test (0.1–100 s^{-1}, 1% strain) conducted at 37 °C. When IH-MG-5 is compared to IH-MG-1 and IH-MG-3, it has the most excellent MG-CHO-2 value and, thus, the highest initial viscosity value. The viscosity

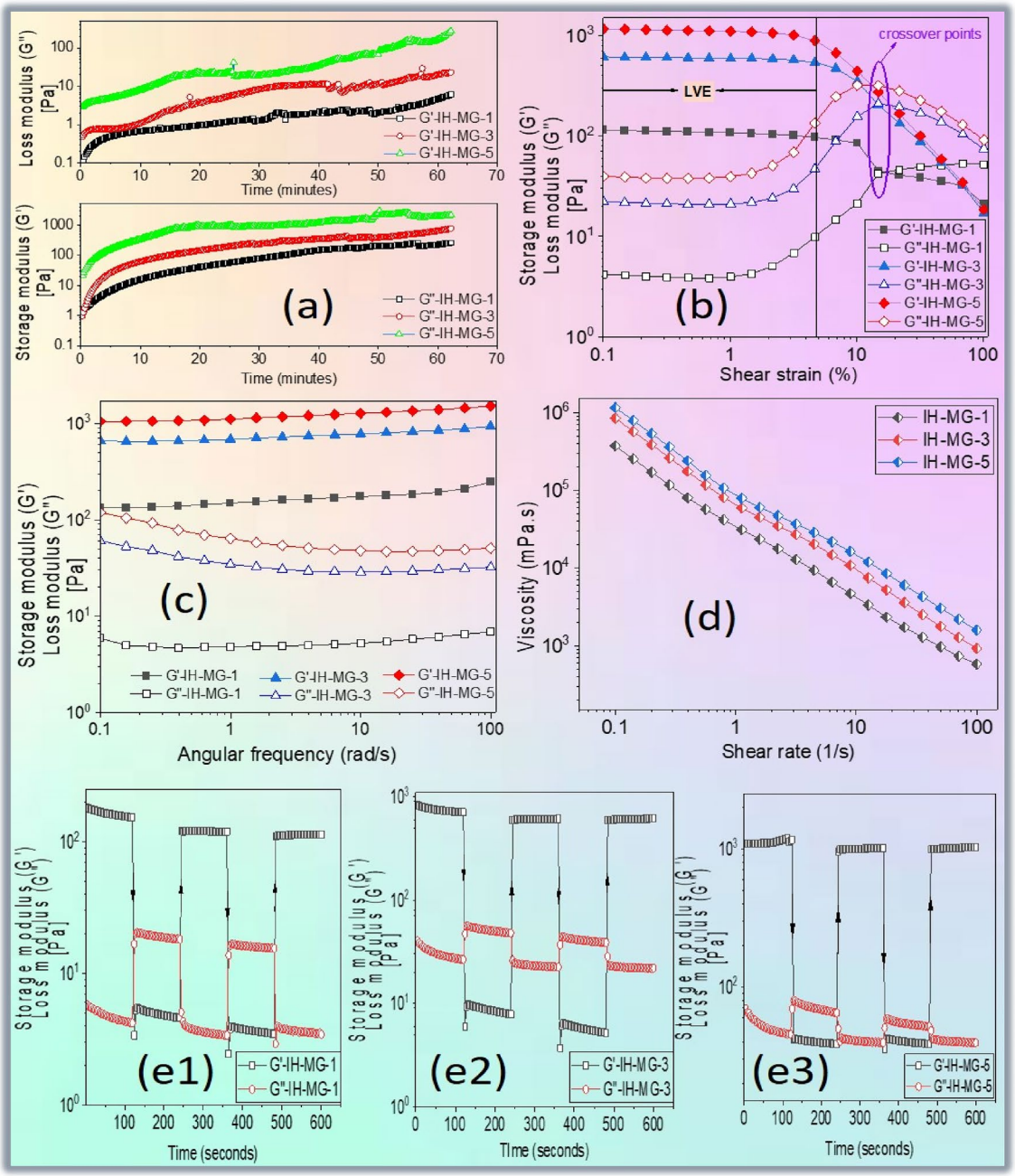


Fig. 4 **a** Isothermal time sweep study **b** Amplitude sweep study **c** Frequency sweep study **d** Flow sweep behavior of IH-MG-1, IH-MG-3, and IH-MG-5. Step-strain measurement of **e1** IH-MG-1 **e2** IH-MG-3 **e3** IH-MG-5

was found to decrease with increasing shear rate (1,151,700 to 1592 mPa.s for IH-MG-5, 842,490 to 926.3 mPa.s for IH-MG-3, and 371,700 to 578.45 mPa.s for IH-MG-1), indicating that the IHs exhibits shear thinning behavior. This behavior could be caused by the Schiff base linkage breaking at a high shear rate (Fig. db). The shear thinning behavior of injectable hydrogel further supports the data on syringeability, injectability, and self-healing properties.

The linear viscoelastic region (LVE) of the IHs was measured using the amplitude sweep test at multiple strain levels ranging from 0.1 to 100% at 37 °C (Fig. 4b). Using the amplitude sweep test, the LVE of IHs was found between 0.1 and 5% of the strain values. A strain value of 1% was chosen for all gelation-related kinetic investigations. At different strain values, it was found that the IHs had more excellent G′ values than G″ values. The crossovers for the IHs appeared at 14.8%, indicating that the hydrogel was converting from a solid to a liquid. Thus, IHs possess a higher amount of MG-CHO-2, resulting in more excellent elasticity, which is necessary for these injectable hydrogels to resist the administration-linked strains during their injectable formulation.

The variable angular frequency shear in the range of $0.1-100$ rad s^{-1} was delivered to the IHs at a constant amplitude (1%) to perform a frequency sweep test (Fig. 4c). The figure shows that these IHs demonstrated mechanical robustness with more excellent G′ values than G″ values throughout all investigated frequency ranges. These injectable hydrogels are both elastic and stiff, as observed by the more excellent value of G′ than G″. Because the IH-MG-5 has a higher crosslinking density, which limits the mobility of chains within the hydrogel matrix, it exhibited higher values of G′ (1518 Pa) than the IH-MG-1 (247 Pa) and IH-MG-3 (934 Pa). The frequency sweep test's G′ values showed a similar pattern to the amplitude sweep. Balitaan et al. synthesized an injectable hydrogel using a Schiff base crosslink between acrylamide-modified β -chitin and alginate dialdehyde. The amplitude sweep test, frequency sweep test, and viscosity measurement were conducted, and results showed that injectable hydrogel possesses better mechanical properties with shear thinning characteristics [33].

A series of rheological recovery tests were carried out to quantify the self-healing properties of IHs. This behavior was evaluated by oscillatory step-strain measurement. The G' value had a high magnitude at low strain (1%), and in all IHs, as the strain shifted to high values (200%), the G' value was found to be closest to G". This resulted from the structure deforming at high strain values. The behaviors were stable over multiple cycles, indicating the self-healing capability and stability of the dynamic Schiff base and dynamic coordination network. After all cycles of alternate strain (low to high), the highest recovery was found in the case of IH-MG-5 (95%, **Fig. 4e1**), followed by IH-MG-3 (78.33%, **Fig. 4e2**) and IH-MG-1 (63%, **Fig. 4e3**) for a given interval of time as shown in Eq. 5.

$$\text{Recovery}\% = 100 - \frac{initialG' - finalG'}{initialG'} \times 100 \tag{5}$$

Jin Qu et al. synthesized an injectable hydrogel using N-carboxyethyl chitosan and dibenzaldehyde-terminated poly(ethylene glycol) via Schiff base chemistry in which self-healing analysis was examined qualitatively and quantitatively [34].

3.6 Morphological evaluation by SEM

The surface morphology of the IHs synthesized with varied amounts of MG-CHO-2 (1, 3, and 5% w/v) was investigated using the SEM (Fig. 5). The SEM images were captured at 1300 X and 1800 X for all IHs. The SEM pictures

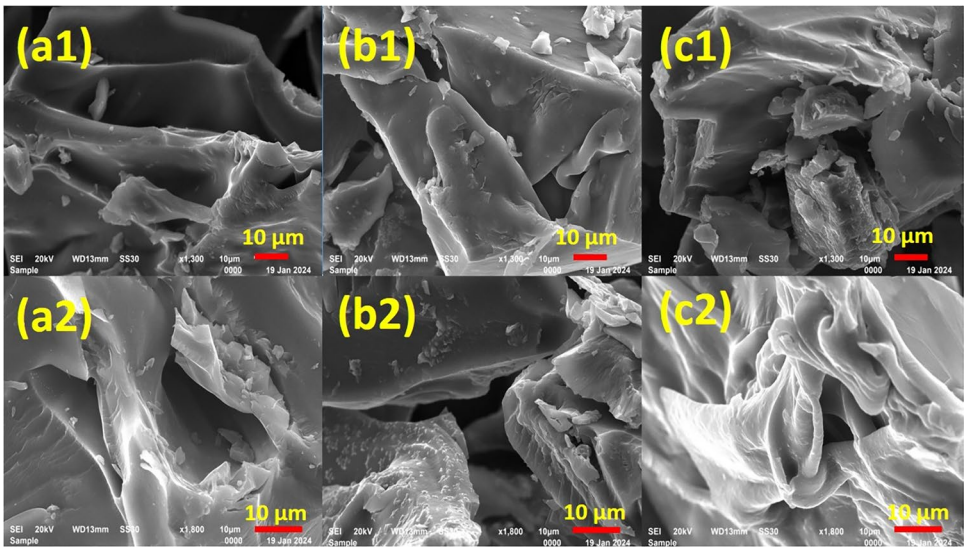


Fig. 5 SEM images at 1300X **a1** IH-MG-1 **b1** IH-MG-3 **c1** IH-MG-5 and 1800X **a2** IH-MG-1 **b2** IH-MG-3 **c2** IH-MG-5

demonstrate the relationship between the structure and the composition of the IHs.

Image J software was used to calculate the average diameter of the pores. When compared with IH-MG-3 (21 µm, **Fig. 5b1, 5b2**) and IH-MG–1 (34 µm, **Fig. 5a1, 5a2**), the IH-MG–5 shows a tight network structure and a comparatively smaller pore size (14 µm, **Fig. 5c1, 5c2**). The SEM images additionally show that the IHs exhibit a continuous three-dimensional porous network at different MG-CHO-2 concentrations (1%, 3%, and 5% w/v). The porosity quickly decreased when the MG-CHO-2 concentration in the IHs increased. The linked pores of the IHs network might provide enough vacancies for the loading of drugs and also help with their release. Shuangqing Wang et al. fabricated an injectable hydrogel using Schiff-based chemistry composed of phenylboronic acid grafted polylysine and oxidized dextran; SEM study revealed a porous structure having size ranges from 209 to 37 µm [35].

3.7 Swelling study of IHs

The swelling behavior of synthesized IHs is investigated in PBS at three pH levels (pH 5.5, pH 7.4, and pH 8.5).

Swelling study from the dry form of IH (Method 1) - The swelling ratio decreases quickly with the increasing quantity of MG-CHO-2 at low pH levels (pH 5.5). The IH-MG-1 exhibits a swelling ratio of about 810%. Figure 6a shows that the IH-MG-3 and IH-MG-5 had swelling ratios of around 714% and 555%, respectively. The swelling ratio values of IH-MG-1, IH-MG-3, and IH-MG-5 in physiological pH (7.4) are 514, 460, and 420%, respectively (Fig. 6b). The swelling ratios are also calculated in alkaline pH (8.5) where the swelling ratios for IH-MG-1, IH-MG-3, and IH-MG-5 were found to be 784, 692, and 528% respectively.

Swelling study from gel form of IH (Method 2) – In an acidic environment (pH 5.5), the swelling ratio decreases rapidly as the concentration of MG-CHO-2 increases. The IH-MG-1 exhibits a swelling ratio of roughly 611%. Conversely, the swelling ratios of the hydrogels IH-MG-3 and IH-MG-5 are around 535 and 483%, respectively (Fig. 6d). At physiological pH (7.4), Fig. 6e displays the swelling ratios for IH-MG-1, IH-MG-3, and IH-MG-5, which are 484, 442, and 401%, respectively. For swelling ratios in alkaline pH (8.5), the values for IH-MG-1, IH- MG -3, and IH- MG-5 are 557, 501, and 452%, respectively (Fig. 6f).

The reason is that when the concentration of MG-CHO-2 rises (from 1 to 5% w/v), the IHs network structure becomes more compact. This leads to a more severe constraint on the mobility of the polymer chains, which minimizes their exposure to water. Because CMCh's amino groups are hydrophilic, they hydrate with water quickly and rapidly. In contrast, the imine group, found in the IHs, is assumed to be hydrophobic. The increased concentration of MG-CHO-2 in synthesized injectable hydrogels may result in a more significant concentration of hydrophobic moieties. This slows the IH swelling rate by reducing its ability to form hydrogen bonds with water. Consequently, in both dry and gel form across all pH ranges (5.5, 7.5,

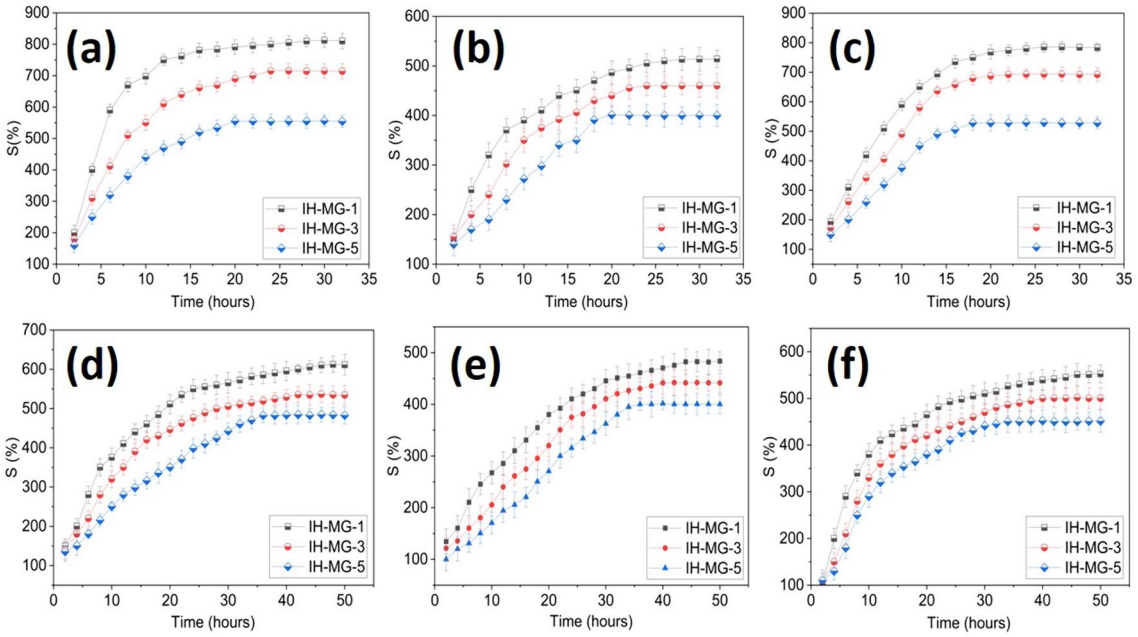


Fig. 6 Swelling behavior of IH-MG-1, IH-MG-3, and IH-MG-5 in PBS in dry form at **a** pH 5.5 **b** pH 7.4 **c** pH 8.5 and in gel form at **d** pH 5.5 **e** pH 7.4 **f** pH 8.5

and 8.5), the injectable hydrogel (IH-MG-5), which has a higher concentration of MG-CHO-2, exhibits a lower swelling ratio than IH-MG-1 and IH-MG-3. Furthermore, as pH rises (from 5.5 to 7.4), the swelling ratios of IH-MG-1, IH-MG-3, and IH-MG-5 (in both dry and gel form) decrease and then increase again to pH 8.5, indicating their pH-sensitive swelling behavior. At pH 5.5, an acidic pH, the amino groups in CMCh undergo protonation, leading to electrostatic repulsion between intra- and interpolymeric chains. As a result, the IHs network's ruptured structure enhances water absorption. However, at pH 7.4 (higher pH), deprotonation within the amine group of CMCh causes a decrease in the swelling ratio in IHs. Higher swelling ratios can be caused by further increasing the pH to 8.5, which could lead to a loose network topology in IHs (dry and gel form). Therefore, the study mentioned above demonstrates that the concentration of MG-CHO-2 and the pH of the swelling media affect the swelling ratios of the dry and gel forms of IHs. Kumar et al. fabricated an injectable hydrogel using *Moringa oleifera* gum and carboxymethyl chitosan crosslinked through the Schiff base mechanism; the swelling study was conducted for dry and gel forms of injectable hydrogels. At pH 5.5, more drug release was observed than at pH 7.4 [36].

3.8 Drug release behavior

The IHs drug release behavior was assessed in PBS at pH 5.5, 7.4, and 8.5 at 37 °C using ciprofloxacin HCl as the model drug. Skin, urinary tract, and respiratory tract bacterial infections are treated with the antibiotic drug ciprofloxacin HCl. Two distinct methods were employed, as stated in the description of the experimental section. The drug release pattern over time is shown in Fig. 7.

Drug release from dried injectable hydrogel (Method-1) – At low pH (pH 5.5), the Cipro release increases rapidly as MG-CHO-2 concentrations rise. With the least amount of MG-CHO-2, the IH-MG-1 yields a cipro release of 13,110 µg 100 mL^{-1} g^{-1}. Conversely, Fig. 7a shows that the IH-MG-3 and IH-MG-5 release cipro at approximately 11,130 and 9125 µg 100 mL^{-1} g^{-1}, respectively. At physiological pH (7.4), the Cipro release for IH-MG-1, IH-MG-3, and IH-MG-5 is 9566, 8011, and 6896 µg 100 mL^{-1} g^{-1}, respectively (Fig. 7b). IH-MG-1, IH-MG-3, and IH-MG-5 have cipro release values in alkaline pH of 12,326, 10,210, and 8895 µg 100 mL^{-1} g^{-1},

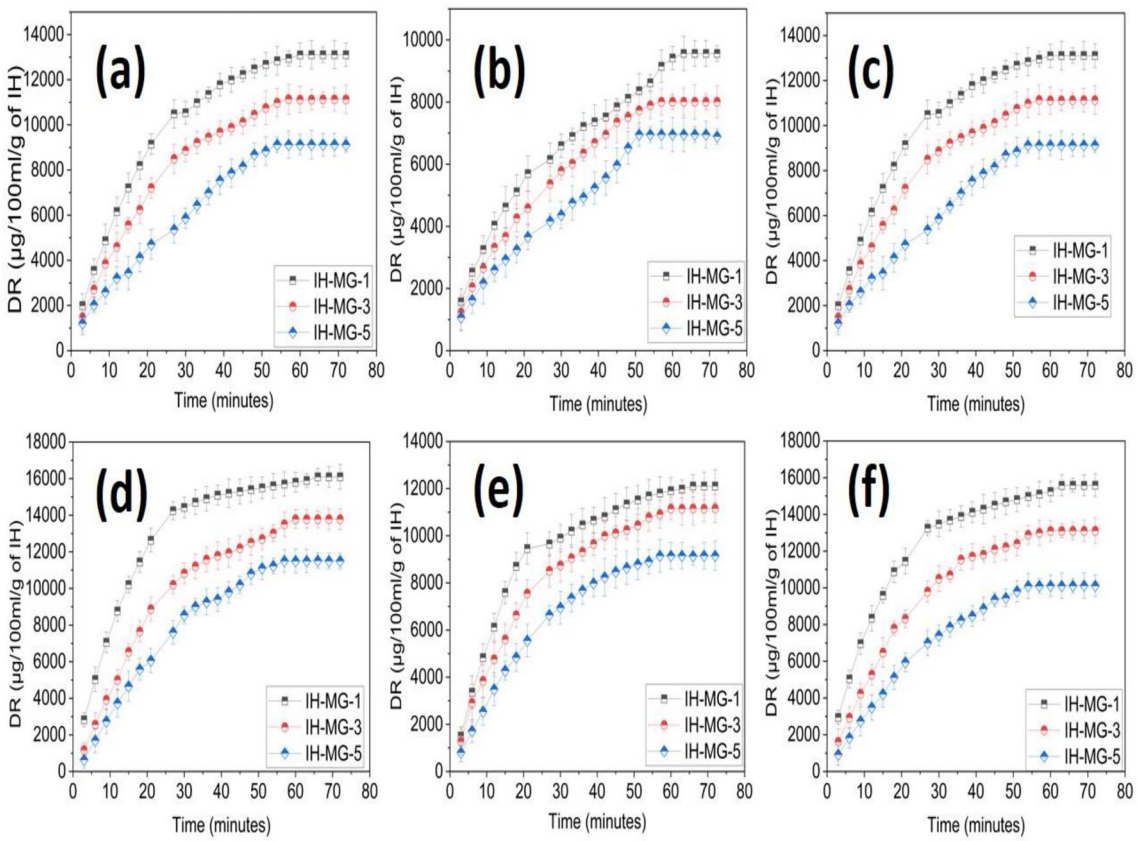


Fig. 7 Drug release behavior of IHs in PBS in dry form **a** at pH 5.5 **b** pH 7.4 **c** pH 8.5 and in gel form **d** at pH 5.5 **e** at pH 7.4 **f** at pH 8.5

respectively (Fig. 7c). A relatable research article published by Huishuang Yin et al. in which they fabricated an injectable hydrogel using Schiff base chemistry between oxidized microcrystalline cellulose from pineapple peel were crosslinked with carboxymethyl chitosan and rutin release studies show a pH-dependent release behavior [37].

Drug release from gel form of hydrogel (Method-2):

The IH-MG-1, which contains the least concentration of MG-CHO-2, has a cipro release of around 16,101 µg 100 mL^{-1} g^{-1}. Moreover, Fig. 7d shows that the cipro release values for IH-MG-3 and IH-MG-5 are 13,787 µg 100 mL^{-1} g^{-1} and 11,500 µg 100 mL^{-1} g^{-1}. For IH-MG-1, IH-MG-3, and IH-MG-5, the cipro release values at physiological pH (7.4) are 12110, 11160, and 9150 µg 100 mL^{-1} g^{-1}, correspondingly (Fig. 7e). The values for IH-MG-1, IH-MG-3, and IH-MG-5 for cipro release in alkaline pH (8.5) are 15590, 13111, and 10112 µg 100 mL^{-1} g^{-1}, respectively (Fig. 7f). This can be explained by the fact that the hydrogel network structure gets more compressed when the concentration of MG-CHO-2 rises (from 1 to 5% w/v), leading to an increase in crosslink density (0.7 to 1.2 g cm^{-3}). This reduces exposure to the release of Cipro by drastically restricting the mobility of the polymer chains. The hydrophilic nature of the amino groups in CMCh makes it easy for them to hydrate with water immediately. The imine group, which is thought to have hydrophobic moieties, is present in the IHs. On the other hand, injectable hydrogels may have more hydrophobic moieties due to the higher concentration of MG-CHO-2 in them. As a result, the IH's capacity to create hydrogen bonds with water diminishes, slowing the release of cipro. Because of this, the IH-MG-5 has a larger concentration of MG-CHO-2 than IH-MG-1 and IH-MG-3, and it also shows a lower cipro release in both dry and gel form across all pH ranges (5.5, 7.5, and 8.5). Furthermore, as pH rises (5.5 to 7.4), the cipro release of IH-MG-1, IH-MG-3, and IH-MG-5 (dry and gel form) reduces, suggesting a pH-sensitive behavior.

The amino groups of CMCh undergo protonation at pH 5.5, an acidic pH, which causes electrostatic repulsion between intra and interpolymeric chains. Cipro release is, therefore, improved by the ruptured structure of the injectable hydrogel network. Nevertheless, deprotonation within the amine group of CMCh reduces the release of Cipro in IH at pH 7.4 (higher pH). An even higher pH of 8.5 may cause a looser network structure in IHs (dry and gel form), leading to a greater cipro release—the drug released more slowly from gel form than from the dry form of IHs in all pH ranges. Additionally, method 2 is more practical than method 1 in the real world, where the injectable hydrogels will be used for the wound site. During the healing phase, the drug's gradual release will aid in the decrease of microbial development at the wound site. Therefore, the previously stated tests show that the pH of the solution and the quantity of MG-CHO-2 present are essential factors in the cipro release of IHs in both their dry and gel forms. Kumar et al. synthesized an injectable hydrogel using a Schiff base crosslink between *Moringa oleifera* gum and carboxymethyl chitosan; the drug release study was conducted in a dry and gel form of IH. More drug release was observed at pH 5.5 than 7.4 [36].

The most realistic explanation of the process of antibiotic drug release using injectable hydrogels is given by the K-P model (Table 4). Diffusion that is erosion- and diffusion-controlled is referred to as non-Fickian or anomalous diffusion when the release exponent value "*n*" is between 0.5 and 1. Fickian diffusion is obtained when $n \leq 0.5$, and supercase transport 2 is indicated by $n \geq 1$ [26]. The R^2 value for each injectable hydrogel's release behavior is determined to be greater than 0.99. All synthesized IHs had an "*n*" value greater than 0.5 and less than 1. Hence, the "*n*" value verified that all IHs displayed non-Fickian or anomalous diffusion over the whole pH range in both dry and gel form. Furthermore, the diffusion rate exceeds the hydrogel matrix's relaxation rate in every instance.

3.9 Antibacterial assay

This assessed the IHs bactericidal properties using *S. aureus* and *E. coli* for 24 h at 37 °C. Figure 8 and Table 5 display the antimicrobial test results of positive control

Table 4 K-P kinetic model parameters for IH-MG-1, IH-MG-3, and IH-MG-5 in dry and gel form in PBS at pH 5.5, 7.4, and 8.5

pH of PBS	Physical state of IH	Sample code	Parameters associated with the K-P model		
			kkp	*n*	R^2
pH 5.5	Dry	IH-MG-1	7.69	0.72	0.99
		IH-MG-3	6.30	0.76	0.99
		IH-MG-5	6.51	0.67	0.99
	Gel	IH-MG-1	9.76	0.68	0.99
		IH-MG-3	3.98	0.87	0.99
		IH-MG-5	3.00	0.86	0.99
pH 7.4	Dry	IH-MG-1	9.27	0.60	0.99
		IH-MG-3	8.08	0.65	0.99
		IH-MG-5	7.99	0.61	0.99
	Gel	IH-MG-1	6.05	0.85	0.99
		IH-MG-3	6.43	0.78	0.99
		IH-MG-5	4.22	0.87	0.99
pH 8.5	Dry	IH-MG-1	5.57	0.85	0.99
		IH-MG-3	4.11	0.88	0.99
		IH-MG-5	5.58	0.79	0.99
	Gel	IH-MG-1	10.88	0.63	0.99
		IH-MG-3	6.45	0.79	0.99
		IH-MG-5	3.84	0.88	0.99

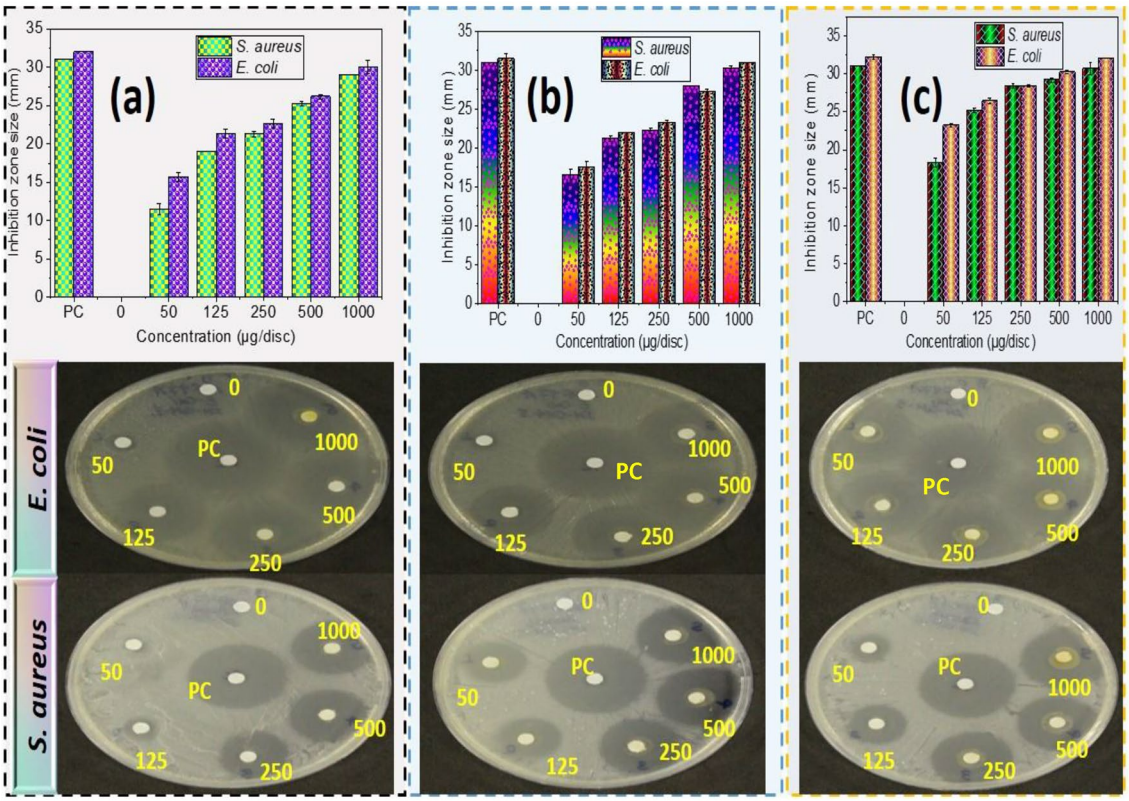


Fig. 8 Antibacterial activity of **a** IH-MG-1 **b** IH-MG-3 **c** IH-MG-5

Table 5 The antibacterial activity of IH-MG-1, IH-MG-3, and IH-MG-5

	Sample name	*S. aureus*						*E. coli*					
		+ ve control	Concentration of IHs (µg/disc)					+ ve control	Concentration of IHs (µg/disc)				
			50	125	250	500	1000		50	125	250	500	1000
Inhibition zone size (mm)	IH-MG-1	28	11	13	22	24	24	27	17	23	24	24	24
	IH-MG-3	28	17	20	23	24	24	29	14	21	22	26	25
	IH-MG-5	28	10	18	21	24	25	25	22	24	24	23	25

along with IH-MG-1, IH-MG-3, and IH-MG-5 (50 to 1000 µg/disc). All IHs exhibited inhibitory effects against each bacterium across all concentrations up to 24 h at 37 °C. This phenomenon was ascribed to chitosan, which could induce the leakage of proteinaceous and other intracellular components from bacteria. Chitosan, containing amino groups and exhibiting a positive electrical charge, traps bacteria with negative charges via electrostatic forces, thus effectively inhibiting bacterial growth. A group of researchers published a research article based on Schiff base chemistry between oxidized konjac glucomannan and chitosan; all synthesized injectable hydrogels showed an antibacterial nature [38].

3.10 Hemolysis assay

Fig. S4 shows that the injectable hydrogels have a similar color to PBS, while the RBC lysis group is bright red. Table S2 and Fig. 9 represent the hemolysis rates of IHs across all concentrations (1 to 1000 µg mL^{-1}). The hemolysis rates of IHs with different concentrations are all < 4%, which meets the standard of hemolysis rates of biomaterials. Therefore, IH-MG-1, IH-MG-3, and IH-MG-5 demonstrated a negligible hemolysis capability for all concentrations (1 to 1000 µg mL^{-1}).

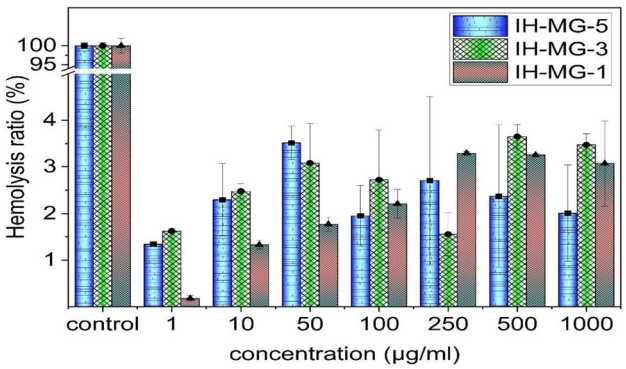


Fig. 9 Statistical representation of hemolysis ratio of IHs

3.11 In vitro cytotoxicity assay

IHs must be biocompatible with the host cells to serve as biomaterials for drug delivery applications. Therefore, an L-929 fibroblast cell line was used to test the cytotoxicity of synthesized injectable hydrogel. The IH-MG-1, IH-MG-3, and IH-MG-5 show negligible toxicity in the MTT experiment, with cell viability of 53%, 58%, and 62% up to 500 µg/mL, respectively (Fig. 10 a). This could be because the MG-CHO-2 contains unreacted carbonyl groups that may interact with the proteins in the cells. The concentration of MG-CHO-2 was found to cause a modest decrease in cell viability. The synthesized injectable hydrogel's inverted phase microscopic images are displayed in Fig. 10b to e2. The outcomes demonstrate that the cytotoxicity of all injectable hydrogels is compatible with L929 fibroblast cell lines. A similar study was conducted by Jitendra K. in which they synthesized an injectable using Schiff base crosslinks between oxidized neem gum and carboxymethyl chitosan, in vitro cytotoxicity results show that synthesized injectable hydrogel was found to be biocompatible against L-929 fibroblast cell lines [39].

3.12 In vitro hydrolytic degradation in PBS at pH 7.4

Synthesized IHs' biodegradability was investigated in PBS (pH 7.4) at 37 °C for 30 days. The weight loss resulting from the hydrolysis of Schiff base bonds in these IHs is evidence of their breakdown, as demonstrated by studies on in vitro hydrolytic degradation.

Hydrolytic degradation from dried IHs (Method-1): In roughly 30 days, IHs with the highest percentage of MG-CHO-2 (IH-MG-5) degraded the least (53.12%). On the other hand, the IH with the lowest amount of MG-CHO-2 (IH-MG-1) was discovered to be the most degraded (70.21%). Furthermore, IH-MG-3 degraded 60.48% in 30 days (Fig. 11a).

Hydrolytic degradation from gel form of IHs (Method 2): In 30 days, IHs—which contains the highest percentage of MG-CHO-2 (IH-MG–5)—observed a 66% degradation. In contrast, IH, with the lowest MG-CHO-2 (IH-MG-1) percentage, observed a 92.45% deterioration in 30 days. Furthermore, the concentration of MG-CHO-2 (IH-MG-3) is in the middle of these two degrades by 80% in PBS (pH 7.4) within 30 days (Fig. 11b).

The relationship between the crosslinking density and the degradation rate can explain this result. The crosslinking density and concentration of MG-CHO-2 significantly influence the hydrolytic breakdown of these IHs. As the amount of MG-CHO-2 increased, the crosslinking density increased, and the rate of IH breakdown decreased. Jia Liu et al. fabricated an injectable hydrogel using Schiff base chemistry between sericin and oxidized dextran and found a slow degradation rate (~ 50% after 60 days) [40].

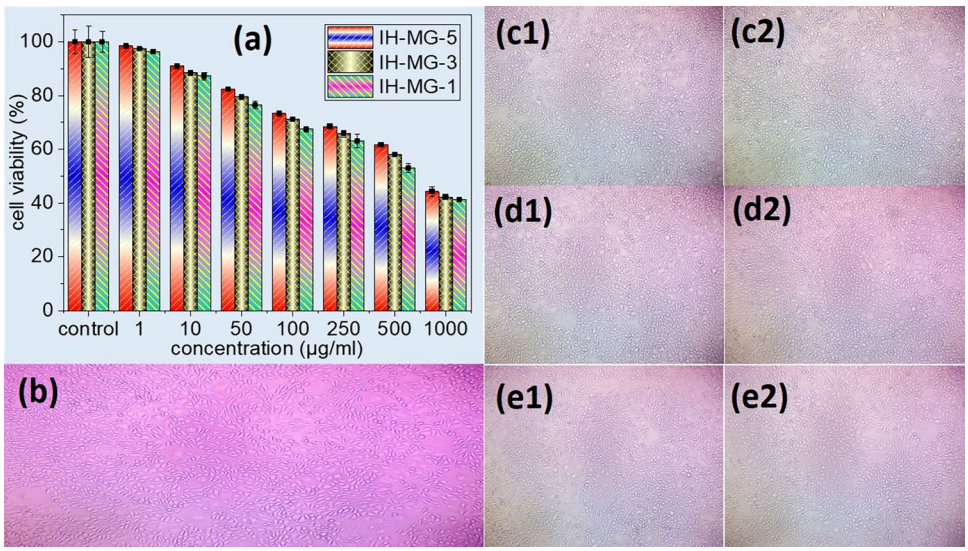


Fig. 10 **a** Cytotoxicity of injectable hydrogels. Inverted phase microscopic images of **b** control. IH-0.5–1 treated L-929 cell line at **c1** 1 µg mL^{-1} **c2** 500 µg mL^{-1}. IH-0.5–3 treated L-929 cell line at **d1**1 µg mL^{-1} **d2** 500 µg mL^{-1}. IH-0.5–5 treated L-929 cell line at **e1** 1 µg mL^{-1} **e2** 500 µg $mL.^{-1}$

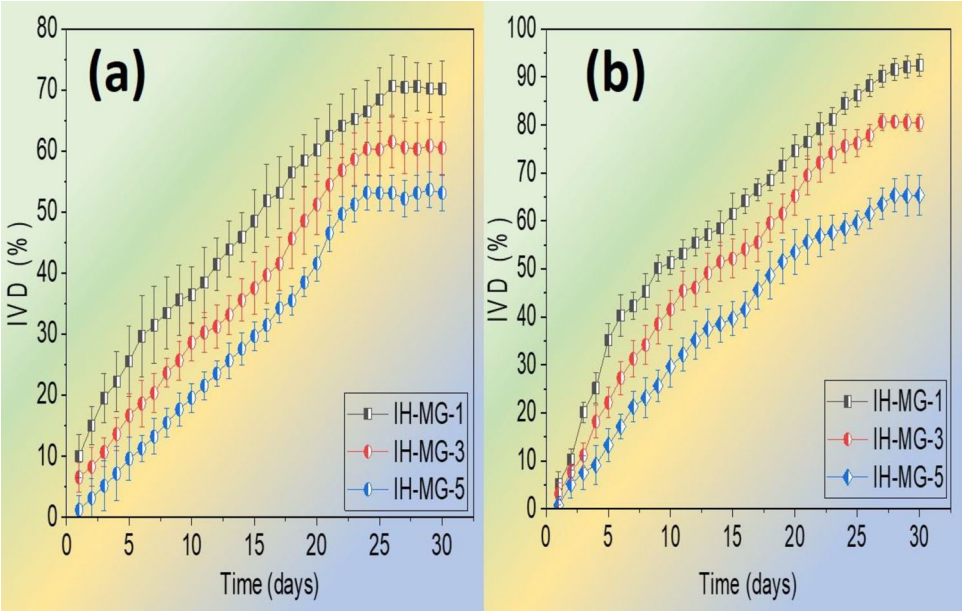


Fig. 11 Hydrolytic degradation behavior of IHs in PBS at pH 7.4 **a** dry form **b** gel form

4 Conclusions

For the first time, self-healing polysaccharide-based injectable hydrogels comprising oxidized mesquite gum with multi-aldehyde groups and carboxymethyl chitosan are fabricated via dynamic Schiff base linkage. The optimized sodium periodate concentration was 2% w/v, which causes mesquite gum's vicinal hydroxyl group to change into an aldehyde group. Since the synthesized IHs have excellent viscoelastic and shear-thinning properties, they are perfect for syringeability and injectability evaluation. IHs have a pH-dependent swellability and drug-release behavior strongly associated with crosslink density. Using ciprofloxacin HCl as a model drug, the IHs demonstrated that the drug release at pH 5.5 was higher in PBS than at pH 7.4, followed by further increases in pH 8.5. The gel form of IHs is a slow-release drug that can be used for wound healing. The non-Fickian diffusion mechanism governs the release of the drug. Hemolysis assay and in vitro cytotoxicity evaluation confirm the biocompatibility of synthesized IHs with antibacterial properties against gram-positive and gram-negative bacteria. Furthermore, synthesized IHs have been shown to gradually degrade in PBS (pH 7.4) at 37 °C. This suggests that they may be helpful for the progressive release of antibiotics at the sites of wounds.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13233-024-00300-7.

Acknowledgements The author greatly acknowledges Delhi Technological University, Bawana Road, Delhi-110042.

Author Contributions This study's conception and design, material preparation, data collection, and analysis were performed by Jitendra Kumar. Review and supervision of the research work was done by Prof. Roli Purwar.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability Data will be made available at the request of the authors.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval Not required.

References

1. L. Zhao, Y. Zhou, J. Zhang, H. Liang, X. Chen, H. Tan, Natural Polymer-Based Hydrogels: From Polymer to Biomedical Applications. Pharmaceutics. **15**(10), 2514 (2023). https://doi.org/10.3390/pharmaceutics15102514
2. S. Sharma, M. Bhende, A. Goel, A review: polysaccharide-based hydrogels and their biomedical applications. Polym. Bull. **81**(10), 8573–8594 (2024). https://doi.org/10.1007/s00289-023-05130-8
3. Y. Gao, T.L. Zhang, H.J. Zhang, J. Gao, P.F. Yang, A Promising Application of Injectable Hydrogels in Nerve Repair and Regeneration for Ischemic Stroke. Int. J. Nanomedicine **19**, 327–345 (2024). https://doi.org/10.2147/IJN.S442304
4. M. Liu, X. Zeng, C. Ma, H. Yi, Z. Ali, X. Mou, S. Li, Y. Deng, N. He, Injectable hydrogels for cartilage and bone tissue engineering. Bone Res. **5**, 17014 (2017). https://doi.org/10.1038/boneres.2017.14
5. C. Mo, L. Xiang, Y. Chen, Advances in Injectable and Self-healing Polysaccharide Hydrogel Based on the Schiff Base Reaction. Macromol. Rapid Commun. **42**(10), 2100025 (2021). https://doi.org/10.1002/marc.202100025
6. S.A.B. Chávez, S.A. Alcalá, Z.M. Almarhoon, A. Turgumbayeva, E.S. Gürer, M. De Los, D. Campos-Echeverria, H. Cortés, A.R.

Montero, M.L.D.P. Audelo, J.S. Rad, G.L. Gómez, Novel ultra-stretchable and self-healing crosslinked poly (ethylene oxide)-cationic guar gum hydrogel. J. Biol. Eng. (2023). https://doi.org/10.1186/s13036-023-00376-2
7. F. Hisham, M.H.M. Akmal, F. Ahmad, K. Ahmad, N. Samat, Biopolymer chitosan: Potential sources, extraction methods, and emerging applications. Ain Shams Eng. J. **15**(2), 102424 (2024). https://doi.org/10.1016/j.asej.2023.102424
8. Y. Yang, L. Xu, J. Wang, Q. Meng, S. Zhong, Y. Gao, X. Cui, Recent advances in polysaccharide-based self-healing hydrogels for biomedical applications. Carbohydr. Polym. **283**, 119161 (2022). https://doi.org/10.1016/j.carbpol.2022.119161
9. I.A. Duceac, S. Coseri, Biopolymers and their derivatives: Key components of advanced biomedical technologies. Biotechnol. Adv. **61**, 108056 (2022). https://doi.org/10.1016/j.biotechadv.2022.108056
10. J. Singh, P.K. Dutta, Preparation, antibacterial and physicochemical behavior of chitosan/ofloxacin complexes. Int. J. Polym. Mater. Polym. Biomater. **59**(10), 793–807 (2010). https://doi.org/10.1080/00914037.2010.483219
11. I.A. Duceac, S. Coseri, Chitosan Schiff-Base Hydrogels—A Critical Perspective Review. Gels. **8**(12), 779 (2022). https://doi.org/10.3390/gels8120779
12. F. Zhang, S. Zhang, R. Lin, S. Cui, X. Jing, S. Coseri, Injectable multifunctional carboxymethyl chitosan/hyaluronic acid hydrogel for drug delivery systems. Int. J. Biol. Macromol. **249**, 125801 (2023). https://doi.org/10.1016/j.ijbiomac.2023.125801
13. L. Upadhyaya, J. Singh, V. Agarwal, R.P. Tewari, Biomedical applications of carboxymethyl chitosans. Carbohydr. Polym. **91**(1), 452–466 (2013). https://doi.org/10.1016/j.carbpol.2012.07.076
14. L. Upadhyaya, J. Singh, V. Agarwal, A.C. Pandey, S.P. Verma, P. Das, R.P. Tewari, Efficient water soluble nanostructured ZnO grafted O-carboxymethyl chitosan/curcumin-nanocomposite for cancer therapy. Process Biochem. **5**, 678–688 (2015). https://doi.org/10.1016/j.procbio.2014.12.029
15. H. Mirhosseini, B.T. Amid, A review study on chemical composition and molecular structure of newly plant gum exudates and seed gums. Food Res. Int. **46**(1), 387–398 (2012). https://doi.org/10.1016/j.foodres.2011.11.017
16. S. Henciya, P. Seturaman, A.R. James, Y.H. Tsai, R. Nikam, Y.C. Wu, H.U. Dahms, F.R. Chang, Biopharmaceutical potentials of Prosopis spp (Mimosaceae, Leguminosa). J. Food Drug Anal. (2017). https://doi.org/10.1016/j.jfda.2016.11.001
17. A.M.P. Torres, C.N.S. Dominguez, K.B. Bernabe, P.Y.C. Garcia, J.A.R. Perez, E.N.G. Treviño, H.G. Blanco, M.S. Dominguez, Green Synthesis of Mesquite-Gum-Stabilized Gold Nanoparticles for Biomedical Applications: Physicochemical Properties and Biocompatibility Assessment. Polymers (Basel). **15**(17), 3533 (2023). https://doi.org/10.3390/polym15173533
18. C. Rodriguez, V. Padilla, K. Lozano, A. Villarreal, L. Materon, R. Gilkerson, Development and Characterization of Forcespinning® Mesquite Gum Nanofibers. SSRN Electron. J. (2022). https://doi.org/10.2139/ssrn.4132959
19. R.R. Rodríguez, H.E. Andrews, N.M. Hernández, C.L. Calleros, E.J.V. Carter, Mesquite gum/chitosan insoluble complexes: relationship between the water state and viscoelastic properties. J. Dispers. Sci. Technol. **40**(89), 1345–1352 (2019). https://doi.org/10.1080/01932691.2018.1513848
20. K.H. Bae, L.-S. Wang, M. Kurisawa, Injectable biodegradable hydrogels: progress and challenges. J. Mater. Chem. B. **1**(40), 5371 (2013). https://doi.org/10.1039/c3tb20940g
21. M.S. Aghajanzadeh, R. Imani, M.H. Nazarpak, In situ forming aldehyde-modified xanthan/gelatin hydrogel for tissue engineering applications: synthesis, characterization, and optimization. J. Mater. Sci. **58**(35), 14187–14206 (2023). https://doi.org/10.1007/s10853-023-08878-6
22. M.B. Santos, C.H.C. dos Santos, M.G. de Carvalho, C.W.P. de Carvalho, E.E.G. Rojas, Physicochemical, thermal and rheological properties of synthesized carboxymethyl tara gum (Caesalpinia spinosa). Int. J. Biol. Macromol. **134**, 595–603 (2019). https://doi.org/10.1016/j.ijbiomac.2019.05.025
23. S. Sun, H. Cao, H. Su, T. Tan, Preparation and characterization of a novel injectable in situ cross-linked hydrogel. Polym. Bull. **62**, 699–711 (2009). https://doi.org/10.1007/s00289-009-0048-9
24. R. Chen, Y. Hao, S. Francesco, X. Mao, W.C. Huang, A chitosan-based antibacterial hydrogel with injectable and self-healing capabilities. Mar. Life Sci. Technol. **6**, 115–125 (2023). https://doi.org/10.1007/s42995-023-00211-z
25. S. Nair, N.S. Remya, S. Remya, P.D. Nair, A biodegradable in situ injectable hydrogel based on chitosan and oxidized hyaluronic acid for tissue engineering applications. Carbohydr. Polym. **85**(4), 838–844 (2011). https://doi.org/10.1016/j.carbpol.2011.04.004
26. A.H. Pandit, N. Mazumdar, K. Imtiyaz, M.M.A. Rizvi, S. Ahmad, Self-Healing and Injectable Hydrogels for Anticancer Drug Delivery: A Study with Multialdehyde Gum Arabic and Succinic Anhydride Chitosan. ACS Appl. Bio Mater. **3**(12), 8460–8470 (2020). https://doi.org/10.1021/acsabm.0c00835
27. M.B.M. Trejo, M.S. Domínguez, Mesquite gum as a novel reducing and stabilizing agent for modified tollens synthesis of highly concentrated ag nanoparticles. Materials (Basel). **9**(10), 817 (2016). https://doi.org/10.3390/ma9100817
28. F. Cheng, Y. Wu, H. Li, T. Yan, X. Wei, G. Wu, J. He, Y. Huang, Biodegradable N, O-carboxymethyl chitosan/oxidized regenerated cellulose composite gauze as a barrier for preventing postoperative adhesion. Carbohydr. Polym. **207**, 180–190 (2019). https://doi.org/10.1016/j.carbpol.2018.10.077
29. S. Yan, T. Wang, X. Li, Y. Jian, K. Zhang, G. Li, J. Yin, Fabrication of injectable hydrogels based on poly(l-glutamic acid) and chitosan. RSC Adv. **7**(28), 17005–17019 (2017). https://doi.org/10.1039/c7ra01864a
30. V. Mourya, N.N. Inamdara, A. Tiwari, Carboxymethyl Chitosan And Its Applications. Adv. Mater. Lett. **1**(1), 11–33 (2010). https://doi.org/10.5185/amlett.2010.3108
31. H. Guo, S. Huang, X. Yang, J. Wu, T.B. Kirk, J. Xu, A. Xu, W. Xue, Injectable and Self-Healing Hydrogels with Double-Dynamic Bond Tunable Mechanical, Gel-Sol Transition and Drug Delivery Properties for Promoting Periodontium Regeneration in Periodontitis. ACS Appl. Mater. Interfaces **13**(51), 61638–61652 (2021). https://doi.org/10.1021/acsami.1c18701
32. Y. Liang, X. Zhao, P.X. Ma, B. Guo, Y. Du, X. Han, pH-responsive injectable hydrogels with mucosal adhesiveness based on chitosan-grafted-dihydrocaffeic acid and oxidized pullulan for localized drug delivery. J. Colloid Interface Sci. **536**, 224–234 (2019). https://doi.org/10.1016/j.jcis.2018.10.056
33. J.N.I. Balitaan, C.D. Hsiao, J.-M. Yeh, K.S. Santiago, Innovation inspired by nature: Biocompatible self-healing injectable hydrogels based on modified-β-chitin for wound healing. Int. J. Biol. Macromol. **162**, 723–736 (2020). https://doi.org/10.1016/j.ijbiomac.2020.06.129
34. J. Qu, X. Zhao, P.X. Ma, B. Guo, pH-responsive self-healing injectable hydrogel based on N-carboxyethyl chitosan for hepatocellular carcinoma therapy. Acta Biomater. **58**, 168–180 (2017). https://doi.org/10.1016/j.actbio.2017.06.001
35. S. Wang, Y. Liu, Q. Sun, B. Zeng, C. Liu, L. Gong, H. Wu, L. Chen, M. Jin, J. Guo, Z. Gao, W. Huang, Triple Cross-linked Dynamic Responsive Hydrogel Loaded with Selenium Nanoparticles for Modulating the Inflammatory Microenvironment via PI3K/Akt/NF-κB and MAPK Signaling Pathways. Adv. Sci. **10**(31), 2303167 (2023). https://doi.org/10.1002/advs.202303167

36. J. Kumar, R. Purwar, Self-Healing, Biocompatible injectable hydrogel based on multialdehyde moringa oleifera gum and carboxymethyl chitosan: a suitable platform for drug delivery in wound healing application. ChemistrySelect (2024). https://doi.org/10.1002/slct.202400309
37. H. Yin, P. Song, X. Chen, Q. Huang, H. Huang, A self-healing hydrogel based on oxidized microcrystalline cellulose and carboxymethyl chitosan as wound dressing material. Int. J. Biol. Macromol. **221**, 1606–1617 (2022). https://doi.org/10.1016/j.ijbiomac.2022.09.060
38. H. Chen, J. Cheng, L. Ran, K. Yu, B. Lu, G. Lan, F. Dai, F. Lu, An injectable self-healing hydrogel with adhesive and antibacterial properties effectively promotes wound healing. Carbohydr. Polym. **201**, 522–531 (2018). https://doi.org/10.1016/j.carbpol.2018.08.090
39. J. Kumar, R. Purwar, *Schiff-based self-healing injectable hydrogel from oxidized Azadirachta indica and carboxymethyl chitosan for biomedical application* (Polym J, Iran, 2024)
40. J. Liu, C. Qi, K. Tao, J. Zhang, J. Zhang, L. Xu, X. Jiang, Y. Zhang, L. Huang, Q. Li, H. Xie, J. Gao, X. Shuai, G. Wang, Z. Wang, L. Wang, Sericin/Dextran Injectable Hydrogel as an Optically Trackable Drug Delivery System for Malignant Melanoma Treatment. ACS Appl. Mater. Interfaces **8**(10), 6411–6422 (2016). https://doi.org/10.1021/acsami.6b00959

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.

Authors and Affiliations

Jitendra Kumar[1] · Roli Purwar[1]

✉ Roli Purwar
roli.purwar@dtu.ac.in

[1] Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University, Delhi, India

Leaf Analysis for Disease Identification using Transfer learning

Rohit Beniwal
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
rohitbeniwal@dtu.ac.in

Sachin Sachin
Department of computer science and engineering
Delhi Technological University
Delhi, India
happy123gsmhappy@gmail.com

Satyam Satyam
Department of computer science and engineering
Delhi Technological University
Delhi, India
yadavsatyam384@gmail.com

Saroj Kumar
Department of computer science and engineering
Delhi Technological University
Delhi, India
sarojkumar33jmd@gmail.com

2024 International Conference on Circuit, Systems and Communication (ICCSC) | 979-8-3503-6530-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCSC62074.2024.10616838

***Abstract*—The escalating challenges in agricultural efficiency due to plant diseases require advanced solutions for early detection and administration. Conventional strategies, despite being viable to a degree, need the exactness, speed, and adaptability required to relieve these challenges effectively. This research presents a novel approach to plant disease identification leveraging the control of transfer learning. By focusing on basic crops such as wheat, rice, corn, and sugarcane, this study points to revolutionizing leaf investigation for disease location, advertising a critical jump toward precision agriculture. Utilizing a comprehensive dataset comprising pictures of solid and unhealthy clears out, we prepared and assessed models for each trim, accomplishing exceptional exactnesses: 98.33% for corn and 97.37% for wheat, and over 90% for rice and sugarcane. These come about to emphasize the adequacy of transfer learning in improving the model capacity to generalize over different plant infections, in this manner empowering more precise and dependable malady discovery. The suggestions of our discoveries amplify past scholastic intrigue, advertising down-to-earth arrangements for agriculturists and rural experts around the world, subsequently supporting economical rural homes in the confrontation of worldwide nourishment security challenges.**

***Keywords*—Escalating, Detection, Transfer Learning, Scholastic, Intrigue, Nourishment.**

I. INTRODUCTION

The purpose is to develop an automatic classification system that can differentiate normal and abnormal crops using leaf images. Focusing on corn, rice, wheat, and sugarcane because they contain the following common diseases among the crops enumerated above: Common Rust, Gray Leaf Spot, and Northern Leaf Blight for corn; Brown Spot, Leaf Blast, and Neck Blast for rice; Brown Rust and Yellow Rust for wheat; Red Rot and Bacterial Blight for sugarcane. The importance of this task is to have the power to transform the agricultural disease control system by becoming a pathological discovery tool for crop development, thereby allowing the highest quality and quantity per crop.

This task poses a number of challenges. One of them is that the appearance of diseases can vary altogether depending on the disease arrangement, natural components, and crop varieties, showing inconsistency in disease introduction. Furthermore, multiple illnesses display visually similar symptoms, complicating the process of distinguishing and increasing the chances of mislabeling. Ultimately, even the look of healthy leaves changes across the plant variety, and age and general factors add a layer of complexity to the classification process. The recommended action plan to address these issues must use a high-order image processing and machine learning process. Özden shows the novelty of transfer learning with pre-trained models and the Synthetic Minority Over-sampling Technique on the example of corn diseases to achieve satisfactory accuracy [1]. Verma et al. [2] reported that a CNN-based method that achieved a 97% accuracy rate in recognizing potato leaf diseases holds promise for lowering economic losses and boosting agricultural sustainability. Tomar and Chaurasia [3] analyzed intelligent methods for detecting sugarcane diseases with image processing and deep learning and discussed the advantages of automatic detection systems in enhancing crop productivity and recovering economic losses. Deep learning models, especially CNN, are suitable for this situation because they can eliminate and learn intricate patterns in images.

Transfer learning can help move forward with the model execution by utilizing pre-trained frameworks on broad datasets and altering them to the particular assignment of crop disease classification.

II. LITERATURE SURVEY

In agricultural technology, the transformation to the use of deep learning and machine learning for plant disease diagnosis represents a big change. The consequence has been improved accuracy and speed in identifying diseases that affect crops such as potatoes, which are important for human food. Overall, they all show how powerful CNNs can be in detecting diseases within agriculture, providing an optimistic step towards managing diseases with better efficiency than ever. Current progress in detecting crop diseases via deep learning techniques is remarkable, covering a wide range of crops.

In addition to the above research work, Shah et al. [4] in 2022 developed a deep learning-based web application to differentiate potato leaf diseases [4]. The result shows that with a MobileNet, the highest accuracy is achieved at

979-8-3503-6530-6/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:11:41 UTC from IEEE Xplore. Restrictions apply.

96.12%. With CNNs, Wasalwar et al. [5] reported a record of concern regarding the classification of potato leaf diseases with a precision of 99%. Another novel approach to identifying wheat disease was introduced by Cheng et al. [6] using position attention CNNs, a variant of ResNet that can accurately classify diseases with 96.4% precision. Bi et al. [7] presented an optimized model for the identification of corn leaf diseases, which has shown high accuracy and precision in the disease with their model. Militante et al. [8] demonstrated both the potential and performance capabilities of deep learning models toward disease detection using 96.5% across varied crop strains. Sugarcane disease detection by using a unique approach, CNNs, was presented by Upadhye et al. [9], and it became critical for precision farming and making appropriate decisions in the agricultural sector, as was previously demonstrated. Mangrule and Afreen [10] also created a three-stage approach for forecasting sugarcane disease. It was possible to achieve high accuracy in identifying the disease with various percentage validation rates by using the color and texture characteristics with the existing hybrid classification. Sugarcane foliar disease classification was excelled by Daphal and Koli [11], and their study paved the way for precision agriculture via transfer learning, with VGG-16 and ResNet models as plum ingredients.

III. RESEARCH APPROACH

Transfer learning is a technology that uses pre-trained neural network models to transfer knowledge between domains. It involves selecting a source model from a broad dataset, such as the Potato Leaves Disease that we build and the VGG16 model, and reusing it for new classification tasks. The process involves modifying the yield layers to suit the new task, adjusting the pre-trained network's weights, selecting which layers to fine-tune or freeze, and adding custom layers. Transfer learning can be used in practical applications to distinguish between healthy and infected crops based on leaf examination.

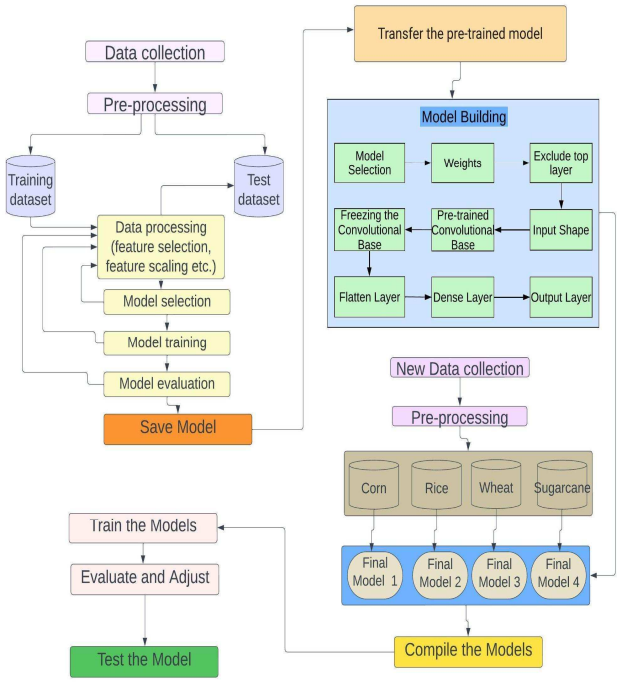


Fig. 1. Methodology

A. Dataset Collection

In our experimental setup, the dataset collection is organized into two specific stages, each centering on assorted goals and leveraging special sources of information. The pre-built Potato Leaves Disease Classification model centers exclusively on potato foliage, comprising 1,558 pictures separated into three categories concerning particular conditions: early blight, healthy foliage, and late blight [12]. The second phase broadens the scope to include a diverse array of crops, with 11,172 images across 14 classes [13]. Fig. 2 shows the summary of the dataset that is used in this experiment.

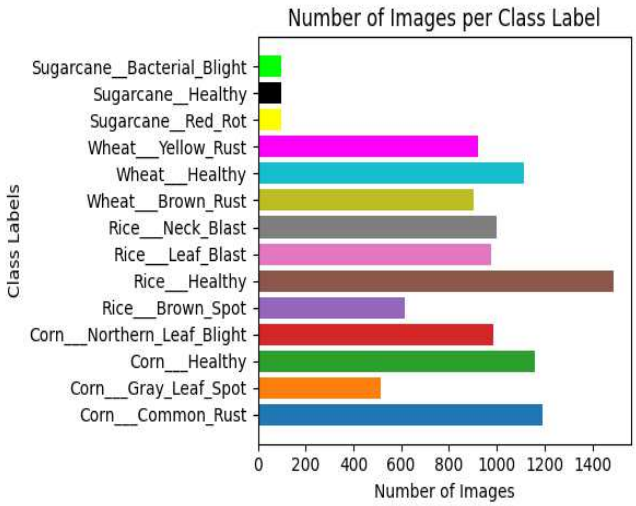


Fig. 2. Number of Images per Class Label

The dataset also contains images of potatoes, but we are not using it here as we already used it in the previous phase. This dataset is designed to train a more comprehensive model with transfer learning techniques to classify both healthy and infected plants across multiple species: corn, rice, wheat, and sugarcane.

B. Pre-built models

1) Potato Leaves Disease Classification MODEL

The potato leaves disease classification is done through variable deep learning using images, making a difference for the farmers to track the crop into maturity and fruition. Eraj and Uddin [14] developed CNN-based methods to classify early-stage potato diseases for seven classifiers with the Keras API with up to 94% accuracy. Bonik et al. [15] found 94.2 % accuracy on disease detection of a potato leaf through the CNN method; there has been a milestone in production to automate disease detection. Prasher et al. [16] achieved up to 97.6% accuracy on the CNN model using RMSProp, Adam, and a stochastic process gradient for potato leaf disease prediction. CNNs have been shown to detect and classify potato blight effectively, as evidenced by Kukreja et al. [17], who achieved a 94.77% accuracy in multi-classification.

The model that we used as a convolutional base for Plant Pathogen Detector A is a CNN tailored for image classification, focusing on potato disease detection. Model preprocessing layers include image resizing, scaling and data augmentation. Core architecture featuring convolutional and pooling layers for feature extraction that uses ReLU activation and a kernel size of (3, 3). 3D feature maps convert into 1D feature vectors through the help of flattening layers

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:11:41 UTC from IEEE Xplore. Restrictions apply.

with ReLU activation. It concludes with dense layers for classification, built using softmax activation to provide clear class probabilities. The model gives 99.15% accuracy in distinguishing healthy potato leaves from those with late and early blight.

2) VGG16 MODEL

The Plant Pathogen Detector B uses the VGG16 Model as a convolutional base. VGG16 is an image recognition model, a deep CNN with max-pooling layers after five blocks of thirteen convolutional layers. It employs 3x3 filters for ReLU activation that increase channels from 64 to 512. The model contains three fully connected layers; the first two are 4096 units each for feature analysis, while the third has 1000 units corresponding to ImageNet's classes. Softmax activation function on its last layer produces a probability distribution over 1000 classes. With approximately one hundred thirty-eight million parameters, VGG16's strength comes from being deep and having the ability to extract features despite its simplicity. This simplicity and well-organized usage of 3x3 convolutional layers at depths make VGG16 suitable for benchmarking advanced studies and recognition of images.

C. Pre-processing

Our models utilize different processing methods to improve dataset quality and model training speed. These techniques include changing the size of input images to be uniform; this ensures that our model converges quickly. Additionally, augmentation improves methods by flipping and rotating images at random, aiding in a model's generalization capacity while simultaneously preventing overfitting. In terms of computation time optimization for neural networks, data is divided into batches of size 32 and split into train sets (80%), validation sets (10%), and test sets (10%) for each crop. For performance improvements, caching, shuffling, and prefetching are used to streamline the data pipeline and ensure accelerated training. These processing techniques form part of what characterizes the optimized performance of a network, preparing it for proper generalization as far as unseen information is concerned through efficient model training.

D. Experimental setup

The above steps help to preprocess the dataset to facilitate effective and efficient model building, optimizing the neural network performance as far as generalizing from the training set to unseen data. So, we load our pre-built model and VGG16 model in this phase.

1) Import Pre-built model as convolutional base

The pre-built model is loaded using the load_model function from the Keras library, indicating the path to the trained model within the Google Colab environment. The model's architecture consists of sequential layers for preprocessing, convolutional layers for feature extraction, and dense layers for classification, with the final layer using softmax activation. To enable transfer learning, all the layers in the convolutional base except the last four are frozen, which retains their weights. For the classification task, custom layers were added, including a flattened layer and a dense layer with ReLU activation. The total parameters of 233,924 are approximately divided into trainable (140676) and non-trainable (93248) ones. The experimental setup aims to modify a pre-trained potato leaves disease classification model using transfer learning to adapt to a new dataset and task.

2) Import the VGG16 model as a convolutional base

The VGG16 model is a strong convolutional base for transfer learning that uses its pre-trained architecture and weights from ImageNet to extract features for new image classification tasks. Fig.3 shows the architecture of the model used in Plant Pathogen Detector B.

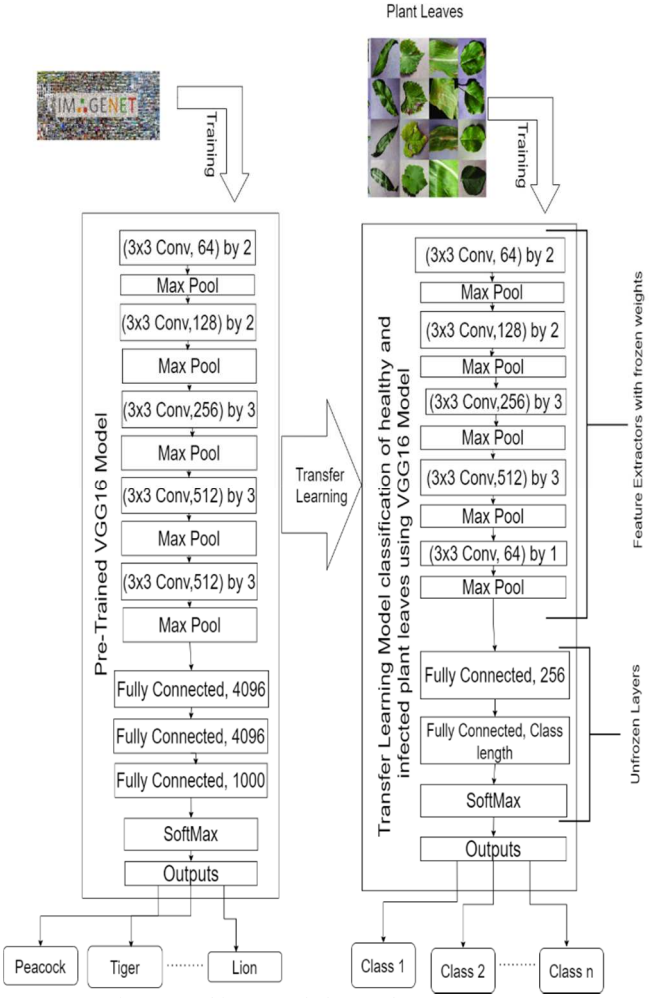


Fig. 3. Architecture of Plant-Pathogen Detector B

It makes use of an input shape of 256x256 pixel RGB images and does not include the top fully connected layers so it can adjust to process images of different sizes. To bring the VGG16 model into line with a new classification task, its convolutional base is frozen. There are customized layers added to the model for this specific classification task, which include a flattened layer that converts 2D feature maps into a 1D vector, a dense layer having ReLU activation function as memory storage of characteristics and an output layer having sigmoid activation function for binary or multi-label classifications. Thus, we now have a model with a total of 23,104,323 parameters. Out of them, 8,389,635 trainable ones were adapted to the fresh dataset, and another 14,714,688 were non-trainable ones, coming from the pre-trained VGG16 model. By combining these two types of parameters, customizable models aim at improving performance and efficiency when faced by specific image classification tasks.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:11:41 UTC from IEEE Xplore. Restrictions apply.

E. Compilation of the Model

Evaluation metrics, optimizer, and loss function are used in the compilation of the model. The Adam optimizer is a good choice for noisy or missing gradients because of its adjustable learning rate and high-dimensional space performance. The Sparse Categorical Cross entropy loss function is employed when doing multi-class classification tasks using integer labels. A straight proportion of valid predictions is used to evaluate how accurate the model performs. TensorFlow/Keras uses the fit() function to implement the training process, minimizing loss and avoiding overfitting.

IV. RESULT AND EVALUATION

The dataset has been trained on two models based on different convolution bases. The performance evaluation of these models involves a variety of metrics such as training and validation accuracy & loss graph, Accuracy, Precision, Recall, F1 Score, and ROC curve. The overall performance of each model with respect to the crop are as shown below.

A. Transfer Model built with Potato Leaves Disease Classification Model

The prepared model was tested on a different segment of dataset for each corresponding crop. Table I shows the performance of Plant Pathogen Detector A.

TABLE I. PERFORMANCE METRICS OF PLANT PATHOGEN DETECTOR A

MODEL	Accuracy	Precision	Recall	F1 Score	Loss
Corn	0.915	0.843	0.819	0.825	0.2453
Rice	0.812	0.762	0.705	0.715	0.539
Sugarcane	0.766	0.741	0.738	0.725	0.583
Wheat	0.901	0.885	0.873	0.871	0.2434

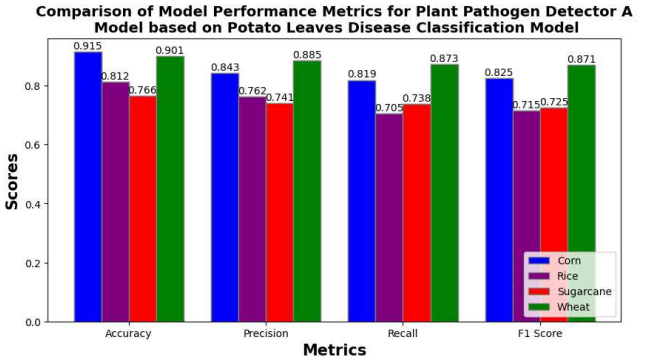


Fig. 4. Combined result of 4 crops with reference to Accuracy, Precision, Recall, F1 Score

Fig.4 shows the corn classification model demonstrates high accuracy and overall strong performance across all metrics. While accuracy is lower in rice, precision remains high, indicating reliable disease predictions. Sugarcane shows moderate accuracy, highest precision, but lower recall and F1 score, which tell missed instances of disease. The wheat model performs similarly to corn, with a lower precision but balanced precision and recall, as indicated by the F1 score.

Fig.5 shows corn graphs with fluctuations, which indicate that there are inconsistencies in the learning process; however, the loss graph demonstrates early rapid learning.

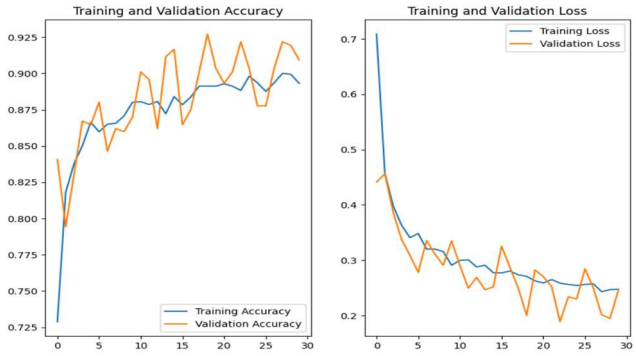


Fig. 5. Evaluation of Corn Model with accuracy and loss graph

Rice has an accuracy chart with less fluctuation as compared to corn, indicating steadier learning as shown in figure 6. The loss graph shows a steady decrease, signifying reliable learning.

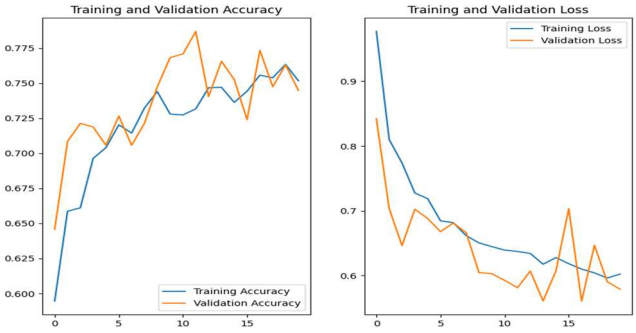


Fig. 6. Evaluation of Rice Model with accuracy and loss graph

Fig.7 shows that the sugarcane model needs further improvement compared to others in order to reduce errors.

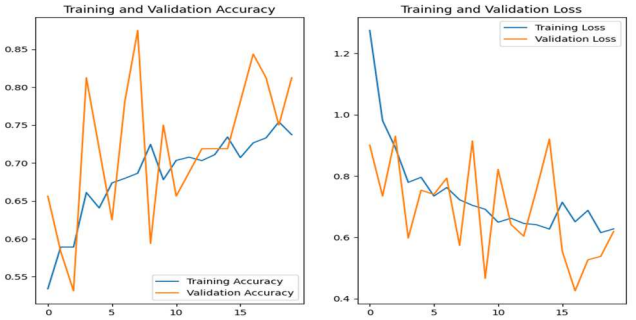


Fig. 7. Evaluation of Sugarcane Model with accuracy and loss graph

For wheat, an increase in performance can be seen from the accuracy graph in Fig.8, but it shows behavior that might indicate the sensitivity of certain instances for some data.

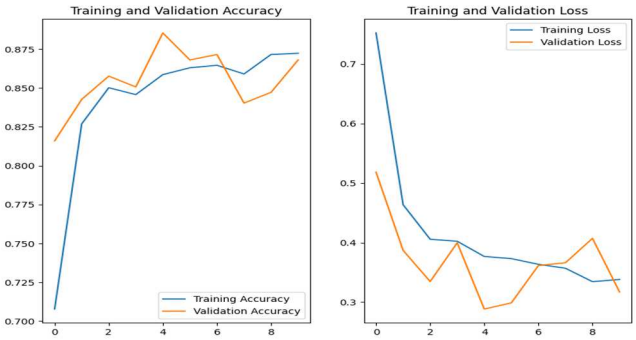


Fig. 8. Evaluation of Wheat Model with accuracy and loss graph

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:11:41 UTC from IEEE Xplore. Restrictions apply.

The overall performance of models can be concluded with the ROC curve shown in Fig.9. The corn model performs better than other models with high AOU values for Common Rust, Healthy leaves, Northern Leaf Blight and Gray Leaf Spot. Rice also has good prediction ability, with Neck Blast and Health Leaves, but Brown Spot and Leaf Blast are harder to identify, with AUCs of 0.88 and 0.87. Sugarcane disease detection is less effective but still fine for Healthy leaves. Bacterial Blight and Red Rot have AUCs of 0.85 and 0.79. Wheat disease detection is highly effective, particularly for Healthy leaves, Brown Rust, and Yellow Rust.

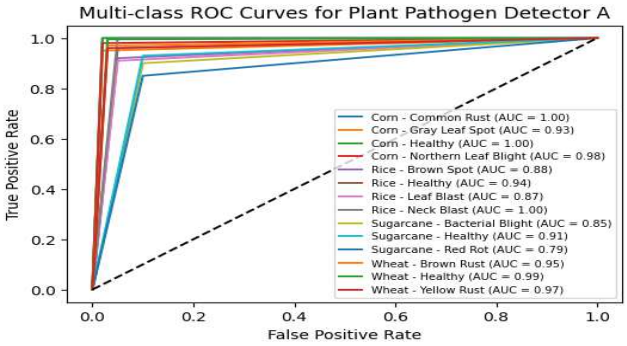


Fig. 9. Combine result for these 4 crops with reference to ROC Curve

B. Transfer Model build with VGG16 Model

The testing result of Plant Pathogen Detector B which is based on VGG16 is shown in Table II .

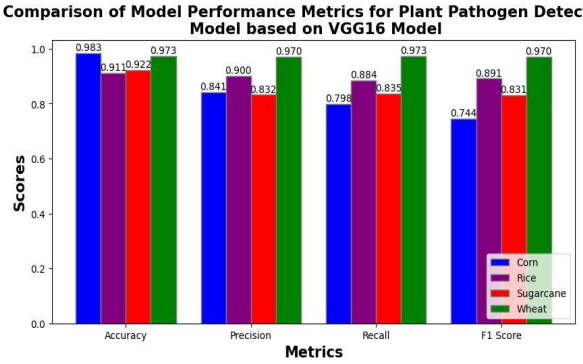


Fig. 10. Combine result of 4 crops with reference to Accuracy, Precision, Recall, F1 Score

The chart in Fig.10 indicates that the model performs best with rice across all metrics, with particularly strong scores in accuracy and recall. Corn has the lowest accuracy but a high recall score. Wheat and sugarcane have similar precision and F1 scores but differ in their accuracy and recall, with wheat having a significantly lower recall. These results show that it performs better than Plant Pathogen Detector A, the model with the Potato Leaves Disease Classification Model as the base for transfer learning model.

TABLE II. PERFORMANCE METRICS OF PLANT PATHOGEN DETECTOR B

MODEL	Accuracy	Precision	Recall	F1 Score	Loss
Corn	0.983	0.841	0.798	0.744	0.07
Rice	0.911	0.9	0.884	0.891	0.2987
Sugarcane	0.922	0.832	0.835	0.831	0.9525
Wheat	0.973	0.97	0.973	0.97	0.5209

Fig.11 shows the corn model demonstrates an almost perfect accuracy rate with the validation accuracy. The loss for both training and validation decreases sharply and then stabilizes, indicating good convergence without significant overfitting.

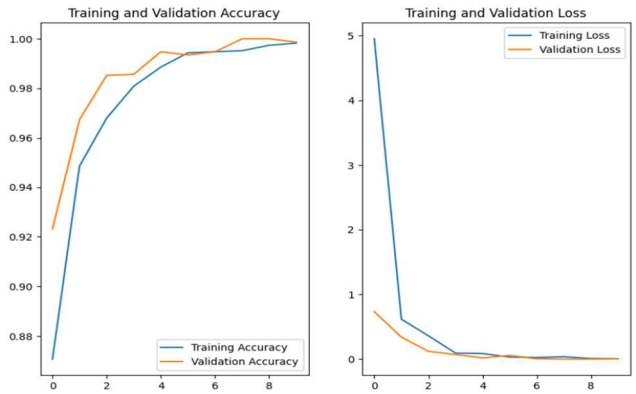


Fig. 11. Evaluation of Corn Model with accuracy and loss graph

Here in Fig.12 the training accuracy quickly ascends to around 95%, but the validation accuracy lags, suggesting overfitting for the rice model.

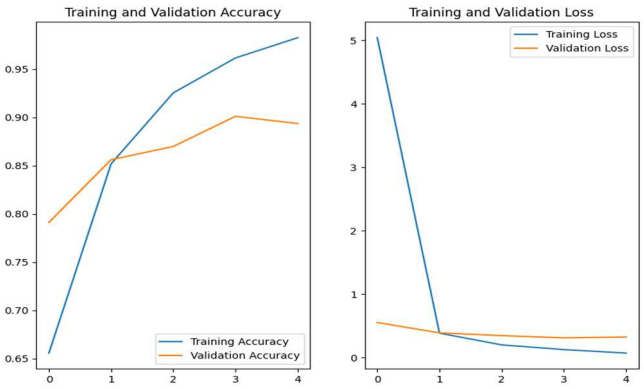


Fig. 12. Evaluation of Rice Model with accuracy and loss graph

The accuracy for sugarcane increases swiftly and reaches a peak then fluctuates as shown in Fig.13. The loss mirrors this pattern with an initial steep decline before reaching a plateau which suggests room for improvement.

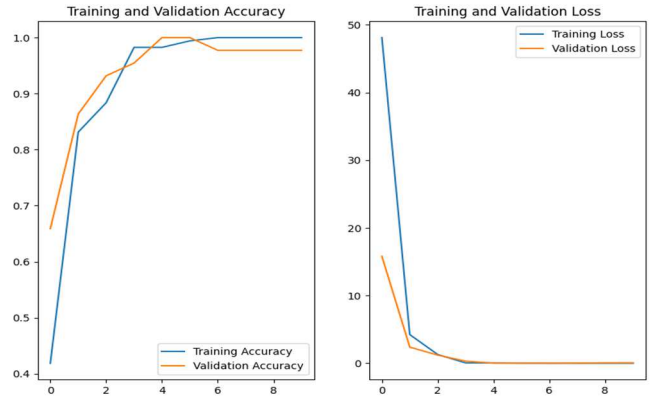


Fig. 13. Evaluation of sugarcane Model with accuracy and loss graph

The wheat model shows a sharp increase in accuracy, achieving near-perfect scores as given in Fig.14. The loss decreases significantly with training and validation loss converging, which indicates a well-fitting model.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:11:41 UTC from IEEE Xplore. Restrictions apply.

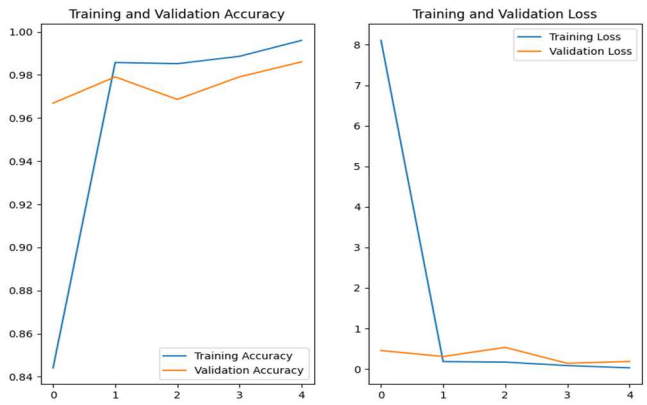


Fig. 14. Evaluation of Wheat Model with accuracy and loss graph

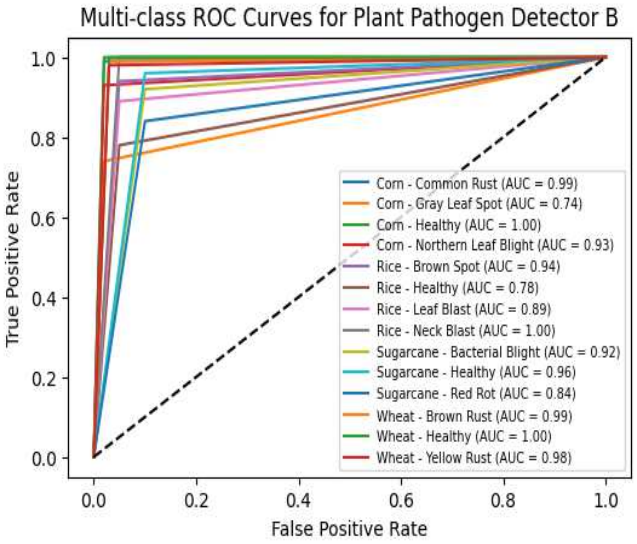


Fig. 15. Combined result for these 4 crops with reference to ROC Curve

The ROC chart of Plant-Pathogen Detector B is shown in Fig.15. Model is effective in identifying common rust and healthy leaves for corn, but struggles with gray leaf spot and northern leaf blight. Rice classification results show excellence in differentiating neck blast, leaf blast, and brown spot, but they are less effective in identifying healthy rice. Sugarcane detection is robust for bacterial blight and healthy conditions, with moderate scores for red rot. Wheat ROC curves are impressive, nearly perfect for brown rust and yellow rust, and perfect for identifying healthy wheat.

V. Conclusion and Future Scope

The VGG16 architecture and pre-built Potato Leaves Disease Classification model have been utilized as a transfer learning model for the classification of healthy and infected plant leaves over different species, including corn, rice, wheat and sugarcane. The evaluation proposes that the model performs exceptionally well in recognizing crop conditions. Plant Pathogen Detector B model shows accuracy higher than 90% for all four crops, whereas Plant Pathogen Detector A model shows accuracy above 90% for two crops and above 75% for other crops. This is mainly due to the difference in the dataset size of base models. Despite that fact, model A performs well in the classification of disease plants because it trains on similar disease classification problems. In conclusion, the "Leaf Analysis for Disease Identification using Transfer learning" project represents a critical step forward in the integration of artificial intelligence into rural home practices. As we look to the future, the scope of this venture can be extended to the incorporation of more crop species and diseases, subsequently broadening its effect. Continuous enhancements in this model accuracy and advancement of the user-friendly interfaces will improve its selection and utility in the field.

References

[1] Özden, C. (2023). Corn Disease Detection Using Transfer Learning. Black Sea Journal of Engineering and Science, 387-393

[2] Verma, R., Mishra, R., Gupta, P., & Trivedi, S. (2023, March). CNN based Leaves Disease Detection in Potato Plant. In 2023 6th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1-5). IEEE.

[3] Tomar, G., & Chaurasia, N. (2022, November). A Comprehensive Review: Sugarcane Disease Detection Using Intelligent Approaches. In 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-5). IEEE.

[4] Shah, H., Thakkar, H., & Dharmadhikari, S. C. (2022, December). Potato Leaf Disease Detection using Sequential Models. In 2022 5th International Conference on Advances in Science and Technology (ICAST) (pp. 297-301). IEEE.

[5] Wasalwar, Y. P., Bagga, K. S., Joshi, V. K., & Joshi, A. (2023, April). Potato Leaf Disease Classification using Convolutional Neural Networks. In 2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP) (pp. 1-5). IEEE.

[6] Cheng, S., Cheng, H., Yang, R., Zhou, J., Li, Z., Shi, B., ... & Ma, Q. (2023). A high performance wheat disease detection based on position information. Plants, 12(5), 1191.

[7] Bi, C., Xu, S., Hu, N., Zhang, S., Zhu, Z., & Yu, H. (2023). Identification method of corn leaf disease based on improved Mobilenetv3 model. Agronomy, 13(2), 300.

[8] Militante, S. V., Gerardo, B. D., & Dionisio, N. V. (2019, October). Plant leaf detection and disease recognition using deep learning. In 2019 IEEE Eurasia conference on IOT, communication and engineering (ECICE) (pp. 579-582). IEEE.

[9] Upadhye, S. A., Dhanvijay, M. R., & Patil, S. M. (2023). Sugarcane Disease Detection Using CNN-Deep Learning Method: An Indian Perspective

[10] Mangrule, R., & Afreen, K. R. (2023). Automated sugarcane crop disease forecasting with colour and texture features. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 1-15.

[11] Daphal, S. D., & Koli, S. M. (2021, November). Transfer Learning approach to Sugarcane Foliar disease Classification with state-of-the-art Sugarcane Database. In 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA) (pp. 1-4). IEEE.

[12] Basak, S. K., Tasnim, N., & Rahaman, M. A. (2023). Latest & Update Potato Leaf Diseases Dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/63476638

[13] Sharma, S. (2023). Five Crop Diseases Dataset [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/shubham2703/five-crop-diseases-dataset/data

[14] Eraj, S. U., & Uddin, M. N. (2023, June). Early stage Potato Disease Classification by analyzing Potato Plants using CNN. In 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM) (pp. 1-6). IEEE.

[15] Bonik, C. C., Akter, F., Rashid, M. H., & Sattar, A. (2023, January). A Convolutional Neural Network Based Potato Leaf Diseases Detection Using Sequential Model. In 2023 International Conference for Advancement in Technology (ICONAT) (pp. 1-6). IEEE.

[16] Prasher, S., Nelson, L., & Jagdish, M. (2023, May). Potato leaf disease prediction using RMSProp, Adam and SGD optimizers. In 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT) (pp. 343-347). IEEE.

[17] Kukreja, V., Baliyan, A., Salonki, V., & Kaushal, R. K. (2021, August). Potato blight: deep learning model for binary and multi-classification. In 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 967-672). IEEE.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:11:41 UTC from IEEE Xplore. Restrictions apply.

2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

MMSD-CAF: MultiModal Sarcasm Detection using CoAttention and Fusion Mechanisms

Jayant Miglani
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
jayantmiglani14@gmail.com

H A Pavan Kashyap
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
ha.pavan.kashyap@gmail.com

Rahul Katarya
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
rahuldtu@gmail.com

Madhav Maheshwari
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
madhavmaheshwari951@gmail.com

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10626994

***Abstract*—A lot of posts and comments across social media platforms including but not limited to Twitter, Instagram intend sarcasm. Sarcasm is a figurative language not just reliant on the text but also on various contextual cues. In our work herewith, we leverage the usage of multimodal cues in the form of textual and image data. The goal is to capture the relationship of sarcasm between the two modalities through deep learning techniques. We evolve our model over three iterations using different embedding representations for text and images while utilising different fusions of text embeddings from FLAVA text encoder and BERT model while image embeddings from FLAVA image encoder and VGG19 model. We leverage mechanisms of co-attention, multi-head attention and keyless attention to capture the contextual cues. Moreover, we utilise the publicly available MMSD2.0 dataset all across. The results are evaluated on the test data in terms of accuracy, precision, recall and F1-score. In the following paper, we show that our proposed novel model, hence named MMSD-CAF, using FLAVA text, image and multimodal learned embeddings along with co-attention and keyless attention-based fusion is able to produce better results than the established models.**

Keywords—Attention mechanism, BERT, Deep learning, FLAVA, Multimodal fusion, Sarcasm Detection

I. INTRODUCTION

Sarcasm is a parabolic language term that denotes total contrast between what is said and what is actually meant. Recently, due to the swift progress of social networking platforms, the domain of sarcasm detection has attracted increased research attention. Accurate sarcasm detection has numerous applications in areas like sentiment analysis, dialogue systems, online moderation and social media monitoring. While traditional approaches [1] have primarily focussed on purely textual data, sarcasm often employs multimodal cues [2], which makes it imperative to come up with effective models that can derive essence from multimodal data to form a robust contextual understanding. In the approach put forward and discussed in this paper, we focus squarely on textual and image data to propose a deep learning model based on effective incongruity determination.

Recent work on multimodal sarcasm detection utilises various state-of-the-art pre-trained multimodality models such as CLIP [3], BLIP [4], ALIGN [5] and FLAVA [6] used herewith. [7] shows the use of attention networks for multimodal sarcasm detection. We hereby conduct experiments on and probe across various exhaustive multimodal fusion techniques as well as different attention mechanisms such as co-attention, cross-attention and keyless attention to try to capture intricate information representations. By jointly modelling and fusing multimodal signals along with leveraging hybrid attention approaches, we aim to balance expressiveness and computational efficiency in capturing cross-modal interactions. The following are our major contributions:

- We propose a novel framework named: MMSD- CAF for multimodal sarcasm detection which makes use of embeddings from FLAVA text, image and multimodal encoders.
- We utilise co-attention and keyless attention-based multimodal fusion mechanisms to capture meaningful relationships between textual and visual modalities.
- Comparison of the results achieved by our model with previously established models in the scope of our work on the basis of classification metrics *viz.* accuracy, precision, recall and F1-score.

This is the flow that the rest of the manuscript will be framed in. In section II we go through an overview of and conduct a literature review of the past work. We then move on to explain the theoretical basis of modules utilised in section III. In section IV, we gradually define and optimise our proposed novel model over the course of 3 iterations, observing what change leads to what performance gain. Finally, in section V, we go on to quantitatively compare our model against published work with regards to performance on the same dataset.

II. RELATED WORK

Misra et al. [8] proposed a new dataset for sarcasm detection on text modality consisting of sarcastic news headlines written in a formal language and with less noise. It

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:57:03 UTC from IEEE Xplore. Restrictions apply.

proposed a hybrid model that involved fusion of embeddings obtained from CNNs and BiLSTM architectures with attention on top of it. Work done by Schifenella et al. [9] is one of the earliest publications on multimodal sarcasm detection that involves the fusion of textual and visual encodings. The dataset leveraged consisted of sarcastic tagged posts across various social media platforms such as Facebook, Instagram and Tumblr. The authors proposed a two-pronged approach to tackle the problem- the first method concatenated textual and semantic features extracted from training on an external dataset. The second method utilised pre-trained ImageNet parameters for visual feature extraction, alongside text vectors also being passed to the visual neural network. The results from the paper point towards a positive indication of the concept of fusion of text and image modalities. Tay et al. in [10] proposed a novel multi-dimensional approach focusing on intra-attention for recognising intermodal incongruity. In [11], Cai et al. proposed a benchmark dataset for multimodal sarcasm detection and a method that views data into 3 modalities namely- text, image and image attributes. They further put forward a hierarchical fusion based model that instead of simply concatenating the feature vectors, performs modality fusion inspired by attention mechanism that focuses on congruent parts of image and text for context extraction. The model obtains ResNet derived image embeddings, GloVe derived text embeddings and performs hierarchical fusion over the 3 modality branches that are considered and operated upon independently. This approach resulted in being much better optimised in training and more effective in better capturing the incongruity. Liang et al. [12] introduced a cross-modal graph convolutional network, which leverages cross-modal graph abstraction to represent relations between multimodal features. Their approach perceives the problem of sarcasm detection as the identification of incongruities between the modalities and explicitly drawing ironic relations. They utilise object detection to recognize crucial visual regions instead of considering the whole image. Then, they connect only the important visual regions to correlated textual tokens to capture incongruous sentiment in sarcastic expressions. Their model uses the BERT [14] base model for text encoding, the ViT(vision transformer) [15] for image encoding, connects the relevant parts of the features through cross-modal graph and finally tries to learn dependencies through multi-modal fusion on concatenated tensors, utilising graph convolutional networks with attention mechanism to capture incongruity. Liu et al. in [16], tackled the problem using multi-head cross attention mechanism and congruity dependent on graph attention networks. The authors encompassed a feature extraction layer through transformer based BERT for text and ViT for images followed by Atomic-Level and Composition-Level Cross-Modal Congruity and Knowledge Enhancement layers. [12], [16] succeeded in showing that graph based networks can better abstract the features and hence obtained improved results by accurately modelling the sarcasm relation among multiple modalities. Qin et al. in [17] proposed a novel dataset for multimodal sarcasm detection which we use in our model. This dataset builds upon the dataset put forward by [11], with enhancements including the removal of spurious cues and appropriate reannotation. They also propose a methodology using the pre-trained CLIP model to capture sarcasm from 3 different views namely image, text and image-text fusion view. The CLIP model, being contrastive by design, has the ability to highlight distinctions and similarities between different modalities within the same context, which turns out to be ideal for incongruity identification. The features are extracted using the CLIP text and image encoders. The image-text fusion view is fed into a transformer encoder followed by the application of keyless attention. The use of contrastive design followed by feature-fusion captures the subtle incongruities leading to the most improved results. We draw inspiration from [6], [17] and approach embedding fusion by progressively probing across effective methods, drawing down on the concept of keyless attention as described by [18]. Further, we leverage the learnings of [19], [22] by utilising a module of hierarchical co-attention ladder which is inspired by the human conduct of progressively reading and forming representations.

III. PRELIMINARY

This section describes the underlying concepts of various topics employed in the implementation and development of our model.

A. Attention Mechanisms

The attention mechanism allows a model to understand how different words in a sentence are related to each other. It captures the relation of a given word with the words on either side of it in the input sentence. We use multi-head attention and keyless attention in this scenario. Multi-head attention [20] operates on query (Q), key (K) and value (V) matrices which are the input embeddings (text or image). Their dimensions are (sequence length, d_{model}). d_{model} is the embedding length for each word. The query, key and value matrices undergo linear projections h times, where h is the number of heads. The linear projections are of dimensions d_k, d_k and d_v respectively. $d_k = d_v = d_{model}/h$. The attention function is applied on each of the projected versions, concatenated and again projected to obtain the final attention outputs. Scaled dot product attention followed by softmax activation as defined in (1) is used to compute the weights on values. Each W_i^j is a trainable weight matrix. The multi-head attention is computed as in (3).

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2)$$

$$Multihead\ Attention(Q,K,V) = concat(head_1,\dots,head_h)W^o \quad (3)$$

Keyless attention [18] computes attention weights using just the input embedding matrix without key and value matrices. The input representations a_i are projected into a 1-dimensional output using a trainable linear layer. Attention weight matrix for the input embedding representation is computed by applying the softmax function over the output of the linear layer as follows. The output attention vector c is computed in (6) with parameters- a_i being the embedding representation; w is the learnable parameter to compute the weight matrix of each a_i in (4); λ_i is the attention weight for each input representation as computed in (5).

$$e_i = w^T a_i \quad (4)$$

$$\lambda_i = \frac{exp(e_i)}{\sum_{j=1}^{n} exp(e_j)} \quad (5)$$

$$c = \sum_{i=1}^{n} \lambda_i a_i \quad (6)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:57:03 UTC from IEEE Xplore. Restrictions apply.

B. BERT

BERT [14] stands for Bidirectional Encoder Representations from Transformer. It utilises the encoder mechanism of the transformer which typically consists of multiple encoder layers stacked together. The input embedding is a concatenation of token embeddings, sentence embeddings and positional embeddings. It is a pre-trained model and uses two training strategies.

- MLM (Masked Language Modelling): 15% of the words in the sequences are masked and the MLM is tasked with the prediction of the words that are masked. The unmasked words serve as a context for predicting the actual value of the masked words. For the prediction of the masked words, the encoder is followed by a classification layer and the resultant vectors are multiplied by the embedding matrix. Then, the probability of each masked word is calculated using softmax activation.
- NSP (Next Sentence Prediction): Given two sentences A and B, the task here is to predict whether sentence B follows sentence A in the original sequence. A CLS token (denoting the start) and a SEP token (denoting the end) are added to each sentence. Sentence embedding and positional embedding are added to each of the tokenized words.

BERT can be fine-tuned for classification tasks by applying a classification layer that follows the transformer output.

C. VGG-19

The VGG-19 [13] architecture is a form of convolutional neural network that is characterised by the use of 3*3 convolutional filters, stacked with max pooling layers along with fully connected layers on flattened outputs. The network is majorly used to obtain feature maps of images for object detection, trained as feature extractor. Certain key highlights of its architecture can be given as:

- Depth : The VGG-19 has a total of 19 trainable layers which include 16 layers of convolution and 3 fully connected layers.
- Convolutional Layers : The convolutional layers use a 3x3 filter field with a stride of 1 and padding to preserve spatial resolution. The kernel convolves the volume of the image/feature map.
- Max- Pooling Layers : VGG-19 comprises of max pooling layers with with filters of size 2*2 and stride of 2 which are used to downsample the feature representation states.
- Fully Connected Layers : VGG-19 also consists of 3 fully connected layers, The first 2 layers consist of 4096 channels each, the third layer consists of 1000 channels which represents the number of output classes in the ImageNet dataset.
- Employs the ReLU activation throughout the network

D. FLAVA

The Foundational Language and Vision Alignment Model [6], [21] is a transformer based reasoning and correlation model designed for efficient multimodal learning across the 3 domains of vision, text and multimodal inputs. The FLAVA architecture comprises majorly of 3 transformer-based components- an image encoder which extracts unimodal image representations, a text encoder which extracts unimodal text representations and a multimodal encoder to get a fused representation of text and image with multimodal reasoning.

- Image encoder: It adopts the ViT- B/16 [15] architecture, where the input to the encoder is image split into patches and linearly embedded.
- Text encoder: A transformer encoder based on the exact same ViT [15] architecture as the image encoder is fed embedded and tokenized text tensors.
- Multimodal encoder: Here the text and image hidden states are fused using a separate transformer model. The text and image hidden states are their learned projections over each hidden state vector. The transformer architecture used here is also based on ViT [15] as the backbone. This allows for effective cross-attention between modal projections.

IV. Proposed Methodology

A. Data Preprocessing

We utilise the MMSD 2.0 dataset (Multi-Modal Sarcasm Detection) defined by the authors in [17]. The dataset consists of text files as well as an image directory. The text files consist of JSON objects with attributes of *image_id, text* and *label.* We create a data dictionary consisting of lists of textual data, raw images and labels for each instance. For the presented research, we sample the data from the 3 given files using the Bernoulli distribution sampling method. The split of total data is composed of: 80% contributing to training data, 10% to validation data and 10% to test data. For processing textual data, we leverage the transformer specific tokenization method, in this case, the *bert_base_cased* tokenizer. For processing images, we first resize them into (3*224*224) and then apply encoder specific normalisation techniques on the pixel values. These processed dictionary objects are further converted to tensors.

B. Model Development

Fig. 1 shows the block diagram of the proposed model architecture. Our model architecture development process proposed herewith builds upon two independently functioning mechanisms on top of text and image embeddings with the multi-modal encoder as the backbone, gradually over three iterations. One of the model branches is based on multi-head co-attention and the second branch is based on a combination of transformer encoder along with fusion using keyless attention as shown in Fig. 1. The number of attention heads in both the model branches is set to 4. ReLU activation function is used for the intermediate linear layers. The loss function and optimizer used are Binary Cross Entropy and Adam respectively.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:57:03 UTC from IEEE Xplore. Restrictions apply.

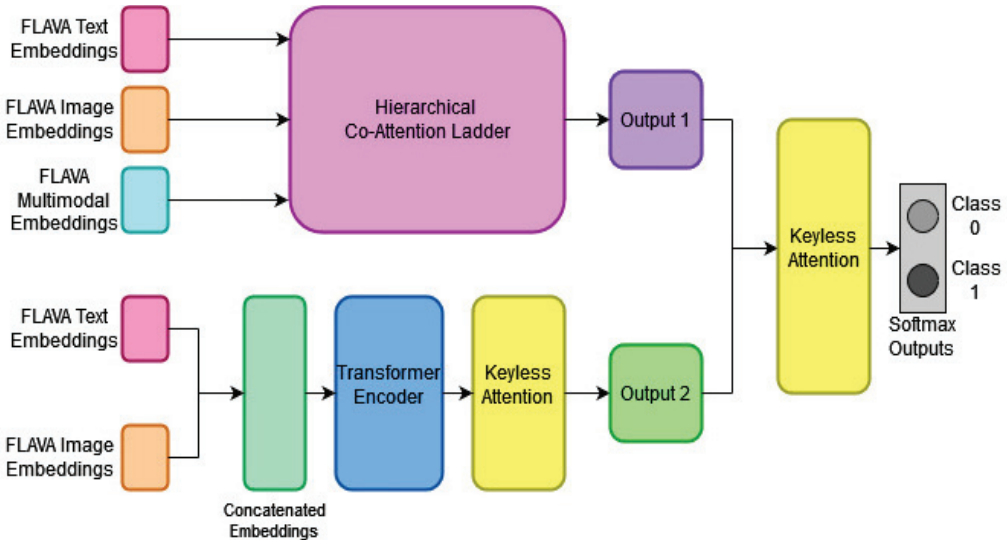


Fig. 1. Architecture of the Proposed Model MMSD-CAF

1) Co-attention layer: This layer is constructed by the parallel connection of two co-attention blocks inspired by [19]. Each co-attention block is based on multi-head attention mechanism as defined in (3) and shown in Fig. 2. It works in a way that, given two modalities A and B, for the first co-attention block, A makes up the query matrix and B makes up the key and value matrices while for the second co-attention block, B makes up the query and A makes up the key and value matrices. The outputs from the two parallel co-attention blocks are concatenated and linearly projected to obtain the fused representation. The inspiration behind using co-attention in our model lies in the fact that while reading up a piece of text-image pair, we tend to first read parts of the text, then extract relevant information from the image to create a fused representation in the mind and keep doing so successively while updating the representation, just as we would look at a tweet- image pair and determine if it is sarcastic or not.

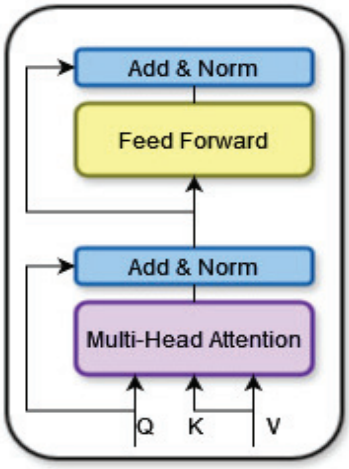


Fig. 2. Architecture of the Co-Attention Block

2) *Hierarchical Co-Attention Ladder:* The co-attention ladder, as shown in Fig. 4, is formed by hierarchically stacking 4 co-attention layers. Each layer results in the formation of a composite representation. This hierarchical stacking of co-attention layers allows an in-depth fusion of the multimodal features of text and visual inputs. The process of fusion is progressive, that is, the output of the current co-attention layer serves as the input to the co-attention layer that immediately follows it. Let us represent the three modalities of textual encodings, visual encodings and multimodal encodings as A, B and C respectively. In the first co-attention layer, we fuse the modalities A and B to obtain a representation Co-Embed1. In the second co-attention layer, the representation Co-Embed1 is fused with modality B to obtain the representation Co-Embed2. Further, in the third progressive layer, the fusion of the representation Co-Embed2 and modality C results in the representation Co-Embed3. In the fourth co-attention layer, the final layer in the hierarchy, we fuse the representation Co-Embed3 with modality C. The representation Co-Embed4 obtained after the fourth co-attention layer is the resultant output of the hierarchical co-attention ladder. The idea behind using such a hierarchical stacking of co-attention blocks lies in the fact that we are able to capture the combined contexts progressively from each modality. Moreover, it facilitates the storing of context states from previous iterations within the hierarchy while capturing newer context in the current state.

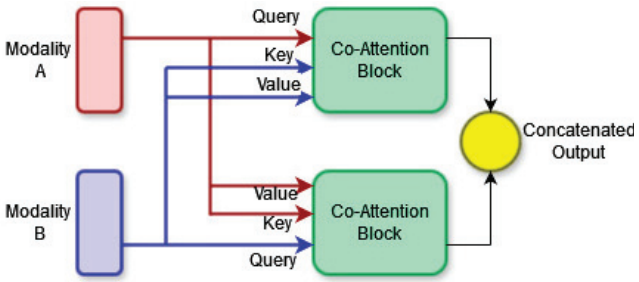


Fig. 3. Co-Attention Layer used in the Iterations

2) Keyless Attention: We employ keyless attention on the concatenated representation of textual and visual embeddings. Keyless attention is computationally less expensive as it computes attention scores in a linear time complexity, thus, within our computational constraints. The underlying method to compute attention score is the dot product method, same as in self attention, but without the use of a key matrix as shown in Fig. 2 and explained in (6).

The intuition behind using the two branches of co-attention layers and keyless attention respectively, is that co-attention layers help to capture the cross-modal

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:57:03 UTC from IEEE Xplore. Restrictions apply.

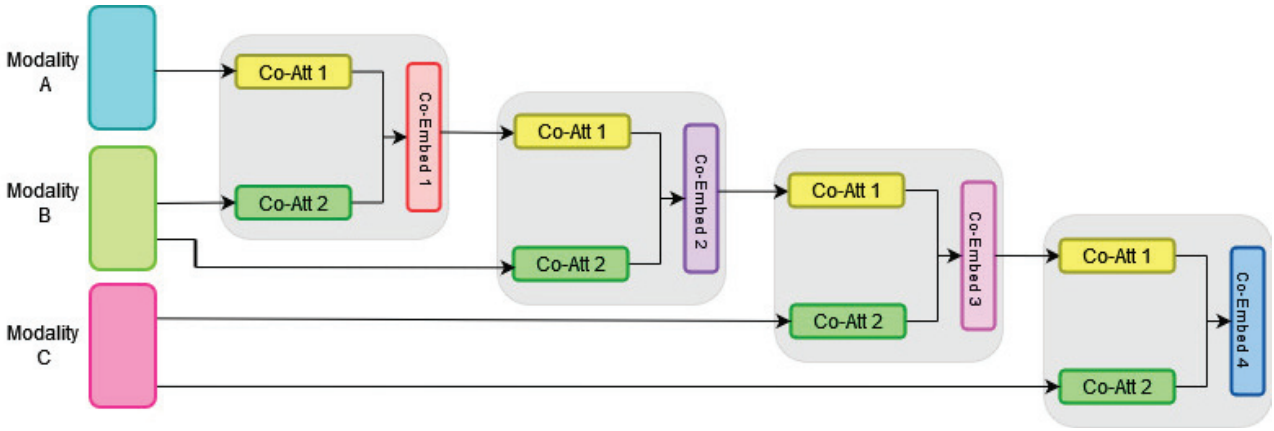


Fig. 4. Architecture of the Hierarchical Co-Attention Ladder

dependencies as embeddings of one modality attend to the embeddings of the other modality. Keyless attention helps the model to understand intra-modal dependencies by training weights for each token or image patch in the input sequence. Thus, the fusion of the two branches helps us effectively model the incongruity between both textual and visual modalities.

In the initial phase of the development process, for the first branch, we obtain text embeddings through BERT model's pooled outputs of size (*batch_size*,768) and image embeddings through pre-trained VGG-19 model of size (*batch_size*,4096). Further, both outputs are linearly projected to a size (*batch_size*,256). These outputs are then similarly fed into a single co-attention layer. For the second branch, we utilise the last hidden state outputs of FLAVA's text encoder of size (*batch_size*,128,768) and image encoder of size (*batch_size*,197,768). We further apply linear projections to convert them into sizes (*batch_size*, 128, 256) and (*batch_size*,197,256) respectively. Further, these representations are concatenated along dim=1, and fed into the transformer encoder. We then apply keyless attention onto the yielded representation and further, text and image pooled representations are added to the output to obtain the result from this branch. The second branch's output obtained is of size (*batch_size*,256), same as that of the first branch. The output from both the branches is fused using a keyless attention module and further passed through a softmax activation layer to obtain class probabilities.

In the second phase, the first branch inputs remain the same as the first iteration. However, these inputs are fed into the hierarchical co-attention module as defined before, to give an output of size (batch_*size*,256). For the second branch, the inputs are now changed to FLAVA text and image encoder's pooled outputs instead of last hidden states. These inputs are concatenated and fed into multi-head attention module. The resultant output from this branch has size (*batch_size*,256). We then apply softmax activations on the outputs of both the branches independently and then sum up the class probabilities obtained for each class.

Finally, we put forward our proposed model **MMSD-CAF (MultiModal Sarcasm Detection using CoAttention and Fusion)** in the final phase of the development process. Here, we throughout utilise image and text embeddings

obtained from the FLAVA model. The first branch is fed pooled output from *text_encoder*, *image_encoder* and *multimodal_encoder*. Each pooled output of size (*batch_size*,768) is linearly projected into a space of (*batch_size*,256). Now the three projected outputs are fed pairwise into a 4 layer hierarchical co-attention ladder, giving an output of size (*batch_size*,256). Similar to the first

iteration, the second branch produces outputs of dimensions (*batch_size*,256). The results from both the branches are fused using keyless attention and output probabilities are calculated using softmax activation.

V. Results and Discussions

A. Dataset

The sample employed for training has been taken from the MMSD2.0 dataset as defined in [17]. The class distribution in the original MMSD2.0 dataset is described in Table I. To overcome our computational constraints, we train our model each time on a 50% subset of the total data. The subsets are sampled from the Bernoulli Distribution.

TABLE I. MMSD2.0 Class Distribution

MMSD2.0	***Train***	***Validation***	***Test***
Instances	19816	2410	2409
Sarcastic	9572	1042	1037
Non-Sarcastic	10240	1368	1372

B. Evaluated Results

After completing the training process over the given model, the trained weight pickle file is put to test. The model performance on the test data is quantitatively measured on the metrics *viz.* accuracy, precision, recall and F1-score. These metrics are further used to examine the delta caused by different modules used. The results of the testing process are hence compared to each other to record differences in performance with different fusion techniques as well as with other published models on the MMSD2.0 dataset.

C. Experimental Results

The performance enhancements over the gradual development of the three iterations is indicated in Table II in terms of accuracy, precision, recall and F1-score. As can be seen from the table, changing the transformer encoder layer in the second branch to multi-head attention on the model in iteration 2, the F1-score increases by 3.93% and the accuracy increases by 4.24% in iteration 2. Further, changing the text

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:57:03 UTC from IEEE Xplore. Restrictions apply.

and image embeddings from BERT model to FLAVA text encoder and VGG19 model to FLAVA image encoder respectively in both the branches on the model in iteration 3, the F1-score increases by 5.95% and the accuracy increases by 6.32% in iteration 3. From the results obtained, it can be inferred that our novel model **(MMSD-CAF)** developed in iteration 3, gives the best results in terms of accuracy, precision, recall and F1-score.

TABLE II. PERFORMANCE COMPARISON

Model	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*
Cai et al. [11]	70.57	64.84	69.05	66.88
Model iter. 1	74.24	75.97	75.82	74.34
Liu et al. [16]	76.50	73.48	71.07	72.25
Model iter. 2	78.48	78.17	78.56	78.27
Liang et al. [12]	79.83	75.82	78.01	76.90
Qin et al. [17]	85.64	80.33	88.24	84.10
MMSD-CAF(*Ours*)	84.80	85.28	83.74	84.22

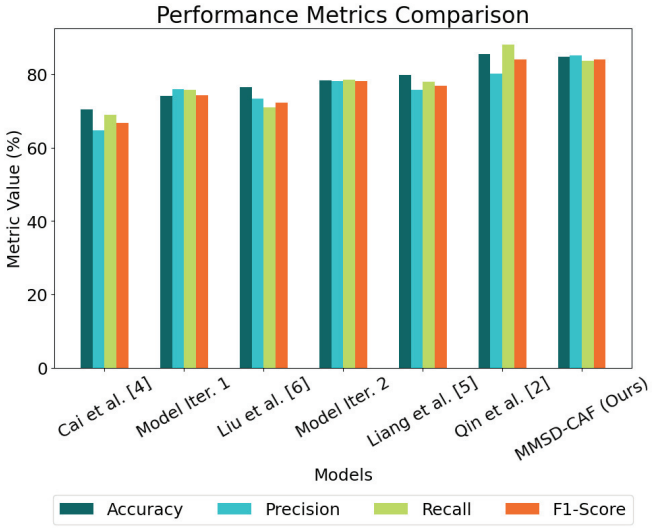


Fig. 5. Performance Metrics Comparison of the Models

B. Comparison with other Works

1) Hierarchical Fusion based Model: The hierarchical fusion model for multimodal sarcasm detection in [11] documents an accuracy of 70.57% and F1-score of 66.88%. We observe that the progressive increments in accuracy across the three model iterations over this model are 3.67%, 7.91% and 14.23% respectively. Similarly, the progressive increments in F1-score across our iterations over this model are 7.46%, 11.39% and 17.34% respectively.

2) Hierarchical Congruity and Knowledge Enhancement Model: The work in [16] reports an accuracy of 76.50% and F1-score of 72.25%. For our model in iteration 2, the accuracy is slightly higher by 1.98%, and also our model achieves a positive delta of 6.02% in terms of F1-score. However, our proposed model MMSD-CAF shows an improvement in both accuracy as well as F1-score by a margin of 8.3% and a significant 11.97% respectively as compared to the said model.

3) Cross-modal Graph Convolutional Network based Model: The model proposed by [12] displays an accuracy of 79.83% and F1-score of 76.90%. We achieve a higher accuracy on our proposed novel model by 4.97% while also bettering the F1-score by 7.32% and outperforming on precision and recall metrics.

4) Multi view CLIP model: The benchmark model provided in [17], achieves an accuracy of 85.64% and F1-score of 84.10%. It can be observed that our proposed model is able to achieve a comparable accuracy and is even able to do better on the F1-score metric by a marginal difference.

VI. CONCLUSION AND FUTURE SCOPE

In the work presented in this paper we progressively build and define a novel framework named **MMSD-CAF** for capturing cross-modal incongruity between textual and visual features. We examine the changes brought by varying the input embeddings, attention structure and embedding fusion methods through iterative enhancements in the qualitative and quantitative understanding of the model. Finally, we present a novel method named MMSD-CAF wherein we train two different embeddings, the first one being a co-attention ladder output of 3 given inputs- multi-modal FLAVA model's text, image and multimodal embeddings respectively. For the second embedding, we concatenate FLAVA's text and image embeddings and pass them into the transformer encoder. Keyless attention is then applied to the text and image representations yielded from the transformer encoder. The fusion of co-attention and keyless attention outputs helps us efficiently capture both cross-modal and intra-modal dependencies, hence, improving upon the previous works. Our proposed novel model achieves an accuracy of 84.80% which comes close to the established baseline model, while bettering it on the F1-score. The model can be useful in contributing to the goal of sarcasm detection on social media platforms and in sentiment analysis tasks. We believe this work can be extended in future to incorporate other modalities including but not limited to emoticons, audio and video cues.

REFERENCES

[1] A. B. Meriem, L. Hlaoua, and L. B. Romdhane, "A fuzzy approach for sarcasm detection in social networks", in *Procedia Computer Science*, vol. 192, pp. 602–611, 2021, doi: 10.1016/j.procs.2021.08.062.

[2] A. Bhat and A. Chauhan, "MultiModal Sarcasm Detection: A Survey," in *2022 IEEE Delhi Section Conference (DELCON)*, New Delhi, India, 2022, pp. 1-7, doi: 10.1109/DELCON54057.2022.9753058.

[3] A. Radford *et al.*, "Learning Transferable Visual Models From Natural Language Supervision," *arXiv:2103.00020*, Feb. 2021.

[4] J. Li, D. Li, C. Xiong, and S. Hoi, "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation," *arXiv:2201.12086*, Feb. 2022.

[5] C. Jia *et al.*, "Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision," *arXiv:2102.05918*, Jun. 2021.

[6] A. Singh *et al.*, "FLAVA: A Foundational Language And Vision Alignment Model," *arXiv:2112.04482v3*, Mar. 29, 2022.

[7] L. Song, Z. Zhao, Y. Ma, Y. Liu and J. Li, "Global-Aware Attention Network for Multi-modal Sarcasm Detection," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Honolulu, Oahu, HI, USA, 2023, pp. 2409-2414, doi: 10.1109/SMC53992.2023.10394483.

[8] R. Misra and P. Arora, "Sarcasm Detection using Hybrid Neural Network," *arXiv:1908.07414v2*, Oct. 2022.

[9] R. Schifanella, P. de Juan, J. Tetreault, and L. Cao, "Detecting sarcasm in multimodal social platforms," in *Proceedings of the 24th ACM international conference on Multimedia,* 2016, pp. 1136–1145, doi: 10.1145/2964284.2964321.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:57:03 UTC from IEEE Xplore. Restrictions apply.

[10] Y. Tay, A. T. Luu, S. C. Hui, and J. Su, “Reasoning with sarcasm by reading in-between,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1010–1020, doi: 10.18653/v1/P18-1093.

[11] Y. Cai, H. Cai, and X. Wan, “Multi-modal sarcasm detection in twitter with hierarchical fusion model,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 2506–2515, doi: 10.18653/v1/P19-1239.

[12] B. Liang, C. Lou, X. Li, M. Yang, L. Gui, Y. He, W. Pei, and R. Xu, “Multi-modal sarcasm detection via cross-modal graph convolutional network,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 1767–1777, doi: 10.18653/v1/2022.acl-long.124

[13] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:*1409.1556, Apr. 10, 2015.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “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)*, 2019, pp. 4171–4186.

[15] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *arXiv:2010.11929 [cs]*, Oct. 2020.

[16] H. Liu, W. Wang, and H. Li, “Towards Multi-Modal Sarcasm Detection via Hierarchical Congruity Modeling with Knowledge Enhancement,” in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing,* 2022, pp. 4995–5006, doi: 10.18653/v1/2022.emnlp-main.333.

[17] L. Qin *et al.*, “MMSD2.0: Towards a Reliable Multi-modal Sarcasm Detection System,” in *Findings of the Association for Computational Linguistics,* 2023, pp. 10834–10845, doi: 10.18653/v1/2023.findings-acl.689.

[18] X. Long *et al.*, “Multimodal Keyless Attention Fusion for Video Classification,” vol. 32, no. 1, Apr. 2018.

[19] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, “Multimodal Fusion with Co-Attention Networks for Fake News Detection,” in *Findings of the Association for Computational Linguistics: ACL-IJCNLP,* 2021, pp. 2560–2569, doi: 10.18653/v1/2021.findings-acl.226.

[20] A. Vaswani *et al.*, “Attention Is All You Need,” *arXiv:*1706.03762v7, Jun. 12, 2017.

[21] Mozhdeh Gheini, X. Ren, and J. May, “Cross-Attention is All You Need: Adapting Pretrained Transformers for Machine Translation,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 1754–1765, doi: 10.18653/v1/2021.emnlp-main.132.

[22] A. Bhat and A. Chauhan, "A Deep Learning based approach for MultiModal Sarcasm Detection," in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India, 2022, pp. 2523-2528, doi: 10.1109/ICAC3N56670.2022.10074506.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:57:03 UTC from IEEE Xplore. Restrictions apply.

Modelling Visual Semantics via Image Captioning to extract Enhanced Multi-Level Cross-Modal Semantic Incongruity Representation with Attention for Multimodal Sarcasm Detection

Sajal Aggarwal, Ananya Pandey, Dinesh Kumar Vishwakarma*
Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Bawana Road, Delhi-110042, India
sajalaggarwal_it20b11_08@dtu.ac.in , ananyapandey_2k21phdit08@dtu.ac.in, dinesh@dtu.ac.in*

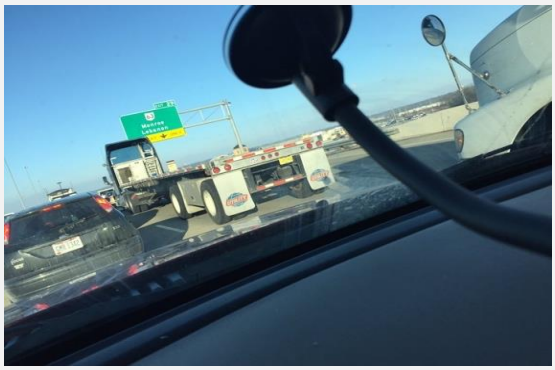
Abstract
Sarcasm is a type of irony, characterized by an inherent mismatch between the literal interpretation and the intended connotation. Though sarcasm detection in text has been extensively studied, there are situations in which textual input alone might be insufficient to perceive sarcasm. The inclusion of additional contextual cues, such as images, is essential to recognize sarcasm in social media data effectively. This study presents a novel framework for multimodal sarcasm detection that can process input triplets. Two components of these triplets comprise the input text and its associated image, as provided in the datasets. Additionally, a supplementary modality is introduced in the form of descriptive image captions. The motivation behind incorporating this visual semantic representation is to more accurately capture the discrepancies between the textual and visual content, which are fundamental to the sarcasm detection task. The primary contributions of this study are: (1) a robust textual feature extraction branch that utilizes a cross-lingual language model; (2) a visual feature extraction branch that incorporates a self-regulated residual ConvNet integrated with a lightweight spatially aware attention module; (3) an additional modality in the form of image captions generated using an encoder-decoder architecture capable of reading text embedded in images; (4) distinct attention modules to effectively identify the incongruities between the text and two levels of image representations; (5) multi-level cross-domain semantic incongruity representation achieved through feature fusion. Compared with cutting-edge baselines, the proposed model achieves the best accuracy of 92.89% and 64.48%, respectively, on the Twitter multimodal sarcasm and MultiBully datasets.
Keywords- *Multimodal Sarcasm Detection, Multimodal learning, Image captioning, Transformers, Multi-head self-attention*

1 Introduction

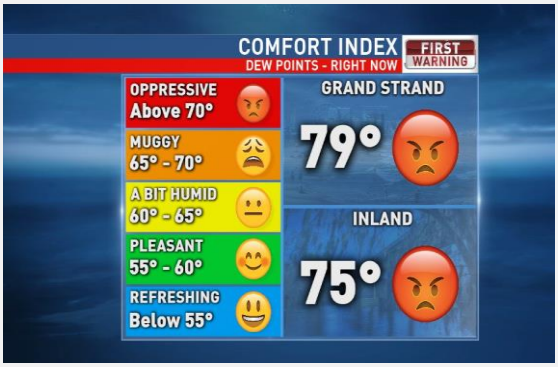
Sarcasm is a kind of wit used to poke fun at or make disparaging remarks about a person or situation. The literal interpretation of sarcastic remarks frequently suggests the polar opposite of their intended meaning. This makes sarcasm detection a somewhat onerous task. For instance, “Oh, I love it when people chew loudly. It’s such a pleasant sound.” is a remark wherein words indicative of a positive sentiment, namely ‘love’ and ‘pleasant’, are used to express an unfavourable opinion for the ‘unpleasant’ behaviour of ‘chewing loudly’. Sarcasm can be communicated by an amalgamation of verbal and nonverbal cues, including but not limited to alterations in vocal tone, exaggeration of a particular word, elongation of a syllable, or a neutral facial expression while delivering a seemingly positive remark. Traditional methods for detecting sarcasm have primarily relied on analyzing either textual [1], [2] or audio inputs [3] alone. However, the fact that sarcasm is fundamentally characterized by linguistic and contextual incompatibility has necessitated the shift to multimodal datasets [4]–[7] so as to capture meaningful information from multiple modalities and contextual history.

*Corresponding Author
1 Delhi Technological University, Bawana Road, Delhi-110042, India

The sarcasm in the tweet “I love commuting”, for example, is lost if the accompanying photograph of a congested traffic situation is disregarded. The satirical tone of the tweet can only be fully appreciated when both text and image are taken into account. Ignoring the visual cues could lead one to mistakenly interpret the comment as sincere, missing the satirical undertones entirely. **Figure 1** provides additional clarification on the necessity of incorporating multimodal inputs in order to achieve accurate sarcasm detection. This is demonstrated through the inclusion of instances from two benchmark datasets. Hence, to improve the complexity and accuracy of an automated system for detecting sarcasm, it is crucial to go beyond the limitations of textual data and incorporate additional data from different modalities that capture a wide range of information, enabling the identification of sarcasm.



I love commuting



Could the air get any more refreshing today?

(A)



Thank you for driving carefully



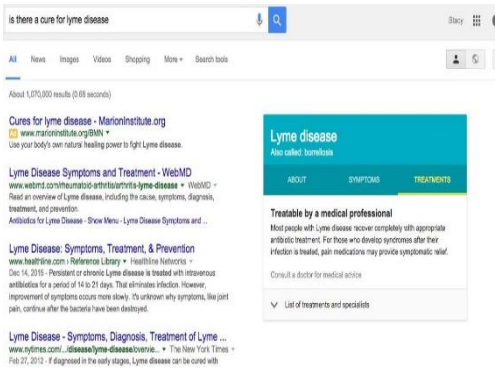
Bruce jenner’s cat

(B)

Figure 1 Examples from the Twitter multimodal sarcasm (A) and MultiBully (B) dataset demonstrate the limitations of relying solely on textual data to accurately convey sarcastic intent in certain instances.

Studies have been conducted to extract image attributes [6] or adjective-noun pairs (ANPs) [8] from images for multimodal sarcasm detection in image-text pairs. The objective of these studies is to incorporate an additional modality to establish a correlation between textual and visual information, thereby enhancing the understanding of visual semantics. However, the incorporation of such elements in the analysis can, at times, lead the sarcasm detection model astray. This occurs when these elements generate attributes or objects that fail to capture the fundamental essence of the visual semantics. In other words, they may refer to objects or attributes that do not effectively support the sarcasm detection task or highlight the inherent incongruity between the visual and the text. As a result, this constraint has the capacity to impede the efficacy of sarcasm detection. In order to upgrade the accuracy and coherence of image descriptions and to bring out the disconnect between textual and visual content, we introduce a novel framework that utilises a pre-trained encoder-decoder framework to generate

detailed image captions. These captions serve as an additional modality for detecting sarcasm in multimodal contexts. **Figure 2** serves to enhance the clarity of differentiating between the utilisation of image attributes and image captions. It accomplishes this by presenting examples that demonstrate how visual attributes, in contrast to captions, fail to adequately capture significant information pertaining to the image, thereby hindering the detection of sarcasm.



Text: Please tell me more about the magic cure you found for # lymedisease

Attributes: white, scene, sign, book, screen

Caption: A google search page shows a website with Lyme disease

Sarcasm

Text: Chocolate chip cookies

Attributes: dark, brown, nuts, black, food

Caption: A stack of chocolate chip cookies with a chocolate drizzle



Non-Sarcasm

Figure 2 In contrast to attributes, image captions in the above examples can improve visual semantics identification for sarcasm detection. The image captioning module used in this study has strong reading comprehension, allowing it to produce accurate descriptions even for images with text.

Major Contributions

The major contributions of the present study are laid out as follows:

1. A cross-lingual language model is employed as the textual tokenizer and encoder to acquire the feature representations for both the text data and the image captions. The efficacy of this model is notable as it has undergone extensive pretraining on a wide array of data encompassing more than 100 languages, including Hindi. This is particularly advantageous as the MultiBully dataset consists of code-mixed English-Hindi statements.
2. In contrast to the majority of prior research studies [6], [9]–[14], which typically utilize a variant of the ResNet architecture for generating image feature representations, the proposed framework employs a more robust self-regulated residual ConvNet [15] architecture to encode images into feature vectors. In order to enhance the obtained feature maps and prioritize the most important regions in the image, we incorporate a lightweight spatially aware attention module into our image feature extraction branch.

3. As opposed to other studies [6], [8], [14], which utilize either image attributes or objects as an additional modality, our proposed approach offers a more efficient method by integrating descriptive captions generated through a generative image-to-text transformer model. This integration allows for the incorporation of an additional source that emphasizes the discrepancy between modalities, ultimately enhancing the accuracy of sarcasm detection.
4. To encapsulate the contextual incongruity between the input text and the high-level image description generated through captions, a co-attention layer is integrated on top of the transformer-based textual feature extraction and caption feature extraction branches. The inter-modality discrepancy between the textual and low-level image representations is effectively addressed through the utilization of a multi-head self-attention layer. Therefore, we are able to capture the incongruity between the textual and visual data in two distinct manners, operating at separate levels. To the best of our knowledge, we are the first to propose a resilient and improved multi-level representation of inter-modal semantic incongruity based on descriptive image captions for multimodal sarcasm detection.
5. Extensive empirical investigations on two publicly accessible datasets [6], [16] demonstrate that the proposed framework exhibits a consistent enhancement when compared to baseline approaches and achieves a competitive level of performance in comparison to several established multimodal sarcasm detection techniques.

Organization: The ensuing sections of this document are organised in the following format: **Section 2** classifies the prior research carried out in the domain of sarcasm detection according to the various data modalities employed. In this section, various subsections elaborate on the cutting-edge techniques utilized for sarcasm recognition in both unimodal and multimodal contexts. **Section 3** provides an in-depth explanation of the proposed framework for detecting sarcasm using image-text pairs. **Section 4** provides an overview of the dataset characteristics for the two datasets employed in this research. **Section 5** offers a detailed description of the experimental setups and presents the findings obtained from the application of the proposed framework. **Section 6** provides a summary of the current study and offers suggestions for potential areas of investigation in future research.

2 Literature Review

2.1 Unimodal Sarcasm Detection

In this section, several state-of-the-art sarcasm detection approaches based on a single modality have been discussed.

2.1.1 Textual Data

Rule-based methodologies endeavor to recognize sarcasm in textual data by means of discerning particular pieces of evidence. According to Maynard and Greenwood [17], the sentiment expressed through hashtags serves as a pivotal determinant for detecting sarcasm. Hashtags are frequently employed by authors of tweets to accentuate sarcasm. Consequently, if the emotion conveyed by a hashtag failed to align with the overall content of the tweet, it was suggested that the tweet was intended to convey sarcasm. For statistical sarcasm detection, the majority of methodologies utilized bag-of-words as their primary feature representation. As demonstrated by Joshi et al. in [18], a crucial component of sarcasm detection is the use of word embeddings as a basis for similarity evaluation. Sarcasm detection studies based on conventional techniques extracted a myriad of features, including emoticons, unigrams [2], n-grams [19], POS sequences [20], semantic similarity [21], sentiment [22], and word frequency [23].

With the increasing prevalence of deep learning-based architectures in the realm of NLP, researchers have employed these methodologies to tackle the intricate challenge of automatic sarcasm detection [18], [24]–[29]. These studies delved into the intricate nuances of contextual elements, such as the complex interaction between the content of tweets and user reactions. In [25], Amir et al. introduced a novel architecture that utilized CNNs to acquire user embeddings alongside utterance-based embeddings to acquire user-specific contextual information. In [26], Poria et al. examined the use of deep convolutional networks as a means of discerning sarcasm in tweets. Goel et al. [30] trained an ensemble model comprising CNN, LSTM and GRU to effectively identify sarcastic statements in two benchmark datasets. Misra and Arora [29] proposed an attention-based hybrid neural network architecture to detect sarcasm in news headlines. Given that an ample supply of annotated data necessitates a substantial expenditure of human effort, Wang et al. [31] examined sarcasm detection from an unsupervised standpoint. The authors proposed a masking and generation paradigm to extract inconsistencies and acquire knowledge pertaining to the expression of sarcasm.

2.1.2 Acoustic Data

Tepperman et al. [3] conducted an experiment to examine the recognition of sarcasm in spoken dialogue. They analyzed 131 instances of the phrase “yeah right”, which were obtained from the Switchboard and Fisher corpora [32], [33]. The authors investigated multiple cues, encompassing prosodic, spectral, and contextual factors, and provided evidence that the inclusion of prosodic features is dispensable in the detection of sarcasm as long as spectral and contextual cues are considered.

2.2 Multimodal Sarcasm Detection

Multimodal sarcasm detection seeks to identify the sarcastic expression across several modalities, particularly image-text pairs [4] or videos [7], as opposed to relying solely on text-based or audio-based sarcasm recognition. The following subsections provide an elaboration of various approaches utilized for the detection of sarcasm in multimodal contexts.

2.2.1 Text and Visual Data

Sarcasm detection systems that depend solely on textual analysis may struggle to distinguish between genuine and sarcastic speech. Given the prevalence of text-image amalgamations within contemporary social media platforms, multimodal approaches that capture the disparity between the two modalities seem to be more promising for sarcasm prediction. Schifanella et al. [4] were the first to approach sarcasm identification as a multimodal classification problem by compiling a dataset comprising text-image posts from three social media platforms - Twitter, Tumblr and Instagram. They performed sarcasm recognition by concatenating image and text features that were either manually handcrafted or deep-learning-based. Using a combination of text, numeric, and visual information taken from Facebook posts, Das and Clark [34] trained multiple ML classifiers to identify sarcasm. Sangwan et al. [5] created two multimodal sarcasm recognition datasets: the Silver-Standard dataset, consisting of 10K sarcastic and non-sarcastic Instagram posts each, categorized on the basis of hashtags, and the Gold-Standard dataset, which includes 1600 randomly selected sarcastic posts from the first dataset and annotated first using only the text modality, and then reannotated using both modalities. The authors developed an RNN-based deep learning system to identify sarcasm by capitalizing on the interdependence of text and visuals.

An extensive dataset for multimodal Twitter sarcasm detection was introduced by Cai et al. in [6]. The authors put forward a hierarchical fusion framework that combined text, image, and attribute feature representations extracted using BiLSTM, ResNet-50, and ResNet-101 models, respectively. Pandey and Vishwakarma [35], [36] employed attention mechanisms in

conjunction with convolutional neural networks to carry out sentiment analysis on multimodal datasets comprising image-text pairs. In [10], Wang et al. utilized the pre-trained BERT and ResNet models for extracting the text and image feature representations, respectively and established a connection between the vector spaces of BERT and ResNet using a bridge layer. In order to represent the differences between the text and the images, the authors used multi-head attention and devised a 2D-intra-attention mechanism to draw focus to the disparities. Xu et al. [8] proposed a novel D&R Net (Decomposition and Relation Network) to model cross-modality incongruity and semantic association for sarcasm detection. While the decomposition network focused on contextual contrast between the text and image in high-level spaces, the relation network used a variant of multi-head attention to capture the degree of semantic association between the input text and adjective-noun pairs extracted from the image.

Zhao et al. [11] utilized BERT and Bi-GRU to extract text features and pre-trained ResNet-152 to extract image features. They then employed textual and visual attention mechanisms to derive context vectors, prioritizing specific keywords and image patches for enhanced focus. The coupled-attention model, as proposed, proceeded to calculate fusion memory vectors for k iterations. In this process, a memory vector, indexed by iteration i, stored the combined visual and textual information that had been gathered up to that particular iteration. Ultimately, the final memory vector underwent processing via a single-layer softmax classifier in order to detect sarcasm.

In [37], Liang et al. explored the use of graph convolution networks to jointly learn the sentiment incongruity within and across the text and image modalities as captured by in-modal and cross-modal graphs, respectively. In [38], Liang et al. presented a novel cross-modal graph convolutional network (CMGCN) to identify key parts of an image and then correlate them with the textual tokens to detect sarcasm.

Liu et al. [39] presented an innovative neural model that incorporated logic-based techniques to detect instances of multimodal rumors and sarcasm. The model utilized interpretable logic clauses to effectively articulate the reasoning process involved in the target task. A new framework called MuLOT (Multimodal Learning using Optimal Transport) was proposed by Pramanick et al. in [12] to detect sarcasm and humor. This framework uses self-attention to consider associations within a single modality, optimal transport to utilize cross-modal relationships, and multimodal attention fusion to account for inter-dependencies between modalities.

A multimodal meme dataset - “MultiBully”, annotated with labels for bullying, sentiment, sarcasm, and emotion, was developed and provided by Maity et al. in [16]. The authors concluded that the proposed multimodal multitask frameworks for recognition of cyberbullying and the three supplemental classification problems (sarcasm detection, sentiment analysis, and emotion recognition) prevailed over their unimodal and single-task variants, thus boosting the accuracy of cyberbullying detection.

2.2.2 Text and Acoustic Data

Bedi et al. [40] proposed a new Hindi-English code-mixed dataset, MaSaC, for multimodal sarcasm detection in conversational dialogue comprising 15K utterances from 400 scenes of the Indian TV comedy show ‘Sarabhai v/s Sarabhai’. The authors also developed MSH-COMICS, an attention-based multimodal classification model consisting of a hierarchical utterance-level attention module in order to acquire a more detailed linguistic representation of each statement.

2.2.3 Text, Visual and Acoustic Data

The first video dataset for sarcasm detection, MUStARD, was released by Castro et al. in [7]. The authors collected and annotated 690 audio-visual utterances from popular TV shows and

used an SVM classifier to predict sarcasm using a combination of textual, audio and visual features. Chauhan et al. [41] used emotion and sentiment for sarcasm recognition after updating the MUStARD dataset to incorporate relevant emotion and sentiment labels. By adding fresh utterances from similar sources, Ray et al. [42] doubled the size of the extended MUStARD dataset provided by Chauhan et al. Labels for arousal, valence, and sarcasm type were added to each phrase, and some incorrectly assigned emotions were also fixed. This revised dataset was released by the authors under the name MUStARD++. The dataset was benchmarked using different combinations of pre-trained feature extraction algorithms and multimodal fusion techniques. Chauhan et al. [43] introduced an enhanced version of the MUStARD dataset, known as SEEmoji MUStARD, wherein each utterance was annotated with a pertinent emoji, along with the associated emotion and sentiment conveyed by the emoji.

2.3 Research Limitations and Motivation

Despite the extensive study of sarcasm recognition in textual data, multimodal sarcasm detection is still a developing discipline. To evaluate their approach, most studies based on multimodal sarcasm recognition in image-text pairs use the Twitter multimodal sarcasm detection dataset. In order to stimulate more research on similar datasets, this paper demonstrates its findings on both the Twitter dataset and another less-studied dataset, MultiBully. The incorporation of this supplementary dataset in our research prompted us to utilize a multilingual variant of a transformer module to handle code-mixed statements. This is in contrast to previous studies that predominantly utilized the standard BERT model to extract features for English-language statements in the Twitter dataset. The majority of studies discussed in **section 2.2.1** have employed one or the other variant of the ResNet architecture as the conventional approach for visual feature extraction. However, there has been a lack of exploration into alternative methods for enhancing the robustness of visual feature extraction. Inspired by this observed pattern, we propose a more robust visual feature extraction pipeline. This pipeline incorporates a self-regulated ConvNet combined with a spatially aware attention module. This integration imposes minimal computational burden while simultaneously improving the quality of the extracted feature maps. Image attributes have been utilized in multimodal sarcasm detection research to incorporate an extra modality for improving the detection of incongruity. The inclusion of supplementary external knowledge is crucial for emphasizing the disparity between the textual message and the visual data. However, it is imperative that this knowledge source is more comprehensive and provides clearer explanations. In certain instances, attributes of images can provide meaningful information, particularly when the images depict simple scenes and standard objects. However, these attributes are insufficient in cases where the image holds a deeper meaning, particularly when there is text embedded within the image that must be understood in order to accurately interpret the semantic information conveyed by the image. Motivated by this shortcoming, this study suggests utilizing a more reliable external source of knowledge, which involves using descriptive image captions generated by an advanced encoder-decoder architecture. This architecture is capable of understanding text embedded in images and producing meaningful captions that accurately represent the visual semantics. This presents an additional opportunity for us to capture the differentiation between the given text and image, thereby improving the precision of automated multimodal sarcasm detection.

3 Proposed Methodology

In this section, we present an extensive account of the proposed framework for multimodal sarcasm detection in image-text pairs. **Figure 3** depicts the basic building blocks of the proposed architecture, which include the following: 1) A transformer-based textual feature extraction branch to capture the latent representation of the text modality; 2) An attentional

visual feature extraction branch comprising [15] combined with a lightweight attention module to uncover the image modality's latent representation; 3) An encoder-decoder architecture to generate descriptive caption corresponding to each image, followed by conversion to feature representation via a cross-lingual language model; and 4) Two distinct intermodal attention modules in order to effectively capture and model the inconsistencies that exist between textual information and two different levels of image representation. The pseudocode for the proposed approach has been outlined in **Table 1**.

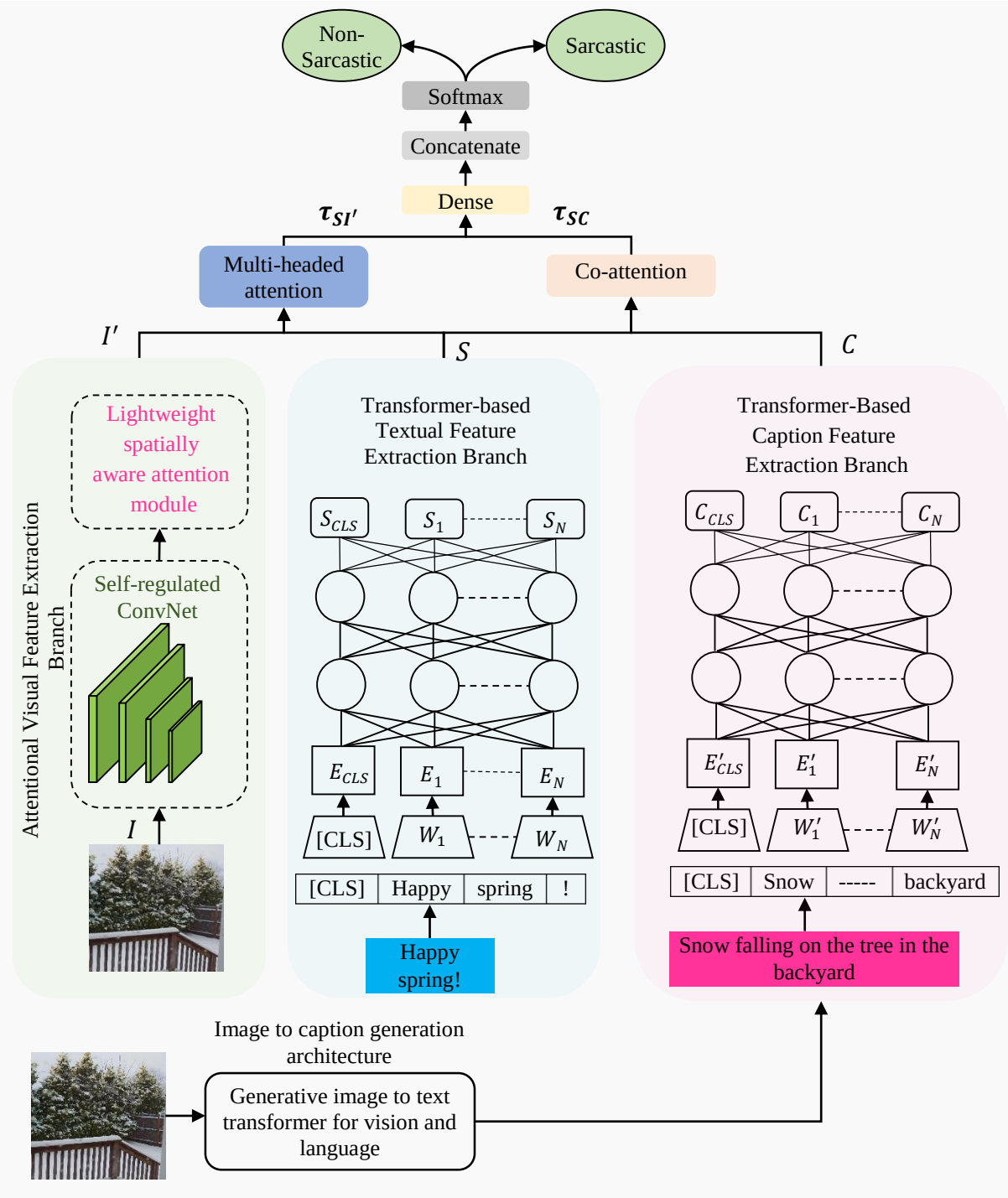


Figure 3 Proposed framework

Table 1 Algorithm for the proposed model

Algorithm-1: Enhanced Multi-Level Cross-Domain Incongruity Representation using Image Captions for Multimodal Sarcasm Detection
Input: Multimodal inputs, $M = \{W, I, W'\}$, where $W, I,$ and W' denote text, image, and caption correspondingly
Output: Sarcasm detection label, $Y \in \{non-sarcastic, sarcastic\}$

for $\mathbb{E} \leftarrow 1$ to epochs do
$\mathrm{W}_{tok} = \{[\mathrm{CLS}], w_1, w_2, w_3, \dots, w_t\} \leftarrow W = \{w_1, w_2, w_3, \dots, w_t\}$ word sequence tokenization using **Eq. (2)**
$S \leftarrow \mathrm{W}_{tok} = \{[\mathrm{CLS}], w_1, w_2, w_3, \dots, w_t\}$ text feature representation using **Eq. (3)**
$ConvNet_{self-reg}(I) \leftarrow I$ image feature representation using **Eq. (4)**
$ConvNet_{self-reg}(I)_{transformed} \leftarrow ConvNet_{self-reg}(I)$ projection of image features into same dimensions as text features using **Eq. (5)**
$I' \leftarrow ConvNet_{self-reg}(I)_{transformed}$ spatially selective feature maps using **Eq. (6)**
$W' \leftarrow$ I using generative image-to-text transformer model
$\mathrm{W}'_{tok} = \{[\mathrm{CLS}], w'_1, w'_2, w'_3, \dots, w'_u\} \leftarrow \mathrm{W}' = \{w'_1, w'_2, w'_3, \dots, w'_u\}$ caption tokenization using **Eq. (7)**
$C \leftarrow \mathrm{W}'_{tok} = \{[\mathrm{CLS}], w'_1, w'_2, w'_3, \dots, w'_u\}$ caption feature representation using **Eq. (8)**
$\tau_{SI'} \leftarrow (S, I')$ text-image incongruity representation using **Eq. (9), Eq. (10), Eq. (11),** and **Eq. (12)**
$\tau_{SC} \leftarrow (S, C)$ text-caption incongruity representation using **Eq. (13), Eq. (14), Eq. (15),** and **Eq. (16)**
$F \leftarrow (\tau_{SI'}, \tau_{SC})$ fused incongruity representation using **Eq. (17)**
$y \in Y \leftarrow F$ model prediction using **Eq. (18)**
end for

3.1 Task Definition

In the context of this study, multimodal sarcasm detection seeks to ascertain whether a particular text-image pair is intended to convey sarcasm or not. In the selected dataset D, each sample $d \in D$ comprises an input text sequence of words, denoted as $W = \{w_1, w_2, w_3, \dots, w_t\}$, along with a corresponding image I. To enhance the comprehension of image semantics, we employ a more effective approach by generating a descriptive caption corresponding to each image. Therefore, our model is capable of accepting additional input as an image caption, W'. Thus, the input triplets accepted by the model can be represented as shown in ***Equation*** (**1**).

$$Multimodal\ Input, M = \{Text\ (W), Image\ (I), Image\ Caption\ (W')\} \quad (\mathbf{1})$$

Our objective is to assign a label $y \in Y$ to each input triplet M, where $Y = \{non-sarcastic, sarcastic\}$.

3.2 Transformer-based Textual Feature Extraction Branch

The cross-lingual language model [44] has been employed to get the textual features from input text sequence of words, $\mathrm{W} = \{w_1, w_2, w_3, \dots, w_t\}$. The model tokenizer, $\mathbb{X}_{tokenizer}$ is utilized as an embedding module to generate text embeddings that encompass extensive semantic information. To achieve this, the input sequence is initially tokenized into a specific format, as shown in ***Equation*** (**2**).

$$\mathrm{W}_{tok} = \mathbb{X}_{tokenizer}\ (W) = \{[\mathrm{CLS}], w_1, w_2, w_3, \dots, w_t\} \quad (\mathbf{2})$$

where, $w_i \in \mathbb{R}^d$ and d denotes the embedding size. The sequential output, S obtained from the use of model encoder, $\mathbb{X}_{encoder}$ [44] for all tokens in the input W_{tok} can be denoted as shown in ***Equation*** (**3**).

$$S = \mathbb{X}_{encoder}(W_{tok}) \quad (3)$$

where, $S \in \mathbb{R}^{T \times d}$, where d = 768 stands for the dimension of each token, and T represents the maximum total sequence length.

With over 2 terabytes of cleaned CommonCrawl data [45], the masked language model [44] has already been pre-trained on over a hundred languages, including Hindi. We chose this particular variant of [46] for our investigation since the majority of the input sentences in the 'MultiBully' meme dataset are code-mixed Hindi-English. By fixing its weaknesses with a technique called sentence piece tokenization, [44] notably outperforms other multilingual models, such as [47], on several cross-lingual benchmarks.

3.3 Attentional Visual Feature Extraction Branch

The ensuing subsections comprehensively elucidate the proposed methodology for capturing vital information from the visual content.

3.3.1 Self-regulated Residual ConvNet

Upon receiving an input image, I with dimensions $I_H \times I_W$, the initial step involves resizing the image to dimensions 224×224. Subsequently, the image is partitioned into regions of size 7×7. Following that, each region is passed through the [15] model, as illustrated in **Figure 4**, to obtain a regional feature representation. After eliminating the terminal FC layer, the output of final convolutional layer corresponding to input image I can be denoted as depicted in ***Equation* (4).**

$$ConvNet_{self-reg}(I) \in \mathbb{R}^{r \times 2048} = \{r_i\}_{i=1}^{49} \quad (4)$$

where $r_i \in \mathbb{R}^{2048}$ represents the 2048-dimensional feature representation corresponding to a single image region.

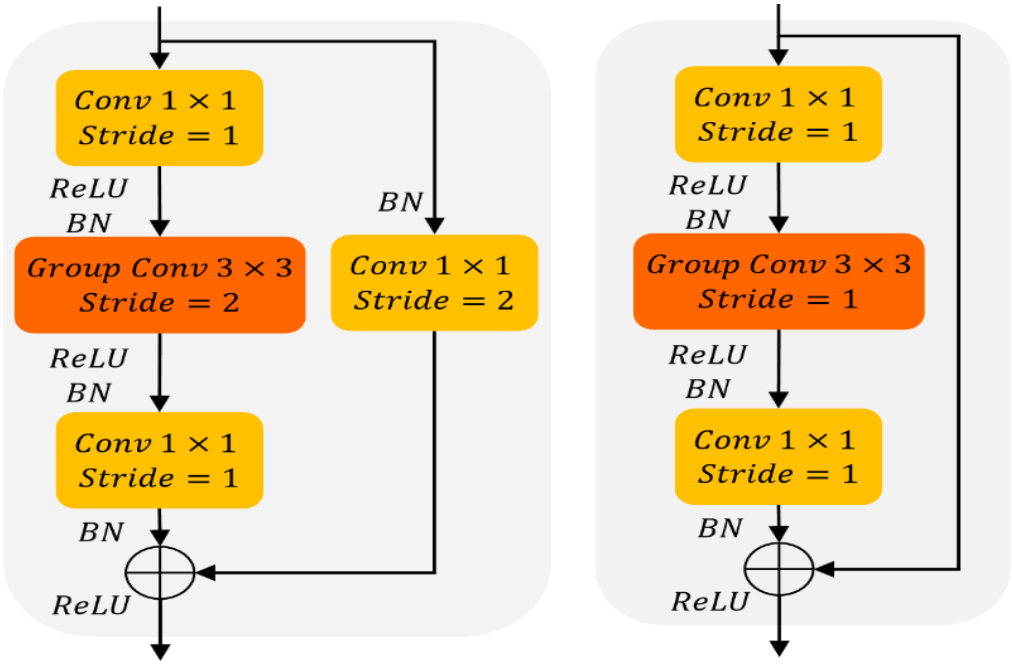


Figure 4 Structures of self-regulated residual ConvNet blocks

The architecture in [15] introduced a regulator module consisting of convolutional RNNs to extract complementary features and control the necessary memory information being passed to

the next block. The input feature from the jth building block, Ij, together with the sequential information, is encoded by a recurrent unit of the RNN (LSTM or GRU) to produce the hidden state H_j. To produce the output O_j, the hidden state is concatenated with I_j and fed to the next convolution operation and recurrent unit.

To ensure that the extracted image features are projected into the same dimensions as those of the text feature representation, the encoded image representation is subjected to a linear transformation as shown in ***Equation*** (**5**).

$$ConvNet_{self-reg}(I)_{transformed} \in \mathbb{R}^{r\times d} = ConvNet_{self-reg}(I)L_V \qquad (\mathbf{5})$$

where, $L_V \in \mathbb{R}^{2048\times \mathrm{d}}$ signifies a trainable parameter, and d corresponds to the dimension of each token of the textual sequence.

3.3.2 Lightweight Spatially Aware Attention Module

Attention modules are extensively employed in tasks related to computer vision. The primary intent of this integration is to assist models in discerning particular areas that necessitate specific focus. Nevertheless, the incorporation of attention mechanisms is somewhat constrained as lightweight networks are unable to effectively manage the associated computational load. Hence, to enhance the accuracy of image feature extraction, this study integrates a simple and flexible attention mechanism [48] that imposes a minimal computational load. The conventional approach of channel attention enables convolutional networks to acquire about the specific aspects that require emphasis during the learning process. As opposed to channel attention, which uses two-dimensional global pooling to reduce a feature tensor to a feature vector and disregards the importance of incorporating positional information, [48] splits global pooling operation into separate one-dimensional feature encoding transformations in the horizontal and vertical directions, so as to efficiently incorporate spatial coordinate data into the resulting attention maps. As a result, precise positioning information can be maintained along one spatial dimension while long-range dependencies can be captured along the other. A pair of attention maps that take into account both direction and location are then encoded from the resulting feature maps to further refine representations of the input feature map's notable features. The integration of the attention module, depicted in **Figure 5**, with blocks in convolutional networks allows for the enhancement of features by zeroing in on the most crucial aspects of the input feature representations, owing to its adaptability and low overhead.

Let the attention module's input feature tensor be $\mathbb{X} \in \mathbb{R}^{\mathbb{C}\times\mathbb{H}\times\mathbb{W}}$, where $\mathbb{C}$ denotes number of channels in $\mathbb{X}$, $\mathbb{H}$ symbolizes its height, and $\mathbb{W}$ stands for its width. Then the output $I' \in \mathbb{R}^{\mathbb{C}\times\mathbb{H}\times\mathbb{W}}$, represents a transformed tensor of the same size as $\mathbb{X}$ but with enhanced representations. To precisely capture inter-channel associations and positional information, the applied attention mechanism operates through a two-step process. This process involves information embedding, where the relevant data is encoded, and attention generation, where the attention weights are computed. The first step performs two transformations, one along each of the two spatial directions, to model the long-range interactions spatially while preserving the positional information. Each channel is encoded along the height and width by applying two average pooling kernels, ($\mathbb{H}$,1) and (1,$\mathbb{W}$). The second phase aims at effectively capturing inter-channel relationships from the aggregated feature maps generated in the preceding stage.

The two feature maps are initially combined and subsequently put through a shared 1×1 convolutional transformation, thereby yielding an intermediate feature map, $\mathbb{f} \in \mathbb{R}^{\frac{\mathbb{C}}{\mathbb{r}}\times(\mathbb{H}+\mathbb{W})}$, that captures spatial information along both horizontal and vertical directions. In this context,

$\mathbb{r}$ denotes the reduction ratio, which serves as the controlling factor for determining the block size. This is followed by factorizing $\mathbb{f}$ into two separate feature maps $\mathbb{f}^{\mathbb{h}} \in \mathbb{R}^{\frac{\mathbb{C}}{\mathbb{r}} \times \mathbb{H}}$, and $\mathbb{f}^{\mathbb{w}} \in \mathbb{R}^{\frac{\mathbb{C}}{\mathbb{r}} \times \mathbb{W}}$, which are then transformed into corresponding tensors $\mathbb{g}^{\mathbb{h}}$ and $\mathbb{g}^{\mathbb{w}}$ with the same number of channels as $\mathbb{X}$ via 1×1 convolutional transformations and sigmoid functions. As depicted in ***Equation* (6),** the output tensor I' of the attention module is calculated using $\mathbb{g}^{\mathbb{h}}$ and $\mathbb{g}^{\mathbb{w}}$ as attention weights.

$$I' = \mathbb{X} \times \mathbb{g}^{\mathbb{h}} \times \mathbb{g}^{\mathbb{w}} \tag{6}$$

where, $\mathbb{X}$ stands for the image feature representation, $ConvNet_{self-reg}(I)_{transformed}$, as derived in ***Equation* (5)**. The attention maps, $\mathbb{g}^{\mathbb{h}}$ and $\mathbb{g}^{\mathbb{w}}$, which possess the ability to perceive direction and sensitivity to position, work in tandem on the input feature map to strengthen the information pertaining to the focal points.

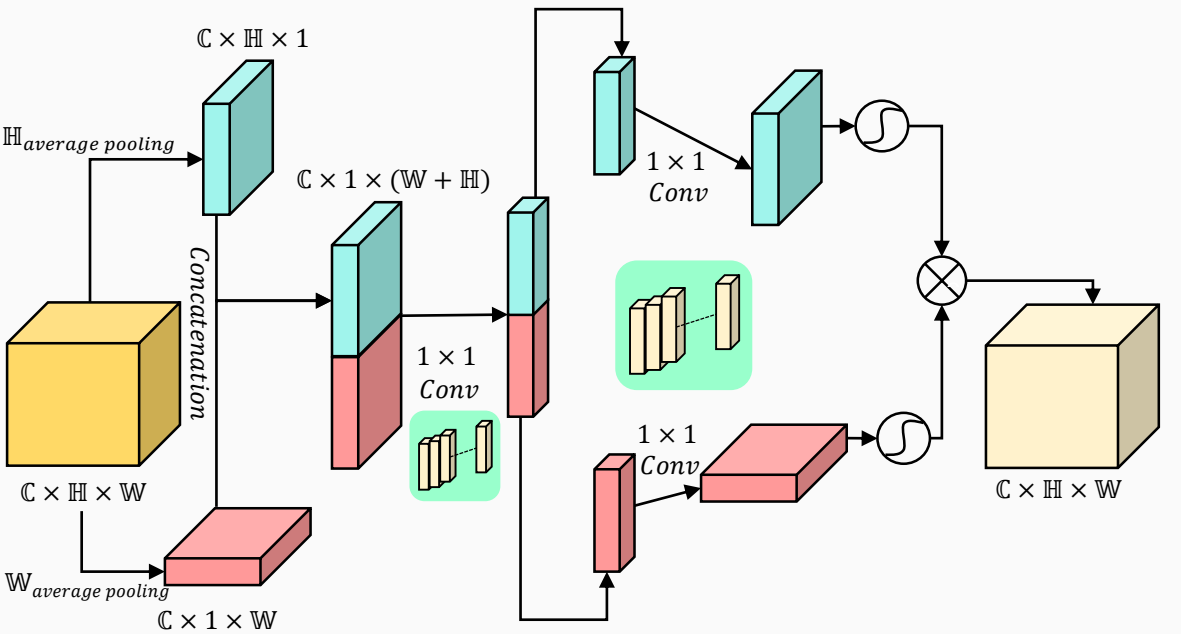


Figure 5 Structure of the lightweight attention module integrated to obtain spatially selective feature maps

3.4 Image Captioning

The following subsections shed light on the architecture utilized for image captioning and the process of feature extraction for the generated captions.

3.4.1 Encoder-Decoder Architecture

The concept of image captioning pertains to the procedure of generating descriptive captions, or textual representations, derived from input images. The generation of a caption necessitates the utilization of both natural language processing and computer vision methodologies [49]. The utilization of the vision encoder-decoder model is conducive to the initialization of an image-to-text model, wherein a diverse array of pre-trained transformer-based vision models, such as [50]–[53], can be employed as the encoder. Similarly, the decoder component can be instantiated using a variety of pre-trained language models, including [46], [47], [54], [55]. For our study, we utilized the generative image-to-text transformer [56] that employed a CLIP/ViT-L/14 [57] for image encoding and a transformer network for decoding. This framework (**Figure 6**) was introduced as a method for consolidating multiple vision-language tasks, namely, image

and video captioning, as well as question answering. It presents a straightforward model that relies on a single vision encoder and a single language decoder.

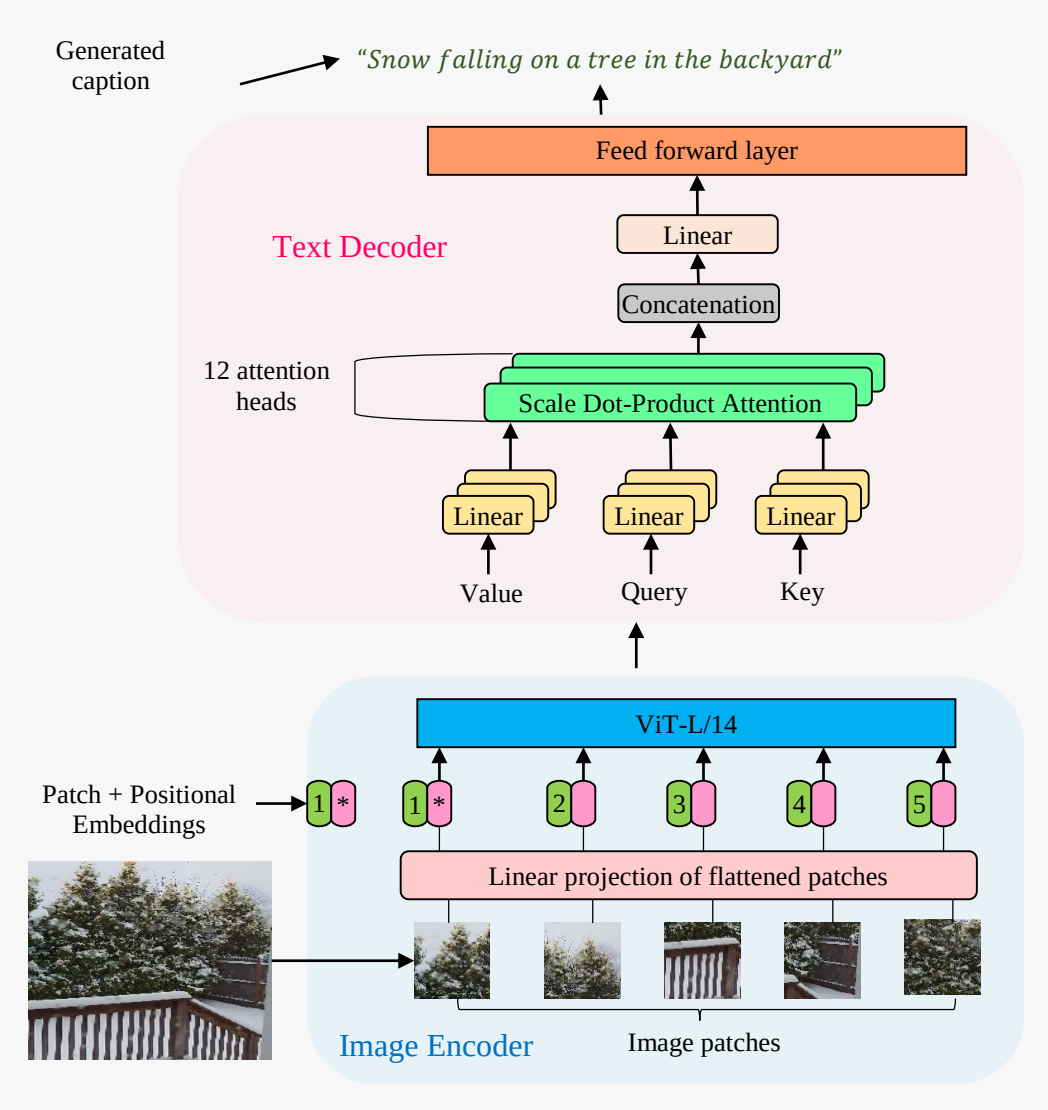


Figure 6 Image captioning encoder-decoder architecture

The vision transformer is a computational framework designed for the purpose of image classification. The model employs an architecture inspired by transformers, which processes image patches. The process involves dividing an image into patches of a predetermined size. These patches are subsequently subjected to linear embedding, followed by the addition of position embeddings. The resultant vector sequence is then inputted into a conventional transformer encoder. GIT_L employs the CLIP/ViT-L/14 vision and language model as its encoder that maps text and images to a shared vector space. CLIP models were designed to investigate the zero-shot generalization of models for image categorization tasks. The GPT-2 model is a transformer architecture that relies on an in-built masking paradigm to make certain that the predictions for a given token are based only on inputs received up to that point in the sequence. The model acquires an internal representation of the English language and is able to eliminate information from subsequent tokens in the prediction phase, allowing for the extraction of useful features for specific downstream applications. The model excels at its pretraining task, in particular, at producing textual outputs in response to inputs. Given its substantial size and the training on a dataset that is both extensive and varied, GPT-2 exhibits

robust performance across a wide range of fields and data sets. In a manner comparable to GPT-2, CLIP has the capability to be employed in various visual classification benchmarks through the straightforward provision of the names of the visual categories that are to be identified, specifically in a zero-shot context. The image encoder, in this case, is a 347M parameter model pre-trained on 20M image-text pairs or 14 M images, which is a combination of COCO [58], SBU [59], CC3M [60], VG [61], and CC12M [62] datasets, for 30 epochs.
The text decoder is a 6-layer transformer network constituted by multiple transformer blocks, each with its own self-attention and feed-forward layers. The decoder network having 12 attention heads, and hidden dimension 768, is randomly initialized and operates by generating text descriptions in an auto-regressive manner. In the initial pretraining phase, the text undergoes tokenization and embedding processes. Subsequently, positional encoding is applied, followed by the inclusion of a norm layer. The transformer module is fed an input consisting of the concatenation of the image features and text embeddings. Beginning with a [BOS] token, the text is decoded until either [EOS] token is found or the maximum number of steps is reached. To facilitate reasoning across image features and text embeddings, a sequential-to-sequential attention mask is utilized. This attention mask ensures that each text token is influenced by the preceding tokens as well as all image tokens while the image tokens themselves attend to one another.
The text that has been extracted using captions contains a substantial amount of information from the image. This information can be utilized to generate supplementary discriminative features for the purpose of detecting sarcasm. In line with the initial textual input, the acquired additional information in the form of image captions also captures valuable knowledge for the purpose of sarcasm detection.

3.4.2 Image Caption Feature Representation

The caption feature representations are derived using the cross-lingual language model, just like the text feature representations as discussed in **section 3.2**. For the input sequence of caption words, $\mathrm{W}' = \{w_1', w_2', w_3', \dots, w_u'\}$, the tokenized sequence is represented as depicted in ***Equation*** (**7**).

$$\mathrm{W}'_{tok} = \mathbb{X}_{tokenizer}(\mathrm{W}') = \{[\mathrm{CLS}], w_1', w_2', w_3', \dots, w_u'\} \quad (\mathbf{7})$$

where $w_i' \in \mathbb{R}^d$ and d denotes the embedding size. The sequential output, C, obtained for all tokens in the input W'_{tok} can be denoted as shown in ***Equation*** (**8**).

$$C = \mathbb{X}_{encoder}(\mathrm{W}'_{tok}) \quad (\mathbf{8})$$

where, $C \in \mathbb{R}^{U \times d}$, where d = 768 stands for the dimension of each token, and U symbolizes the maximum total length of W'_{tok}.

3.5 Multi-level Cross-Domain Incongruity Representation

This section presents a thorough account of the methodologies proposed for capturing multi-level cross-domain discrepancies. The initial subsection focuses on the extraction of incongruity between feature representations obtained from textual and visual feature extraction branches. The subsequent subsection elucidates the modelling of discrepancies arising from the inclusion of the image caption modality.

3.5.1 Text-Image Incongruity

The relationship between each token pair of a sequence is taken into account in the internal representation generated using self-attention mechanism [63]. Given the inherent significance

of discordance in sarcasm, the input tokens will place greater emphasis on image regions that directly contradict them. We take inspiration from this innate capacity for focused attention to create a layer, as illustrated in **Figure 7**, that can detect discrepancies between text and images. This layer in our model is designed to accept text features $S \in \mathbb{R}^{T\times d}$, as queries, and image features $I' \in \mathbb{R}^{r\times d}$, as keys and values. The model can then be directed to compare the query matrices of the text features to the key matrices of all image regions and allocate greater focus towards the ones that exhibit incongruity. The application of self-attention multiple times in parallel allows attending to different token pairs differently. Let us denote the total number of heads in the multi-head attention function as h. Then, for the i[th] attention head, the output m_i, can be computed as shown in ***Equation*** (**9**).

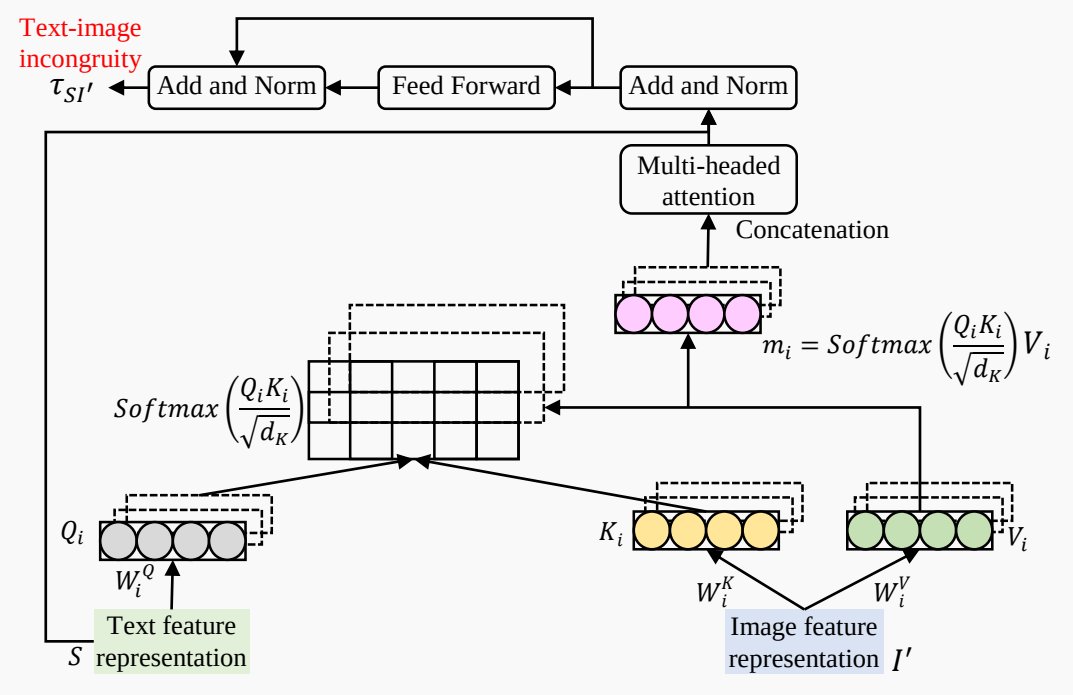


Figure 7 Text-image incongruity

$$m_i(S,I) = Softmax(\frac{(SW_i^q)(I'W_i^k)^{\mathbb{T}}}{\sqrt{d_k}})(I'W_i^v) \quad (\mathbf{9})$$

where $d_k \in \mathbb{R}^{\frac{d}{h}}$, $m_i(S,I') \in \mathbb{R}^{T\times d_k}$, $\mathbb{T}$ symbolizes the matrix transpose operation, and $\{W_i^q, W_i^k, W_i^v\} \in \mathbb{R}^{d\times d_k}$ stand for the learnable query, key and value weight matrices. The multi-head attention is calculated using the output features of each head, m_i, as shown in ***Equation*** (**10**).

$$Multi-head\ attention\ (S,I') = (m_1 \oplus m_2 \oplus m_3 \oplus \ldots \oplus m_h)W_O \quad (\mathbf{10})$$

where $W_O \in \mathbb{R}^{d\times d}$, and $\oplus$ denotes vector concatenation. Followed by a two-layer MLP and a residual connection, the text-image discrepancy captured can be obtained as illustrated in ***Equation*** (**11**).

$$Incongurity\ (S,I') = LNorm\left(S + MLP\left(Multi-head\ attention\ (S,I')\right)\right) \quad (\mathbf{11})$$

where $Incongurity\ (S,I') \in \mathbb{R}^{T\times d}$. The final disparity representation, $\tau_{SI'} \in \mathbb{R}^d$ is obtained via ***Equation*** (**12**).

$$\tau_{SI'} = [CLS] - encoding\ (Incongurity\ (S, I')) \quad (\mathbf{12})$$

3.5.2 Text-Caption Incongruity

We utilize a co-attention mechanism, as demonstrated in the field of visual question answering by Lu et al. in [64], to effectively address the disparities between the original text and the generated image captions. As shown in **Figure 8**, the co-attention dual input comprises the feature representations of the text and the image captions, as obtained in ***Equation*** **(3)** and ***Equation*** **(8)**, respectively. The initial step involves the computation of affinity matrix, $A \in \mathbb{R}^{T \times U}$ through the utilization of a bi-linear transformation W, as shown in ***Equation*** **(13)**. The motive of this transformation is to accurately encapsulate the interrelation between the text and the captions of the accompanying images.

$$A = \tanh(SWC^{\mathbb{T}}) \quad (\mathbf{13})$$

Here, $S \in \mathbb{R}^{T \times d}$ denotes the text features, $C \in \mathbb{R}^{U \times d}$ denotes the image caption features (T and U denote the maximum size of these features correspondingly, and d symbolizes the hidden size of [44]), and $W \in \mathbb{R}^{d \times d}$ stands for a learnable parameter comprising weights. The affinity matrix that has been acquired serves to transform the attention space of the text into the attention space of the caption. In order to highlight the words that most significantly contribute to the incongruity characterizing the sarcastic expression, the affinity matrix is maximized over the text feature's locations to obtain caption attention, a_c, as illustrated in ***Equation*** **(14)**.

$$a_c[u] = max_i(A_{i,u}) \quad (\mathbf{14})$$

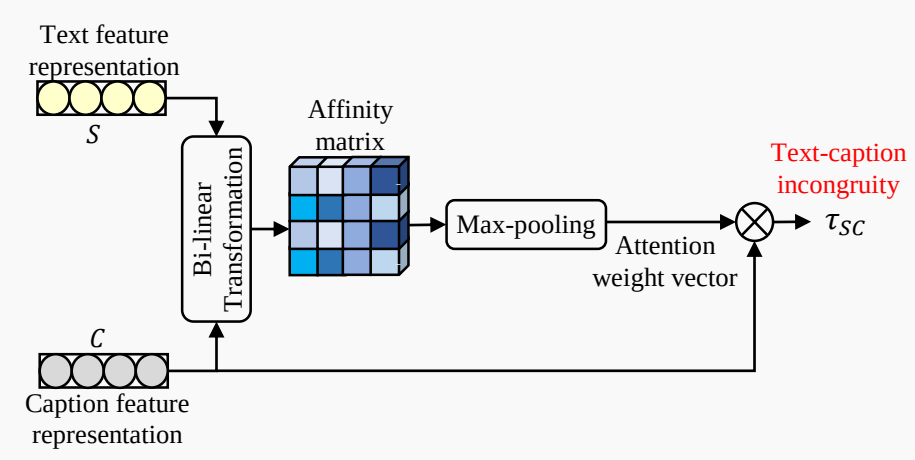


Figure 8 Text-caption incongruity

The attention weight vector $v \in \mathbb{R}^U$ is thus calculated by applying a column-wised max-pooling operation kernel of dimensions $T \times 1$ over affinity matrix A, as depicted in ***Equation*** **(15)**.

$$v = MaxPool\ (A) \quad (\mathbf{15})$$

The final text-image caption incongruity representation used to compute the model's prediction is obtained as $\tau_{SC} \in \mathbb{R}^d$, using ***Equation*** **(16)**.

$$\tau_{SC} = vC \quad (\mathbf{16})$$

3.5.3 Final Fused Incongruity Representation and Model Prediction

After capturing both, the incongruity between the text and low-level image feature representation, $\tau_{SI'}$, and that between the text and high-level image description in the form of generated captions, τ_{SC}, we concatenate the two disparity descriptors to obtain fusion vector, F for final model predictions, as shown in ***Equation*** (**17**), where $\oplus$ signifies vector concatenation.

$$F = \tau_{SI'} \oplus \tau_{SC} \quad (\mathbf{17})$$

The fusion vector, F is put through the softmax function for classification purpose. The final prediction y is computed as shown in ***Equation*** (**18**), where $W_t \in \mathbb{R}^{2d}$ is a trainable parameter and b denotes the bias.

$$y = Softmax(W_t F + b) \quad (\mathbf{18})$$

4 Dataset Description

Our model for multimodal sarcasm detection is evaluated using two publicly available datasets constructed by Cai et al. [6] and Maity et al. [16].

4.1 Twitter Dataset (Cai et al. [6])

Every individual instance within this dataset comprises a combination of textual and corresponding visual information that has been acquired from the social media platform Twitter. Specific hashtag queries, such as #sarcasm, #sarcastic, #irony, #ironic, and so forth, were employed to gather positive samples which pertain to instances of sarcasm. Conversely, the negative samples, denoting non-sarcastic content, were obtained from tweets lacking the aforementioned hashtags. The dataset has been partitioned into three distinct subsets: the training set, which comprises 80% of the data; the development set, which accounts for 10%; and the test set, which also represents 10% of the dataset. The pertinent information has been presented in a tabular format, as outlined in **Table 2**.

Table 2 Twitter multimodal sarcasm detection dataset statistics

	Training	**Development**	**Testing**
Sarcastic	8,642	959	959
Non-Sarcastic	11,174	1,451	1,450
Total	19,816	2,410	2,409

4.2 MultiBully (Maity et al. [16])

The authors released a dataset collected from publicly available sources such as Twitter and Reddit to examine the influence of sentiment, emotion, and sarcasm in the detection of cyberbullying from multimodal memes within a code-mixed language context. They incorporated three supplementary tasks, namely sarcasm detection, sentiment analysis, and recognition of emotions, aiming to achieve better accuracy with respect to the primary objective, i.e. cyberbullying detection. For this study, we focus only on the subtask of sarcasm detection from multimodal data, i.e. text-image pairs. **Table 3** provides an overview of the statistics regarding the number of sarcastic and non-sarcastic memes held by the train, development, and test sets.

Table 3 MultiBully dataset statistics

	Training	**Development**	**Testing**
Sarcastic	1,545	201	429
Non-Sarcastic	2,552	384	743

Total	4,097	585	1,172

5 Experimental Results

This section provides elaborate information regarding the experimental settings and hyperparameters utilized in the proposed framework, as well as performance evaluations in comparison to baseline models.

5.1 Experimental Settings and Hyperparameters

Our model is implemented in PyTorch [65] using Google COLAB Pro Plus, which featured an NVIDIA V100 GPU, CUDA version 11.2, driver version 460.32.03, 40 GB of graphics memory, 100 GB of hard disk space, and 80 GB of RAM. We used softmax as a classifier for recognizing sarcasm in image-text pairs. The loss function employed to optimize the presented model was the L2-regularized binary cross-entropy, as represented in ***Equation*** (**19**).

$$J_{BCE}^{L2} = -\frac{1}{Train\ size}\sum_{i=1}^{Train\ size} y^i \log(\hat{y}_i) + (1-y^i)\log(1-\hat{y}_i) + \lambda \sum_{i=1}^{Train\ size} w_i^2 \qquad (\mathbf{19})$$

In the present study, the Adam optimizer [66] was employed as the optimization algorithm to update all the trainable parameters of the model. The constituted model was trained with a learning rate of 5×10^{-5}, using a batch size of 32. Additionally, a warmup rate of 0.1 was used to control the variation of learning rate and facilitate the optimization process. To mitigate the issue of overfitting, we implemented the dropout technique with a probability value of 0.5. Further, to save time and computational resources, an early stopping patience value of 5 was enforced. By employing this technique, we were able to effectively monitor the training process and terminate it once the model's performance began to plateau, thereby ensuring that the model does not become overly specialized to the training data and could exhibit good generalization capabilities towards novel data. The proposed model was fine-tuned for 15 epochs.

5.2 Evaluation Metrics

In accordance with the evaluation methodology employed by [6], we employ four widely used performance metrics, namely Accuracy (A), Precision (P), Recall (R), and F1-score (F1), to evaluate the efficacy of the proposed model. The mathematical formulae of these metrics are provided in **Table 4**.

Table 4 Performance metrics

Performance metric	Mathematical formula
Accuracy (*A*)	$\frac{True\ Pos. + True\ Neg.}{True\ Pos. + True\ Neg. + False\ Pos. + False\ Neg.}$
Precision (*P*)	$\frac{True\ Pos.}{True\ Pos. + False\ Pos.}$
Recall (*R*)	$\frac{True\ Pos.}{True\ Pos. + False\ Neg.}$
F-1 Score (*F*1)	$\frac{2 \times P \times R}{P + R}$

5.3 Baselines

The presented framework is being compared to the baselines that are discussed below.

Text-only

These models made the final predictions solely on the basis of textual data, including **TextCNN** [72], **SIARN** [68], **SMSD** [69], **MIARN** [68], **BiLSTM** [70], **BERT** [47], and **TextGraph** [37].

Visual-only
These models utilized only the visual information for detecting sarcasm, including **Image** (vectors obtained after ResNet pooling layer) [6], **ViT** [51], **ResNet-152** [71], and **ImageGraph** [37].

Multimodal
The models mentioned in the text incorporate both text and image data, and they have the capability to incorporate additional inputs from other modalities, depending on the specific framework being used. The **HFM** [6] framework was designed to combine text, image, and attribute features in a hierarchical manner. The **Bridge** [10] model utilized a bridge layer to map the image feature representation into the BERT vector space. It also addressed the differences between images and text by incorporating multi-head attention. The **Res-BERT** [67] concatenated both text and image features to detect sarcasm, while **Att-BERT** [67] employed separate attention modules to capture disparities within and between different modalities. The **D&R Net** [8] approach involved incorporating ANPs, text, and images to analyze the contextual contrast between text and images, as well as the level of semantic association. The Coupled Attention framework, introduced in **CAN** [11], aims to prioritize certain keywords and regions in text and image data. The authors of **InCrossMGs** [37] utilized graph convolution networks to analyze the sentiment incongruity present in text and images. They achieved this by constructing both in-modal and cross-modal graphs. In their study, **CMGCN** [38] employed object recognition techniques to identify important sections of an image. They then established connections between textual tokens and these identified parts to effectively detect sarcasm. **MuLOT** [12] captures internal relationships within a single modality by employing self-attention. Additionally, it leverages optimal transport techniques to take advantage of cross-modal associations. Logic-based techniques were utilized in order to identify occurrences of multimodal rumors and sarcasm in **LogicDM** [39].

5.4 Experimental Results

Table 5 and Table 6 present a comparative analysis between our proposed methodology and other cutting-edge benchmarks. Our proposed solution presented in this study demonstrates superior performance compared to the existing state-of-the-art approaches across each of the four evaluation metrics, namely $Accuracy$, $Recall$, $Precision$, and $F1-score$. Our model indicates a significant enhancement of 2.07% in $Accuracy$ and 0.51% in $F1-score$ compared to the current state-of-the-art MuLOT (with OCR) [12] for Twitter dataset and 1.49% in $Accuracy$ and 0.47% in $F1-score$ compared to Maity et al. [16] for MultiBully dataset. This outcome serves as empirical evidence supporting the efficacy of our proposed model.
It can be observed from **Table 5** and **Table 6** that the independent treatment of images or text does not yield satisfactory results in addressing the sarcasm detection problem. From a qualitative perspective, it can be observed that models relying solely on images tend to exhibit lower performance compared to models relying solely on text. This is due to the fact that an image alone lacks the necessary information to accurately identify the presence of sarcasm. Furthermore, it is evident that unimodal approaches exhibit inferior performance in sarcasm identification, highlighting the need for multimodal approaches to be employed for more effective sarcasm detection. It can be inferred from **Table 5** and **Table 6** that our proposed approach exhibits superior performance in comparison to other multimodal approaches by effectively capturing the disparities between the image and text pairs. **Figure 9** and **Figure 10**

depict a visual representation for the comparative analysis between our proposed framework and other multimodal methods based on various evaluation metrics.

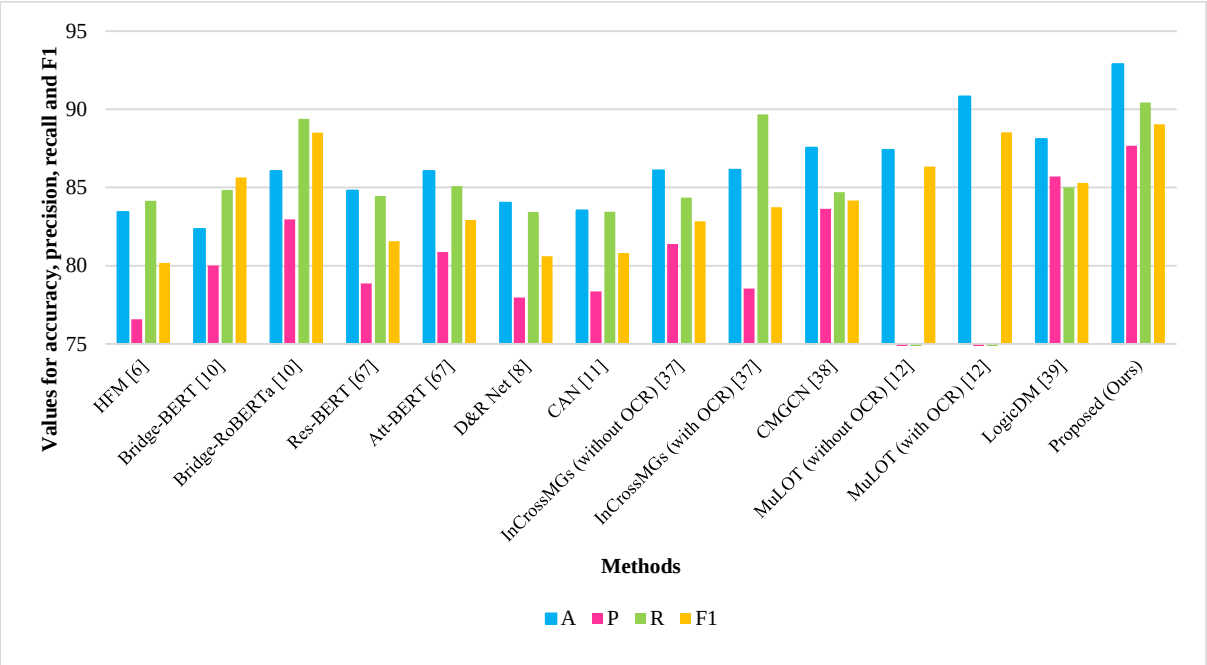


Figure 9 Comparison of accuracy, precision, recall and F1 of our proposed model with the baseline approaches for multimodal Twitter dataset

Table 5 Performance comparison with state-of-the-art unimodal and multimodal baselines for Twitter multimodal sarcasm detection dataset

Modality	Method	*A*	*P*	*R*	*F1*
Text	TextCNN [72]	80.03	74.29	76.39	75.32
	SIARN [68]	80.57	75.55	75.70	75.63
	SMSD [69]	80.90	76.46	75.18	75.82
	MIARN [68]	82.48	79.67	75.18	77.36
	BiLSTM [70]	81.90	76.66	78.42	77.53
	BERT [47]	83.85	78.27	82.27	80.22
	TextGraph [37]	84.23	78.86	82.48	80.63
	Cross-lingual language model (Ours)	**88.32**	**87.03**	**87.61**	**87.32**
Image	Image [6]	64.76	54.41	70.80	61.53
	ViT [51]	67.83	57.93	70.07	63.43
	ResNet-152 [71]	72.60	65.11	67.15	66.11
	ImageGraph [37]	73.89	63.24	**82.17**	**71.47**
	Self-regulated ConvNet + Lightweight Attention (Ours)	**74.11**	**63.62**	72.98	67.98
Multimodal	HFM [6]	83.44	76.57	84.15	80.18
	Bridge-BERT [10]	82.35	80.01	84.84	85.64
	Bridge-RoBERTa [10]	86.05	82.95	89.39	88.51
	Res-BERT [67]	84.80	78.87	84.46	81.57
	Att-BERT [67]	86.05	80.87	85.08	82.92
	D&R Net [8]	84.02	77.97	83.42	80.60
	CAN [11]	83.53	78.36	83.45	80.82
	InCrossMGs (without OCR) [37]	86.10	81.38	84.36	82.84
	InCrossMGs (with OCR) [37]	86.13	78.54	89.68	83.74
	CMGCN [38]	87.55	83.63	84.69	84.16
	MuLOT (without OCR) [12]	87.41	-	-	86.33
	MuLOT (with OCR) [12]	90.82	-	-	88.52
	LogicDM [39]	88.10	85.70	85.00	85.30
	Proposed (Ours)	**92.89**	**87.67**	**90.43**	**89.03**

Table 6 Performance comparison with state-of-the-art unimodal and multimodal baselines for MultiBully dataset

Modality	Method	*A*	*P*	*R*	*F1*
Text	m-BERT-GRU [16]	59.72	-	-	59.12
	Cross-lingual language model (Ours)	**63.83**	**61.72**	**62.69**	**62.20**
Image	ResNet [16]	59.39	-	-	57.79
	Self-regulated ConvNet + Lightweight Attention (Ours)	**62.91**	**61.07**	**61.98**	**61.53**
Multimodal	Maity et al. [16]	62.99	-	-	63.80
	Proposed (Ours)	**64.48**	**63.69**	**64.87**	**64.27**

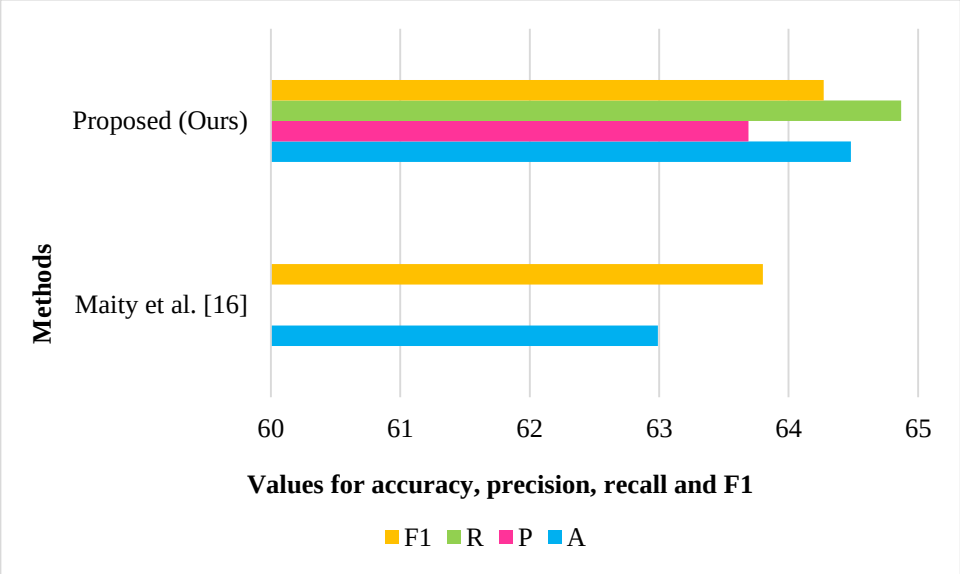


Figure 10 Comparison of accuracy, precision, recall and F1 of our proposed model with the baseline approaches for MultiBully dataset

5.5 Ablation Study

In this section, a set of ablation trials are carried out on the two datasets, namely the Twitter multimodal sarcasm dataset and the MultiBully dataset. The objective is to thoroughly evaluate the effectiveness of each proposed component for the multimodal input. As per the suggested framework, we have omitted the lightweight spatially aware attention module from the attentional visual feature extraction branch, referred to as the *proposed model w/o visual attention*. Additionally, we have also eliminated the incongruity representation $\boldsymbol{\tau_{SI'}}$ from the proposed model, denoted as the *proposed model w/o* $\boldsymbol{\tau_{SI'}}$. Likewise, the removal of $\boldsymbol{\tau_{SC}}$ from the proposed model has been represented as *proposed model w/o* $\boldsymbol{\tau_{SC}}$. The results obtained from the ablation trials are succinctly displayed in **Table 7**.

Table 7 Ablation Study

Approaches	Twitter multimodal sarcasm detection dataset				MultiBully dataset			
	A	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*
Proposed model w/o visual attention	89.33	85.16	87.42	86.28	62.02	59.14	59.86	59.49
Proposed model w/o $\boldsymbol{\tau_{SI'}}$	80.12	79.68	80.83	80.25	52.04	50.86	51.94	51.39
Proposed model w/o $\boldsymbol{\tau_{SC}}$	83.67	82.04	82.99	82.51	53.89	49.10	51.6	50.32

Proposed (Ours)	**92.89**	**87.67**	**90.43**	**89.03**	**64.48**	**63.69**	**64.87**	**64.27**

From **Table 7** it can be observed that the exclusion of the attention module from the visual branch of our proposed model hinders the overall predictive capability. Also, the elimination of incongruity representations also leads to a decline in the model's performance, indicating that the inclusion of multi-head and co-attention to capture incongruity between textual and image features is crucial for the detection of sarcasm. Based on the results mentioned above, it can be inferred that our proposed model demonstrates superior ability in capturing the disparity between visual and textual modalities, consequently facilitating effective acquisition of the underlying sarcasm.

6 Conclusion and Future Scope

In this study, we propose a novel approach for multimodal sarcasm recognition by utilizing image captions to create a multi-tier cross-modal incongruity representation. The inclusion of captions in this particular context provides a valuable opportunity to improve our understanding of the meaning conveyed by images and draw attention to the difference in emotional tone between the message conveyed in a tweet and the visual content that accompanies it. This methodology improves the efficacy of the suggested framework in identifying instances of sarcasm within two openly available datasets. The experimental results indicate that our proposed framework consistently improves upon state-of-the-art baselines for multimodal sarcasm detection. The effectiveness of each module within the provided framework is additionally supported by the ablation study. In future investigations, our objective is to analyze the feasibility of adapting our model to accommodate supplementary multimodal tasks. Future research endeavors may delve into improved fusion methodologies in order to better capture the interconnections among various modalities. Furthermore, researchers may also aim to integrate low-parameter or domain-adaptive models into their studies. This is due to the fact that models with a reduced number of trainable parameters provide notable benefits by facilitating real-time deployment.

References

[1] D. Davidov, O. Tsur, and A. Rappoport, 'Semi-Supervised Recognition of Sarcasm in Twitter and Amazon', in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, Uppsala, Sweden: Association for Computational Linguistics, Jul. 2010, pp. 107–116. Accessed: Apr. 15, 2023. [Online]. Available: https://aclanthology.org/W10-2914

[2] R. González-Ibáñez, S. Muresan, and N. Wacholder, 'Identifying Sarcasm in Twitter: A Closer Look', in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA: Association for Computational Linguistics, Jun. 2011, pp. 581–586. Accessed: Apr. 15, 2023. [Online]. Available: https://aclanthology.org/P11-2102

[3] J. Tepperman, D. Traum, and S. Narayanan, 'Yeah right: Sarcasm recognition for spoken dialogue systems', Sep. 2006. doi: 10.21437/Interspeech.2006-507.

[4] R. Schifanella, P. de Juan, J. Tetreault, and L. Cao, 'Detecting Sarcasm in Multimodal Social Platforms', in *Proceedings of the 24th ACM international conference on Multimedia*, Oct. 2016, pp. 1136–1145. doi: 10.1145/2964284.2964321.

[5] S. Sangwan, M. S. Akhtar, P. Behera, and A. Ekbal, 'I didn't mean what I wrote! Exploring Multimodality for Sarcasm Detection', in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9206905.

[6] Y. Cai, H. Cai, and X. Wan, 'Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model', in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 2506–2515. doi: 10.18653/v1/P19-1239.

[7] S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, and S. Poria, 'Towards Multimodal Sarcasm Detection (An _Obviously_ Perfect Paper)', in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 4619–4629. doi: 10.18653/v1/P19-1455.

[8] N. Xu, Z. Zeng, and W. Mao, ‘Reasoning with Multimodal Sarcastic Tweets via Modeling Cross-Modality Contrast and Semantic Association’, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, Jul. 2020, pp. 3777–3786. doi: 10.18653/v1/2020.acl-main.349.

[9] Z. Yin and F. You, ‘Multi-Modal Sarcasm Detection in Weibo’, in *2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT)*, Jun. 2021, pp. 740–743. doi: 10.1109/ISCIPT53667.2021.00156.

[10] X. Wang, X. Sun, T. Yang, and H. Wang, ‘Building a Bridge: A Method for Image-Text Sarcasm Detection Without Pretraining on Image-Text Data’, in *Proceedings of the First International Workshop on Natural Language Processing Beyond Text*, Online: Association for Computational Linguistics, Nov. 2020, pp. 19–29. doi: 10.18653/v1/2020.nlpbt-1.3.

[11] X. Zhao, J. Huang, and H. Yang, ‘CANs: Coupled-Attention Networks for Sarcasm Detection on Social Media’, in *2021 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9533800.

[12] S. Pramanick, A. Roy, and V. M. P. Johns, ‘Multimodal Learning using Optimal Transport for Sarcasm and Humor Detection’, in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2022, pp. 546–556. doi: 10.1109/WACV51458.2022.00062.

[13] X. Guo, A. Kot, and A. W.-K. Kong, ‘Pace-adaptive and Noise-resistant Contrastive Learning for Multimodal Feature Fusion’, *IEEE Trans. Multimed.*, pp. 1–12, 2023, doi: 10.1109/TMM.2023.3252270.

[14] F. Geng, H. Yang, C. Wu, and J. Li, ‘Multimodal sentiment analysis based on multi-head self-attention and convolutional block attention module’, in *2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, Apr. 2022, pp. 262–266. doi: 10.1109/AEMCSE55572.2022.00059.

[15] J. Xu, Y. Pan, X. Pan, S. Hoi, Z. Yi, and Z. Xu, ‘RegNet: Self-Regulated Network for Image Classification’, *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–6, 2022, doi: 10.1109/TNNLS.2022.3158966.

[16] K. Maity, P. Jha, S. Saha, and P. Bhattacharyya, ‘A Multitask Framework for Sentiment, Emotion and Sarcasm aware Cyberbullying Detection from Multi-modal Code-Mixed Memes’, in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, in SIGIR ’22. New York, NY, USA: Association for Computing Machinery, Jul. 2022, pp. 1739–1749. doi: 10.1145/3477495.3531925.

[17] D. G. Maynard and M. A. Greenwood, ‘Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis’, *LREC 2014 Proceedings*, Mar. 06, 2014. http://www.lrec-conf.org/proceedings/lrec2014/index.html (accessed Jul. 11, 2023).

[18] A. Joshi, V. Tripathi, K. Patel, P. Bhattacharyya, and M. Carman, ‘Are Word Embedding-based Features Useful for Sarcasm Detection?’, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 1006–1011. doi: 10.18653/v1/D16-1104.

[19] A. Reyes and P. Rosso, ‘Making objective decisions from subjective data: Detecting irony in customer reviews’, *Decis. Support Syst.*, vol. 53, no. 4, pp. 754–760, Nov. 2012, doi: 10.1016/j.dss.2012.05.027.

[20] P. Liu, W. Chen, G. Ou, T. Wang, D. Yang, and K. Lei, ‘Sarcasm Detection in Social Media Based on Imbalanced Classification’, in *Web-Age Information Management*, F. Li, G. Li, S. Hwang, B. Yao, and Z. Zhang, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, pp. 459–471. doi: 10.1007/978-3-319-08010-9_49.

[21] I. Hernández-Farías, J.-M. Benedí, and P. Rosso, ‘Applying Basic Features from Sentiment Analysis for Automatic Irony Detection’, in *Pattern Recognition and Image Analysis*, R. Paredes, J. S. Cardoso, and X. M. Pardo, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 337–344. doi: 10.1007/978-3-319-19390-8_38.

[22] M. Bouazizi and T. Ohtsuki, ‘Sarcasm Detection in Twitter: “All Your Products Are Incredibly Amazing!!!” - Are They Really?’, in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2015, pp. 1–6. doi: 10.1109/GLOCOM.2015.7417640.

[23] F. Barbieri, ‘Italian irony detection in Twitter : a first approach’, *Proc. First Ital. Conf. Comput. Linguist. CLiC-It 2014 Fourth Int. Workshop EVALITA 2014 9-11 Dec. 2014 Pisa*, pp. 28–32, 2014.

[24] D. Bamman and N. Smith, ‘Contextualized Sarcasm Detection on Twitter’, *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 9, no. 1, Art. no. 1, 2015, doi: 10.1609/icwsm.v9i1.14655.

[25] S. Amir, B. C. Wallace, H. Lyu, P. Carvalho, and M. J. Silva, ‘Modelling Context with User Embeddings for Sarcasm Detection in Social Media’, in *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 167–177. doi: 10.18653/v1/K16-1017.

[26] S. Poria, E. Cambria, D. Hazarika, and P. Vij, 'A Deeper Look into Sarcastic Tweets Using Deep Convolutional Neural Networks', *CoRR*, vol. abs/1610.08815, 2016, Accessed: Jul. 11, 2023. [Online]. Available: http://arxiv.org/abs/1610.08815
[27] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, 'Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm', in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017, pp. 1615–1625. doi: 10.18653/v1/D17-1169.
[28] A. Mohan, A. M. Nair, B. Jayakumar, and S. Muraleedharan, 'Sarcasm Detection Using Bidirectional Encoder Representations from Transformers and Graph Convolutional Networks', *Procedia Comput. Sci.*, vol. 218, pp. 93–102, Jan. 2023, doi: 10.1016/j.procs.2022.12.405.
[29] R. Misra and P. Arora, 'Sarcasm detection using news headlines dataset', *AI Open*, vol. 4, pp. 13–18, Jan. 2023, doi: 10.1016/j.aiopen.2023.01.001.
[30] P. Goel, R. Jain, A. Nayyar, S. Singhal, and M. Srivastava, 'Sarcasm detection using deep learning and ensemble learning', *Multimed. Tools Appl.*, vol. 81, no. 30, pp. 43229–43252, Dec. 2022, doi: 10.1007/s11042-022-12930-z.
[31] R. Wang *et al.*, 'Masking and Generation: An Unsupervised Method for Sarcasm Detection', in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, in SIGIR '22. New York, NY, USA: Association for Computing Machinery, Jul. 2022, pp. 2172–2177. doi: 10.1145/3477495.3531825.
[32] Godfrey, John J. and Holliman, Edward, 'Switchboard-1 Release 2'. Linguistic Data Consortium, p. 14610176 KB, 1993. doi: 10.35111/SW3H-RW02.
[33] Cieri, Christopher, Graff, David, Kimball, Owen, Miller, Dave, and Walker, Kevin, 'Fisher English Training Speech Part 1 Transcripts'. Linguistic Data Consortium, Dec. 15, 2004. doi: 10.35111/W4BK-9B14.
[34] D. Das and A. J. Clark, 'Sarcasm detection on Facebook: a supervised learning approach', in *Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct*, in ICMI '18. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 1–5. doi: 10.1145/3281151.3281154.
[35] A. Pandey and D. K. Vishwakarma, 'VABDC-Net: A framework for Visual-Caption Sentiment Recognition via spatio-depth visual attention and bi-directional caption processing', *Knowl.-Based Syst.*, vol. 269, p. 110515, Jun. 2023, doi: 10.1016/j.knosys.2023.110515.
[36] A. Pandey and D. K. Vishwakarma, 'Attention-based Model for Multi-modal sentiment recognition using Text-Image Pairs', in *2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Feb. 2023, pp. 1–5. doi: 10.1109/ICITIIT57246.2023.10068626.
[37] B. Liang, C. Lou, X. Li, L. Gui, M. Yang, and R. Xu, 'Multi-Modal Sarcasm Detection with Interactive In-Modal and Cross-Modal Graphs', in *Proceedings of the 29th ACM International Conference on Multimedia*, in MM '21. New York, NY, USA: Association for Computing Machinery, Oct. 2021, pp. 4707–4715. doi: 10.1145/3474085.3475190.
[38] B. Liang *et al.*, 'Multi-Modal Sarcasm Detection via Cross-Modal Graph Convolutional Network', in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 1767–1777. doi: 10.18653/v1/2022.acl-long.124.
[39] H. Liu, W. Wang, and H. Li, 'Interpretable Multimodal Misinformation Detection with Logic Reasoning', in *Findings of the Association for Computational Linguistics: ACL 2023*, Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 9781–9796. Accessed: Jul. 26, 2023. [Online]. Available: https://aclanthology.org/2023.findings-acl.620
[40] M. Bedi, S. Kumar, M. S. Akhtar, and T. Chakraborty, 'Multi-modal Sarcasm Detection and Humor Classification in Code-mixed Conversations', *IEEE Trans. Affect. Comput.*, pp. 1–1, 2021, doi: 10.1109/TAFFC.2021.3083522.
[41] D. S. Chauhan, D. S R, A. Ekbal, and P. Bhattacharyya, 'Sentiment and Emotion help Sarcasm? A Multitask Learning Framework for Multi-Modal Sarcasm, Sentiment and Emotion Analysis', in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, Jul. 2020, pp. 4351–4360. doi: 10.18653/v1/2020.acl-main.401.
[42] A. Ray, S. Mishra, A. Nunna, and P. Bhattacharyya, 'A Multimodal Corpus for Emotion Recognition in Sarcasm', in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, Marseille, France: European Language Resources Association, Jun. 2022, pp. 6992–7003. Accessed: Jun. 30, 2023. [Online]. Available: https://aclanthology.org/2022.lrec-1.756
[43] D. S. Chauhan, G. V. Singh, A. Arora, A. Ekbal, and P. Bhattacharyya, 'An emoji-aware multitask framework for multimodal sarcasm detection', *Knowl.-Based Syst.*, vol. 257, p. 109924, Dec. 2022, doi: 10.1016/j.201.2022.109924.

[44] A. Conneau *et al.*, 'Unsupervised Cross-lingual Representation Learning at Scale', in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, Jul. 2020, pp. 8440–8451. doi: 10.18653/v1/2020.acl-main.747.
[45] 'Common Crawl'. https://commoncrawl.org/ (accessed Jul. 16, 2023).
[46] Y. Liu *et al.*, 'RoBERTa: A Robustly Optimized BERT Pretraining Approach'. arXiv, Jul. 26, 2019. doi: 10.48550/arXiv.1907.11692.
[47] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pretraining 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)*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
[48] Q. Hou, D. Zhou, and J. Feng, 'Coordinate Attention for Efficient Mobile Network Design', in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 13708–13717. doi: 10.1109/CVPR46437.2021.01350.
[49] L. Xu, Q. Tang, J. Lv, B. Zheng, X. Zeng, and W. Li, 'Deep image captioning: A review of methods, trends and future challenges', *Neurocomputing*, vol. 546, p. 126287, Aug. 2023, doi: 10.1016/j.neucom.2023.126287.
[50] Z. Liu *et al.*, 'Swin Transformer: Hierarchical Vision Transformer using Shifted Windows', in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9992–10002. doi: 10.1109/ICCV48922.2021.00986.
[51] A. Dosovitskiy *et al.*, 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale', presented at the International Conference on Learning Representations, Oct. 2020. Accessed: Jul. 12, 2023. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy
[52] W. Wang *et al.*, 'Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions', in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 548–558. doi: 10.1109/ICCV48922.2021.00061.
[53] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, 'Training data-efficient image transformers & distillation through attention', in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, Jul. 2021, pp. 10347–10357. Accessed: Jul. 26, 2023. [Online]. Available: https://proceedings.mlr.press/v139/touvron21a.html
[54] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, 'ALBERT: A Lite BERT for Self-supervised Learning of Language Representations', presented at the Eighth International Conference on Learning Representations, Apr. 2020. Accessed: Jul. 16, 2023. [Online]. Available: https://iclr.cc/virtual_2020/poster_H1eA7AEtvS.html
[55] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, 'Language Models are Unsupervised Multitask Learners'.
[56] J. Wang *et al.*, 'GIT: A Generative Image-to-text Transformer for Vision and Language'. arXiv, Dec. 15, 2022. doi: 10.48550/arXiv.2205.14100.
[57] A. Radford *et al.*, 'Learning Transferable Visual Models From Natural Language Supervision', in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, Jul. 2021, pp. 8748–8763. Accessed: Jul. 17, 2023. [Online]. Available: https://proceedings.mlr.press/v139/radford21a.html
[58] 'Microsoft COCO: Common Objects in Context | SpringerLink'. https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48 (accessed Jul. 26, 2023).
[59] V. Ordonez, G. Kulkarni, and T. Berg, 'Im2Text: Describing Images Using 1 Million Captioned Photographs', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2011. Accessed: Jul. 17, 2023. [Online]. Available: https://papers.nips.cc/paper_files/paper/2011/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html
[60] P. Sharma, N. Ding, S. Goodman, and R. Soricut, 'Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset For Automatic Image Captioning', in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 2556–2565. doi: 10.18653/v1/P18-1238.
[61] R. Krishna *et al.*, 'Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations', *Int. J. Comput. Vis.*, vol. 123, no. 1, pp. 32–73, May 2017, doi: 10.1007/s11263-016-0981-7.
[62] S. Changpinyo, P. Sharma, N. Ding, and R. Soricut, 'Conceptual 12M: Pushing Web-Scale Image-Text Pretraining To Recognize Long-Tail Visual Concepts', in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA: IEEE, Jun. 2021, pp. 3557–3567. doi: 10.1109/CVPR46437.2021.00356.
[63] A. Vaswani *et al.*, 'Attention is All you Need', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Jul. 12, 2023. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[64] J. Lu, J. Yang, D. Batra, and D. Parikh, ‘Hierarchical Question-Image Co-Attention for Visual Question Answering’, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2016. Accessed: Jul. 16, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2016/hash/9dcb88e0137649590b755372b040afad-Abstract.html

[65] A. Paszke *et al.*, ‘PyTorch: An Imperative Style, High-Performance Deep Learning Library’, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2019. Accessed: Jul. 16, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html

[66] D. P. Kingma and J. Ba, ‘Adam: A Method for Stochastic Optimization’, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. Accessed: Jul. 16, 2023. [Online]. Available: http://arxiv.org/abs/1412.6980

[67] H. Pan, Z. Lin, P. Fu, Y. Qi, and W. Wang, ‘Modeling Intra and Inter-modality Incongruity for Multi-Modal Sarcasm Detection’, in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Online: Association for Computational Linguistics, Nov. 2020, pp. 1383–1392. doi: 10.18653/v1/2020.findings-emnlp.124.

[68] Y. Tay, A. T. Luu, S. C. Hui, and J. Su, ‘Reasoning with Sarcasm by Reading In-Between’, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 1010–1020. doi: 10.18653/v1/P18-1093.

[69] G. Kumari, A. Das, and A. Ekbal, ‘Co-attention based Multimodal Factorized Bilinear Pooling for Internet Memes Analysis’, in *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*, National Institute of Technology Silchar, Silchar, India: NLP Association of India (NLPAI), Dec. 2021, pp. 261–270. Accessed: Jul. 01, 2023. [Online]. Available: https://aclanthology.org/2021.icon-main.31

[70] S. Hochreiter and J. Schmidhuber, ‘Long Short-Term Memory’, *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[71] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[72] Y. Kim, ‘Convolutional Neural Networks for Sentence Classification’, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1746–1751. doi: 10.3115/v1/D14-1181.

[1] **Sampreetananda Bharali**
[2] **Shubham**
[3] **Sarthak Joshi**
[4] **M. Zunaid**

Optimization of Bypass Flow Introduction Angle in Dual-Throat Nozzle

Journal of Electrical Systems

Abstract: - This research paper investigates the optimal angle for introducing a bypass jet in rocket engines for thrust vectoring. Numerical simulations using a finite-volume method were performed to test a range of bypass jet angles and their effect on flow performance. Results indicate that the optimal bypass jet angle is between 40 and 50 degrees, with larger angles leading to decreased thrust. These findings demonstrate the potential for bypass jet introduction to improving thrust vectoring in rocket engines, especially in the supersonic regime. Such insights may inform the development of more efficient and effective rocket engines for various applications, including satellite launches, space exploration, and defense. Overall, this study represents a valuable contribution to the field of rocket propulsion and may drive advances in engine design and performance.

Keywords: Nozzle; Thrust vectoring; CFD; Computational Analysis; BDTN; Bypass nozzle.

I. INTRODUCTION

The design of efficient and high-performing rocket propulsion systems has been a crucial aspect of aerospace engineering for several decades. The development of rocket engines requires careful consideration of various factors, including thrust performance, fuel consumption, and operational safety. One of the critical factors that affect the thrust performance of rocket engines is the direction and intensity of the exhaust jet.

Thrust vectoring is a technique of changing the direction of the exhaust jet to control the rocket's flight path. This technique is particularly useful in situations where precise control of a rocket's trajectory is required, such as when a spacecraft is tethered, maneuvering, or landing. It can help recover or even control the stall regions of a flight along with providing STOL capabilities [1]. Thrust vectoring can be achieved by a variety of methods, including mechanical vanes, fluid jets, and nozzle formation. One promising technology that has received increasing attention in recent years is bypass jet launching. Although there are other methods to use fluidic end effectors to control an aircraft as highlighted by Wilde et al[2], this paper focuses on thrust vectoring.

Introducing a bypass jet involves introducing a secondary gas jet into the exhaust stream to change the direction and intensity of the exhaust jet. This technique can be used to achieve thrust vectoring by pointing the secondary beam in a specific direction. The introduction of bypass jets also improves exhaust gas mixing, leading to better combustion and higher thrust performance. However, the effectiveness of the bypass jet introduction for thrust vectoring depends on several factors, including the angle of introduction and the flow velocity of the secondary jet.

Therefore, determining the optimal entry angle of the bypass jet is critical to maximizing the effectiveness of thrust vectoring. Several studies have investigated the effect of entry angle on the thrust vectoring capability of a rocket engine. However, these studies have produced mixed results and there is still no consensus on the optimum angle of the introduction of the shunt beam thrust vector.

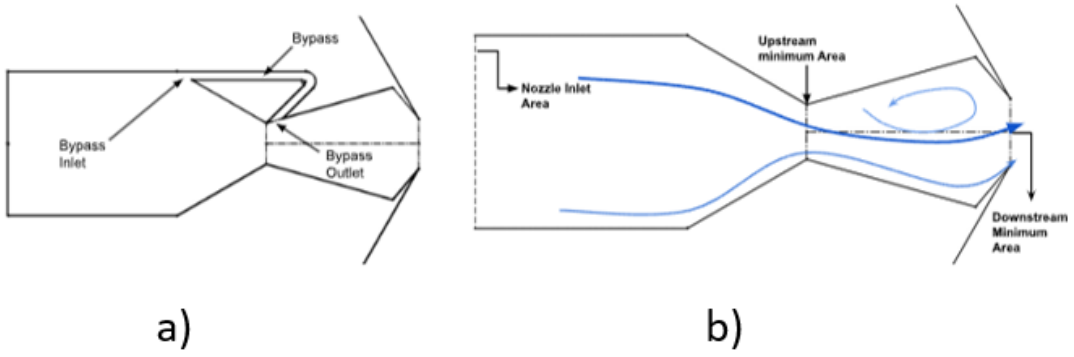


Fig. 1. a) Bypass dual throat nozzle b) Expected fluid flow lines [1]

[1] Student, Mechanical Engineering Department, Delhi Technological University, New Delhi, India. sampreetbharali@gmail.coms
[2] Student, Mechanical Engineering Department, Delhi Technological University, New Delhi, India. shmalik97@gmail.com
[3] Student, Mechanical Engineering Department, Delhi Technological University, New Delhi, India. sarthakjoshi2001@gmail.com
[4] Assistant Professor, Mechanical Engineering Department, Delhi Technological University, New Delhi, India. mzunaid3k@gmail.com
Copyright © JES 2024 on-line : journal.esrgroups.org

1.1 Literature Review

In the research paper titled "Experimental and numerical investigations of a bypass dual throat nozzle" by Gu et al.[3], the authors found that bypass flow can increase the thrust vectoring efficiency, and the thrust vectoring effect was enhanced with a decrease in the bypass slot height.

Gal-Or's book[4], "Vectored Propulsion Supermaneuverability and Robot Aircraft," discusses the concept of vector propulsion and its applications in supermaneuverability and robot aircraft. The book covers various aspects of vector propulsion and its implementation in different aircraft.

Deere et al[5], a paper titled, "Summary of Fluidic Thrust Vectoring Research Conducted at NASA Langley Research Center",discusses the different fluidic thrust vectoring concepts, their advantages, and their applications in aircraft.

Karen A. Deere et al[6], provides valuable insights into the design and performance of thrust vectoring nozzles for supersonic aircraft. The proposed dual throat nozzle is shown to be a promising alternative to traditional mechanical thrust vectoring systems and the results of the CFD simulations provide valuable guidance for future research in this area.

P.J. Yagle et al[7], in a paper titled "Demonstration of Fluidic Throat Skewing for Thrust Vectoring in Structurally Fixed Nozzles" provides a review of the existing literature on thrust vectoring and fluidic thrust vectoring, highlighting the benefits and limitations of these approaches. They then describe the design and performance of their proposed fluidic throat skewing system and present the results of experiments that were conducted to evaluate its performance. The authors find that their fluidic throat skewing system can effectively vector the thrust in a fixed nozzle and that it performs similarly to a mechanically actuated nozzle in terms of thrust vectoring angle and efficiency.

Shin et al[8] provides valuable insights into the performance and potential of the dual throat nozzle for use in thrust vectoring applications. The study's findings suggest that the nozzle has several advantages over other designs, and could be a valuable addition to the range of available thrust vectoring nozzles. Further research and development in this area may lead to improvements in the performance and control of jet engines.

In the research paper titled "Experimental study of an axisymmetric dual throat fluidic thrust vectoring nozzle for supersonic aircraft application" by Flamm et al[9]. The authors found that the nozzle exhibited excellent thrust vectoring capability and reduced drag in the off-design conditions.

Shi et al.[10], in the paper, "Investigation on flow characteristics of SVC nozzles," presents numerical simulations to study the effects of various design parameters on the flow characteristics of SVC nozzles.

Anderson et al.'s paper[11], investigates the feasibility of hybrid fluidic/mechanical thrust vectoring for fixed-exit exhaust nozzles. The authors conducted experimental tests and found that hybrid fluidic/mechanical thrust vectoring is feasible and can provide better performance compared to conventional thrust vectoring methods.

Mason and Crowther's paper[12], "Fluidic Thrust Vectoring for Low Observable Air Vehicles," The authors present a fluidic thrust vectoring concept that can be used to reduce the infrared signature of the aircraft. They also discuss the potential of fluidic thrust vectoring for reducing the radar cross-section of the aircraft.

In their paper, "Visualization and analysis on the thrust vectoring control in three-dimensional dual-throat nozzles", Wu et. al.[13], provides a comprehensive analysis of the use of TVC in three-dimensional dual-throat nozzles, using a combination of numerical simulations and experimental data. The study contributes to the understanding of the flow physics and thrust performance of TVC nozzles, and provides valuable insights for the design and optimization of such systems.

Afridi et.al.[14], in their paper "Multi-objective nozzle design optimization for maximum thrust vectoring performance" investigates the effect of various parameters such as nozzle bypass angle, convergence angle, and bypass width on the performance of the bypass dual throat nozzle. The results show that bypass width and bypass angle significantly affected thrust vectoring.

The paper titled "Computational study of axisymmetric divergent bypass dual throat nozzle" by Wang et. Al[15], provides a comprehensive analysis of the flow characteristics and performance of an axisymmetric divergent bypass dual throat nozzle, using numerical simulations. The study contributes to the understanding of the design and optimization of such nozzles and provides valuable insights for the development of advanced aircraft propulsion systems

These recent studies demonstrate the potential of bypass jet introduction for improving the thrust performance and control of rocket engines. However, there is still a need for further research to determine the optimal angle of introduction for bypass jet thrust vectoring. In this paper, we aim to build on these recent studies by investigating the optimal angle of bypass jet introduction for thrust vectoring using CFD simulations.

In this paper, we aim to investigate the optimal angle of bypass jet introduction for thrust vectoring using computational fluid dynamics (CFD) simulations. The study will involve varying the angle of introduction and flow rate of the secondary jet and evaluating the effect on the rocket engine's thrust vectoring capability and overall thrust performance. The results of this study will provide valuable insights into the effectiveness of thrust vectoring for introducing bypass jets and informing the design of future rocket propulsion systems. Additionally, the study will demonstrate the potential of CFD simulations to assess rocket engine performance and develop optimized designs.

1.2. Structure

The nozzle used to analyze BDTN is shown in Fig. 2. The cavity divergence angle is 15.69 degrees and the cavity convergence angle is 50 degrees. The nozzle throat radius is 9mm and the cavity is 26mm. The bypass incident angle is 45 degrees. Table 1 lists other geometric variables such as expansion rate (A2/A1), shunt width (We), and cavity bottom (Rz).

Table 1. Geometric Values of BDTN

Geometric Variable	Values
R_2/R_1	1.22
A_2/A_1	1.49
W_e	3.7mm
R_2	11mm
θ	30deg, 40 deg, 45 deg, 50deg, 60 deg, 70deg, 80deg, 90deg

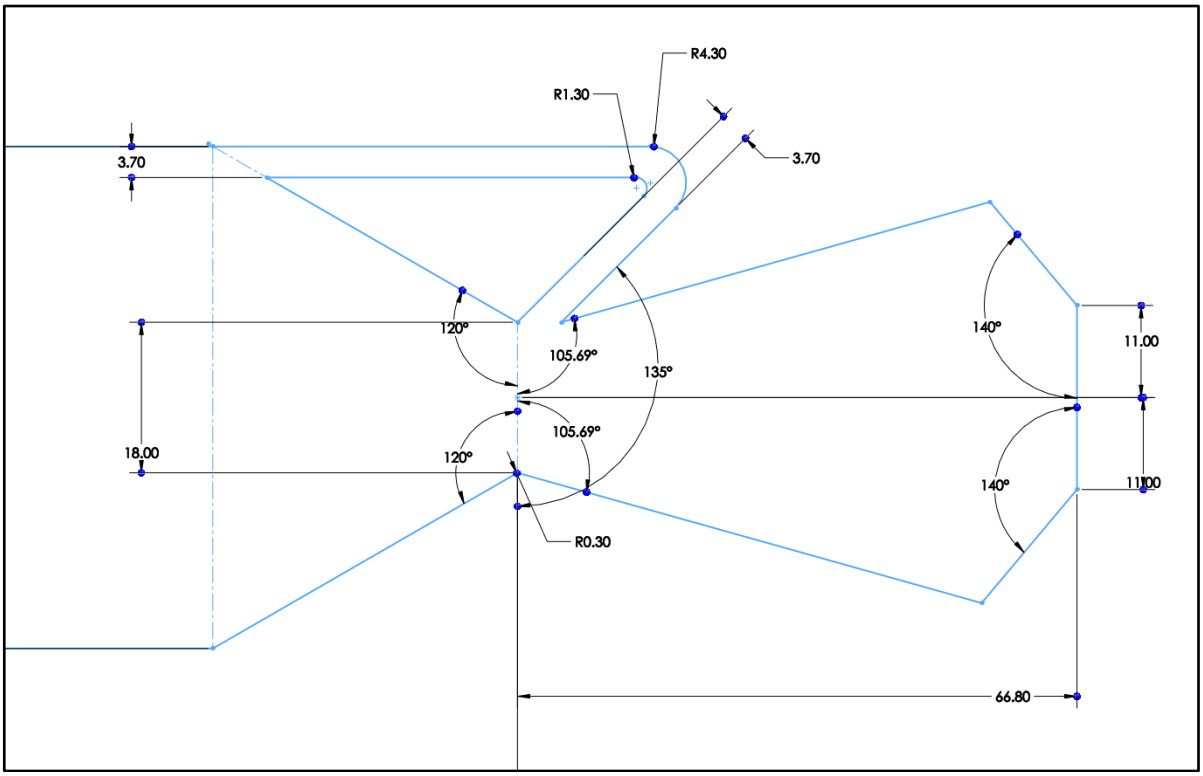


Fig. 2. BDTN geometric dimensions;

II. COMPUTATIONAL FLUID DYNAMICS

2.1 CFD Model Calibration

The Dual Throat Bypass Nozzle was designed in the Design Modeller environment of Ansys Workbench. The analysis would be on a 2D nozzle model. Next, the meshing operations were carried out with the general mesh size set as 2mm. An edge-sizing operation was performed along the central axis of the nozzle with a mesh size of 1mm, hard enforced. The total number of mesh nodes was 386974.

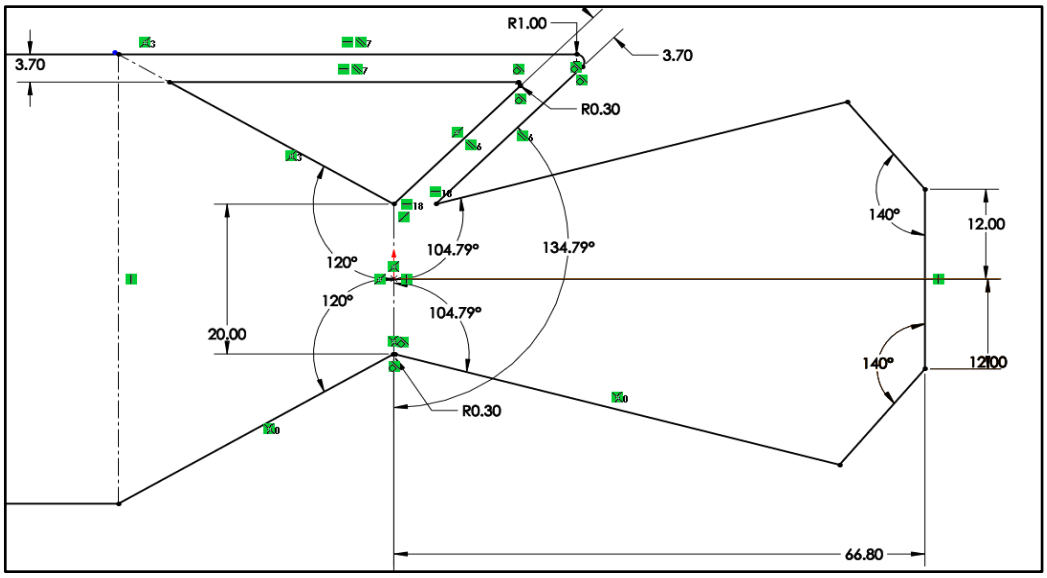


Fig.3. Model for CFD method calibration

For the analysis setup in Fluent, the following configurations were used:
Launcher configurations: Double precision, parallel solver with 4 processes and ACT enabled

- Energy equations: On
- Solver method: K-ε model (2-equation model)
- Conditions: Realizable, Standard wall model
- Air density model: Ideal gas
- Viscosity model: Sutherland
- Inlet type: Pressure Inlet
- Initial pressure at inlet: 405300 Pa
- Gauge pressure at inlet: 101325 Pa
- Output type: Pressure Output
- Reference values: Standard, Computed from the inlet
- Solution methods:
  - Turbulent Dissipation Rate: Second Order Upwind
  - Gradient: Least Squares Cell-based
  - Flow: Second Order Upwind
  - Turbulent KE: Second Order Upwind
- Residuals convergence criterion: 10-3 for all values
- Initialization: Standard, relative cell initialization from the inlet

The CFD setup uses mesh topology that is refined near boundary areas and uses edge sizing along the central axis for more accurate values along the axis, especially near throat regions. Dimensions of the calibration model are shown in fig. 3. The comparison of the data points extracted from the analysis to the experimental data points recovered from previous literature (Hamedi-Estakhrsar et al[16], Gu et al[1]) is shown in Fig.4. As it shows to correctly simulate the flow details, a standard realizable wall function with the k-e turbulent model will be used.

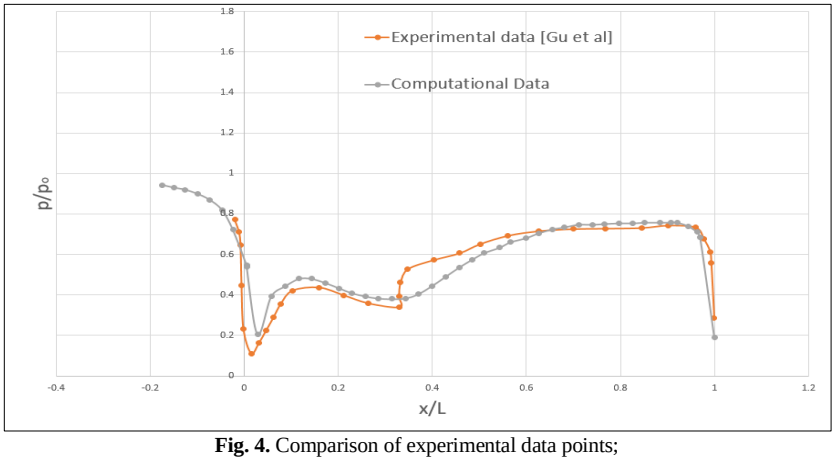


Fig. 4. Comparison of experimental data points;

2.2. Grid Independance Test

The meshing procedure followed a standard mesh size of 2mm. An edge-sizing operation was cascaded with the X-axis as the principal axis. The element size for edge sizing is 1mm, hardbound behavior. Both the general mesh element size and edge element sizing have been considered as the input parameters for the grid independence test. The maximum mach speed at nozzle outlet has been considered as the output parameter to monitor the effect of mesh element density:

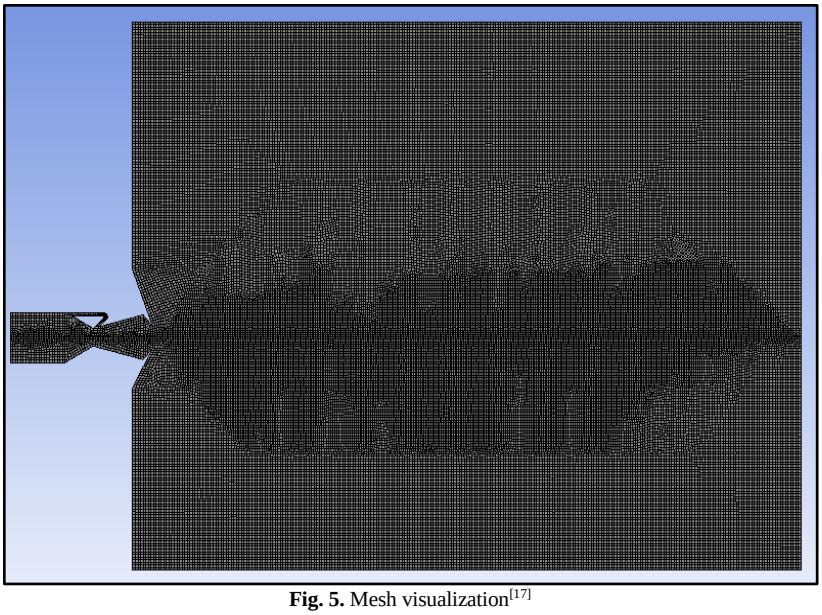


Fig. 5. Mesh visualization[17]

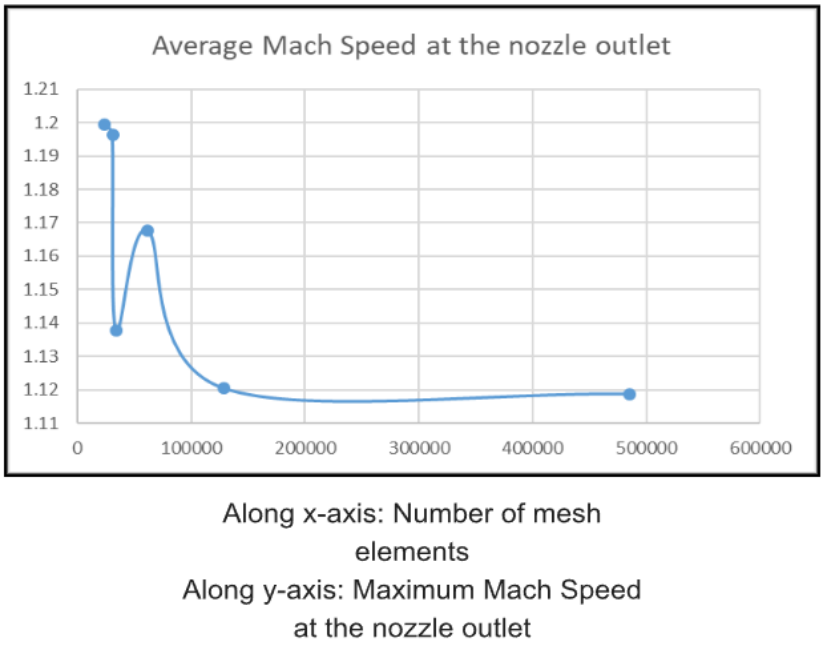


Fig. 6. Graphical representation of the grid independence study

The maximum mach number of the jet stream leaving the nozzle at the outlet has been calculated for multiple meshing configurations, with the number of mesh elements varying from 25000 to 500000. The variation of values have been plotted. From the above graphs, it is clear that the values stabilize considerably when number of elements exceed 100000. So a mesh with total of 128584 elements has been selected to get accurate and consistent results while not exceeding computational requirements of the system.

III. RESULTS AND DISCUSSION

The nozzle bypass introduction angle has been tested for 30º, 40º, 45º, 50º, 60º, 70º, 80º, and 90º with the horizontal. The Nozzle Pressure Ratio (NPR) is kept at 4 and the area ratio is 1.49. The Mach number and pressure contours are shown in figures 8 and 9. From it, we can clearly see that the maximum mass flow rate is achieved when the bypass introduction is done at an angle between 40º and 50º with the horizontal. Further fine-tuned analysis shows that the optimal angle for flow introduction is at an angle of 45º with the horizontal.

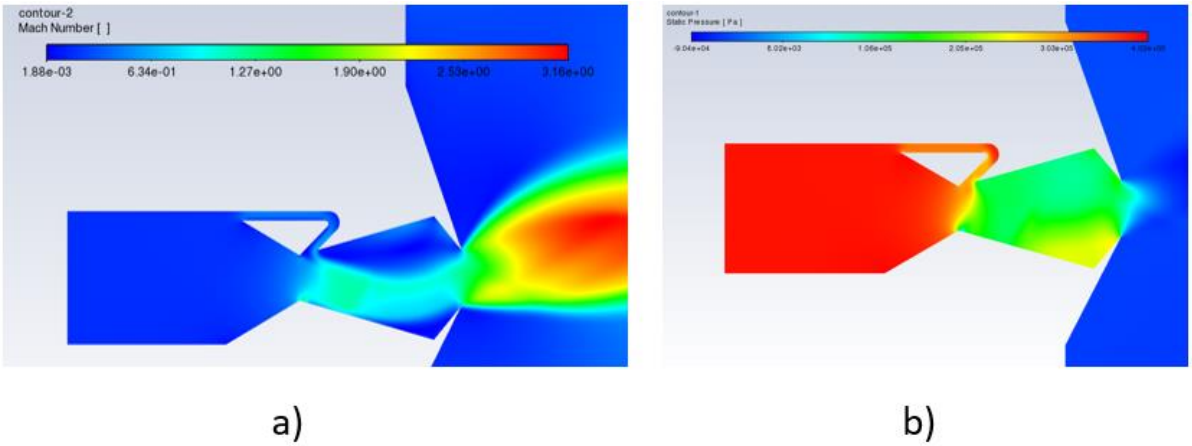


Fig. 7. Contours of (a) Mach number at bypass jet introduction angle (θ) = 45° (b) Pressure at bypass jet introduction angle (θ) = 45°

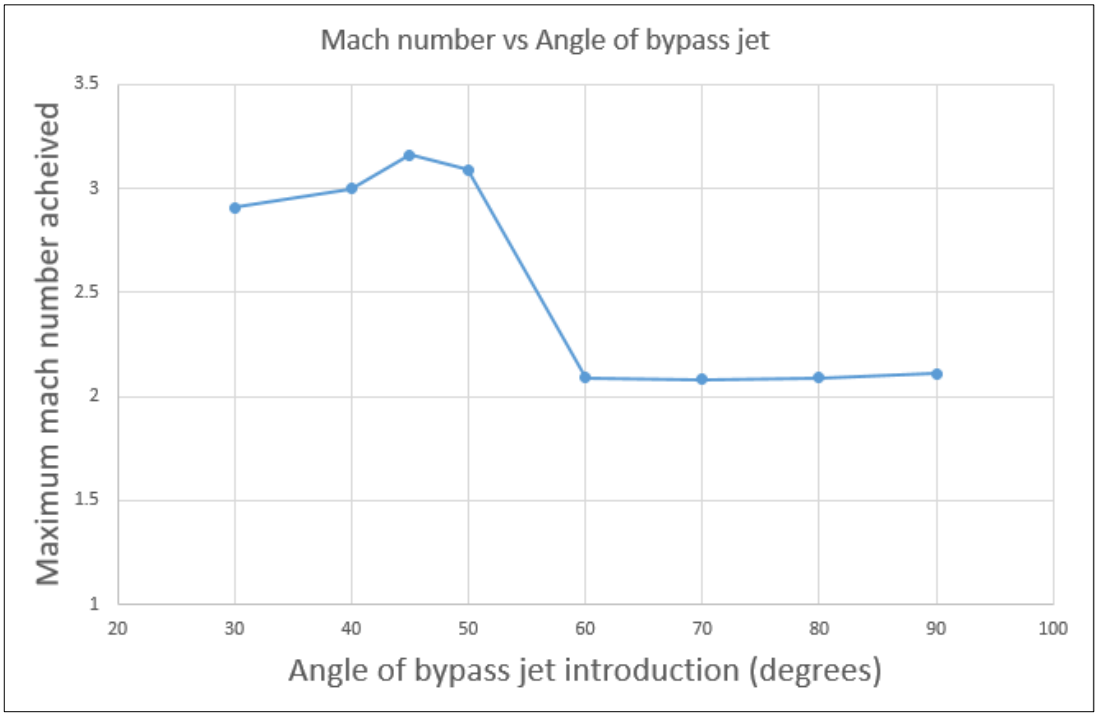


Fig. 8. Plot of maximum Mach number achieved vs the angle of bypass flow introduction

The initial angle from 30° shows the increment of jet velocity with the increase in the angle of introduction. The velocity peaks between 40° and 50°. Post 50°, the velocity reduces and stabilizes at around 60°. The Mach number values stay between 2 and 2.1 till the angle of introduction reaches 80°. Beyond that, the Mach number rises slightly till the angle is 90°. Fig. 8 shows the variation in the maximum Mach number with respect to the angle of flow introduction. The highest flow velocity was achieved by an angle of 45^0 where the maximum recorded mach number of the fluid was 3.16. The pressure contours and velocity contours for the same are shown in fig.7.

IV. CONCLUSION

The paper successfully explored and calibrated the computational method for a flow analysis of a dual throat bypass nozzle for thrust vectoring and compared it to the experimental values achieved by previous research. The experiment yielded a high degree of similarity to the experimental setup and results were validated by other publications.

The experiment to find the most effective bypass angle was conducted by creating multiple geometries and meshing then using the same approach as the calibration model. Each geometry was analyzed at NPR 4 an area ratio of 1.49. The only variable was the bypass flow introduction angle.

The comparison of various angles of bypass stream introduction as per performance parameters shows that the introduction of the flow at 45° results in minimum loss of jet velocity for the given yaw angle divergence at a nozzle pressure ratio of 4. The maximum output velocity of the fluid at this angle is 45°, and the average nozzle pressure at the outlet for the same is 103083.1 Pa.

V. ACKNOWLEDGEMENTS

The authors would like to thank the Department of Mechanical Engineering of Delhi Technological University for the support and resources provided to carry out this research.

REFERENCES

[1] J.W. Shi, Z.X. Wang, L. Zhou, X.L. Sun Investigation on flow characteristics of SVC nozzles J. Appl. Fluid Mech., 11 (2) (2018), pp. 331-342

[2] P.I.A. Wilde, W.J. Crowther, A. Buonanno, A. Savvaris Aircraft Control Using Fluidic Maneuver Effectors AIAA Paper 2008-6406 (2008)

[3] Gu, Rui, Jinglei Xu, and Shuai Guo. "Experimental and numerical investigations of a bypass dual throat nozzle." *Journal of Engineering for Gas Turbines and Power* 136.8 (2014).

[4] Gal-Or, Book.,1990, Vectored Propulsion Supermaneuverability and Robot Aircraft, Springer-Verlag, New York.

[5] Deere, K. A.,2003, "Summary of Fluidic Thrust Vectoring Research Conducted at NASA LangleyResearchCenter," AIAAPaper No. 2003-3800. 10.2514/6.2003-3800

[6] Deere, Karen, et al. "Computational study of an axisymmetric dual throat fluidic thrust vectoring nozzle for a supersonic aircraft application." *43rd AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*. 2007.

[7] P.J. Yagle, D.N. Miller, K.B. Ginn, J.W. Hamstra Demonstration of fluidic throat skewing for thrust vectoring in structurally fixed nozzles J. Eng. Gas Turbines Power, 123 (3) (2001), pp. 502-507

[8] Shin, Choon Sik, et al. "A computational study of thrust vectoring control using dual throat nozzle." *Journal of Thermal Science* 19 (2010): 486-490.

[9] Flamm, Jeffrey, et al. "Experimental study of an axisymmetric dual throat fluidic thrust vectoring nozzle for supersonic aircraft application." *43rd AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*. 2007.

[10] Shi, Jingwei, et al. "Investigation on flow characteristics of SVC nozzles." *Journal of Applied Fluid Mechanics* 11.2 (2018): 331-342.

[11] Anderson, Christopher, et al. "Investigation of hybrid fluidic/mechanical thrust vectoring for fixed-exit exhaust nozzles." *33rd Joint Propulsion Conference and Exhibit*. 1997.

[12] M.S. Mason, W.J. Crowther Fluidic Thrust Vectoring for Low Observable Air Vehicles AIAA Paper 2004-2210 (2004)

[13] Wu, Kexin, Tae Ho Kim, and Heuy Dong Kim. "Visualization and analysis on the thrust vectoring control in three-dimensional dual-throat nozzles." Journal of Visualization 24 (2021): 891-915.

[14] Afridi, Saadia, and Tariq Amin Khan. "Multi-objective nozzle design optimization for maximum thrust vectoring performance." Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering 237.3 (2023): 587-599.

[15] Wang, Y., Xu, J., Huang, S., Lin, Y., & Jiang, J. (2019). Computational study of axisymmetric divergent bypass dual throat nozzle. Aerospace Science and Technology, 86, 177-190.

[16] Hamedi-Estakhrsar MH, Mahdavy-Moghaddam H. Experimental evaluation and numerical simulation of performance of the bypass dual throat nozzle. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering. 2021;235(7):768-781. doi:10.1177/0954410020959886

[17] Aniket. (2020, November 19). You will need to manually split the faces [Comment on the online forum post *How can I make a uniform meshing in the bypass area of this dual throat nozzle?*]. Ansys forum. https://forum.ansys.com/forums/topic/how-can-i-make-a-uniform-meshing-in-the-bypass-area-of-this-dual-throat-nozzle/

© 2024. This work is published under https://creativecommons.org/licenses/by/4.0/legalcode(the“License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.

Journal Pre-proof

Overview of pro-inflammatory and pro-survival components in neuroinflammatory signalling and neurodegeneration

Shefali Kardam, Rashmi K Ambasta, Pravir Kumar



PII: S1568-1637(24)00283-6

DOI: https://doi.org/10.1016/j.arr.2024.102465

Reference: ARR102465

To appear in: *Ageing Research Reviews*

Received date: 10 July 2024
Revised date: 7 August 2024
Accepted date: 18 August 2024

Please cite this article as: Shefali Kardam, Rashmi K Ambasta and Pravir Kumar, Overview of pro-inflammatory and pro-survival components in neuroinflammatory signalling and neurodegeneration, *Ageing Research Reviews,* (2024) doi:https://doi.org/10.1016/j.arr.2024.102465

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.

© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Journal Pre-proof

Overview of pro-inflammatory and pro-survival components in neuroinflammatory signalling and neurodegeneration

Shefali Kardam[a], Rashmi K Ambasta[b], Pravir Kumar[a,**,1]

[a]Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University (Formerly DCE), Delhi 110042

[b]Department of Biotechnology and Microbiology, SRM University, Sonepat, India

[*]*Author to whom correspondence should be addressed:*
Pravir Kumar, PhD
Professor, Department of Biotechnology
Dean International Affairs
Molecular Neuroscience and Functional Genomics Laboratory
*Former Dean***,** Delhi Technological University (Formerly Delhi College of Engineering)
Former Faculty, Tufts University School of Medicine, Boston, MA, USA
Room# FW4TF3, Mechanical Engineering Building
Shahbad Daulatpur, Bawana Road, Delhi 110042; Phone: +91- 9818898622
Email: pravirkumar@dtu.ac.in; kpravir@gmail.co
https://www.ncbi.nlm.nih.gov/myncbi/1DW465XG9bp5p/bibliography/public/
https://scholar.google.co.in/citations?user=WVLI4i4AAAAJ&hl=en
ORCID ID: 0000-0001-7444-2344; **Scopus ID:** 14831447800; **ISI Research ID:** B-2164-2015

Abstract: Neurodegenerative diseases (NDDs) are identified by the progressive deterioration of neurons and a subsequent decline in cognitive function, creating an enormous burden on the healthcare system globally. Neuroinflammation is an intricate procedure that initiates the immune response in the central nervous system (CNS) and significantly impacts the expansion of NDDs. This study scrutinizes the complicated interaction between neuronal degeneration and neuroinflammation, with an appropriate emphasis on their reciprocal impacts. It also describes how neuroinflammatory reactions in NDDs are controlled by activating certain pro-inflammatory transcription factors, including p38 MAPK, FAF1, Toll-like receptors (TLRs), and STAT3. Alternatively, it evaluates the impact of pro-survival transcription factors, such as the SOCS pathway, YY1, SIRT1, and MEF2, which provide neuroprotective protection against damage triggered by neuroinflammation. Moreover, we study the feasibility of accommodating drug repositioning as a therapeutic approach for treating neuroinflammatory disorders. This suggests the use of existing medications for novel utilization in the treatment of NDDs. Furthermore, the study intends to reveal novel biomarkers of neuroinflammation that contribute fundamental observation for the initial detection and diagnosis of these disorders. This study aims to strengthen therapy interference and augment patient outcomes by combining ongoing data and evaluating novel therapeutic and diagnostic approaches. The goal is to devote the growth of an effective strategy to reducing the impact of neuroinflammation on neuronal protection in NDDs.

Keywords: Neuroinflammation, Neurodegenerative diseases, Pro-inflammatory transcription factors, Drug repositioning, Biomarkers

1. Introduction

Neurodegenerative disorders (NDDs) are complex disorders characterized by the progressive loss of neurons and the gradual deterioration of several parts of the nervous system. As the elderly population increases, the prevalence of these diseases is also on the rise. Genetic alterations cause the malfunctioning of proteins, linking them to Parkinson's disease (PD), Alzheimer's disease (AD), spinal muscular atrophy, Huntington's disease (HD), spinocerebellar ataxia, and prion disease among the most common neurological disorders. The exact etiology of these disorders remains uncertain; however, it is hypothesized that a multifaceted interplay between genetic, epigenetic, and ecological variables may be involved. Currently, no effective therapies can slow, halt, or prevent these diseases. However, research on the molecular mechanisms underlying the pathogenesis of NDDs is ongoing. Neuroglia plays a fundamental role in the pathophysiology of these diseases, and lifestyle changes can enhance their neuroprotective capabilities (Ciaccio and Agnello, 2023,Zaib et al., 2023 and Verkhratsky and Butt, 2023). Neuroinflammation is an umbrella term for inflammation of the nervous system, mainly the central nervous system (CNS). Numerous complex disorders, including neuropsychiatric disorders, cancer, and type 2 diabetes, have been interlinked with it (Asslih et al., 2021).

In NDDs, pro-inflammatory transcription factors, including AD and PD, play a crucial role (Jin et al., 2019). The nucleotide-binding oligomerization domain (NOD), leucine-rich repeat (LRR), and NLR family pyrin domain-containing protein 3 (NLRP3) inflammasome complex is interlinked with pro-inflammatory signalling in PD and AD (Cornut et al., 2020). Inflammasomes are sophisticated molecular complexes that show significant inflammatory properties and are strictly regulated at the transcriptional stage. Neuroinflammation biomarkers are widely recognized as a substantial indication of the progression of NDDs. This emphasizes the fact that inflammatory responses vary constantly and can significantly affect the progression of diseases. Several studies have identified that dysregulation of immune components, including chemokines, cytokines, and reactive oxygen species (ROS), can give rise to neuroinflammatory responses and lead to the damage and death of neurons (Heneka et al., 2015). New investigations on the impact of neuroinflammation in NDDs have offered highlights into the intricate signalling processes and cellular interactions that cause neuroinflammation quickly. Particular cell types, such as astrocytes, microglia, and invading immune cells, have been noted as mediators of the harmful effects of neuroinflammation (Glass et al., 2010).

Advancements in neuroimaging technologies have allowed the direct observation and assessment of neuroinflammation in confined laboratory environments. This has facilitated a comprehensive knowledge regarding the advancement of diseases and their fundamental causes. Research, such as the study "Imaging neuroinflammation in neurodegenerative disorders," has disclosed that biomarkers and the activation of microglia, which are connected to neuroinflammation, can be detected using MRI scans and PET. Additionally, these scans can identify mutations in the functioning of the BBB. This non-invasive evaluation and analysis process has shown remarkable efficiency (Varley et al., 2015).

Journal Pre-proof

Research has shown that the existence of neuroinflammation and the impairment of the BBB might result in neurodegeneration. A recent study has elucidated an improved explanation of the interactions across these pathways and uncovered the fundamental causes of neuron degeneration. The study intends to investigate the mutual interaction between inflammatory cytokines and the various components of the BBB. Moreover, it highlights the detrimental effect of BBB destruction on neuroinflammatory responses and the sustainability of neurons (Sweeney et al., 2018). It is crucial to acquire an extensive understanding of the complex interactions between neurons and immune cells in the nervous system to fully comprehend the significance of neuroinflammation in the development of NDDs. A study has illustrated the contradictory functions of microglia in mediating neuroinflammatory responses, along with how these reactions can either promote neuronal damage or provide neuroprotection in various scenarios (Liddelow and Barres, 2017).

2. Neuroinflammation in Neurodegenerative Diseases

The global rise in life expectancy is indirectly increasing the worldwide socioeconomic burden of neurodegenerative conditions, including AD, ALS, and PD. The exact pathogenic mechanisms of NDDs prevail partially because of a finite understanding of their pathophysiology. Neurodegeneration is believed to be triggered by various factors, such as genetics, environment, and internal factors. Standard pathogenic processes include aberrant protein dynamics, oxidative stress (including ROS), mitochondrial malfunction, and DNA damage (Jellinger, 2010). Neurodegenerative processes involve neuroinflammation, which is a protective mechanism originally intended to eliminate or hinder certain infections in the brain (Wyss-Coray and Mucke, 2002a). Inflammatory responses are advantageous as they facilitate tissue regeneration and eliminate cellular waste. Inflammation is a crucial part of the healing process, but if it continues for an extended period, it can impede the regeneration of injured tissue. These persistent inflammatory responses can be caused by endogenous factors (including protein aggregations and genetic mutations) or environmental factors (specifically infection, trauma, and drugs) (Glass et al., 2010). These persistent inflammatory responses include microglia (Kwon and Koh, 2020). Neurodegeneration refers to the gradual deterioration of neurons, resulting in both functional and mental impairments over an extended period (Campbell et al., 1999). The causes of this condition are not fully understood, but the rate of degeneration tends to increase with age, particularly in the middle to late adult years (Guo et al., 2022a). Neurodegenerative disorders such as AD, MS (Multiple Sclerosis), PD, and ALS following viral infections can directly damage neurons by directly killing or inducing apoptosis (Amor et al., 2010). In MS, the pathologies include BBB permeability, myelin loss, axon damage, glial scar formation, and inflammatory cells, mainly lymphocytes, in the central nervous system (CNS) (Jadidi-Niaragh and Mirshafiey, 2010). Myelin loss in MS leads to clinical symptoms such as neuropathic pain (pain and paralysis), muscle spasms, optic neuritis, and optic neuritis (Chastain et al., 2011).

In the CNS, microglia continuously explore the microenvironment by producing factors that interact with astrocytes (Sofroniew and Vinters, 2010) and neurons (Wyss-Coray and Mucke, 2002b), especially in response to pathogenic invasion or tissue damage, resulting in an

inflammatory response that is further mediated through the immune system to initiate tissue repair (Lull and Block, 2010a). Inflammation in tissue pathology, which may lead to the generation of neurotoxic factors that amplify disease states, suggests the persistence of inflammatory signals or the failure of standard resolution mechanisms (Das Sarma, 2014). Therefore, in neurodegenerative disorders, particular stimuli that provoke inflammation converge on mechanisms that are accountable for detecting, transposing, and enhancing the inflammatory mechanisms that generate neurotoxic mediators (Fig. 1). TNF-α, IL-6, and other pro-inflammatory cytokines can be produced by neurons and glial cells (CHEN et al., 2016).

Microglia's involvement in neuroinflammation is multifaceted, with both beneficial and harmful impacts. Microglia play a crucial role in upholding balance in the CNS by constantly surveying the environment and aiding in the removal of cellular debris and protein aggregates, like amyloid-β plaques in AD, to prevent the accumulation of harmful substances (Linnartz and Neumann, 2013). Microglia are also important in the process of synaptic pruning as the brain develops and throughout adulthood, ensuring the maintenance of healthy neural circuits (Paolicelli et al., 2011), they release neurotrophic molecules like brain-derived neurotrophic factor (BDNF) to promote the survival, growth, and adaptability of neurons (Parkhurst et al., 2013). Moreover, in specific conditions, microglial activation may become detrimental. Persistent activation results in the continuous release of pro-inflammatory cytokines like TNF-α, IL-β, and IL-6, worsening neural damage and playing a role in the advancement of diseases such as PD (McGeer and McGeer, 2008). Furthermore, activated microglia produce high levels of ROS and nitric oxide, causing oxidative stress and damaging neuronal components, which is important in the pathogenesis of NDDs (Block et al., 2007). Moreover, microglial activation can worsen excitotoxicity, which is especially relevant in diseases like ALS, where excessive glutamate release causes the overactivation of NMDA receptors, a subsequent calcium influx, and neuronal damage and death (Dong et al., 2009a).

Glycoprotein nonmetastatic melanoma protein B (GPNMB) exerts a noteworthy influence on neurological disorders such as PD, ALS, and AD. GPNMB exists in microglia next to amyloid-β plaques in AD, perhaps promoting neuroinflammation (Doroszkiewicz et al., 2023a). GPNMB shows specific expression patterns in motor neurons damaged by ALS, attributing its implication to the disease's transcriptional profile. In PD, the GPNMB gene is related to a specific locus on chromosome 7, which enhances the feasibility of inheriting PD. GPNMB has a direct interaction with α-synuclein and plays an essential aspect in the internalization of α-synuclein and the development of pathological changes in neurons (Diaz-Ortiz et al., 2022 and Pizarro Galleguillos et al., 2022). In addition, GPNMB promotes the stability of neurons after neonatal hypoxic-ischemic encephalopathy (NHIE) by using alternative splicing mechanisms (Chen et al., 2022). GPNMB reveals promise as a possible treatment target and a biomarker in distinct NDDs.

A particular type of microglia has been identified and titled "disease-associated microglia" (DAM) as it is compact with NDDs. The transcriptional modification of the microglia was induced by the existence of "apolipoprotein E" (APOE) and "triggering receptor expressed on myeloid cells 2" (TREM2). A study has disclosed that DAM carries a molecular mechanism that increases the synthesis of genes responsible for inflammatory molecules like Ccl2 and Itgax, although simultaneously reducing the production of genes responsible for homeostatic

molecules like Hexb, Cx3cr1, Trem119, and Csf1r. Moreover, it has been verified that DAM boosts the expression of glycoprotein nonmetastatic melanoma protein B, also known as GPNMB (Saade et al., 2021).

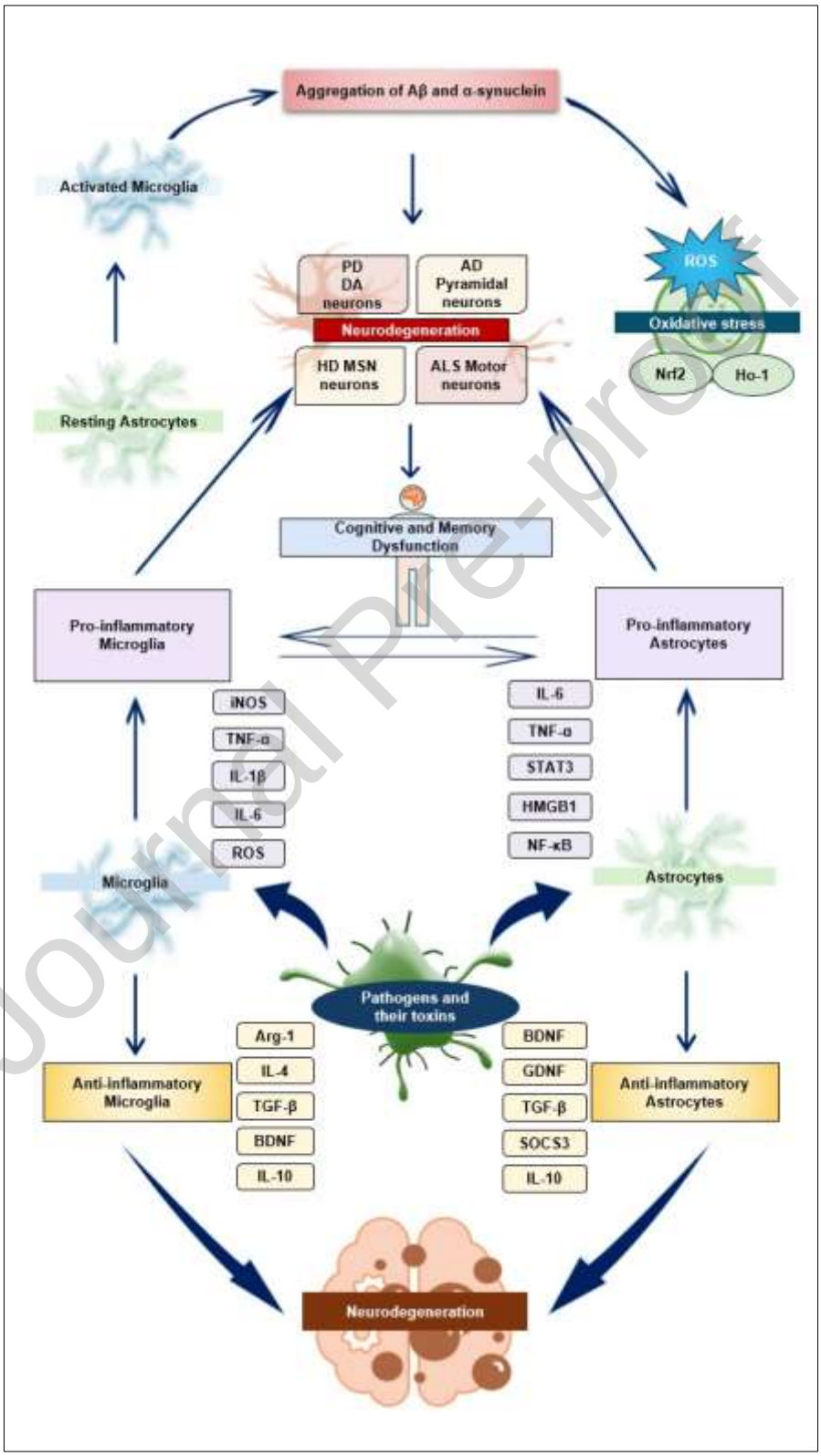


Fig. 1 The figure demonstrates the polarisation of microglia and astrocytes concerning microbial infection and their involvement in the progression of neurodegeneration. Microglia and astrocytes undergo alterations to their

characteristics in response to microbial infection, leading to their activation and polarization. When glial cells become active, they release pro-inflammatory cytokines and neurotoxic substances. These compounds encourage the development of neuroinflammation and disrupt neurons, leading to the development of dementia.

2.1. Mechanisms Specifically Involved

Glial Activation

Activated microglia and astrocytes are essential in neuroinflammation. Upon being activated, these glial cells release pro-inflammatory cytokines (such as TNF-alpha, IL-beta, and IL-6) and chemokines, which may worsen neuronal harm and stimulate neurodegeneration (Lull and Block, 2010b). The activation of glial cells has both positive and negative effects; it is important for removing debris and fighting infections, but prolonged activation can result in the generation of harmful neurotoxic substances that can damage neurons. Moreover, nitric oxide and ROS are also released by glial cells, causing oxidative stress and neuronal damage (Dias et al., 2013,Lee et al., 2021and Lassmann and van Horssen, 2016).

Utilizing methods like anti-inflammatory drugs (e.g., minocycline) or TLR4 antagonists could lessen neuroinflammation and safeguard neurons by controlling persistent microglial and astrocytic activation (Maurya et al., 2021).

Mitochondrial Dysfunction

Many neurodegenerative diseases are characterized by dysfunctional mitochondria. Malfunctioning mitochondria create too much ROS, leading to oxidative stress and harming cell parts such as protein, lipids, and DNA. This oxidative stress may also trigger glial cells, leading to a harmful cycle of inflammation and neurodegeneration (Guo et al., 2022b). Dysfunctional changes in mitochondrial dynamics, such as splitting and merging, are also a factor in the susceptibility and decline of neurons (Youle and van der Bliek, 2012).

Antioxidants like coenzyme Q10 and N-acetylcysteine, as well as substances that boost mitochondrial function such as PGC-1α activators, can help reduce oxidative stress and protect the health of neurons (Reddy and Beal, 2008). Therapies that target the mitochondria, like MitoQ, show promise in preserving mitochondrial function and averting neural injury (Jin et al., 2014).

Protein Aggregation

Irregular protein dynamics, such as folding incorrectly and clumping together, play a key role in neurodegenerative diseases. Examples include amyloid-β plaques in AD, tau tangles in AD, and α-synuclein aggregates in PD leading to microglial activation and resulting inflammatory reactions. These protein aggregates demonstrate the intricate interactions between these processes by acting as both a cause and an effect of neuroinflammation (Doroszkiewicz et al., 2023b). Furthermore, protein aggregates can interfere with autophagy and proteasomal breakdown processes, which might upset cellular homeostasis (Wen et al., 2023).

Treatments like immunotherapy (e.g., anti-tau or anti-amyloid antibodies) or small compounds (e.g., inhibitors of protein aggregation) that target protein aggregation and improve clearance systems can lessen the impact of harmful protein aggregates (Congdon and Sigurdsson, 2018).

Journal Pre-proof

Misfolded protein removal can also be achieved by boosting autophagy with mTOR inhibitors or proteasome function with proteasome activators (Rubinsztein et al., 2011).

Excitotoxicity

Neuronal damage and death result from excitotoxicity, which is brought on by excessive glutamate release and poor glutamate uptake. Glial cells are in charge of preserving glutamate homeostasis, especially astrocytes. Dysfunctional astrocytes in NDDs are unable to eliminate excess glutamate, which leads to an increase in calcium influx, an overactivation of NMDA receptors, and the activation of pathways that cause cell death (Belov Kirdajova et al., 2020). In addition to causing oxidative stress and mitochondrial malfunction, chronic excitotoxicity damages neurons (Dong et al., 2009b).

Excitotoxic neuronal damage can be prevented by agents that promote glutamate absorption by astrocytes (e.g., β-lactam antibodies) or modify glutamate receptors (e.g., NMDA receptor antagonists) (Rothstein et al., 2005). Furthermore, therapeutic potential can be obtained by focusing on downstream signalling pathways that are triggered by excitotoxicity, such as oxidative stress and calcium signalling (Hardingham and Bading, 2010).

2.1. Crosstalk Between Neuroinflammation and Neuronal Degradation

Neuroinflammation and neuronal degeneration are mutually independent. The progression of NDDs is characterized by neuroinflammation, which is recognized by the activation of immune constituents like microglia and the production of pro-inflammatory mediators (Tiwari et al., 2023). An essential factor in the deterioration of neuronal cells in NDDs, including PD, HD, AD, and multiple sclerosis (MS), is neuroinflammation (Dora et al., 2021). Tau proteins tend to form abnormal structures in the brain, which may result in the development of NDDs. Microglial cells are major components in the immune responses within the CNS. These have a significant role in the progression of NDDs. Inadequate communication between neurons and microglia can cause cellular damage, therefore contributing to the progression of tauopathies. The simultaneous occurrence of synucleinopathies and tauopathies is extensively defined, as numerous studies have shown the presence of neurofibrillary tangles (NFTs), including pathological tau in the substantia nigra of patients dealing with PD. This convergence indicates that these neurodegenerative pathologies have similar functions, especially at the level of neuroinflammatory pathways and protein misfolding. Tau proteins play a crucial role in stabilizing microtubules in nerve cells, which is essential for the maintenance of cell structure. Nevertheless, strange interactions across microtubules can result in a group of NDDs known as tauopathies. The aggregation of tau protein causes the formation of NFTs, which impede synaptic communication and contribute to neuronal dysfunction. Additionally, the simultaneous appearance of NFTs and Lewy bodies in distinct brain regions emphasizes the linkage of these disorders (Cortés et al., 2018).

The eradication of functioning neurons is promoted by the formation of inflammatory mediators and the activation of the complement system, microglia, and inflammasome (Baidya et al., 2021). On the other hand, the degradation of neurons can promote neuroinflammation.

Journal Pre-proof

The blood-myenteric barrier (BMB) in colitis is damaged by macrophages, resulting in a loss of its barrier function and the consequent reduction in inflammatory stimuli affecting the myenteric plexus, leading to the development of neuroinflammation (Confettura et al., 2022). Hence, focusing on the neuroinflammatory process to effectively control NDDs from the interlinked relationship between neuroinflammation and neurodegeneration (Confettura et al., 2022).

3. Pro-inflammatory Transcription Factors in NDDs

3.1. STAT3 (Signal Transducer and Activator of Transcription 3) in NDDs

A specific class of transcription factors consists of the signal transducer and activator of transcription (STAT)-3. The discovery of a DNA-binding protein was initially formed in 1994. This protein has an eminent specificity regarding its ability to adhere to the IL-6-responsive region situated in the interior of the promoter region of acute-phase genes. These genes are located in hepatocytes that have been stimulated with IL-6. Moreover, the role of STAT-3 as a protein that interacts with DNA in terms of epidermal growth factor was discovered individually (Aggarwal et al., 2009).

STAT3 has been identified as a crucial factor in NDDs such as schizophrenia, intellectual disabilities, and autism spectrum disorders (ASD). Multiple studies have elucidated the different functions of STAT3 in monitoring neuronal activity and synaptic plasticity, which eventually affects the brain's growth and function.

Transcription factors, like STAT3, are crucial in promoting the growth, invasion, and replication of cancer cells (Johnson et al., 2018). Cholangiocarcinoma is a tremendously metastatic form of cancer characterized by its destructive nature (Loeuillard et al., 2019). The cytokine-induced proliferation and migration of cholangiocarcinoma KKU-100 cells are deteriorated by MYR by inhibiting the STAT3 pathway. Furthermore, MYR suppresses genes linked with downstream metastasis and inflammation, notably MMP-9, cyclooxygenase, intercellular adhesion molecule-1, and inducible nitric oxide synthase (Song et al., 2021).

3.2. FAF1 (Fas-associated Factor 1) in NDDs

FAF1 is a long non-coding RNA surrounded by the Fas death-inducing signalling complex (Song et al., 2009). It plays a crucial part in multiple biological processes, such as programmed cell death, inflammation, cell proliferation, and protein homeostasis (Song et al., 2005 and Ryu et al., 2003). FAF1 is attached with many NDDs, along with PD. It is linked with α-synuclein, a protein that has a connection with PD. Increased expression of FAF1 shows the proliferation of α-synuclein and inhibits its clearance, aggravating PD (Han et al., 2023 andG. Park et al., 2020). FAF1 could have a notable effect on the constant decrease in dopaminergic (DA) exertion in PD. FAF1 is mainly confirmed as a substrate protein for parkin, and it serves as a ubiquitin E3 ligase. When parkin is in deactivated mode, it generates an escalation in the expression of FAF1, resulting in the excessive death of dopaminergic neurons (Sul et al., 2013

andH. S. Park et al., 2020). Postmortem investigation of the brains of individuals having PD discloses a crucial existence of FAF1 expression in the frontal lobe. This expression is powerfully connected to the accumulation of Lewy bodies and the reduction in dopamine transporter expression (Betarbet et al., 2008).

On disclosure to ischemia stress, it was discovered that FAF1 shows prior to JNK1, resulting in the disability of mitochondrial function and the beginning of necrotic cell death pathways. Knocking down FAF1 stops the activation of JNK1 produced by ischemia stress, therefore safeguarding cells as opposed to ischemic cell death (Yu et al., 2018). Moreover, FAF1 is released from SH-SY5Y cells by exocytosis and BFA-resistant secretory pathways, although it is accomplished from nearby cells across endocytosis. Extracellular FAF1 instigates cell death in adjacent cells by simultaneously activating necrotic and apoptotic pathways, encouraging neuronal death through cell-to-cell transmission. FAF1 has been discovered in exosomes, implying its function in regulating exosome production (G. Park et al., 2020).

3.3. TLRs (Toll-like Receptors) in NDDs

TLRs are a family of receptors that identify particular patterns and are located in various types of cells all around the body, like immune cells, neurons, astrocytes, and microglia in the CNS (Dabi et al., 2023). TLRs, also known as type I transmembrane receptors, are proteins that have been kept during the evolutionary history of both invertebrates and vertebrates. TLRs are illustrated by an intracellular toll/IL-1 receptor-like (TIR) domain that initiates downstream signal transduction and an exterior leucine-rich repeat domain that is recommended for identifying the ligands recognized as PAMPs/DAMPs (Piccinini and Midwood, 2010). TLR expression is maintained by alteration in response to infections, cytokines, and environmental stimuli. TLR is expressed in both immunological and non-immune cells. The TLR family is located on the surface of cells or within intracellular regions. It includes 12 members in mice (TLR11–TLR13) and ten members in humans (TLR1–TLR10) (Pascual et al., 2021). Many neurological disorders are connected to damage-associated DAMPs and damaged brain tissue. Therefore, this evokes a reaction from TLRs and begins the activation of supplementary signalling pathways. These processes result in the integration of chemokines and pro-inflammatory cytokines, leading to the neuroinflammation associated with dementia.

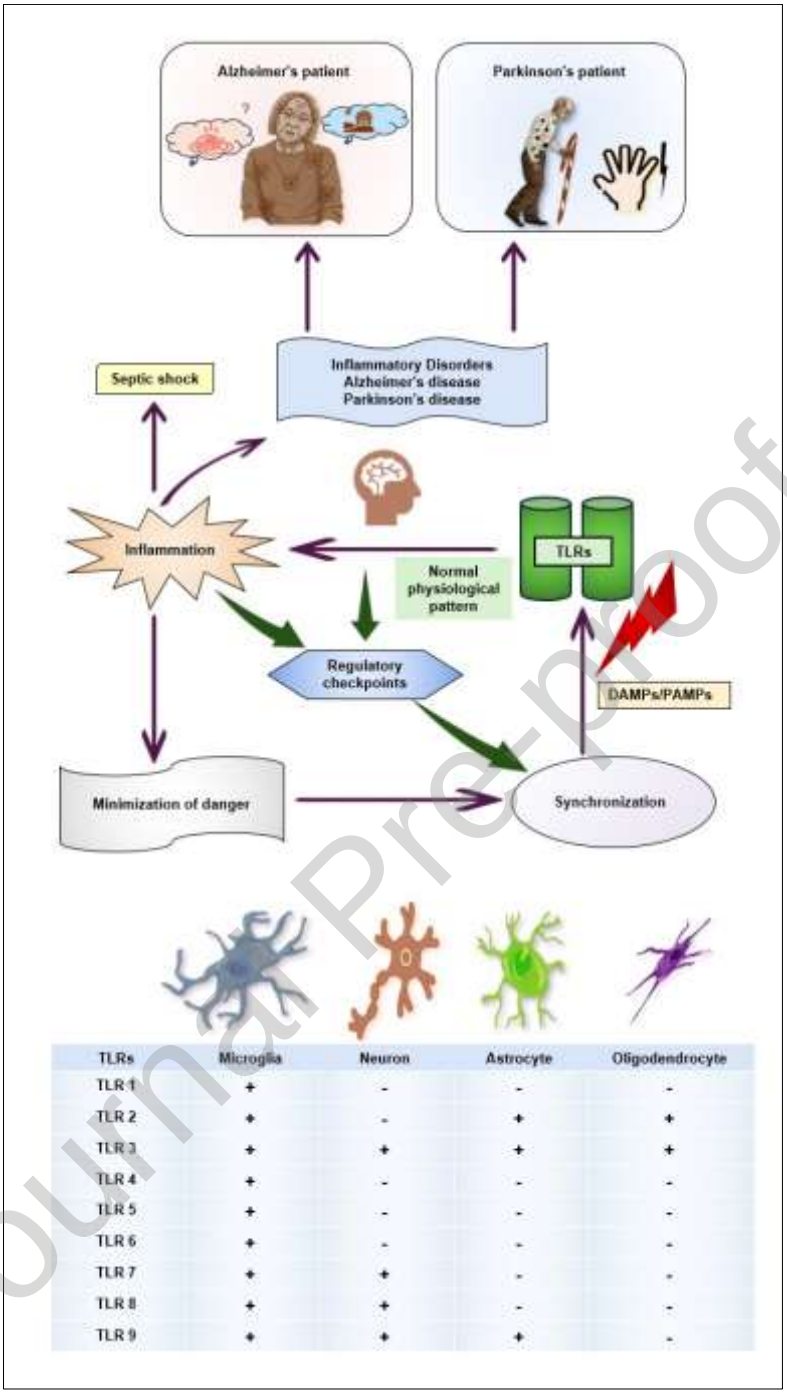
AD is a common neurodegenerative condition, considerable by a progressive deterioration in cognitive capacity. Dementia is initially obtained by it. The pathophysiology of AD consists of numerous pathways, along with neuroinflammation, which is regulated by immune cells like microglia and astrocytes. TLRs are the main components of neuroinflammation in AD. They are responsible for identifying infections and regulating the immune response. TLRs, along with TLR2, TLR4, and other related receptors, have abnormal regulation in the brains of persons afflicted by AD. Although some studies demonstrate that the activation of TLR may enhance the destruction of Aβ and reduce neuroinflammation, other studies state that excessive activation can accelerate the generation of pro-inflammatory cytokines and aggravate neuronal damage. TLR4 has been correlated with both beneficial and adverse effects in the pathophysiology of AD. The activation of microglia is required for the elimination of Aβ, but

Journal Pre-proof

uncontrolled and sustained activation can lead to neurotoxicity and cognitive worsening. Likewise, the absence of TLR2 had different outcomes, as many studies suggest a late accumulation of Aβ and a reduction in inflammation, whereas some indicate a decline in neurobehavioral functioning (Pascual et al., 2021). Research has indicated that genetic mutation of TLR5 may have a negative impact on AD. Additionally, the production of TLR9 has been established to boost the ability of microglial cells to concede Aβ, showing potential protective function (Iribarren et al., 2005). Moreover, further research is mandatory to completely understand the exact mechanism and variable effects of TLRs at multiple stages of AD.

PD is characterized by high levels of α-synuclein and loss of dopaminergic neurons, leading to motor symptoms. TLRs play an important part during PD. Accumulation of α-synuclein activates glial TLR4, promoting pro-inflammatory cytokines and neuronal injury. TLR4 inhibition can be beneficial in reducing neuroinflammation in PD. Considering the correlation between the elevated levels of TLR2 in neurons and the production of inflammatory cytokines along with the accumulation of α-synuclein, it may be concluded that TLR2 plays a crucial role in PD. Furthermore, the shortage of TLR7 and TLR8 in the zone of impact reduces the large number of T-cells that accumulate in that area. Certain types of extracellular TLRs, including TLR2, TLR4, TLR7, and TLR8, have a consequence on the progression of PD. Toll-like receptor 4 (TLR4) has the tendency to enhance the accumulation of Aβ peptide, hence provoking the neurodegeneration that is associated with PD. Moreover, the NLRP3 inflammasome, which is regulated by TLR4, leads to the advancement of PD. Studies have demonstrated that suppressing NLRP3 might help to reduce motor dysfunction and neuronal mortality. A promising approach for discovering novel therapies for PD seeks to target the inflammation that is triggered by TLRs (Pascual et al., 2021). The figure below (Fig. 2) represents the equilibrium between pro-inflammatory and anti-inflammatory responses mediated by TLRs in normal physiological conditions.

Journal Pre-proof



TLRs	Microglia	Neuron	Astrocyte	Oligodendrocyte
TLR 1	+	-	-	-
TLR 2	+	-	+	+
TLR 3	+	+	+	+
TLR 4	+	-	-	-
TLR 5	+	-	-	-
TLR 6	+	-	-	-
TLR 7	+	+	-	-
TLR 8	+	+	-	-
TLR 9	+	+	+	-

Journal Pre-proof

Fig. 2 This figure illustrates the equilibrium between pro-inflammatory and anti-inflammatory responses mediated by TLRs under normal physiological conditions. TLRs detect PAMPs/DAMPs when they penetrate the protective barrier, triggering an inflammatory response. TLRs initiate the activation of inhibitory molecules as a component of the inflammatory reaction, which can also be induced by inflammatory cytokines. The inflammatory response serves to mitigate the threat, whereas negative regulators are in place to limit excessive activation of the immune system to safeguard the host. If there is a lack of proper stimulation of TLRs, failure to propagate signals, or failure to induce inflammatory mediators, it can lead to the presence of bacteria in the bloodstream. On the other hand, if any component is overactivated, it can lead to the development of inflammatory disorders and septic shock. Additionally, the figure illustrates the behavior of TLR family constituents in CNS cells, signifying their involvement in the immunological response of the CNS.

3.4. p38 MAPK (Mitogen-Activated Protein Kinase) in NDDs

Mitogen-activated protein kinases (MAPKs) are a group of protein kinases that predominantly regulate cell survival, proliferation, and differentiation, along with other cellular processes. MAPKs mediate signal transmission from the cell membrane to the nucleus after being activated by external stimuli. There are three distinct subgroups of classical MAPK that have been identified: p38 MAPK, extracellular signal-regulated kinase (ERK), and c-Jun NH2-terminal kinase (JNK). Whereas JNK and p38 MAPK are activated in reaction to environmental challenges such as UV irradiation, osmotic stress, DNA damage, inflammation, and oxidative stress, ERK is primarily activated in response to growth stimuli. MAPKs are activated by the dual specificity kinases known as MAPK kinases (MAP2K), which phosphorylate tyrosine and threonine residues simultaneously. Among the methods to revoke the activation is via MKP (Gräb and Rybniker, 2019). Numerous NDDs, including dementia with Lewy bodies (DLB), PD, and AD, are linked with p38 MAPK (Shi et al., 2023 andIba et al., 2023). Mutations in p38 MAPK isoforms result in neuroinflammation, degradation of synapses, and the buildup of α-synuclein in DLB and PD.

Amyloid-β (Aβ) peptide aggregates and hyperphosphorylated tau protein forms are two neurodegenerative processes that characterize AD. Owing to its role in AD pathogenesis, the p38 MAP kinase pathway has gained attention as a possible target for therapy. Both transgenic AD animal models generating pathogenic amyloid-β precursor protein (APP) and neurons from AD brains have been shown to contain elevated levels of activated p38. P38 in neurons may be activated by Aβ through NMDA receptors. It has been demonstrated that tau pathology and Aβ-induced cognitive impairment can be lessened with p38α inhibition. Unexpectedly, p38α controls extracellular Aβ accumulation and BACE1 protein levels, which in turn affects Aβ processing. Nonetheless, microglia's suppression of p38α implies a function in facilitating amyloid-beta phagocytosis. The activation of the receptor for advanced-glycation end products (RAGE) also activates p38, which helps suppress long-term potentiation (LTP) in AD. In animal models, blocking RAGE inhibits p38 activation, impairment of LTP, and memory impairment (Asih et al., 2020).

Unlike p38α, p38γ functions as a neuroprotective role in AD. It significantly affects the post-synaptic sites and interacts with PSD-95, resulting in the phosphorylation of tau at Threonine-205. By interrupting the PSD-95-tau complex, it effectively hinders the adverse effects of Aβ. The unique function of p38γ in inhibiting Aβ toxicity is isolated from different p38 kinases due to its precise location (Asih et al., 2020). Ideally, targeting the p38 pathway, mainly p38α and p38γ, signifies promising for diagnosing AD by alleviating the negative impact of Aβ on cognitive function and the accumulation of tau. However, further study is crucial to deeply understand the complex association between p38 isoforms, Aβ, tau, and neuronal protection in the progression of AD. Degeneration of DA neurons in the substantia nigra pars compacta is a hallmark of PD, resulting in symptoms that are both motor and non-motor (Antony et al., 2013). Studies have demonstrated that the p38 kinase pathway plays a significant role in the pathophysiology of PD. Data from animal models and samples from PD patients have demonstrated increased p38 activity in dopamine neurons. Although inhibiting p38 has been

shown to have protective effects on these neurons, activating p38 leads to the degeneration of DA neurons through the nuclear localization of p53. Furthermore, the longevity of DA neurons and their resistance to oxidative and inflammatory stress are regulated by p38 activation. Klotho acts as a suppressor of ASK1 downstream, reducing the loss of DA neurons and premature ageing. By phosphorylating Parkin-1, p38 activation in PD also leads to mitochondrial dysfunction and the death of DA neurons. It is interesting to note that PD disease involves serotonergic dysfunction in addition to DA neurons, especially in situations where SNCA gene mutations are present. Since p38α regulates serotonin release and reuptake via serotonin transporter trafficking, it is likely a therapeutic target for symptoms related to serotonergic dysfunction in PD (Asih et al., 2020). A detailed overview of the p38 MAPK pathway is described in (Fig. 3).

Journal Pre-proof

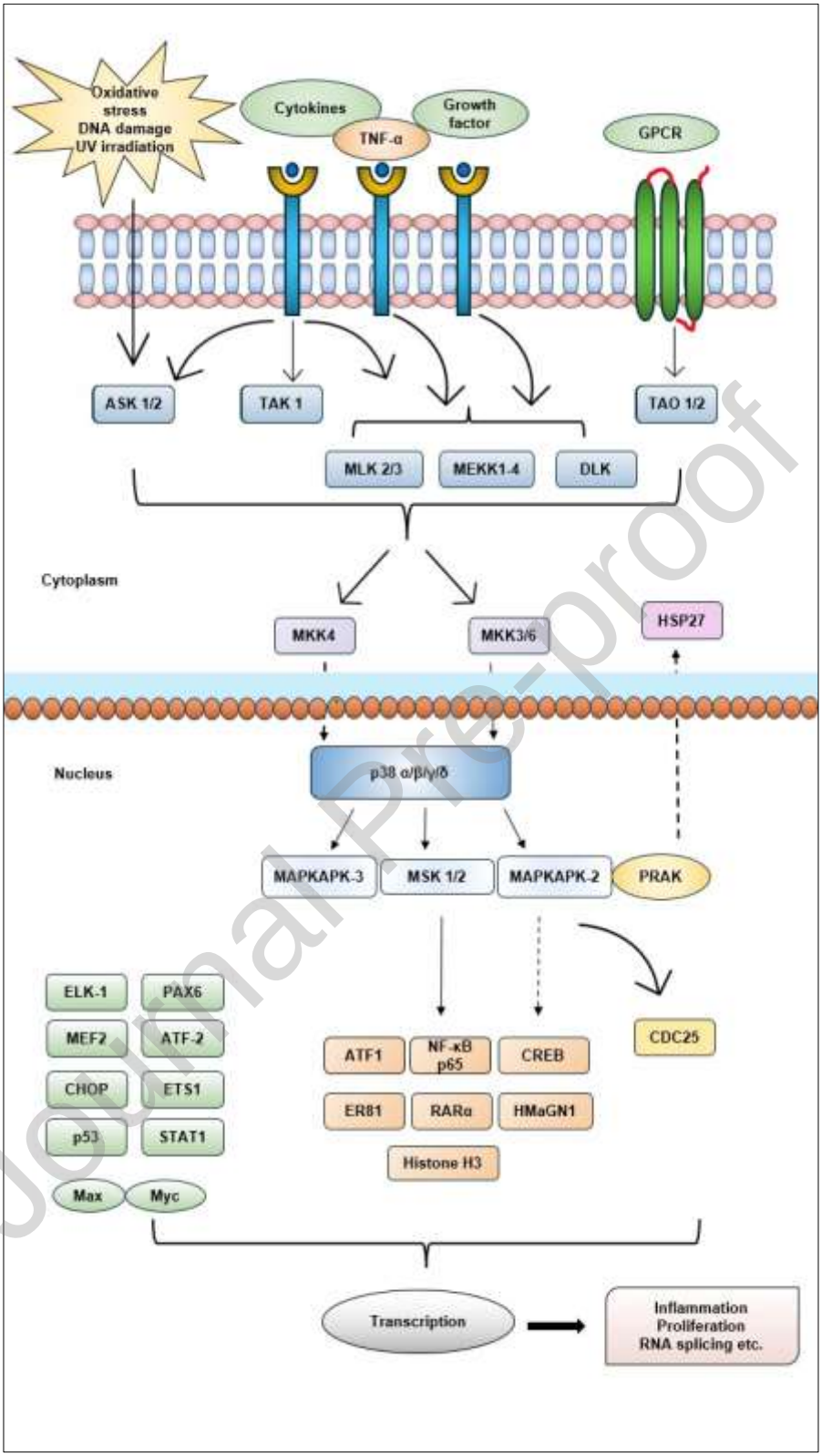


Fig. 3 The p38 MAPK pathway and its activation in both upstream and downstream aspects. The activation of the four members of the p38 MAPK family (p38α, p38β, p38, and p38) occurs in response to UV radiations, external stress, or inflammatory cytokines, resembling the JNK pathway. Upon activation, this pathway triggers the synthesis of pro-apoptotic transcription factors involved in inflammation, proliferation, apoptosis, and RNA splicing.

Journal Pre-proof

4. Pro-survival Transcription Factors in NDDs

4.1. Transcription Factor YY1 (Yin Yang1) in NDDs

A transcription factor known as Yin Yang 1 (YY1) was initially identified in 1991 by three distinct groups, each of whom gave the protein a different name depending on the molecular processes they connected it to. The name NF-E1 was given to YY1 due to its binding capacity with μE1 intron enhancer at the immunoglobulin heavy chain (IgH) (Park and Atchison, 1991). In mammalian cells, YY1 is expressed everywhere. It generates homodimers that are stabilized by RNA binding with limited selectivity (Wai et al., 2016).

The participation of YY1 in neurodevelopmental disorders has been indicated in several studies (Donohoe et al., 1999,Varum et al., 2019 andZurkirchen et al., 2019); however, the fundamental mechanisms involved in these processes are not yet completely comprehended. Conditional knockout (cKO) of Yy1 in animal models during early cortical development led to elevated levels of apoptosis and cell cycle arrest in the neuroepithelium and neural progenitor cells (NPCs) (Zurkirchen et al., 2019and Dong and Kwan, 2020). Nevertheless, the impact on non-player characters (NPCs) was diminished when the conditional knockout (cKO) was initiated during later developmental phases (Zurkirchen et al., 2019). Another work conducted by Varum and colleagues has shown that Yy1 has a stage-dependent impact on neural crest (NC) formation in cKO mice. The early KO decreased many lineages generated from the neural crest (NC). However, the late KO (occurring after embryonic day 11.5) did not show any noticeable difference in phenotype compared to the control (Varum et al., 2019). This indicates that there is a diminished reliance on mechanisms mediated by YY1 during later stages of neuronal development.

4.2. SOCS (Suppressor of Cytokine Signalling Gene) Pathway in NDDs

The protein complex known as suppressor of cytokine signalling (SOCS) has been connected to a wide range of diseases, along with NDDs. SOCS1, a constituent of the SOCS family, has a pivotal function in controlling cytokine signalling and is indispensable in averting excessive inflammation (Doggett et al., 2023). SOCS3 expression in the brain was first discovered in 2000; mRNA identification indicated that SOCS3 is probably ubiquitous in the brain, particularly the hippocampus, the granular layer of the cerebellum, the thalamus, and the basal ganglia (Cao et al., 2018).

Many cytokines use the JAK/STAT pathway for signalling, and suppressors of cytokine signalling (SOCS-1) proteins have been identified as adverse feedback inhibitors for this mechanism. Eight members (SOCS-1 through SOCS-7 and CIS) have been discovered within this protein family (Alexander and Archer, 2003). In general, STAT-dependent elements found in the appropriate promoters act as catalysts for triggering the production of SOCS proteins when cytokine receptors are activated (Auernhammer et al., 1999). SOCS proteins can inhibit subsequent receptor signalling. The SOCS family employs many strategies to do this. Upon

binding to JAKs, SOCS-1 is capable of directly inhibiting kinase activity through its SH2 domain and proximal kinase inhibitory region. However, it is speculated that CIS and SOCS-3 compete for STAT recruiting. SOCS-3 is recruited by phosphotyrosine residues found in the intracellular domain of the cytokine receptor. Further, it acts to decrease JAK activity. The SOCS box is an additional domain located at the C terminus that plays an essential role in the operation of SOCS, together with the core SH2 domain (Hilton et al., 1998). The SOCS box facilitates the interaction between elongins B and C in the E3 ubiquitin ligase complex (Kamura et al., 1998). SOCS may enhance the inhibitory effects of JAK2 by promoting ubiquitination and proteasomal degradation (Fig. 4) (Kamizono et al., 2001).

Journal Pre-proof

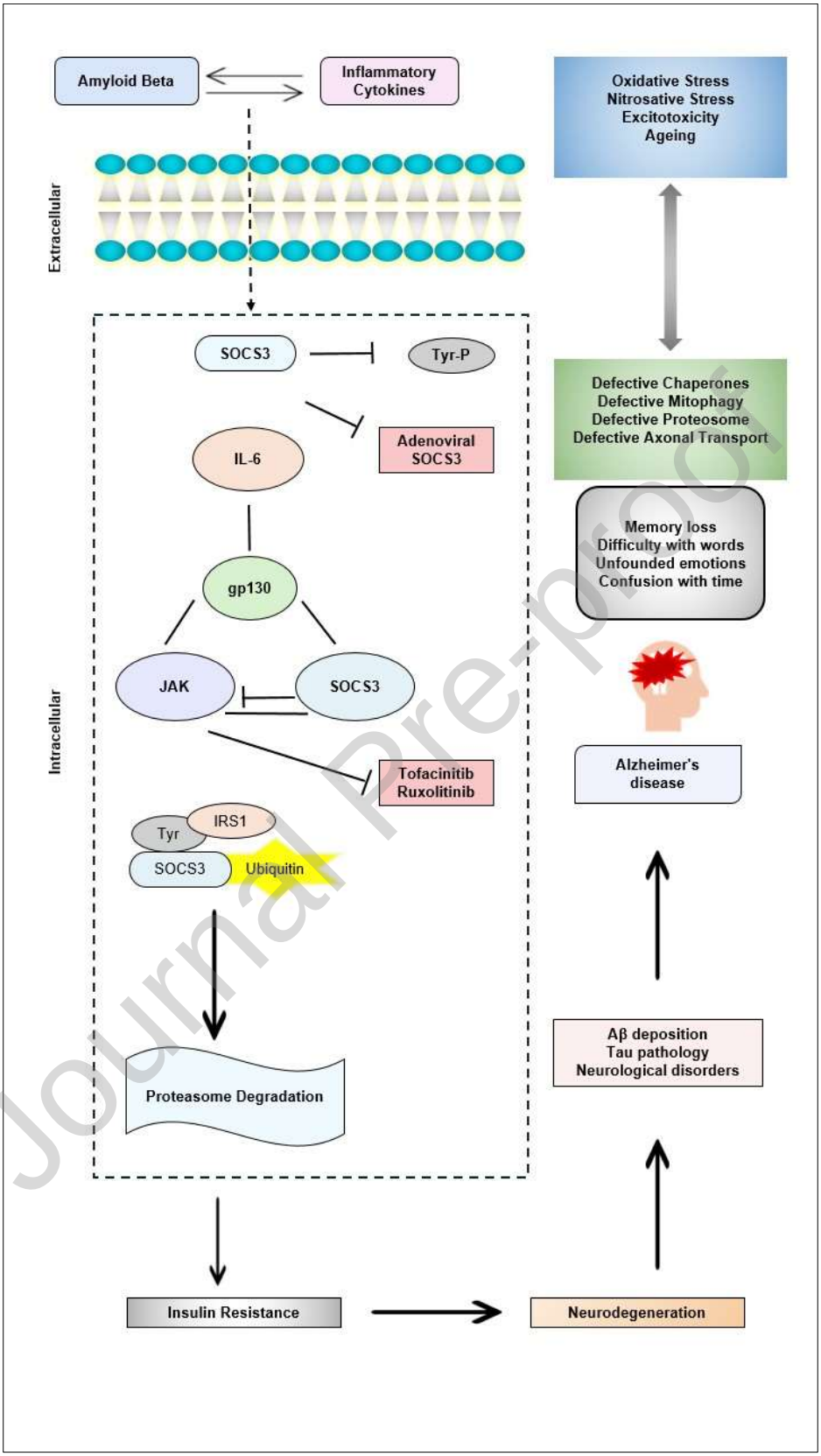


Fig. 4 Neuronal insulin signalling is regulated by SOCS3. SOCS3 expression is increased subsequently to cytokine activation: resulting in the inhibition of insulin signalling via the tyrosine phosphorylation of IRS1. Furthermore, an increase in serine phosphorylation is observed in IRS1, resulting in the emergence of inflammatory responses and neuronal dysfunction.

Journal Pre-proof

4.3. MEF2 (Myocyte Enhancer Factor 2) in NDDs

MEF2, a transcription factor, is essential for monitoring cellular homeostasis and promoting neuronal growth. It performs a vital role in different cellular regulatory networks, including those that control cell division, growth, survival, and apoptosis (Liu et al., 2022). MEF2, which regulates neuronal protection, neuronal growth, and synaptic plasticity, has been associated with an alteration in brain function that causes neuropsychiatric disorders (Lisek et al., 2023). Additionally, the protein MEF2C, which is related to the MEF2 family, has significance in the development of the brain and is focused on understanding neuropsychiatric disorders such as AD, ASD, schizophrenia, and epilepsy (Bu and Cripps, 2022). Moreover, scientific studies have demonstrated that MEF2 plays a vital role in regulating the activation of genes associated with muscle function, highlighting its significance in muscle growth and development (Zhang and Zhao, 2022).

Recent research has proved that MEF2A regulates the response of neurons to oxytocin and substantially impacts regulating mitochondrial activity in these cells. MEF2A signalling has been demonstrated to inhibit mitochondrial function, revealing a connection between this transcription factor and the metabolic condition of neurons. This dual function of MEF2A emphasizes its importance in energy control and cellular signalling, illustrating its potential impacts on neuronal physiology and associated diseases (Meyer et al., 2020).

Brg1 has a critical function in the development of synapses and the regeneration of the synapses mediated by MEF2. Neural plasticity is significantly impacted by its association with MEF2 to regulate activity-induced gene expression. The excretion of Brg1 in neurons leads to a reduction in the number of dendritic spines, which is important in securing synaptic structure. Furthermore, alterations in Brg1 have been associated with neurological dysfunction and diseases within the autism spectrum, emphasizing the crucial function that Brg1 performs in healthy brain development (Zhang et al., 2016).

4.4. SIRT1 (Silent Information Regulator) in NDDs

SIRT1, a member of the sirtuin family of NAD^+- dependent deacetylases, plays a significant role in defending the nervous system and has been associated with several neurodegenerative diseases, along with AD (Shen et al., 2023). SIRT1 has a role in multiple physiological processes, including mitochondrial biogenesis, apoptosis, genomic stability, DNA repair, and antioxidant impact (Yanagisawa et al., 2018 and Alam et al., 2021). According to (Chen et al., 2020), SIRT1 stops the ageing mechanism, therefore expanding lifespan. Furthermore, according to (Peng et al., 2022), SIRT1 activates the formation of synapses and supports synaptic activity. Thus, it has the possibility to slow down the progression of several degenerative brain disorders, like AD and PD (Wang et al., 2010). Research has suggested that the ageing procedure is linked with a reduction in the functioning of SIRT1 and raised sensitivity to degenerative brain diseases. SIRT1 works as an intracellular regulatory protein that modulates the action of mutational genes. SIRT1 controls the transcription of several regulatory proteins occupied in metabolic results (Hu et al., 2019). SIRT1 simplified the

production of peroxisome proliferator-activated receptor gamma coactivator 1 alpha (PGC-1α) and estrogen-related receptor alpha (ERRα), both of which are also called nuclear receptor 1 (Wan et al., 2022). In addition, SIRT1 interconnects with another molecular pathway, notably YRPW motif protein 2 (HEY2), autoimmune regulator (AIRE), miR-132, and poly-ADP-ribose polymerase 1 (PARP-1) (Wahab et al., 2021). SIRT1 degenerates the p53 protein, therefore suppressing apoptosis and cellular death (Khanahmadi et al., 2020). Moreover, SIRT1 enhances autophagy by suppressing the acetylation step, which is crucial for instigating autophagy (Li et al., 2019). SIRT1 modulates the immune response by improving the activation of T-helper cells 17, which have a key role in the development and propagation of autoimmunity (Jia et al., 2018). SIRT1 serves a vital function in the onset of addiction by connecting with histone deacetylase 1 (HDAC1) and the AP-1 promoter in dopaminergic neurons (Engel et al., 2016). Thereby, SIRT1 impacts a wide range of physiological activities associated with DNA damage and oxidative stress. In (Fig. 5) the role of SIRT1 in AD, PD, and HD is shown below.

Journal Pre-proof

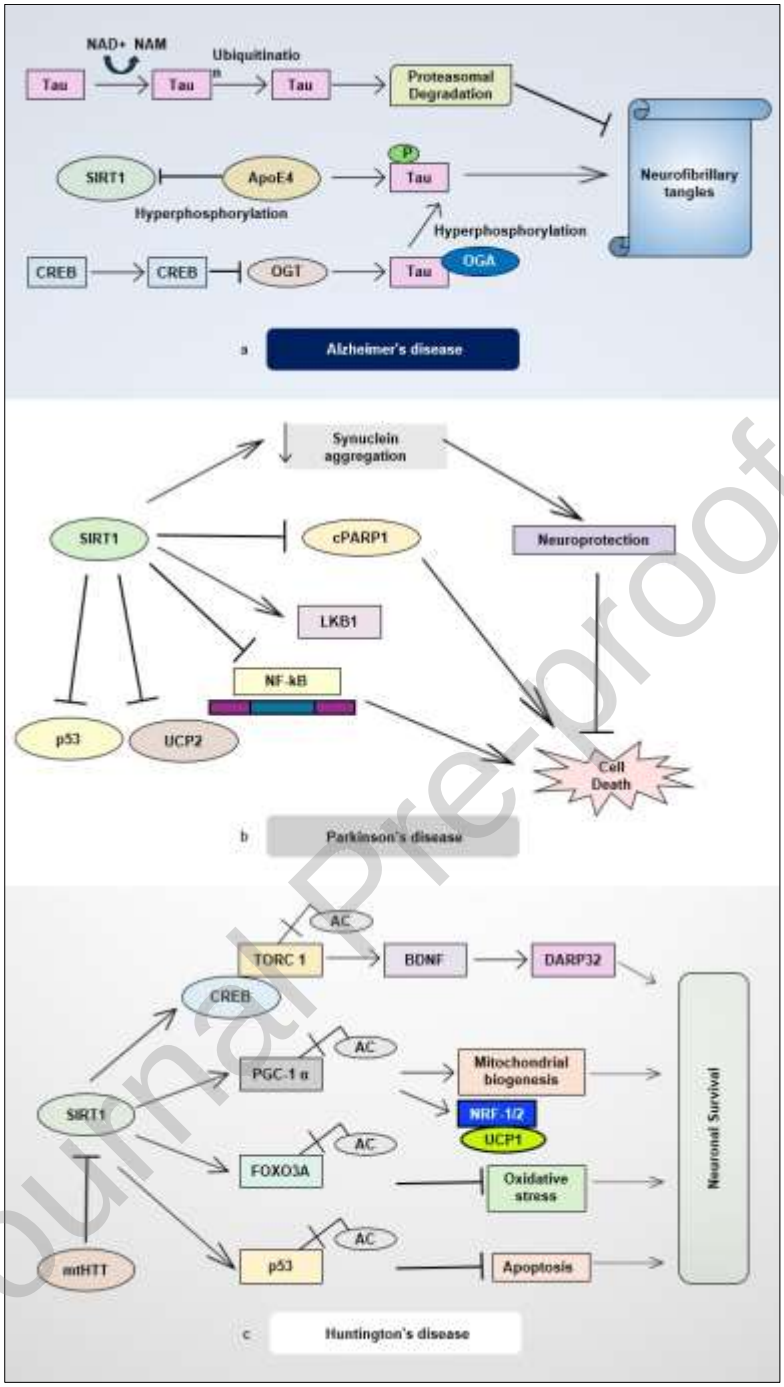


Fig. 5 Role of SIRT1 in neurodegenerative diseases. (a) SIRT1's involvement in the regulation of tau protein. It removes acetyl groups from the tau protein, which enhances its ubiquitination and decreases the formation of neurofibrillary tangles. Suppression of SIRT1 results in excessive phosphorylation of tau, which is caused by elevated functioning of ApoE4 and OGTT. (b) The activity of SIRT1 in PD. The effect of SIRT1 and SIRT2 on cell death in PD is independent of each other. SIRT1 suppresses the expression of c-PARP and NF- κβ, thereby decreasing the formation of protein aggregates in cells. (c) SIRT1 has a pivotal function in Huntington's disease. The mutant Huntington protein hinders the function of SIRT1, disrupting many pathways. One of these avenues encompasses the suppression of deacetylated TORC1. In general, when TORC1 is deacetylated, it interacts with CREB and DARPP32, which enhances the production of BDNF in neurons. Nevertheless, mutant Huntington hinders the deacetylation process in various targets, including TORC1, p53, FOXO3A, and PGC-1α, resulting in the eventual death of cells.

Journal Pre-proof

5. Targeting Neuroinflammatory Disorders with Drugs Repositioning

Neuroinflammatory disorders such as AD and PD have a global impact, affecting around 50 million individuals. The strategy of drug repositioning is emerging as a hopeful approach in the search for novel treatments for these conditions. According to a recent study by (Arbo et al., 2022), multiple studies have demonstrated the efficacy of repurposed drugs in treating various conditions, including NDDs. Neuroinflammation is a notion that remarkably affects numerous diseases, and repurposed drugs having anti-inflammatory properties may undoubtedly serve as advantageous treatments. Probenecid is a well-established medication. It is engaged in enhancing the removal of uric acid, defense of the nervous system, and decline of inflammation, as suggested by (García-Rodríguez et al., 2023). Drug repositioning is a cost-effective approach that can contribute to new opportunities for treating therapeutic NDDs. Such as kinase inhibitors, developed as cancer treatments, have presented powerful neuroprotective effects in the treatment of NDDs, as documented by (Kakoti et al., 2022). Novel drug delivery methods have effortlessly been used in preclinical and clinical trials to treat neuroinflammatory disorders by repurposing pharmaceuticals such as immunomodulators and anti-inflammatory drugs. These statements have demonstrated reduced adverse effects, higher treatment outcomes, and enhanced patient consent in contrast to traditional drug delivery systems (Jung et al., 2021). Table 1 systematically outlines a range of therapeutic drugs, targeted diseases, target tissues, and observed results.

Using existing drugs for other motives provides multiple advantages. Firstly, this approach is systematic in terms of both cost and time, as it allows for medications that have already undergone clinical and preclinical testing. This eradicates the necessity to produce and upgrade new chemicals and conduct thorough safety investigations from the beginning. The safety profiles of repurposed drugs in humans have previously been confirmed, hence decreasing the possibility of unforeseen side effects throughout clinical research. It addresses the utmost challenge in traditional drug development, where competitors usually fail because of safety concerns. In addition, repurposing broadens the range of therapeutic options by discovering novel applications for a previously existing medication that works beyond its original intended usage. This confers innovative possibilities for addressing diseases that now deficit-authorized therapeutic interventions (Mao, 2020 andJarada et al., 2020). (Fig. 6) illustrates the various drug repurposing techniques.

Numerous medications now in use have demonstrated potential for repurposing because of their anti-inflammatory and neuroprotective properties. For example, the antibiotic minocycline has shown neuroprotective effects in preclinical models of AD and PD by lowering neuroprotective effects in preclinical models of AD and PD by lowering neuroinflammation and blocking microglial activation (Noble et al., 2009) (Hua et al., 2022). Another illustration is the commonly prescribed drug metformin, which has demonstrated potential advantages in AD by activating AMP-activated protein kinase (AMPK) and modifying mitochondrial activity (de Marañón et al., 2022 andRena et al., 2017). Riluzole's capacity to control glutamate release and decrease excitotoxicity has led to its initial approval for the treatment of ALS. However, it is also being investigated for potential use in other NDDs (Bensimon et al., 1994). Furthermore,

the effects of the antimalarial medications chloroquine and hydroxychloroquine on neuroinflammation and autophagy in conditions like AD and PD are being investigated (Hu et al., 2022). These illustrations show how medication repositioning may provide novel therapy approaches for NDDs and emphasize the significance of additional research into these repurposing techniques.

Neuroinflammatory disorders show substantial adverse effects, limited efficacy, and extravagant costs. Numerous treatments only give symptomatic relief and do not address the fundamental origin of inflammation or prohibit the disease progression. Adverse effects encompass immunosuppression, gastrointestinal issues, and patient safety and compliance. Moreover, specific medical interventions are costly, posing a challenge for several patients regarding their accessibility. When drugs are repurposed, there may be intellectual property issues if the original patents are still in effect for their initial indications (Kakoti et al., 2022). These challenges highlight the need for new therapies that effectively treat neuroinflammation (Mao, 2020 and Jarada et al., 2020).

Journal Pre-proof

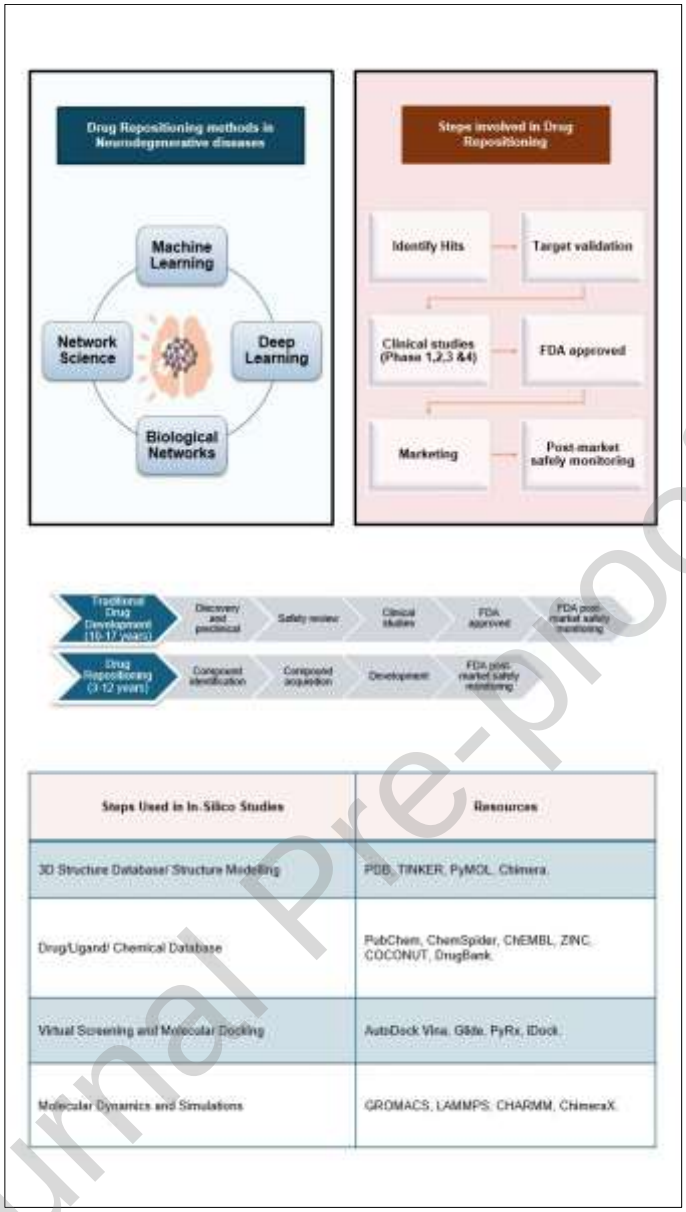


Fig. 6 This figure demonstrates various drug repositioning techniques used in the treatment of NDDs. It outlines the process involved in drug repositioning and provides a comparison between traditional drug development and drug repositioning. In addition, the table that is included illustrates the steps employed in *In-Silico* studies and bioinformatic resources for drug repositioning.

Table 1 Overview of Therapeutic Drugs, Disorders, Target Cells/Tissues, and Results

Therapeutic Drugs	Targeted Disease	Target cells or Tissues	Observed Results	References
Triptolide	*Alzheimer's disease*	Microglia	Triptolide exhibited anti-inflammatory and antioxidant properties, enhanced spatial learning, and reduced the amount of	(Wang et al., 2014)

Therapeutic Drugs	Targeted Disease	Target cells or Tissues	Observed Results	References
			amyloid beta produced in the brain.	
sRAGE-MSCk	*Alzheimer's disease*	Microglia	M2 microglia were present in more significant numbers. Additionally, sRAGE-MSCs successfully reduced RAGE.	(Oh et al., 2018)
Sesame oil	*Alzheimer's disease*	Microglia	Reduced levels of acetylcholine esterase and amyloid beta, as well as the expression of TNF-alpha, IL1-beta, and NF-kappa-B, were all enhanced by sesame oil. It also reduced memory impairment and learning.	(Mohamed et al., 2021)
SCP2-1	*Alzheimer's disease*	Microglia	By reducing the JNK and NF-kappa-B signalling pathways, the M1 polarisation of microglia was suppressed, and cognition was improved.	(Xu et al., 2020)
10-hydroxy decanoic acid	*Alzheimer's disease*	Microglia	N9 and BV2 cell lines have lower levels of NO and iNOS. Along with expressing less pro-inflammatory cytokines and more TREM2, these cells were also more expressed than untreated ones.	(Xu et al., 2020)

Journal Pre-proof

Therapeutic Drugs	Targeted Disease	Target cells or Tissues	Observed Results	References
DMXBA	*Alzheimer's disease*	Microglia	Decreased accumulated amyloid-beta and enhanced memory and spatial learning. It enhanced microglial amyloid-beta phagocytosis activity and activated nAChR.	(Anwar and Rivest, 2020)
DAPPD	*Alzheimer's disease*	Microglia	Enhanced microglial function, reduced inflammatory cytokines including TNF-alpha and IL6, ApoE, and improved memory through the expression of anti-inflammatory genes like TGFb-beta and IL4, and suppressed NF-kappa-B.	(Anwar and Rivest, 2020)
Astaxanthin	*Alzheimer's disease*	Microglia	The M2 microglia's shift to the pro-inflammatory M1 phenotype was suppressed.	(Kim et al., 2020)
Cyclosporine	*Parkinson's disease*		Parkinson's disease mouse model with neural repair cyclosporine therapy.	(Tamburrino et al., 2015)
FK506	*Parkinson's disease*	Various immune cells	FK506 decreased T helper infiltration, cytotoxic T cell infiltration, and the number of macrophages and microglia while	(Van der Perren et al., 2015)

Journal Pre-proof

Therapeutic Drugs	Targeted Disease	Target cells or Tissues	Observed Results	References
			enhancing the survival of dopaminergic neurons.	
Ginsenoside Rg1	*Parkinson's disease*	Microglia	Suppression of pro-inflammatory cytokines and elevation of anti-inflammatory cytokines were correlated with the activation of dopaminergic neurons in substantia nigra glial cells. Moreover, midbrain ginsenoside Rg1 reduced Nf-Kappa-B production by microglia and converted M1 microglia to M2 type.	(Liu et al., 2020)
MCC950	*Parkinson's disease*	Microglia	MCC950 reduced inflammation, alpha-syn accumulation, motor impairments, and nigrostriatal dopamine degradation.	(Gordon et al., 2018)
Pioglitazone	*Parkinson's disease*	Microglia/ macrophages	The substantia nigra of 5 mg/kg pioglitazone-treated monkeys had fewer macrophage and microglia cells than those in the control group.	(Swanson et al., 2011)
Donepezil	*Multiple sclerosis*	Microglia	Influence of immunomodulatio	(Jiang et al., 2013)

Journal Pre-proof

Therapeutic Drugs	Targeted Disease	Target cells or Tissues	Observed Results	References
			n on IL-4 and IFN-gamma	
Ethyl pyruvate	*Multiple sclerosis*	Microglia and astrocytes	Ethyl pyruvate inhibited the infiltration of immune cells into the brain and the quantity of activated microglia and astrocytes.	(Djedović et al., 2017)
PPAR -gamma agonists	*Multiple sclerosis*	Microglia/ astrocytes	They impeded the synthesis of TNF-alpha, IL1, IL6, MCP-1, and nitric oxide.	(Storer et al., 2005)

5.1 Overview of Clinical Trials on Neuroinflammation and Neurodegenerative Diseases

Table 2 serves as an essential resource for clinicians, researchers, and stakeholders with an interest in the association between neuroinflammation and neurodegenerative diseases. This meticulous summary offers detailed information regarding several clinical studies encompassing Clinical Trial ID, Title, Study Type, Status, Phase, and condition under investigation.

Table 2 Clinical Trials in Neuroinflammation

Clinical Trial ID	Title	Study Type	Status	Phase	Condition
NCT05040048	Taxonomy of Neurodegenerative Diseases: Observational Study in Alzheimer's Disease and Parkinson's Disease	Observational	Completed	Not Applicable	Parkinson's Disease and Alzheimer's Disease
NCT03457493	The University of Alabama at Birmingham (UAB) Neuroinflammation in Parkinson's Disease-TSPO-Positron Emission	Interventional	Recruiting	Phase 1, Phase 2	Parkinson's Disease

Journal Pre-proof

	Tomography (PET) Sub-study				
NCT05419453	Microbiome Composition and Function Contributes to Cognitive Impairment and Neuroinflammation in Parkinson's Disease	Observational	Recruiting	Not Applicable	Parkinson's Disease
NCT04062526	Evaluation of Neuroinflammation in Parkinson's Disease Using 18F-NOS PET/CT	Interventional	Active, not recruiting	Early Phase 1	Parkinson's Disease
NCT03633513	Parkinson's Disease Inflammatory Biomarker Profiling	Observational	Terminated	Not Applicable	Parkinson's Disease
NCT02319382	Measure of Microglial Activation in the Brain of Parkinson's Disease Patients With PET	Interventional	Unknown status	Not Applicable	Parkinson's Disease
NCT05807581	Clinical, Molecular and Electrophysiological Profiling of Parkinson's Disease: The Role of Non-pharmacological Therapies	Interventional	Recruiting	Not Applicable	Parkinson's Disease
NCT05781711	Clinical Study to Evaluate the Possible Efficacy of Metformin in Patients with Parkinson's Disease	Interventional	Recruiting	Phase 2	Parkinson's Disease
NCT05854524	Exercise Neuroprotection	Interventional	Not yet recruiting	Not Applicable	Parkinson's Disease

Journal Pre-proof

	in Parkinson's Disease				
NCT02511015	Hereditary Parkinson's Disease Natural History Protocol	Observational	Completed	Not Applicable	Parkinson's Disease
NCT05815524	Physical Activity in Patients with Parkinson's Disease: a "Disease Modifying" Intervention?	Interventional	Recruiting	Not Applicable	Parkinson's Disease
NCT05395624	Safety, PK and, Biodistribution of 18F-OP-801 in Patients With ALS, AD, MS, PD and, Healthy Volunteers	Interventional	Recruiting	Phase 1, Phase 2	Parkinson's Disease Amyotrophic Lateral Sclerosis Alzheimer's Disease Multiple Sclerosis
NCT04855344	Deep Brain Stimulation Therapy and Intestinal Microbiota	Observational	Unknown status	Not Applicable	Deep Brain Stimulation Gastrointestinal Microbiome
NCT05456451	Effectiveness of Noninvasive Vagus Stimulation for Upper Extremity in Parkinson's Disease	Interventional	Unknown status	Not Applicable	Parkinson's Disease Tremor Upper Extremity Dysfunction
NCT05008094	The Epidemiology of Parkinson's Disease in Croatia and the Influence of Genetic Factors and Microbiota on the Progression and Treatment Outcomes of the Disease (GiOPARK)	Observational	Unknown status	Not Applicable	Parkinson's Disease Genetic Disease Microbiota Neuroinflammatory Response
NCT05962957	Clinical Study to Compare the	Interventional	Recruiting	Phase 2	Parkinson's Disease

Journal Pre-proof

	Possible Safety and Efficacy of Pentoxifylline in Patients with Parkinson's Disease Treated with Conventional Treatment				
NCT02377206	Neuroinflammation and Cognitive Decline in Alzheimer's Disease	Interventional	Completed	Early Phase 1	Alzheimer's Disease
NCT04786223	Targeting Neuroinflammation as a Contributing Pathology in Alzheimer's Disease Dementia	Interventional	Enrolling by invitation	Phase 2	Alzheimer's Disease
NCT05077579	Neuroinflammation and Alzheimer's Disease Imaging Biomarkers in Midlife Obesity	Observational	Recruiting	Not Applicable	Alzheimer's Disease
NCT04274998	Evaluation of in Vivo Neuroinflammation in Alzheimer's Disease Using Novel Positron Emission Tomography (PET/CT) Imaging	Interventional	Recruiting	Early Phase 1	Alzheimer's Disease
NCT03548883	Examining Neuroinflammation in Alzheimer's Disease Via Transcriptomic Profiling and Microglia Modeling Using Human Peripheral Blood	Observational	Completed	Not Applicable	Alzheimer's Disease

Journal Pre-proof

	Mononuclear Cells (ENHANCE)				
NCT05378659	Neuroinflammation and Alzheimer's Pathology in Post-operative Cognitive Dysfunction: A Pilot Study	Observational	Unknown status	Not Applicable	Alzheimer's Disease
NCT05468073	Therapeutic Evaluation of Low-dose IL-2-based Immunomodulatory Approach in Patients with Early AD	Interventional	Recruiting	Phase 2	Alzheimer's Disease
NCT05114499	Efficacy and Safety of Donepezil and Sodium Oligomannate in Patients with Mild to Moderate Alzheimer's Disease	Observational	Not yet recruiting	Not Applicable	Alzheimer's Disease
NCT03435861	Effect of Neflamapimod (VX-745) on Brain Inflammation Using Positron Emission Tomography (PET) Scan in Alzheimer's Disease (AD) Patients	Interventional	Completed	Phase 2	Alzheimer's Disease
NCT05911178	Impact of Microglial Activation on Synaptic Density in Alzheimer's Disease	Interventional	Recruiting	Not Applicable	Alzheimer's Disease
NCT04517552	Investigation of Inflammation Using [C-11]-CS1P1	Observational	Active, not recruiting	Not Applicable	Alzheimer's Disease

Journal Pre-proof

NCT03744312	Imaging Inflammation in Alzheimer's Disease With 11C-ER176	Interventional	Completed	Phase 1 Phase 2	Alzheimer’s Disease
NCT05821153	Phase I Trial Using Interleukin-2 (IL-2) to Expand Regulatory T Cells in Patients with Alzheimer's Disease	Interventional	Completed	Phase 1	Alzheimer’s Disease
NCT06096090	A Phase II Clinical Trial of Interleukin-2 (IL-2) in Patients with Mild to Moderate Alzheimer's Disease	Interventional	Recruiting	Phase 2	Alzheimer’s Disease
NCT04388254	A 12-Month, Open-Label Safety Study of Simufilam Followed by a 6-Month Randomized Withdrawal and 6 Additional Months Open-Label in Mild-to-moderate Alzheimer's Disease Patients	Interventional	Completed	Phase 2	Alzheimer’s Disease
NCT05349318	Hyperbaric Oxygen Therapy for Prodromal Alzheimer´s Disease with Cerebrovascular Disease: A Prospective, Randomized, Double-Blind Study	Interventional	Recruiting	Not Applicable	Prodromal Alzheimer’s Disease

Journal Pre-proof

6. Novel Neuroinflammation Biomarkers

Extensive research has been conducted on neuroinflammation biomarkers across a range of neurological conditions. Biomarkers can capture pathophysiological aspects such as the accumulation of activation of microglia, misfolded proteins, synaptic dysfunction, and neuronal death. They have the potential to describe disease phenotypes and monitor treatment outcomes in the CSF compartment (Kapaki et al., 2022a). In one study, it was discovered that patients with AD had higher levels of migration inhibitory factor (MIF), soluble triggering receptor expressed on myeloid cells 1 (sTREM1), and soluble triggering receptor expressed on myeloid cells 2 (sTREM2). Elevated levels of MIF and sTREM1 were observed in AD patients as compared to both healthy controls and patients with mild cognitive impairment (MCI) (Hok-A-Hin et al., 2023). Further study on juvenile systemic lupus erythematosus (j-SLE) has found neopterin, interferon-alpha (IFN-α), and cerebrospinal fluid (CSF) as possible biomarkers for the diagnosis and monitoring of juvenile neuropsychiatric SLE (j-NPSLE) (Kapaki et al., 2022b). Individuals with ASD have been found to show metabolic dysregulations, including dysfunctional mitochondria, oxidative stress, and abnormalities in sphingolipid metabolism. These dysregulations may be present as biomarkers that can help in the diagnosis of ASD (Labouret et al., 2023a). In addition, a study inspecting the unfavorable impact of radiation on the brain identified microscopic molecular indicators in the bloodstream that could be favorable for shielding against radiation and for originating innovative medications to discuss NDDs (Esvap and Ulgen, 2023a). Table 2 shows novel biomarkers for neurological development disorders.

Journal Pre-proof

Table 3 Novel Biomarkers for Neurological Developmental Disorders

S. No.	Novel Biomarkers of NDDs	Neurological Disorder	Reference
1.	Sphingosine-1-phosphate (S1P), Ceramide and Glucosylceramide	Autism spectrum disorder	(Esvap and Ulgen, 2023b)
2.	Neopterin and Interferon-alpha (IFN-α)	Juvenile neuropsychiatric system lupus erythematosus (j-NPSLE)	(Labouret et al., 2023b)
3.	IL-15 and Serum amyloid A	Traumatic brain injury diagnosis	(Yue et al., 2023)
4.	Soluble triggering receptors expressed on myeloid cells-2 (sTREM-2)	Multiple sclerosis	(Azzolini et al., 2022)
5.	MIF, sTREM1, and sTREM2	Alzheimer's disease	(Hok-A-Hin et al., 2023)
6.	[18F] NOS PET	Parkinson's disease	(Doot et al., 2022)
7.	SIRT2, HGF, MMP-10, and CXCL5	Alzheimer's disease	(Gaetani et al., 2021)

7. Conclusion and Future Perspectives

Neuroinflammation is an essential factor in NDDs, as it disrupts the advancement of the diseases and exacerbates neuronal death. The intricate association between neuroinflammation and neuronal degeneration demonstrates the complicated interaction of these disorders and their crucial need for effective treatment avenues. The neuroinflammatory reactions in non-diminished dementia syndrome are maintained by pro-inflammatory transcription factors, such as TLRs, STAT3, MAPK, FAF1, and p38. Nevertheless, certain transcription factors, including YY1, MEF2, SIRT1, and SOCS pathway, have a role in encouraging neuronal survival and reducing neuroinflammation. These factors might be the subject of future inventions.

Drug repositioning is a key approach in the realm of therapeutic development for addressing neuroinflammatory diseases. Repurposing current medications for novel therapeutic purposes minimizes potential hazards and improves the drug development process. Various medications mainly obtained for several diseases have proven successful for treating neuroinflammation,

Journal Pre-proof

and these can be quickly used in clinical applications. Furthermore, identifying novel biomarkers for neuroinflammation offers tremendous potential in the field of early detection, prognosis, and monitoring of NDDs. These biomarkers permit customized therapy approaches that are modified to the individual needs of every patient by providing essential information regarding the disease's course and the treatment's efficacy.

As we think about the future, we must study multiple fundamental avenues to boost our understanding and management of neuroinflammatory disorders. A detailed investigation is mandatory to attain a thorough knowledge of the complicated molecular pathways responsible for neuroinflammation in NDDs. The advancement in technology, which includes single-cell transcriptomics, proteomics, and imaging techniques, will provide ample information on the cellular and molecular mechanisms before the neuroinflammatory reactions. It may be viable to discover improved and more efficient treatments by targeting particular signalling pathways linked with neuroinflammation. Researchers may be capable of disclosing new therapeutic targets for intervention by thorough knowledge of the biological roles of critical chemicals such as chemokines, cytokines, and the activation pathways of macrophages/microglia.

Identifying and focusing on understanding the particular molecular pathway connecting neuroinflammation to NDDs is crucial in the research areas. This emphasis is significant because it assists in revealing the fundamental processes responsible for disease progression, which is vital for creating precise and efficient treatments. A thorough grasp of this molecular interaction will also provide a strong groundwork for delving deeper into related areas, like the function of various microglial phenotypes, the impact of genetic and environmental factors, and possible therapeutic targets. By making progress in this field of research, we can make significant advancements in comprehending and managing NDDs more efficiently.

Abbreviation

NDDs- Neurodegenerative diseases; TLRs- Toll-like receptors; MAPK- Mitogen-activated protein kinase; FAF1- Fass-associated factor 1; STAT3- Signal transducer and activator of transcription 3; YY1- Yin Yang 1; SOCS- Suppressor of cytokine signalling; MEF2- Myocyte enhancer factor 2; SIRT1- Sirtuin 1; PD- Parkinson's disease; AD- Alzheimer's disease; NOD- Nucleotide-binding oligomerization domain; LRR- Leucine-rich repeat; NLRP3- Pyrin domain-containing protein 3; ROS- Reactive oxygen species; BBB- Blood brain barrier; PET- Positron emission tomography; MRI- Magnetic resonance imaging; ALS- Amyotrophic lateral sclerosis; DNA- Deoxyribonucleic acid; MS- Multiple sclerosis; CNS- Central Nervous System; TNF-α- Tumour Necrosis Factor alpha; IL-6- Interleukin 6; GPNMB- Glycoprotein Nonmetastatic Melanoma Protein B; NHIE- Neonatal hypoxic-ischemic encephalopathy; DAM- Disease-associated microglia; TREM2- Triggering receptor expressed on myeloid cells 2; APOE- Apolipoprotein E; HD- Huntington's disease; BMB- Blood-myenteric barrier; ASD- Autism spectrum disorders; RNA- Ribonucleic Acid; JNK1- c-Jun N-terminal kinase 1; BFA- Brefeldin A; TIR- Intracellular toll/IL-1 receptor-like; PAMPs- Pathogen-associated molecular patterns; DAMPs- Damage-associated molecular patterns Aβ- Amyloid beta; ERK- Extracellular signal-regulated kinase; DLB- Dementia with Lewy bodies; APP- Amyloid-β

Journal Pre-proof

precursor protein; NMDA- N-methyl-D-aspartate; BACE1- Beta-secretase 1; RAGE- Receptor for Advanced Glycation end products; LTP- Long-term potentiation; PSD-95- Postsynaptic density protein-95; DA- Dopaminergic; ASK1- Apoptosis signal-regulating kinase 1; SNCA- alpha synuclein, IgH- immunoglobulin heavy chain, CKO- conditional knockout, NPCs- Nuclear pore complex; NC- Neural crest; NAD^+- Nicotinamide adenine dinucleotide; PGC-1α- Peroxisome proliferator-activated receptor-γ coactivator-1α ; ERRα- Estrogen-related receptor alpha; HEY2- YRPW motif protein 2; AIRE- Autoimmune regulator; PARP-1- Poly-ADP-ribose polymerase 1; HDAC1- Histone deacetylase 1; MIF- Migration inhibitory factor; sTREM1- Soluble triggering receptor expressed on myeloid cells 1; MCI- Mild cognitive impairment; j-SLE- Juvenile systemic lupus erythematosus; CSF- Cerebrospinal fluid; j-NPSLE- j-neuropsychiatric SLE; S1P- Sphingosine-1-phosphate; NFTs- Neurofibrillary tangles

ACKNOWLEDGEMENT

We would like to sincerely thank the senior management of Delhi Technological University (DTU).

References

Aggarwal, B.B., Kunnumakkara, A.B., Harikumar, K.B., Gupta, S.R., Tharakan, S.T., Koca, C., Dey, S., Sung, B., 2009. Signal Transducer and Activator of Transcription-3, Inflammation, and Cancer. Ann N Y Acad Sci 1171, 59–76. https://doi.org/10.1111/j.1749-6632.2009.04911.x

Alam, F., Syed, H., Amjad, S., Baig, M., Khan, T.A., Rehman, R., 2021. The interplay between oxidative stress, SIRT1, and reproductive and metabolic functions. Curr Res Physiol 4, 119–124. https://doi.org/10.1016/j.crphys.2021.03.002

Alexander, D., Archer, S., 2003. On the myth of "Anglo-Saxon" financial accounting: A response to Nobes. The International Journal of Accounting 38, 503–504. https://doi.org/10.1016/j.intacc.2003.09.003

Amor, S., Puentes, F., Baker, D., Van Der Valk, P., 2010. Inflammation in neurodegenerative diseases. Immunology 129, 154–169. https://doi.org/10.1111/j.1365-2567.2009.03225.x

Antony, P.M.A., Diederich, N.J., Krüger, R., Balling, R., 2013. The hallmarks of <scp>P</scp> arkinson's disease. FEBS J 280, 5981–5993. https://doi.org/10.1111/febs.12335

Anwar, S., Rivest, S., 2020. Alzheimer's disease: microglia targets and their modulation to promote amyloid phagocytosis and mitigate neuroinflammation. Expert Opin Ther Targets 24, 331–344. https://doi.org/10.1080/14728222.2020.1738391

Arbo, B.D., Schimith, L.E., Goulart dos Santos, M., Hort, M.A., 2022. Repositioning and development of new treatments for neurodegenerative diseases: Focus on neuroinflammation. Eur J Pharmacol 919, 174800. https://doi.org/10.1016/j.ejphar.2022.174800

Asih, P.R., Prikas, E., Stefanoska, K., Tan, A.R.P., Ahel, H.I., Ittner, A., 2020. Functions of p38 MAP Kinases in the Central Nervous System. Front Mol Neurosci 13. https://doi.org/10.3389/fnmol.2020.570586

Journal Pre-proof

Asslih, S., Damri, O., Agam, G., 2021. Neuroinflammation as a Common Denominator of Complex Diseases (Cancer, Diabetes Type 2, and Neuropsychiatric Disorders). Int J Mol Sci 22, 6138. https://doi.org/10.3390/ijms22116138

Auernhammer, C.J., Bousquet, C., Melmed, S., 1999. Autoregulation of pituitary corticotroph SOCS-3 expression: Characterization of the murine SOCS-3 promoter. Proceedings of the National Academy of Sciences 96, 6964–6969. https://doi.org/10.1073/pnas.96.12.6964

Azzolini, F., Gilio, L., Pavone, L., Iezzi, E., Dolcetti, E., Bruno, A., Buttari, F., Musella, A., Mandolesi, G., Guadalupi, L., Furlan, R., Finardi, A., Micillo, T., Carbone, F., Matarese, G., Centonze, D., Stampanoni Bassi, M., 2022. Neuroinflammation Is Associated with GFAP and sTREM2 Levels in Multiple Sclerosis. Biomolecules 12, 222. https://doi.org/10.3390/biom12020222

Baidya, F., Bohra, M., Datta, A., Sarmah, D., Shah, B., Jagtap, P., Raut, S., Sarkar, A., Singh, U., Kalia, K., Borah, A., Wang, X., Dave, K.R., Yavagal, D.R., Bhattacharya, P., 2021. Neuroimmune crosstalk and evolving pharmacotherapies in neurodegenerative diseases. Immunology 162, 160–178. https://doi.org/10.1111/imm.13264

Belov Kirdajova, D., Kriska, J., Tureckova, J., Anderova, M., 2020. Ischemia-Triggered Glutamate Excitotoxicity From the Perspective of Glial Cells. Front Cell Neurosci 14. https://doi.org/10.3389/fncel.2020.00051

Bensimon, G., Lacomblez, L., Meininger, V., 1994. A Controlled Trial of Riluzole in Amyotrophic Lateral Sclerosis. New England Journal of Medicine 330, 585–591. https://doi.org/10.1056/NEJM199403033300901

Betarbet, R., Anderson, L.R., Gearing, M., Hodges, T.R., Fritz, J.J., Lah, J.J., Levey, A.I., 2008. Fas-associated factor 1 and Parkinson's disease. Neurobiol Dis 31, 309–315. https://doi.org/10.1016/j.nbd.2008.05.006

Block, M.L., Zecca, L., Hong, J.-S., 2007. Microglia-mediated neurotoxicity: uncovering the molecular mechanisms. Nat Rev Neurosci 8, 57–69. https://doi.org/10.1038/nrn2038

Bu, L., Cripps, R.M., 2022. Promoter architecture of Drosophila genes regulated by Myocyte enhancer factor-2. PLoS One 17, e0271554. https://doi.org/10.1371/journal.pone.0271554

Campbell, I.L., Krucker, T., Steffensen, S., Akwa, Y., Powell, H.C., Lane, T., Carr, D.J., Gold, L.H., Henriksen, S.J., Siggins, G.R., 1999. Structural and functional neuropathology in transgenic mice with CNS expression of IFN-α1Published on the World Wide Web on 17 March 1999.1. Brain Res 835, 46–61. https://doi.org/10.1016/S0006-8993(99)01328-1

Cao, L., Wang, Z., Wan, W., 2018. Suppressor of Cytokine Signaling 3: Emerging Role Linking Central Insulin Resistance and Alzheimer's Disease. Front Neurosci 12. https://doi.org/10.3389/fnins.2018.00417

Chastain, E.M.L., Duncan, D.S., Rodgers, J.M., Miller, S.D., 2011. The role of antigen presenting cells in multiple sclerosis. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1812, 265–274. https://doi.org/10.1016/j.bbadis.2010.07.008

Chen, C., Zhou, M., Ge, Y., Wang, X., 2020. SIRT1 and aging related signaling pathways. Mech Ageing Dev 187, 111215. https://doi.org/10.1016/j.mad.2020.111215

Journal Pre-proof

Chen, G., Yan, S., Zhang, Jing-Han, Zhang, Ji, Deng, I.B., He, R., 2022. The alternative 3' splice site of GPNMB may promote neuronal survival after neonatal hypoxic–ischemic encephalopathy injury. Ibrain 8, 302–313. https://doi.org/10.1002/ibra.12056

CHEN, W.-W., ZHANG, X., HUANG, W.-J., 2016. Role of neuroinflammation in neurodegenerative diseases (Review). Mol Med Rep 13, 3391–3396. https://doi.org/10.3892/mmr.2016.4948

Ciaccio, M., Agnello, L. (Eds.), 2023. Neurodegenerative Diseases. MDPI. https://doi.org/10.3390/books978-3-0365-7074-7

Confettura, A.D., Cuboni, E., Ammar, M.R., Jia, S., Gomes, G.M., Yuanxiang, P., Raman, R., Li, T., Grochowska, K.M., Ahrends, R., Karpova, A., Dityatev, A., Kreutz, M.R., 2022. Neddylation-dependent protein degradation is a nexus between synaptic insulin resistance, neuroinflammation and Alzheimer's disease. Transl Neurodegener 11, 2. https://doi.org/10.1186/s40035-021-00277-8

Congdon, E.E., Sigurdsson, E.M., 2018. Tau-targeting therapies for Alzheimer disease. Nat Rev Neurol 14, 399–415. https://doi.org/10.1038/s41582-018-0013-z

Cornut, M., Bourdonnay, E., Henry, T., 2020. Transcriptional Regulation of Inflammasomes. Int J Mol Sci 21, 8087. https://doi.org/10.3390/ijms21218087

Cortés, N., Andrade, V., Guzmán-Martínez, L., Estrella, M., Maccioni, R., 2018. Neuroimmune Tau Mechanisms: Their Role in the Progression of Neuronal Degeneration. Int J Mol Sci 19, 956. https://doi.org/10.3390/ijms19040956

Dabi, Y.T., Ajagbe, A.O., Degechisa, S.T., 2023. Toll-like receptors in pathogenesis of neurodegenerative diseases and their therapeutic potential. Immun Inflamm Dis 11. https://doi.org/10.1002/iid3.839

Das Sarma, J., 2014. Microglia-mediated neuroinflammation is an amplifier of virus-induced neuropathology. J Neurovirol 20, 122–136. https://doi.org/10.1007/s13365-013-0188-4

de Marañón, A.M., Díaz-Pozo, P., Canet, F., Díaz-Morales, N., Abad-Jiménez, Z., López-Domènech, S., Vezza, T., Apostolova, N., Morillas, C., Rocha, M., Víctor, V.M., 2022. Metformin modulates mitochondrial function and mitophagy in peripheral blood mononuclear cells from type 2 diabetic patients. Redox Biol 53, 102342. https://doi.org/10.1016/j.redox.2022.102342

Dias, V., Junn, E., Mouradian, M.M., 2013. The Role of Oxidative Stress in Parkinson's Disease. J Parkinsons Dis 3, 461–491. https://doi.org/10.3233/JPD-130230

Diaz-Ortiz, M.E., Seo, Y., Posavi, M., Carceles Cordon, M., Clark, E., Jain, N., Charan, R., Gallagher, M.D., Unger, T.L., Amari, N., Skrinak, R.T., Davila-Rivera, R., Brody, E.M., Han, N., Zack, R., Van Deerlin, V.M., Tropea, T.F., Luk, K.C., Lee, E.B., Weintraub, D., Chen-Plotkin, A.S., 2022. GPNMB confers risk for Parkinson's disease through interaction with α-synuclein. Science (1979) 377. https://doi.org/10.1126/science.abk0637

Djedović, N., Stanisavljevic, S., Jevtić, B., Momčilović, M., Lavrnja, I., Miljković, D., 2017. Anti-encephalitogenic effects of ethyl pyruvate are reflected in the central nervous system and the gut. Biomedicine & Pharmacotherapy 96, 78–85. https://doi.org/10.1016/j.biopha.2017.09.110

Doggett, K., Keating, N., Dehkhoda, F., Bidgood, G.M., Meza Guzman, L.G., Leong, E., Kueh, A., Nicola, N.A., Kershaw, N.J., Babon, J.J., Alexander, W.S., Nicholson, S.E., 2023. The SOCS1 KIR

Journal Pre-proof

and SH2 domain are both required for suppression of cytokine signaling in vivo. Cytokine 165, 156167. https://doi.org/10.1016/j.cyto.2023.156167

Dong, X., Kwan, K.M., 2020. Yin Yang 1 is critical for mid-hindbrain neuroepithelium development and involved in cerebellar agenesis. Mol Brain 13, 104. https://doi.org/10.1186/s13041-020-00643-z

Dong, X., Wang, Y., Qin, Z., 2009a. Molecular mechanisms of excitotoxicity and their relevance to pathogenesis of neurodegenerative diseases. Acta Pharmacol Sin 30, 379–387. https://doi.org/10.1038/aps.2009.24

Dong, X., Wang, Y., Qin, Z., 2009b. Molecular mechanisms of excitotoxicity and their relevance to pathogenesis of neurodegenerative diseases. Acta Pharmacol Sin 30, 379–387. https://doi.org/10.1038/aps.2009.24

Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., Shi, Y., 1999. Targeted Disruption of Mouse Yin Yang 1 Transcription Factor Results in Peri-Implantation Lethality. Mol Cell Biol 19, 7237–7244. https://doi.org/10.1128/MCB.19.10.7237

Doot, R.K., Young, A.J., Nasrallah, I.M., Wetherill, R.R., Siderowf, A., Mach, R.H., Dubroff, J.G., 2022. [18F]NOS PET Brain Imaging Suggests Elevated Neuroinflammation in Idiopathic Parkinson's Disease. Cells 11, 3081. https://doi.org/10.3390/cells11193081

Dora, D., Ferenczi, S., Stavely, R., Toth, V.E., Varga, Z. V., Kovacs, T., Bodi, I., Hotta, R., Kovacs, K.J., Goldstein, A.M., Nagy, N., 2021. Evidence of a Myenteric Plexus Barrier and Its Macrophage-Dependent Degradation During Murine Colitis: Implications in Enteric Neuroinflammation. Cell Mol Gastroenterol Hepatol 12, 1617–1641. https://doi.org/10.1016/j.jcmgh.2021.07.003

Doroszkiewicz, J., Kulczynska-Przybik, A., Dulewicz, M., Borawska, R., Slowik, A., Mroczko, B., 2023a. The cerebrospinal fluid Glycoprotein Nonmetastatic Melanoma Protein B (GPNMB) concentration in Alzheimer's Disease (AD). Alzheimer's & Dementia 19. https://doi.org/10.1002/alz.062120

Doroszkiewicz, J., Kulczyńska-Przybik, A., Dulewicz, M., Borawska, R., Zajkowska, M., Słowik, A., Mroczko, B., 2023b. Potential Utility of Cerebrospinal Fluid Glycoprotein Nonmetastatic Melanoma Protein B as a Neuroinflammatory Diagnostic Biomarker in Mild Cognitive Impairment and Alzheimer's Disease. J Clin Med 12, 4689. https://doi.org/10.3390/jcm12144689

Engel, G.L., Marella, S., Kaun, K.R., Wu, J., Adhikari, P., Kong, E.C., Wolf, F.W., 2016. Sir2/Sirt1 Links Acute Inebriation to Presynaptic Changes and the Development of Alcohol Tolerance, Preference, and Reward. The Journal of Neuroscience 36, 5241–5251. https://doi.org/10.1523/JNEUROSCI.0499-16.2016

Esvap, E., Ulgen, K.O., 2023a. Neuroinflammation, Energy and Sphingolipid Metabolism Biomarkers Are Revealed by Metabolic Modeling of Autistic Brains. Biomedicines 11, 583. https://doi.org/10.3390/biomedicines11020583

Esvap, E., Ulgen, K.O., 2023b. Neuroinflammation, Energy and Sphingolipid Metabolism Biomarkers Are Revealed by Metabolic Modeling of Autistic Brains. Biomedicines 11, 583. https://doi.org/10.3390/biomedicines11020583

Journal Pre-proof

Gaetani, L., Bellomo, G., Parnetti, L., Blennow, K., Zetterberg, H., Di Filippo, M., 2021. Neuroinflammation and Alzheimer's Disease: A Machine Learning Approach to CSF Proteomics. Cells 10, 1930. https://doi.org/10.3390/cells10081930

García-Rodríguez, C., Mujica, P., Illanes-González, J., López, A., Vargas, C., Sáez, J.C., González-Jamett, A., Ardiles, Á.O., 2023. Probenecid, an Old Drug with Potential New Uses for Central Nervous System Disorders and Neuroinflammation. Biomedicines 11, 1516. https://doi.org/10.3390/biomedicines11061516

Glass, C.K., Saijo, K., Winner, B., Marchetto, M.C., Gage, F.H., 2010. Mechanisms Underlying Inflammation in Neurodegeneration. Cell. https://doi.org/10.1016/j.cell.2010.02.016

Gordon, R., Albornoz, E.A., Christie, D.C., Langley, M.R., Kumar, V., Mantovani, S., Robertson, A.A.B., Butler, M.S., Rowe, D.B., O'Neill, L.A., Kanthasamy, A.G., Schroder, K., Cooper, M.A., Woodruff, T.M., 2018. Inflammasome inhibition prevents α-synuclein pathology and dopaminergic neurodegeneration in mice. Sci Transl Med 10. https://doi.org/10.1126/scitranslmed.aah4066

Gräb, J., Rybniker, J., 2019. The Expanding Role of p38 Mitogen-Activated Protein Kinase in Programmed Host Cell Death. Microbiol Insights 12, 117863611986459. https://doi.org/10.1177/1178636119864594

Guo, J., Huang, X., Dou, L., Yan, M., Shen, T., Tang, W., Li, J., 2022a. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. Signal Transduct Target Ther 7, 391. https://doi.org/10.1038/s41392-022-01251-0

Guo, J., Huang, X., Dou, L., Yan, M., Shen, T., Tang, W., Li, J., 2022b. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. Signal Transduct Target Ther 7, 391. https://doi.org/10.1038/s41392-022-01251-0

Han, J.-H., Park, D., Jung, B., Yoo, H.-G., Cho, K.-O., Kim, E., 2023. Abstract 1728: Exosomal delivery of Fas-associated factor 1 (FAF1), a multifunctional tumor suppressor and genuine exosome cargo, effectively antagonizes tumor growth. Cancer Res 83, 1728–1728. https://doi.org/10.1158/1538-7445.AM2023-1728

Hardingham, G.E., Bading, H., 2010. Synaptic versus extrasynaptic NMDA receptor signalling: implications for neurodegenerative disorders. Nat Rev Neurosci 11, 682–696. https://doi.org/10.1038/nrn2911

Heneka, M.T., Carson, M.J., Khoury, J. El, Landreth, G.E., Brosseron, F., Feinstein, D.L., Jacobs, A.H., Wyss-Coray, T., Vitorica, J., Ransohoff, R.M., Herrup, K., Frautschy, S.A., Finsen, B., Brown, G.C., Verkhratsky, A., Yamanaka, K., Koistinaho, J., Latz, E., Halle, A., Petzold, G.C., Town, T., Morgan, D., Shinohara, M.L., Perry, V.H., Holmes, C., Bazan, N.G., Brooks, D.J., Hunot, S., Joseph, B., Deigendesch, N., Garaschuk, O., Boddeke, E., Dinarello, C.A., Breitner, J.C., Cole, G.M., Golenbock, D.T., Kummer, M.P., 2015. Neuroinflammation in Alzheimer's disease. Lancet Neurol. https://doi.org/10.1016/S1474-4422(15)70016-5

Hilton, D.J., Richardson, R.T., Alexander, W.S., Viney, E.M., Willson, T.A., Sprigg, N.S., Starr, R., Nicholson, S.E., Metcalf, D., Nicola, N.A., 1998. Twenty proteins containing a C-terminal SOCS box form five structural classes. Proceedings of the National Academy of Sciences 95, 114–119. https://doi.org/10.1073/pnas.95.1.114

Hok-A-Hin, Y.S., del Campo, M., Boiten, W.A., Stoops, E., Vanhooren, M., Lemstra, A.W., van der Flier, W.M., Teunissen, C.E., 2023. Neuroinflammatory CSF biomarkers MIF, sTREM1, and

Journal Pre-proof

sTREM2 show dynamic expression profiles in Alzheimer's disease. J Neuroinflammation 20, 107. https://doi.org/10.1186/s12974-023-02796-9

Hu, J., Wang, X., Chen, X., Fang, Y., Chen, K., Peng, W., Wang, Z., Guo, K., Tan, X., Liang, F., Lin, L., Xiong, Y., 2022. Hydroxychloroquine attenuates neuroinflammation following traumatic brain injury by regulating the TLR4/NF-κB signaling pathway. J Neuroinflammation 19, 71. https://doi.org/10.1186/s12974-022-02430-0

Hu, X., Lu, Z., Yu, S., Reilly, J., Liu, F., Jia, D., Qin, Y., Han, S., Liu, X., Qu, Z., Lv, Y., Li, J., Huang, Y., Jiang, T., Jia, H., Wang, Q., Liu, J., Shu, X., Tang, Z., Liu, M., 2019. CERKL regulates autophagy via the NAD-dependent deacetylase SIRT1. Autophagy 15, 453–465. https://doi.org/10.1080/15548627.2018.1520548

Hua, Y., Dai, X., Xu, Y., Xing, G., Liu, H., Lu, T., Chen, Y., Zhang, Y., 2022. Drug repositioning: Progress and challenges in drug discovery for various diseases. Eur J Med Chem 234, 114239. https://doi.org/10.1016/j.ejmech.2022.114239

Iba, M., Kim, C., Kwon, S., Szabo, M., Horan-Portelance, L., Peer, C.J., Figg, W.D., Reed, X., Ding, J., Lee, S.-J., Rissman, R.A., Cookson, M.R., Overk, C., Wrasidlo, W., Masliah, E., 2023. Inhibition of p38α MAPK restores neuronal p38γ MAPK and ameliorates synaptic degeneration in a mouse model of DLB/PD. Sci Transl Med 15. https://doi.org/10.1126/scitranslmed.abq6089

Iribarren, P., Chen, K., Hu, J., Gong, W., Cho, E.H., Lockert, S., Uranchimeg, B., Wang, J.M., 2005. CpG-containing oligodeoxynucleotide promotes microglial cell uptake of amyloid β 1–42 peptide by up-regulating the expression of the G-protein-coupled receptor mFPR2. The FASEB Journal 19, 2032–2034. https://doi.org/10.1096/fj.05-4578fje

Jadidi-Niaragh, F., Mirshafiey, A., 2010. Histamine and histamine receptors in pathogenesis and treatment of multiple sclerosis. Neuropharmacology 59, 180–189. https://doi.org/10.1016/j.neuropharm.2010.05.005

Jarada, T.N., Rokne, J.G., Alhajj, R., 2020. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. J Cheminform 12, 46. https://doi.org/10.1186/s13321-020-00450-7

Jellinger, K.A., 2010. Basic mechanisms of neurodegeneration: a critical update. J Cell Mol Med 14, 457–487. https://doi.org/10.1111/j.1582-4934.2010.01010.x

Jia, Y., Li, Z., Cai, W., Xiao, D., Han, S., Han, F., Bai, X., Wang, K., Liu, Y., Li, X., Guan, H., Hu, D., 2018. SIRT1 regulates inflammation response of macrophages in sepsis mediated by long noncoding RNA. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1864, 784–792. https://doi.org/10.1016/j.bbadis.2017.12.029

Jiang, Y., Zou, Y., Chen, S., Zhu, C., Wu, A., Liu, Y., Ma, L., Zhu, D., Ma, X., Liu, M., Kang, Z., Pi, R., Peng, F., Wang, Q., Chen, X., 2013. The anti-inflammatory effect of donepezil on experimental autoimmune encephalomyelitis in C57 BL/6 mice. Neuropharmacology 73, 415–424. https://doi.org/10.1016/j.neuropharm.2013.06.023

Jin, H., Kanthasamy, A., Ghosh, A., Anantharam, V., Kalyanaraman, B., Kanthasamy, A.G., 2014. Mitochondria-targeted antioxidants for treatment of Parkinson's disease: Preclinical and clinical outcomes. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1842, 1282–1294. https://doi.org/10.1016/j.bbadis.2013.09.007

Journal Pre-proof

Jin, W., Qazi, T.J., Quan, Z., Li, N., Qing, H., 2019. Dysregulation of Transcription Factors: A Key Culprit Behind Neurodegenerative Disorders. The Neuroscientist 25, 548–565. https://doi.org/10.1177/1073858418811787

Johnson, D.E., O'Keefe, R.A., Grandis, J.R., 2018. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. Nat Rev Clin Oncol 15, 234–248. https://doi.org/10.1038/nrclinonc.2018.8

Jung, Y.J., Tweedie, D., Scerba, M.T., Kim, D.S., Palmas, M.F., Pisanu, A., Carta, A.R., Greig, N.H., 2021. Repurposing Immunomodulatory Imide Drugs (IMiDs) in Neuropsychiatric and Neurodegenerative Disorders. Front Neurosci 15. https://doi.org/10.3389/fnins.2021.656921

Kakoti, B.B., Bezbaruah, R., Ahmed, N., 2022. Therapeutic drug repositioning with special emphasis on neurodegenerative diseases: Threats and issues. Front Pharmacol 13. https://doi.org/10.3389/fphar.2022.1007315

Kamizono, S., Hanada, T., Yasukawa, H., Minoguchi, S., Kato, R., Minoguchi, M., Hattori, K., Hatakeyama, S., Yada, M., Morita, S., Kitamura, T., Kato, H., Nakayama, K., Yoshimura, A., 2001. The SOCS Box of SOCS-1 Accelerates Ubiquitin-dependent Proteolysis of TEL-JAK2. Journal of Biological Chemistry 276, 12530–12538. https://doi.org/10.1074/jbc.M010074200

Kamura, T., Sato, S., Haque, D., Liu, L., Kaelin, W.G., Conaway, R.C., Conaway, J.W., 1998. The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families. Genes Dev 12, 3872–3881. https://doi.org/10.1101/gad.12.24.3872

Kapaki, E., Vakrakou, A.G., Boufidou, F., 2022a. Novel CSF Biomarkers Tracking Autoimmune Inflammatory and Neurodegenerative Aspects of CNS Diseases. Diagnostics 13, 73. https://doi.org/10.3390/diagnostics13010073

Kapaki, E., Vakrakou, A.G., Boufidou, F., 2022b. Novel CSF Biomarkers Tracking Autoimmune Inflammatory and Neurodegenerative Aspects of CNS Diseases. Diagnostics 13, 73. https://doi.org/10.3390/diagnostics13010073

Khanahmadi, M., Manafi, B., Tayebinia, H., Karimi, J., Khodadadi, I., 2020. Downregulation of Sirt1 is correlated to upregulation of p53 and increased apoptosis in epicardial adipose tissue of patients with coronary artery disease. EXCLI J 19, 1387–1398. https://doi.org/10.17179/excli2020-2423

Kim, R.-E., Shin, C.Y., Han, S.-H., Kwon, K.J., 2020. Astaxanthin Suppresses PM2.5-Induced Neuroinflammation by Regulating Akt Phosphorylation in BV-2 Microglial Cells. Int J Mol Sci 21, 7227. https://doi.org/10.3390/ijms21197227

Kwon, H.S., Koh, S.-H., 2020. Neuroinflammation in neurodegenerative disorders: the roles of microglia and astrocytes. Transl Neurodegener 9, 42. https://doi.org/10.1186/s40035-020-00221-2

Labouret, M., Costi, S., Bondet, V., Trebossen, V., Le Roux, E., Ntorkou, A., Bartoli, S., Auvin, S., Bader-Meunier, B., Baudouin, V., Corseri, O., Dingulu, G., Ducrocq, C., Dumaine, C., Elmaleh, M., Fabien, N., Faye, A., Hau, I., Hentgen, V., Kwon, T., Meinzer, U., Ouldali, N., Parmentier, C., Pouletty, M., Renaldo, F., Savioz, I., Rozenberg, F., Frémond, M.-L., Lepelley, A., Rice, G.I., Seabra, L., Benoist, J.-F., Duffy, D., Crow, Y.J., Ellul, P., Melki, I., 2023a. Juvenile Neuropsychiatric Systemic Lupus Erythematosus: Identification of Novel Central

Neuroinflammation Biomarkers. J Clin Immunol 43, 615–624. https://doi.org/10.1007/s10875-022-01407-1

Labouret, M., Costi, S., Bondet, V., Trebossen, V., Le Roux, E., Ntorkou, A., Bartoli, S., Auvin, S., Bader-Meunier, B., Baudouin, V., Corseri, O., Dingulu, G., Ducrocq, C., Dumaine, C., Elmaleh, M., Fabien, N., Faye, A., Hau, I., Hentgen, V., Kwon, T., Meinzer, U., Ouldali, N., Parmentier, C., Pouletty, M., Renaldo, F., Savioz, I., Rozenberg, F., Frémond, M.-L., Lepelley, A., Rice, G.I., Seabra, L., Benoist, J.-F., Duffy, D., Crow, Y.J., Ellul, P., Melki, I., 2023b. Juvenile Neuropsychiatric Systemic Lupus Erythematosus: Identification of Novel Central Neuroinflammation Biomarkers. J Clin Immunol 43, 615–624. https://doi.org/10.1007/s10875-022-01407-1

Lassmann, H., van Horssen, J., 2016. Oxidative stress and its impact on neurons and glia in multiple sclerosis lesions. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1862, 506–510. https://doi.org/10.1016/j.bbadis.2015.09.018

Lee, K.H., Cha, M., Lee, B.H., 2021. Crosstalk between Neuron and Glial Cells in Oxidative Injury and Neuroprotection. Int J Mol Sci 22, 13315. https://doi.org/10.3390/ijms222413315

Li, K., Liu, T.-X., Li, J.-F., Ma, Y.-R., Liu, M.-L., Wang, Y.-Q., Wu, R., Li, B., Shi, L.-Z., Chen, C., 2019. rhEPO inhibited cell apoptosis to alleviate acute kidney injury in sepsis by AMPK/SIRT1 activated autophagy. Biochem Biophys Res Commun 517, 557–565. https://doi.org/10.1016/j.bbrc.2019.07.027

Liddelow, S.A., Barres, B.A., 2017. Reactive Astrocytes: Production, Function, and Therapeutic Potential. Immunity. https://doi.org/10.1016/j.immuni.2017.06.006

Linnartz, B., Neumann, H., 2013. Microglial activatory (immunoreceptor tyrosine-based activation motif)- and inhibitory (immunoreceptor tyrosine-based inhibition motif)-signaling receptors for recognition of the neuronal glycocalyx. Glia 61, 37–46. https://doi.org/10.1002/glia.22359

Lisek, M., Przybyszewski, O., Zylinska, L., Guo, F., Boczek, T., 2023. The Role of MEF2 Transcription Factor Family in Neuronal Survival and Degeneration. Int J Mol Sci 24, 3120. https://doi.org/10.3390/ijms24043120

Liu, B., Ou, W.-C., Fang, L., Tian, C.-W., Xiong, Y., 2022. Myocyte Enhancer Factor 2A Plays a Central Role in the Regulatory Networks of Cellular Physiopathology. Aging Dis 0. https://doi.org/10.14336/AD.2022.0825

Liu, J., Zhao, M., Zhang, Z., Cui, L., Zhou, X., Zhang, W., Chu, S., Zhang, D., Chen, N., 2020. Rg1 improves LPS-induced Parkinsonian symptoms in mice via inhibition of NF-κB signaling and modulation of M1/M2 polarization. Acta Pharmacol Sin 41, 523–534. https://doi.org/10.1038/s41401-020-0358-x

Loeuillard, E., Fischbach, S.R., Gores, G.J., Rizvi, S., 2019. Animal models of cholangiocarcinoma. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1865, 982–992. https://doi.org/10.1016/j.bbadis.2018.03.026

Lull, M.E., Block, M.L., 2010a. Microglial Activation and Chronic Neurodegeneration. Neurotherapeutics 7, 354–365. https://doi.org/10.1016/j.nurt.2010.05.014

Journal Pre-proof

Mao, X.-Y., 2020. Drug Repurposing in Neurological Diseases: Opportunities and Challenges, in: Drug Repurposing - Hypothesis, Molecular Aspects and Therapeutic Applications. IntechOpen. https://doi.org/10.5772/intechopen.93093

Maurya, S.K., Bhattacharya, N., Mishra, S., Bhattacharya, A., Banerjee, P., Senapati, S., Mishra, R., 2021. Microglia Specific Drug Targeting Using Natural Products for the Regulation of Redox Imbalance in Neurodegeneration. Front Pharmacol 12. https://doi.org/10.3389/fphar.2021.654489

McGeer, P.L., McGeer, E.G., 2008. Glial reactions in Parkinson's disease. Movement Disorders 23, 474–483. https://doi.org/10.1002/mds.21751

Meyer, M., Kuffner, K., Winter, J., Neumann, I.D., Wetzel, C.H., Jurek, B., 2020. Myocyte Enhancer Factor 2A (MEF2A) Defines Oxytocin-Induced Morphological Effects and Regulates Mitochondrial Function in Neurons. Int J Mol Sci 21, 2200. https://doi.org/10.3390/ijms21062200

Mohamed, E.A., Ahmed, H.I., Zaky, H.S., Badr, A.M., 2021. Sesame oil mitigates memory impairment, oxidative stress, and neurodegeneration in a rat model of Alzheimer's disease. A pivotal role of NF-κB/p38MAPK/BDNF/PPAR-γ pathways. J Ethnopharmacol 267, 113468. https://doi.org/10.1016/j.jep.2020.113468

Noble, W., Garwood, C.J., Hanger, D.P., 2009. Minocycline as a potential therapeutic agent in neurodegenerative disorders characterized by protein misfolding. Prion 3, 78–83. https://doi.org/10.4161/pri.3.2.8820

Oh, S., Son, M., Choi, J., Lee, S., Byun, K., 2018. sRAGE prolonged stem cell survival and suppressed RAGE-related inflammatory cell and T lymphocyte accumulations in an Alzheimer's disease model. Biochem Biophys Res Commun 495, 807–813. https://doi.org/10.1016/j.bbrc.2017.11.035

Paolicelli, R.C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., Giustetto, M., Ferreira, T.A., Guiducci, E., Dumas, L., Ragozzino, D., Gross, C.T., 2011. Synaptic Pruning by Microglia Is Necessary for Normal Brain Development. Science (1979) 333, 1456–1458. https://doi.org/10.1126/science.1202529

Park, G., Kim, B.-S., Kim, E., 2020. A novel function of FAF1, which induces dopaminergic neuronal death through cell-to-cell transmission. Cell Communication and Signaling 18, 133. https://doi.org/10.1186/s12964-020-00632-8

Park, H.S., Song, Y.S., Moon, B.S., Yoo, S.-E., Lee, J.M., Chung, Y.-T., Kim, E., Lee, B.C., Kim, S.E., 2020. Neurorestorative Effects of a Novel Fas-Associated Factor 1 Inhibitor in the MPTP Model: An [18F]FE-PE2I Positron Emission Tomography Analysis Study. Front Pharmacol 11. https://doi.org/10.3389/fphar.2020.00953

Park, K., Atchison, M.L., 1991. Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta), that binds to the immunoglobulin kappa 3' enhancer and the immunoglobulin heavy-chain mu E1 site. Proceedings of the National Academy of Sciences 88, 9804–9808. https://doi.org/10.1073/pnas.88.21.9804

Parkhurst, C.N., Yang, G., Ninan, I., Savas, J.N., Yates, J.R., Lafaille, J.J., Hempstead, B.L., Littman, D.R., Gan, W.-B., 2013. Microglia Promote Learning-Dependent Synapse Formation through Brain-Derived Neurotrophic Factor. Cell 155, 1596–1609. https://doi.org/10.1016/j.cell.2013.11.030

Pascual, M., Calvo-Rodriguez, M., Núñez, L., Villalobos, C., Ureña, J., Guerri, C., 2021. Toll-like receptors in neuroinflammation, neurodegeneration, and alcohol-induced brain damage. IUBMB Life 73, 900–915. https://doi.org/10.1002/iub.2510

Peng, X., Wang, J., Peng, J., Jiang, H., Le, K., 2022. Resveratrol Improves Synaptic Plasticity in Hypoxic-Ischemic Brain Injury in Neonatal Mice via Alleviating SIRT1/NF-κB Signaling–Mediated Neuroinflammation. Journal of Molecular Neuroscience 72, 113–125. https://doi.org/10.1007/s12031-021-01908-5

Piccinini, A.M., Midwood, K.S., 2010. DAMPening Inflammation by Modulating TLR Signalling. Mediators Inflamm 2010, 1–21. https://doi.org/10.1155/2010/672395

Pizarro Galleguillos, B.M., Mohamed, W., Periñan, M.T., 2022. The Glycoprotein <scp> GPNMB </scp> : A Secret Ingredient for Understanding Parkinson's Disease Etiology? Movement Disorders 37, 2342–2342. https://doi.org/10.1002/mds.29264

Reddy, P.H., Beal, M.F., 2008. Amyloid beta, mitochondrial dysfunction and synaptic damage: implications for cognitive decline in aging and Alzheimer's disease. Trends Mol Med 14, 45–53. https://doi.org/10.1016/j.molmed.2007.12.002

Rena, G., Hardie, D.G., Pearson, E.R., 2017. The mechanisms of action of metformin. Diabetologia 60, 1577–1585. https://doi.org/10.1007/s00125-017-4342-z

Rothstein, J.D., Patel, S., Regan, M.R., Haenggeli, C., Huang, Y.H., Bergles, D.E., Jin, L., Dykes Hoberg, M., Vidensky, S., Chung, D.S., Toan, S.V., Bruijn, L.I., Su, Z., Gupta, P., Fisher, P.B., 2005. β-Lactam antibiotics offer neuroprotection by increasing glutamate transporter expression. Nature 433, 73–77. https://doi.org/10.1038/nature03180

Rubinsztein, D.C., Mariño, G., Kroemer, G., 2011. Autophagy and Aging. Cell 146, 682–695. https://doi.org/10.1016/j.cell.2011.07.030

Ryu, S.W., Lee, S.J., Park, M.Y., Jun, J. il, Jung, Y.K., Kim, E., 2003. Fas-associated factor 1, FAF1, Is a member of Fas death-inducing signaling complex. Journal of Biological Chemistry 278, 24003–24010. https://doi.org/10.1074/jbc.M302200200

Saade, M., Araujo de Souza, G., Scavone, C., Kinoshita, P.F., 2021. The Role of GPNMB in Inflammation. Front Immunol 12. https://doi.org/10.3389/fimmu.2021.674739

Shen, L.-L., Sun, H.-Y., Wang, H.-Q., 2023. [Therapeutic potential of targeting SIRT1 for the treatment of Alzheimer's disease]. Sheng Li Xue Bao 75, 99–107.

Shi, Z., Zhu, L., Song, Z., Liu, T., Hao, D., 2023. Roles of p38 <scp>MAPK</scp> signalling in intervertebral disc degeneration. Cell Prolif 56. https://doi.org/10.1111/cpr.13438

Sofroniew, M. V., Vinters, H. V., 2010. Astrocytes: biology and pathology. Acta Neuropathol 119, 7–35. https://doi.org/10.1007/s00401-009-0619-8

Song, E.J., Yim, S.-H., Kim, E., Kim, N.-S., Lee, K.-J., 2005. Human Fas-Associated Factor 1, Interacting with Ubiquitinated Proteins and Valosin-Containing Protein, Is Involved in the Ubiquitin-Proteasome Pathway. Mol Cell Biol 25, 2511–2524. https://doi.org/10.1128/MCB.25.6.2511-2524.2005

Journal Pre-proof

Song, J., Park, J.K., Lee, J., Choi, Y., Ryu, K., Kim, J., Kim, E., Lee, K., Jeon, Y., Kim, E.E., 2009. Structure and interaction of ubiquitin-associated domain of human Fas-associated factor 1. Protein Science 18, 2265–2276. https://doi.org/10.1002/pro.237

Song, X., Tan, L., Wang, M., Ren, C., Guo, C., Yang, B., Ren, Y., Cao, Z., Li, Y., Pei, J., 2021. Myricetin: A review of the most recent research. Biomedicine & Pharmacotherapy 134, 111017. https://doi.org/10.1016/j.biopha.2020.111017

Storer, P.D., Xu, J., Chavis, J., Drew, P.D., 2005. Peroxisome proliferator-activated receptor-gamma agonists inhibit the activation of microglia and astrocytes: Implications for multiple sclerosis. J Neuroimmunol 161, 113–122. https://doi.org/10.1016/j.jneuroim.2004.12.015

Sul, J.-W., Park, M.-Y., Shin, J., Kim, Y.-R., Yoo, S.-E., Kong, Y.-Y., Kwon, K.-S., Lee, Y.H., Kim, E., 2013. Accumulation of the parkin substrate, FAF1, plays a key role in the dopaminergic neurodegeneration. Hum Mol Genet 22, 1558–1573. https://doi.org/10.1093/hmg/ddt006

Swanson, C.R., Joers, V., Bondarenko, V., Brunner, K., Simmons, H.A., Ziegler, T.E., Kemnitz, J.W., Johnson, J.A., Emborg, M.E., 2011. The PPAR-γ agonist pioglitazone modulates inflammation and induces neuroprotection in parkinsonian monkeys. J Neuroinflammation 8, 91. https://doi.org/10.1186/1742-2094-8-91

Sweeney, M.D., Sagare, A.P., Zlokovic, B. V., 2018. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. Nat Rev Neurol. https://doi.org/10.1038/nrneurol.2017.188

Tamburrino, A., Churchill, M.J., Wan, O.W., Colino-Sanguino, Y., Ippolito, R., Bergstrand, S., Wolf, D.A., Herz, N.J., Sconce, M.D., Björklund, A., Meshul, C.K., Decressac, M., 2015. Cyclosporin promotes neurorestoration and cell replacement therapy in pre-clinical models of Parkinson's disease. Acta Neuropathol Commun 3, 84. https://doi.org/10.1186/s40478-015-0263-6

Tiwari, Vinod, Uniyal, A., Tiwari, Vineeta, Thakur, V., Rani, M., Akhilesh, 2023. Delineating the Neuroinflammatory Crosstalk in Neurodegeneration and Probing the Near Future Therapeutics, in: Traditional Medicine for Neuronal Health. BENTHAM SCIENCE PUBLISHERS, pp. 24–46. https://doi.org/10.2174/9789815040197123010005

Van der Perren, A., Macchi, F., Toelen, J., Carlon, M.S., Maris, M., de Loor, H., Kuypers, D.R.J., Gijsbers, R., Van den Haute, C., Debyser, Z., Baekelandt, V., 2015. FK506 reduces neuroinflammation and dopaminergic neurodegeneration in an α-synuclein-based rat model for Parkinson's disease. Neurobiol Aging 36, 1559–1568. https://doi.org/10.1016/j.neurobiolaging.2015.01.014

Varley, J., Brooks, D.J., Edison, P., 2015. Imaging neuroinflammation in Alzheimer's disease and other dementias: Recent advances and future directions. Alzheimer's & Dementia 11, 1110–1120. https://doi.org/10.1016/j.jalz.2014.08.105

Varum, S., Baggiolini, A., Zurkirchen, L., Atak, Z.K., Cantù, C., Marzorati, E., Bossart, R., Wouters, J., Häusel, J., Tuncer, E., Zingg, D., Veen, D., John, N., Balz, M., Levesque, M.P., Basler, K., Aerts, S., Zamboni, N., Dummer, R., Sommer, L., 2019. Yin Yang 1 Orchestrates a Metabolic Program Required for Both Neural Crest Development and Melanoma Formation. Cell Stem Cell 24, 637-653.e9. https://doi.org/10.1016/j.stem.2019.03.011

Verkhratsky, A., Butt, A.M., 2023. Neurodegenerative diseases, in: Neuroglia. Elsevier, pp. 563–598. https://doi.org/10.1016/B978-0-12-821565-4.00012-2

Journal Pre-proof

Wai, D.C.C., Shihab, M., Low, J.K.K., Mackay, J.P., 2016. The zinc fingers of YY1 bind single-stranded RNA with low sequence specificity. Nucleic Acids Res gkw590. https://doi.org/10.1093/nar/gkw590

Wan, B., Huang, L., Jing, C., Li, Y., Jiao, N., Liang, M., Jiang, S., Yang, W., 2022. Zearalenone promotes follicle development through activating the SIRT1/PGC-1α signaling pathway in the ovaries of weaned gilts. J Anim Sci 100. https://doi.org/10.1093/jas/skac058

Wang, J., Fivecoat, H., Ho, L., Pan, Y., Ling, E., Pasinetti, G.M., 2010. The role of Sirt1: At the crossroad between promotion of longevity and protection against Alzheimer's disease neuropathology. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1804, 1690–1694. https://doi.org/10.1016/j.bbapap.2009.11.015

Wang, Q., Xiao, B., Cui, S., Song, H., Qian, Y., Dong, L., An, H., Cui, Y., Zhang, W., He, Y., Zhang, J., Yang, J., Zhang, F., Hu, G., Gong, X., Yan, Z., Zheng, Y., Wang, X., 2014. Triptolide treatment reduces Alzheimer's disease (AD)-like pathology through inhibition of BACE1 in a transgenic mouse model of AD. Dis Model Mech 7, 1385–1395. https://doi.org/10.1242/dmm.018218

Wen, J.-H., He, X.-H., Feng, Z.-S., Li, D.-Y., Tang, J.-X., Liu, H.-F., 2023. Cellular Protein Aggregates: Formation, Biological Effects, and Ways of Elimination. Int J Mol Sci 24, 8593. https://doi.org/10.3390/ijms24108593

Wyss-Coray, T., Mucke, L., 2002a. Inflammation in Neurodegenerative Disease—A Double-Edged Sword. Neuron 35, 419–432. https://doi.org/10.1016/S0896-6273(02)00794-8

Wyss-Coray, T., Mucke, L., 2002b. Inflammation in Neurodegenerative Disease—A Double-Edged Sword. Neuron 35, 419–432. https://doi.org/10.1016/S0896-6273(02)00794-8

Xu, M., Wang, J., Zhang, X., Yan, T., Wu, B., Bi, K., Jia, Y., 2020. Polysaccharide from Schisandra chinensis acts via LRP-1 to reverse microglia activation through suppression of the NF-κB and MAPK signaling. J Ethnopharmacol 256, 112798. https://doi.org/10.1016/j.jep.2020.112798

Yanagisawa, S., Baker, J.R., Vuppusetty, C., Koga, T., Colley, T., Fenwick, P., Donnelly, L.E., Barnes, P.J., Ito, K., 2018. The dynamic shuttling of SIRT1 between cytoplasm and nuclei in bronchial epithelial cells by single and repeated cigarette smoke exposure. PLoS One 13, e0193921. https://doi.org/10.1371/journal.pone.0193921

Youle, R.J., van der Bliek, A.M., 2012. Mitochondrial Fission, Fusion, and Stress. Science (1979) 337, 1062–1065. https://doi.org/10.1126/science.1219855

Yu, C., Kim, B., Park, M., Do, Y.-J., Kong, Y.-Y., Kim, E., 2018. FAF1 mediates necrosis through JNK1-mediated mitochondrial dysfunction leading to retinal degeneration in the ganglion cell layer upon ischemic insult. Cell Communication and Signaling 16, 56. https://doi.org/10.1186/s12964-018-0265-7

Yue, J.K., Kobeissy, F.H., Jain, S., Sun, X., Phelps, R.R.L., Korley, F.K., Gardner, R.C., Ferguson, A.R., Huie, J.R., Schneider, A.L.C., Yang, Z., Xu, H., Lynch, C.E., Deng, H., Rabinowitz, M., Vassar, M.J., Taylor, S.R., Mukherjee, P., Yuh, E.L., Markowitz, A.J., Puccio, A.M., Okonkwo, D.O., Diaz-Arrastia, R., Manley, G.T., Wang, K.K.W., Badjatia, N., Foreman, B., Gopinath, S., Grandhi, R., Jha, R.M., Lingsma, H.F., Madden, C., Madhok, D.Y., McCrea, M.A., Merchant, R., Nelson, L.D., Ngwenya, L.B., Robertson, C.S., Rodgers, R.B., Satris, G.G., Schnyer, D.M., Valadka, A.B., van Essen, T.A., Zafonte, R., 2023. Neuroinflammatory Biomarkers for Traumatic Brain Injury

Diagnosis and Prognosis: A TRACK-TBI Pilot Study. Neurotrauma Rep 4, 171–183. https://doi.org/10.1089/neur.2022.0060

Zaib, S., Javed, H., Khan, I., Jaber, F., Sohail, A., Zaib, Z., Mehboob, T., Tabassam, N., Ogaly, H.A., 2023. Neurodegenerative Diseases: Their Onset, Epidemiology, Causes and Treatment. ChemistrySelect 8. https://doi.org/10.1002/slct.202300225

Zhang, Z., Cao, M., Chang, C.-W., Wang, C., Shi, X., Zhan, X., Birnbaum, S.G., Bezprozvanny, I., Huber, K.M., Wu, J.I., 2016. Autism-Associated Chromatin Regulator Brg1/SmarcA4 Is Required for Synapse Development and Myocyte Enhancer Factor 2-Mediated Synapse Remodeling. Mol Cell Biol 36, 70–83. https://doi.org/10.1128/MCB.00534-15

Zhang, Z., Zhao, Y., 2022. Progress on the roles of MEF2C in neuropsychiatric diseases. Mol Brain 15, 8. https://doi.org/10.1186/s13041-021-00892-6

Zurkirchen, L., Varum, S., Giger, S., Klug, A., Häusel, J., Bossart, R., Zemke, M., Cantù, C., Atak, Z.K., Zamboni, N., Basler, K., Sommer, L., 2019. Yin Yang 1 sustains biosynthetic demands during brain development in a stage-specific manner. Nat Commun 10, 2192. https://doi.org/10.1038/s41467-019-09823-5

Statements & Declarations

Funding

"The authors declare that no funds, grants, or other support were received during the preparation of this manuscript."

Declaration of Competing Interests

The authors declare no competing interests.

Author Contributions

Conceived and designed by P.K. Materials collected and artwork done by S.K. Critical evaluation and analysis of data done R.K.A. and P.K. Manuscript written by S.K., R.K.A. and P.K. All authors read the manuscript and agreed to submit.

Data Availability

No datasets were generated or analyzed during the current study.

Journal Pre-proof

Journal Pre-proof

Author's contribution

Conceived and designed by P.K. Materials collected and artwork done by S.K., Critical evaluation and analysis of data done by R.K.A, P.K. Manuscript written by S.K., R.K.A. and P.K. All authors read the manuscript and agreed to submit it.

Declaration of Competing Interest

The authors declare no conflict and competing interest.

Data Availability

No data was used for the research described in the article.

Highlights

- Neuroinflammation profoundly affects neurodegenerative diseases (NDDs) progression.
- Pro-inflammatory transcription factors trigger neuroinflammatory responses.
- Pro-survival transcription factors provide neuroprotection against inflammation.
- Drug repositioning is a potential NDD treatment strategy.
- Identifying biomarkers aids early diagnosis and intervention.

Journal Pre-proof

Journal Pre-proof



Physicochemical characterization of novel okra mucilage/ hyaluronic acid-based oral disintegrating films for functional food applications

Krizma Khatreja, Deenan Santhiya

PII: S0141-8130(24)05438-2

DOI: https://doi.org/10.1016/j.ijbiomac.2024.134633

Reference: BIOMAC 134633

To appear in: *International Journal of Biological Macromolecules*

Received date: 27 May 2024

Revised date: 25 July 2024

Accepted date: 8 August 2024

Please cite this article as: K. Khatreja and D. Santhiya, Physicochemical characterization of novel okra mucilage/hyaluronic acid-based oral disintegrating films for functional food applications, *International Journal of Biological Macromolecules* (2024), https://doi.org/10.1016/j.ijbiomac.2024.134633

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.

© 2024 Published by Elsevier B.V.

Journal Pre-proof

Physicochemical characterization of novel okra mucilage/hyaluronic acid-based oral disintegrating films for functional food applications

Krizma Khatreja[a], Deenan Santhiya[a*]

[a] Department of Applied Chemistry, Delhi Technological University, Bawana Road, Delhi-110 042, Delhi, India

Correspondence[*] to Dr. Deenan Santhiya

Email: deenan.santhiya@dce.ac.in, Phone No. +91-9958580295

Abstract

Oral disintegrating films (ODFs) offer a patient-friendly approach with enhanced convenience and rapid onset of action over various health benefits. ODFs are fabricated for geriatric, pediatric, and individuals facing swallowing challenges. The present work aims to fabricate and characterize ODFs mainly composed of okra mucilage (OM), hyaluronic acid (HA), vitamin-C-loaded bioactive glass nanoparticles (VBG NPs), and clove essential oil. A bio-inspired method was employed to synthesize VBG NPs using fructose template. The nutrient analysis of OM depicted that it is a rich source of protein, carbohydrates, magnesium, and flavonoids (quercetin), accounting for its antioxidant activity. The physicochemical characteristics of the ODFs studied using contact angle measurement, surface pH, opacity, and *in vitro* disintegration time revealed that ODFs disintegrated rapidly in simulated saliva. The neutral surface pH of ODFs indicates their non-irritant behavior to the oral mucosa. VBG NPs and essential oil (EO) addition enhance the thermal and mechanical properties. Further, EO infusion in the film matrix resulted in the porous and antibacterial nature of the functional film as revealed by FE-SEM micrographs and antibacterial disk diffusion assay respectively. The obtained novel nutrient-rich ODF is hemocompatible with a hemolysis rate (HR%) <5% and suitable for functional food applications.

Keywords: Oral disintegrating film, Functional food, Mouth ulcers, Okra mucilage, Hyaluronic acid film.

1 Introduction:

The term ‘functional food’ refers to foods that provide health benefits beyond the basic nutritional value. If these foods are included in a regular diet, they promote proper growth

and development by decreasing the risk of chronic diseases [1]. A sharp rise in functional food publications in the 21st century shows awareness of personal health amongst the population. As a result, a significant hike in the functional food market is also seen [2]. A broad assemblage of naturally occurring bioactive compounds extracted from plant and animal sources are considered worthy of their best application as a functional food in preserving human health[3].

Ulcers constitute one of the most widespread diseases prevalent among elderly people or geriatrics[4]. The high incidence of this ailment within this particular age bracket could stem from the weakening of mucosal membranes[5]. A long-term subjection of these membranes to various factors, such as annoyance caused by drugs or foods, repercussions of autoimmune and systemic diseases, and the consumption of Non-steroidal anti-inflammatory drugs (NSAIDs) are known for their role in mucosal permeability[6]. Common causes of mouth ulcers, also known as aphthous ulcers, may encompass tissue trauma or injury. Tissue injury arises from biting or oral devices, infections due to the consumption of certain drinks and foods, as well as the use of toothpaste or oral rinses. A recent research on the relation between recurrent aphthous ulcers and dietary factors has revealed that deficiencies in vitamin C [7,8], vitamin D, vitamin B12, ferritin, and folic acid, might play a role in contributing to ulceration [9]. Vitamin C a powerful antioxidant reduces inflammation. Though some local anesthetics including topical gels, creams, ointments, and mouthwashes may relieve ulcers, to date, there is no medication for a permanent cure. Therefore, the strategy for addressing ulcers revolves around reducing pain, promoting the healing process, minimizing both the number and size, and prolonging the recurrence [10]. Hence, innovative and effective strategies are needed to address these ulcerative troubles.

Ongoing research is exploring medications for mouth ulcers and emphasizing mucoadhesive gels as well as patches for sustained drug delivery to the ulcer sites. For example, investigation on mucoadhesive patches incorporated with *Glycyrrhiza glabra* extract[11], *Ginger officinale* extract[12], several polysaccharides[13] including pectin[14], chitosan[15] and others, cinnamaldehyde from cinnamon bark oil[16] are presented for mouth ulcer treatment. A carotenoid found in the stigmas of the saffron flower, known as crocin, possesses anti-inflammatory and anti-oxidative properties. The formulation of hydrogels containing crocin-loaded niosomes offered an efficient topical drug delivery system for addressing aphthous stomatitis[17]. Yahom buccal mucoadhesive tablets were developed using a direct compression process[4] and nanofibrous mesh-based multilayered mucoadhesive patches.

These patches were also infused with medication possessing anti-inflammatory, antioxidant, antibacterial properties, and pain relief[18]. However, it is important to note that these mucoadhesive preparations are not without potential side effects. The side effects may include local irritation at the application site, allergic reactions, sensations of dry mouth or increased thirst, alterations in taste perception, and restrictions in eating and drinking [19]. Therefore, recent advances in the medication for mouth ulcers involve oral disintegrating formulations in the form of films and tablets [20].

Oral disintegrating/dissolving formulations help in the swift absorption of active components into the oral mucosa and facilitate easy consumption for people who experience dysphagia[21]. Nurokind a commercially available fast-disintegrating oral strip designed to treat vitamin B12 deficiency is an interesting example of these formulations. An orally disintegrating tablet formulated using Tadalafil-loaded Limonene was examined for its gastric ulcer protective effect in a rat model [22]. *Cordia verbenacea* extract integrated into starch and hydroxypropyl methylcellulose matrix[23], Hyaluronic acid (HA) -supplemented vitamin D oral films[24] are other recent examples of oral disintegrating films (ODFs) for biomedical applications. Several reports suggest the use of okra mucilage to mask the taste of bitter and unpalatable drugs [25]. The present study aimed to prepare flexible multifunctional ODFs made up of okra mucilage (OM), hyaluronic acid (HA), bioactive glass (BGs), and clove essential oil (EO) as a functional food for mouth ulcers. The clove EO acts as a cross-linker and also serves as an anti-inflammatory, antibacterial, and numbing agent for mouth ulcers [26].

Lady finger or okra *(Abelmoschus esculentus)* belongs to the Malvaceae family which grows in warm temperate, subtropical, and tropical climates. OM contains vitamins, minerals, iron, and flavonoids such as quercetin (also reported in the nutrient analysis report Table 1) which results in antioxidant, anti-ulcer, and anti-inflammatory properties [27]. The extract also greatly accelerates angiogenesis, cell migration, and re-epithelization at chronic wound sites. Therefore formulations derived from okra were reported to be an excellent candidate for tissue engineering applications and also a potential nominee for its applications in wound dressing[28]. In recent findings, artificial saliva resembling the human salivary formulation was prepared using okra mucilage and showed beneficial effects for xerostomia patients[29].

Hyaluronic acid (HA) plays an active role in promoting cell growth and re-epithelization by aiding in the multiplication of basal keratinocytes and activating and self-regulating

inflammatory responses. As a result, several mucoadhesive buccal formulations containing HA have been developed and studied for their efficacy in treating aphthous ulcers [30,31]. Recent advances report the use of HA-based oral films for relieving dry mouth. Further suggesting the incorporation of active functional compounds into the HA-based films for easy and convenient uptake specially designed for people facing swallowing challenges [32].

Bioactive glasses (BGs) despite their varied research in dental and orthopedic applications are also well known for their efficiency in soft tissue regeneration. The ability of BG to bind biologically to soft tissues promotes the growth of epithelial cells, lessens inflammation at the injured site, and promotes angiogenesis. This has attracted researchers for their role in soft tissue regeneration [33]. A recent study providing insights on the use of 45S5 bioglass nanoparticles for ulcers [34].

The objectives of the present study are to synthesize and evaluate a novel ODF named OHVBGEO composed of OM, HA, VBG NPs, and clove EO as a cross-linker. The rationale behind selecting the constituents of ODF is based on their potential benefits for providing nutrition and alleviating pain related to mouth ulcers. The bioactive glass nanoparticles (BG NPs) are synthesized by the bio-inspired method using fructose as a template. The motivation behind using fructose, an organic natural sugar as a template is to functionalize BG NPs. These NPs are further loaded with vitamin C (VBG NPs) and thoroughly characterized using BG NPs as control. The rationale for using as-received OM for film casting is to ensure the usage of all its beneficial components from the okra pods. The nutritional analysis of extracted OM is done using standard test methods. Films made up of OM and HA (OHF) and OM, HA, and VBG NPs (OHVBG) are used as controls. ODFs are characterized using FTIR, XRD, FE-SEM/EDS, contact angle measurement, thermal and mechanical studies. The films are also tested for surface pH, opacity, water vapor permeability, and *in-vitro* disintegration time. The antioxidant assay of the functional film is done using the DPPH method and the biosafety is studied by hemolysis assay.

2 Materials and Methods

2.1 Materials

For mucilage extraction, fresh okra pods were obtained from the local market near Rohini, New Delhi. BG NPs synthesized using precursors namely sodium acetate (NaAc) (C.N. S2889), triethyl phosphate (TEP) (C.N. 821141), tetraethyl orthosilicate (TEOS) (C.N. 86578), calcium acetate (CaAc) (C.N. 1086334) and template fructose (MW = 180.16, purity

≥ 99%) (C.N. F0127) were obtained from Sigma-Aldrich, USA. Vitamin C (MW = 176.12, purity 99%) (C.N. A92902) for loading BG NPs, Hyaluronic acid sodium salt (750-1000 KDa) (C.N. 53163), and clove oil (C.N. C8392) were further obtained from Sigma-Aldrich, USA. For the antibacterial disk diffusion assay the bacterial strains namely *Staphylococcus aureus* (ATCC 29213) and *Escherichia coli* (ATCC 25922) were purchased from CSIR-National Chemical Laboratory (NCL) Pune, India. Other chemicals employed in all the experimentation were high-purity analytical reagent (AR) grade. Every experimental procedure made the use of Milli Q water.

2.2 Methodology

2.2.1 Mucilage extraction from fresh okra pods

For extracting mucilage from the okra pods, a reported procedure[35] was followed with slight modifications. To remove the dirt, okra pods were soaked in running water for 15 min and washed with de-ionized water. The dried pods were then cut into slices of half an inch and immersed in Milli Q water (1:1 w/v ratio of okra pods and water) and stirred for 2 h followed by 12 h incubation to ensure the nutrients from the pods were thoroughly embraced into a mucilage form. Then the mucilage was extracted with the help of a clean muslin cloth as shown in scheme 1 and collected in the falcon tube, sealed with parafilm, and stored at 4°C in a refrigerator until further use. 5 mL of as-received OM was used to fabricate ODFs. The obtained OM was analyzed for its nutrition profile as displayed in Table 1.

2.2.2 Preparation of bioactive glass nanoparticles (BG and VBG)

BG NPs were synthesized using a previously reported bio-inspired route[36]. The fructose template (1mg/mL) was dissolved in 10 mM TRIZMA buffer at pH 8 under a continuous stirring at physiological temperature, followed by the addition of BG precursors namely TEOS (92.9 g/L), TEP (10 g/L), NaAc (63.6 g/L) and CaAc (42.1 g/L) each at an interval of 30 min in the order as mentioned above. After that, the reaction mixture was maintained in a silicon oil bath for the next 24 h at 37°C, followed by centrifugation and washing with 70% v/v ethanol as shown in scheme 1. The obtained BG NPs were dried in an oven and further loaded with vitamin C in a 1:4 w/w ratio of BG NPs to vitamin C as reported by Bharti *et al.* [37] and named VBG NPs. For loading vitamin C, 100 mg/mL of vitamin C was dissolved in Milli Q water for 30 min and 25 mg/mL BG NPs were suspended in the solution [38]. The suspension was further incubated for 24 h in an orbital shaker for effective loading of vitamin C on BG NPs. The sample was centrifuged, washed, and dried in an oven for the next 48 h and until further use stored in a desiccator.

For calculating the loading efficiency (%) of vitamin C on BG NPs, the filtrate was collected after vitamin loading followed by centrifugation and analyzed spectrophotometrically at 261 nm corresponding to the maximum absorbance peak for vitamin C [39].

$$\text{Loading efficiency (\%)} = \frac{Vit_{total} - Vit_{free}}{Vit_{total}} \times 100 \quad (1)$$

where Vit_{total} denotes the total amount of vitamin taken and Vit_{free} is the amount of free vitamin in the filtrate.

2.2.3 Fabrication of ODFs

The solvent casting method was employed to fabricate ODFs. To synthesize **OHVBGEO** (functional film), a homogeneous suspension was made by adding 1mg/mL VBG NPs in 5 mL of as-received OM with constant stirring for 30 min at RT, followed by the addition of 0.2% HA, as depicted in scheme 1. After the complete mixing of HA, clove EO was added as a crosslinker. An emulsion was formulated with a 1:1 v/v ratio of oil and water (60 µL each) and 120 µL of this emulsion was incorporated into the film matrix and swirled for 30 min for complete crosslinking. After obtaining a homogeneous filmogenic suspension, degasification was done. This was followed by casting in a 5 cm diameter PP petri plate, drying overnight at 45°C in an oven, and peeled from PP after complete drying. **OHF** (composed of 5 mL OM and 0.2% HA) and **OHVBG** (composed of 5 mL OM, 0.2% HA, and 1mg/mL VBG) were also prepared as control films. Thickness measurements (using a digital caliper) for OHF, OHVBG, and OHVBGEO films were observed to be 0.09 ± 0.012 mm, 0.14 ± 0.037 mm, and 0.18 ± 0.022 mm respectively. The observed thickness variations in the above films despite maintaining a similar composition of required contents may be attributed to the variation in the distribution of VBG NPs or infusion of EO. The corresponding molecular structures of OM (polysaccharide [40] and protein (lectin [41]) and HA are shown in Fig. 4.

2.3 Particle size and zeta potential

Particle size analysis of BG/VBG NPs was done using high-resolution transmission electron microscopy (HR-TEM, TALOS, Thermo Scientific, USA) operating at 200 kV. ImageJ bundled with 64-bit Java 1.8.0_172 software was used to determine the particle size distribution. Malvern Zetasizer based on the light scattering principle was used to study the zeta potential of NPs. Before the measurements, both BG and VBG NPs were dispersed using ethanol as a dispersion medium.

Journal Pre-proof

2.4 Physical characteristics of the procured ODFs

To assess the quality and performance of synthesized ODFs, several physical characteristics of the films were studied as follows:

2.4.1 Opacity

Opacity measurements were done using a Digital opacity tester (International equipments, Mumbai). A solitary source of light passed via the sample and the system measured relative transmission on a scale of 0 to 100 opacity units, which represents perfectly transparent to purely opaque materials respectively. Calibrations for the machine were done using two standards, one being minimally opaque (transparent plastic sheet) representing white standard and the other being 100% opaque (black carbon paper) which is black standard. The opacity values were calculated [42] using the following equation (2) in which O_{black} represents the opacity of the film using the black standard whereas O_{white} is the opacity using the white standard. The measurements were done in triplicates for all the film samples.

$$\text{Opacity} = \frac{O_{black}}{O_{white}} \times 100 \quad (2)$$

2.4.2 Water vapor permeability

Water vapor permeability (WVP) is the measure of the tendency of water vapor to pass through the film. The cup method is a common technique used to measure WVP in films, measurements were done following the procedure described by Zhou *et al*. [43]. Each sample was cut into a circle with a known diameter and parafilm was used to secure it to the cup holes. 5 g anhydrous calcium chloride (0% RH) was added into the cups sealed with the sample films and the complete system was kept in the desiccator containing water (100% RH). At a regular interval of 2 h weight of the cups was measured for at least up to 12 h. All the measurements were done at RT (25°C ± 2°C) and in triplicates. To calculate WVP the following equation (3) was used:

$$\text{WVP} = \frac{\Delta M \times D}{A \times T \times \Delta P} \quad (3)$$

where **Δ** M represents the change in cup weight (g), D is the thickness of the film (cm), T represents measuring time (s), A is the measuring area (cm^2) and **Δ** P represents the difference in the water vapor pressure across the two sides of the film sample (kPa). Therefore, WVP is expressed in the units (g/cm · s · kPa).

Journal Pre-proof

2.4.3 Contact angle measurements

The measurements were carried out at 20°C ± 2°C using a contact angle measuring system (KRÜSS G10, Germany) with KRÜSS ADVANCE software, by sessile drop method [44]. For all the formulations, contact angle measurements were conducted on the side that was exposed to air during the drying phase of film fabrication. A drop of Milli Q water (5µL) was dropped onto the surface of film (3.0 cm × 2.0 cm) with the help of a precision micro syringe, further, after 10 s the image was recorded. The mean value obtained from right and left contact angles were calculated and reported.

2.4.4 Surface pH

The surface pH measurement of ODFs was done according to Tedesco *et al.*[21], briefly, phosphate buffer saline (PBS) was prepared (800 mg NaCl, 20 mg KCl, 24 mg KH_2PO_4, and 144 mg $NaHPO_4$ per 100 mL DI), which mimics the human saliva. The ODF (3.0 cm × 2.0 cm) was immersed in a Petri plate containing half an mL of the buffer solution for 30 s. The bulb of a calibrated pH meter (Toshcon Industries Pvt. Ltd.) electrode was touched on the surface of the film for a minute and the reported pH value was noted. The measurements were repeated thrice and expressed as mean ± SD.

2.5 In-vitro disintegration time

The disintegration time of ODFs was calculated using the slide frame method as described in [21]. Film with dimensions of 3 cm × 2 cm was fixed on the slide frame and the whole system was further placed on a petri plate. A 200 µL drop of distilled water was placed on the film surface and the time required for the ODF to break down and create a hole was tracked. The time determination represents the *in vitro* disintegration time for the film. All the measurements were done at RT and in triplicates.

2.6 FE-SEM and Elemental Analysis

To study the surface morphology, all the procured samples were examined using field emission scanning electron microscopy (FE-SEM, Zeiss GeminiSEM 500 Thermal field emission type), operating at an accelerating voltage of 20 kV. The samples were coated using gold sputtering which involves bombarding a gold target with energetic ions and this coating enhances the conductivity of the non-conducting samples. The elemental analysis was done using energy dispersive spectroscopy (EDS) with an X flash 6|60 detector component integrated into the Quanta 650 (FEI) system.

Content uniformity of VBG NPs suspended in the film

Journal Pre-proof

FE-SEM images at a magnification of 500 X were recorded for five different sections (right, left, middle, top, and bottom) punched from the OHVBGEO film as shown in (supplementary data) Fig. S1. To further support this analysis, the UV-Spectroscopic method was used for which OHVBGEO film was taken at 3 different diameters (1 cm, 3 cm, and 5 cm), followed by dissolving them in DI and recording the absorbance spectra (190 nm-600 nm) for all the sizes of the film [45]. OHF was used as a blank.

2.7 Fourier transform Infrared spectroscopy

All the samples were crushed using a laboratory-grade mortar and pestle and mixed with potassium bromide powder (KBr) at a ratio of 1:100, followed by pelletization using hydraulic press. The spectra were recorded using a NICOLET 380 FTIR spectrometer, operating with a resolution of 4 cm^{-1} within the range of 4000 – 400 cm^{-1}.

2.8 X-ray diffraction measurements

To figure out the solid nature of the samples, XRD spectra were recorded using a Bruker, D8 Discover X-ray diffractometer with an X-ray source of Cu. Each sample was scanned over a 2θ range of 10° to 70° at a scanning rate of 0.2°/min, operating at 20 kV and a current of 10 mA using Cu-Kα radiations.

2.9 Thermal stability analysis

2.9.1 TGA

Thermogravimetric analysis was done using TGA (TGA 4000, Perkin Elmer, USA), samples weighed 5 mg were heated from 0°C-1000°C (NPs)/ 0°C-800°C (ODFs) at a heating rate of 10°C/min in N_2 environment with a gas flow of 50 mL/min.

2.9.2 DSC

Differential scanning calorimetry measurements were performed using Perkin Elmer DSC 8000, USA. DSC measures the heat flow in or out of a sample as it undergoes controlled heating. The samples weighed 3mg were subjected to a heat flow of 3°C/min at a temperature range of 0°C-700°C in N_2 atmosphere at a flow rate of 50 mL/min and a cooling gas line at a flow rate of 2 mL/min was used.

2.10 Mechanical properties

2.10.1 Rheological measurements

Rheology assists in understanding the film-forming process by evaluating the viscosity and viscoelastic properties of the film formulation. For all measurements, a freshly prepared film suspension was used [46]. The specimens were degassed before placing into the parallel plate

(PP-40, measuring distance of 1 mm) in the rheometer (MCR 702, advanced rheometer, Anton Paar, Austria). The shear-thinning behaviour was measured as a function of shear viscosity and shear rate in the range of 0.1-100 s^{-1}. Amplitude sweep studies were carried out to evaluate the material's response by varying strain amplitudes (0.01 to 1000 strain %) while keeping frequency constant (10 rad/s). This helped to analyze the crossover point of storage (G') and loss (G") moduli, which denote the sol-gel transition of the film. The flow behaviour and mechanical deformation of the films were also measured through frequency sweep at constant strain amplitude (0.1 % strain) obtaining G' and G" as a function of angular frequency (ω) in the range of 1-100 rad/s.

2.10.2 Tensile testing

A universal testing machine (ZwickRoell, Z010, USA) with a force range of 10 kN, was used to evaluate the elastic modulus, elongation at break (EB), and tensile strength (TS), for ODFs following ASTM D882-10 [24]. ODFs were cut into 50mm×5mm dimensions and the moving speed of the grip was adjusted to 12.5 mm/min. The maximum tensile force (W_{max} (N)) and the strain (%) values were recorded. All the above parameters were calculated using the following equations:

$$\text{Tensile strength (MPa)} = \frac{W_{max}}{A_0} \quad (4)$$

$$\text{Elongation at break (\%)} = \frac{\Delta L}{L_0} \times 100 \quad (5)$$

$$\text{Young's modulus / Elastic modulus} = \sigma_2 - \sigma_1 \,/\, \varepsilon_2 - \varepsilon_1 \quad (6)$$

Where W_{max} is the maximum tensile force before the fracture, A_0 is the cross-sectional area of the film, and **Δ** L and L_0 represent the change in the gauge length at the moment of rupture and initial gauge length respectively. Young's or elastic modulus is the ratio between applied stress (σ_1 and σ_1) and applied strain (ε_1 and ε_2) representing the slope of the stress-strain curve.

2.11 Antimicrobial Assay

In vitro antimicrobial activity of the various samples was conducted using the disk diffusion method. To assess the effectiveness of the samples against a wide spectrum of bacterial pathogens, both Gram-positive *Staphylococcus aureus,* a well-documented causative agent of numerous post-operative wound infections, and Gram-negative *Escherichia coli,* a prominent pathogen associated with urinary tract infections (UTIs) were studied[47], adhering to a standard protocol described in [48]. Over the night, bacteria cultures were cultivated in

nutrient broth at a controlled temperature of 37°C. Seven agar plates were prepared for each bacterial strain in the laminar flow hood, onto which 20 µL bacterial inoculum was meticulously spread. The samples (5 mm^2 ODFs, 5 mg NPs, and 60 µL EO (control)) were aseptically placed at the centre of the agar plates, followed by incubation at 37°C for a duration of 18 h - 20 h. The concentrations of the above samples were chosen based on their respective concentrations in the standard 5 cm diameter functional film (OHVBHEO). The presence of any antibacterial agent within the sample led to the development of a clear zone of inhibition, the diameter of which was subsequently measured using a digital caliper.

2.12 Antioxidant property

To validate the presence of antioxidative agents in OM, the DPPH˙ (2,2-Diphenyl-1-picrylhydrazyl) method as described by Sridhar *et al.* [49] was followed with slight modifications. 100 µM DPPH solution was prepared in methanol and stored in dark condition. Various dilutions of as-received OM were done using de-ionized water (DI) in v/v ratios of 1:1 (representing 1 mL OM: 1 mL DI), similarly 2:1, 3:1, 4:1, 5:1, and 6:1. Ascorbic acid (standard) was also diluted with DI in a similar ratio as OM. OHVBGEO prepared using 5 mL OM was also dissolved in 1 mL DI to study enhanced RSA due to the suspended vitamin-C loaded BG NPs. 2 mL of each dilution was mixed with 2 mL volume of the above prepared DPPH solution followed by vigorous shaking and incubation for the next 30 min in darkness. 2 mL methanol mixed with the same volume of DPPH solution was used as a control for each measurement. Absorbance at 515 nm (maximum absorbance value for DPPH) against a blank was noted using a UV-visible spectrophotometer (Eppendorf UV-vis Spectrophotometer (Bio Spectrometer basic model)). All the measurements were done in triplicates and results were represented as mean ± SD. The following equation (7) was used for calculating the radical scavenging activity (RSA%)

$$\text{RSA\%} = \frac{A_{control} - A_{sample}}{A_{control}} \times 100 \quad (7)$$

where $A_{control}$ represents the absorbance of ascorbic acid and A_{sample} is the absorbance of the OM/OHVBGEO.

2.13 Hemocompatibility assay

Hemocompatibility of the fabricated samples was studied according to the standard procedure ISO 10 993-5 [50]. Nearly 4 mL of human fresh blood was collected and preserved in heparinized tubes. Phosphate buffer saline (PBS) was prepared using autoclaved water and adjusted to physiological pH (pH=7.4). All the samples (5 cm diameter films, 5 mg BG and

VBG, 5 mL OM, 120 µL EO, 100 mg/mL vitamin C, and 0.2% HA) were dipped in 10 mL prepared PBS and kept in incubator shaker series (120 rpm) at 37°C for the next 30 min. The concentrations of the above samples were chosen based on their respective concentrations in the standard 5 cm diameter functional film (OHVBHEO). Diluted blood was prepared by mixing 2.0 mL of whole blood with 2.5 mL of PBS. 0.2 mL of the diluted blood was added into all the tubes post-shaking, followed by further incubation for 1 h at 37°C. Physiological saline was used as a negative control and distilled water as a positive control. After incubation for 1 h, the tubes were centrifuged at 1000 rpm for 10 min and the absorbance of the supernatant was measured using an ultraviolet spectrophotometer (Eppendorf UV-vis Spectrophotometer (Bio Spectrometer basic model) at 545 nm which corresponds to the maximum absorbance value for blood hemoglobin. All the measurements were done in triplicates and equation (8) was used to calculate the hemolysis rate (HR%).

$$\mathrm{HR\%} = \frac{OD_{\,specimen} - OD_{\,negative}}{OD_{\,positive} - OD_{\,negative}} \times 100 \quad (8)$$

where $OD_{\,specimen}$ denotes the absorbance value for the sample, $OD_{\,negative}$ and $OD_{\,positive}$ represent the absorbance for negative (10 mL PBS with 0.2 mL diluted blood) and positive control (10 mL water with 0.2 mL diluted blood) respectively.

2.14 Statistical Analysis

Statistical analysis of all the experimental data was done using one-way analysis of variance (ANOVA) from Origin Lab 2023 software at a significance level of p-value <0.05. Every experiment was repeated thrice (n=3) and values are expressed as mean ± SD.

3 Results and Discussions

3.1 Characterization of BG and VBG NPs

FTIR:

FTIR spectra (2000 cm^{-1} to 400 cm^{-1}) of fructose, vitamin C, BG, and VBG NPs are shown in Fig. 1 (a). A characteristic 5-membered heterocyclic ring of fructose molecule showed vibrations at 1638 cm^{-1}, 1502 cm^{-1}, and 1409 cm^{-1}. A carbohydrate peak with maximal absorption at 1060 cm^{-1} was also seen due to the C—O and C—C group vibrations [51].

The FTIR spectrum of fructose templated **BG** showed characteristic peaks of bioglass network at 1086 cm^{-1} (broad), 968 cm^{-1}, 796 cm^{-1} and 468 cm^{-1}. In detail, peaks originating at 968 cm^{-1} were reported due to the P—O—C, P—O—Si, and P—H stretching, and the peak at 658 cm^{-1} (P—O bending) showed the existence of the phosphate group in the silicate

network. Some of the intensive peaks observed at 1086 cm^{-1} (asymmetric Si—O—Si and Si—O—C stretching), 796 cm^{-1} (Si—O—Si bending), and 468 cm^{-1} (Si—O—Si bending) confirmed bioactive glass formation [38]. Vibrations of a 5-membered heterocyclic ring of native fructose at 1638 cm^{-1}, 1502 cm^{-1}, and 1409 cm^{-1} were seen at higher wave numbers at 1641 cm^{-1}, 1559 cm^{-1}, and 1416 cm^{-1} respectively in the case of fructose templated BG. The characteristic carbohydrate peak at 1060 cm^{-1} of free fructose must have overlapped with 1086 cm^{-1}, which was assigned previously for Si—O—Si stretching vibration for BG. The existence of the characteristic peaks of fructose in BG confirms the presence of fructose as template molecules in the BG network. The reported higher wavenumber shifts of template fructose peaks in the BG sample compared to native fructose confirm the physical and chemical interactions of fructose molecules with the BG network through O−H, −C=O, and a 5-membered ring of fructose molecules. The results presented are in line with our previous findings [52].

Characteristic peaks of vitamin C (Fig.1 (a) corresponding to the lactone ring system were reported at 1752 cm^{-1} (C=O), 1660 cm^{-1} (C=C), 1322 cm^{-1} (C−O−C), 1020 cm^{-1} (C−O−H), 860 cm^{-1} (C−C ring stretching) and 752 cm^{-1} (C−O in-plane deformation) [53].

In the case of **VBG**, most of the characteristic BG peaks were reported more or less at similar wavenumbers. For example, Si−O−Si bond vibrations were seen at 1077 cm^{-1} (broad), 794 $cm^{-1,}$ and 471 cm^{-1} corresponding to P—O—C, P—O—Si respectively, P—H stretching was seen at 958 cm^{-1} and P—O bending was also reported at 603 cm^{-1}. Additional peaks reported for VBG compared to BG at 1760 cm^{-1} (C=O) and 1630 cm^{-1} (C=C) must have arrived from the loaded vitamin C molecules (Fig.1 (a). A peak at 1660 cm^{-1} due to C=O vibration present in the lactone ring in the vitamin C spectrum reportedly shifted to a lower wavenumber at 1630 cm^{-1} in VBG indicating weak physical interactions between vitamin C and BG NPs after loading. Peaks corresponding to template fructose molecules (5-membered heterocyclic ring (1638 cm^{-1}, 1502 cm^{-1}, and 1409 cm^{-1}) and carbohydrate peak (1060 cm^{-1}) were overlapped with characteristic peaks of vitamin C as well as BG in VBG spectrum. FTIR findings confirmed that VBG contained both template fructose molecules as well as loaded vitamin C.

XRD:

The XRD diffractograms of **BG** and **VBG** are displayed in Fig.1 (b). The XRD analysis on NPs showed a broad hump in the 2θ range of 10 ° to 20 °. The lack of distinct diffraction maxima confirms the amorphous nature of BG and VBG NPs that might hold a nanoscale

range of particle size. The expected nanoscale range of particle size may result in a decrease in the number of crystallographic planes contributing to diffraction with broad humps [54]. A distinct property of amorphous bioactive glass is the appearance of large diffraction maxima between 10° to 35° 2θ degrees was also reported elsewhere [38]. The inclusion of template fructose in BG and VBG may facilitate the random deposition of precursors of bioactive glass and could be a possible reason for their amorphous nature as reported by Dey *et al.* [36].

Zeta Potential measurement:

The zeta potential measurement revealed -20.4 mV and -20.6 mV negative surface charges (Fig. 1 (c) for **BG** and **VBG** NPs respectively. The negative potential may be due to the presence of phosphate and silicate ions within the glass structure as confirmed by FTIR analysis (Fig.1 (a). As per the previous findings the negative surface charge helps in cell attachment and promotes angiogenesis by facilitating the interaction of NPs with cells and tissues [55]. The slight increase in the surface charge of VBG may be due to the anionic character of the loaded vitamin. The PDI value (0.68 and 0.49 for BG and VBG NPs respectively) along with the zeta potential obtained from DLS studies, significantly revealed that the VBG NPs are less agglomerated in comparison to BG NPs. **Vitamin C loading efficiency** of VBG was calculated to be 60.4 % ± 0.82 %.

Morphology and Elemental Analysis:

FE-SEM micrographs for **BG** and **VBG** as depicted in Fig. 2 (a & b respectively) reveal an irregular and agglomerated spherical shape of the NPs showing a dense morphology. Fig. 2 (c) – (f) shows the **EDS** spectra and mapping for both NPs and confirms the presence of Si, Ca, Na, and P being their principal elements [36]. **TEM** micrographs of **BG** and **VBG NPs** are depicted in (Fig. 3 (a) and (b). The particle size distribution of NPs was evaluated using the Image-J analysis (Fig. 3 (e) and (f). NPs were spherical and the mean diameter of BG and VBG was determined to be 19.01 ± 2.80 nm and 17.57 ± 2.46 nm respectively, the findings align with the previous reports [38]. The appearance of diffused rings in the selected area electron diffraction (SAED) pattern of NPs (Fig.3 (c) and (d) confirms the amorphous nature of both the NPs and is in good agreement with XRD analysis (Fig.1 (b). The amorphous nature of NPs promises enhanced solubility of BGs, which leads to greater bioavailability and improved oral delivery [36].

3.2 Physical characteristics of the ODFs

OHF, OHVBG, and OHVBGEO were subjected to various physical characterizations to assess the corresponding quality and performance.

3.2.1 *Opacity*

ODF opacity measurement is essential to ensure the stability of vitamin C in suspended VBG NPs, consumer compliance, and its quality [56]. The addition of VBG NPs and EO in the film increased opacity (%) values (Table 2). A 14.06 % and 31.5 % increase in the opacity values of OHVBG and OHVBGEO respectively, compared to OHF (11.24 % ± 0.56 %) may be attributed to the fact that the suspended NPs might prevent light transmission and enhance the opacity of the film. This observation follows Hao Lu *et al*. findings [57], which states that the opacity of embedded NPs in the starch film was twice that of the control. The inclusion of EO in OHVBGEO resulted in a further increment in the opacity values. This is possibly due to the intensification of the light scattering by the dispersed EO droplets (refractive index 1.525) throughout the film matrix [42]. The light-sensitive compounds in the film are best preserved by the film's opaqueness [56]. Since vitamin C is susceptible to oxidative degradation when exposed to light. The increased opaqueness of OHVBGEO may prevent the degradation of this vitamin and retain its quality.

3.2.2 *Water vapor permeability*

Water vapor permeability (WVP) is directly related to the ODF quality, as higher WVP increases the evaporation of water molecules and results in film decay [58]. Table 2 shows the increased WVP value for OHVBG [$(7.15 \pm 0.209) \times 10^{-4}$ g/cm · s · kPa] and OHVBGEO [$(10.81 \pm 0.482) \times 10^{-4}$ g/cm · s · kPa] films as compared to OHF [$(4.6 \pm 0.251) \times 10^{-4}$ g/cm · s · kPa]. Although the infused clove EO in OHVBGEO is hydrophobic, the increased WVP is related to EO's impact on the film's microstructure. FE-SEM micrograph in Fig. 6 (c) reveals the formation of pores in the OHVBGEO that facilitate the passage of water vapor. According to Zhou *et al*. [43], this may be due to the lipid droplets from EO creating discontinuities in the polymer network. A similar trend was reported in [42]. Kaur *et al.* [48] reported a decrease in WVP for the zein film incorporated with cinnamon and clove EOs. The authors related this behaviour to the hydrogen bonding between the zein network and EOs that limit the bond formation with water molecules and lower the amount of water absorption.

3.2.3 *Contact angle measurements*

This parameter provides insights into wettability referred to as surface hydrophobicity/ hydrophilicity of a material that plays a crucial role in evaluating how rapidly the ODFs disintegrate [59]. Fig. 5 (c) depicts the lowest contact angle (59.08° ± 5.16°) value for OHVBGEO compared to OHF (61.1° ± 6.21°) and OHVBG (95.03° ± 11.03°). The observation deviates from the expected outcome, as OHVBGEO film was presumed to show

some hydrophobic character due to the incorporated EO. The porous nature of this functional film (OHVBGEO), as revealed by the FE-SEM micrograph (Fig. 6 (c) could be a possible explanation for this intensified hydrophilicity leading to a falloff in the contact angle value [44]. Contact angle values less than 90° indicate the hydrophilic behaviour of ODFs, in which due to molecular attractions, the drop spreads on its surface, resulting in a low contact angle. Whereas, the values greater than 90° depict ODF's hydrophobic character, as it leads to the liquid drop retraction from its surface. All the ODFs are hydrophilic due to the incorporation of a very hydrophilic polymer i.e., HA, and these findings are in good agreement with the *in-vitro* disintegration time.

3.2.4 Surface pH

The surface pH of ODFs plays a predominant role as they notably influence patient acceptability by affecting the taste, comfort, and most importantly potential irritation to the oral mucosa. Table 2 depicts the surface pH values for the obtained ODFs. The range of salivary pH lies between 5.8 -8.4, therefore the surface pH for ODFs should also have a value near neutral [21]. The pH values recorded for OHF (6.72 ± 0.2), OHVBG (7.77 ± 0.2), and OHVBGEO (7.06 ± 0.1), revealed that there is a small change in surface pH with the addition of VBG NPs and EO as compared to OHF. The inclusion of VBG NPs which are slightly basic due to the calcium, phosphate, and silicate ions. As a result, there is a 13.5 % and 4.8 % increase in the surface pH value for OHVBG and OHVBGEO respectively when compared to OHF (without VBG NPs and EO). Further clove EO being moderately acidic, the pH value for OHVBGEO shifts marginally towards acidic nature as compared to OHVBG (without EO). Nevertheless, the pH remained close to neutral (pH = 7.0) in all the cases.

3.3 In-vitro disintegration time

The rapid disintegration of ODFs helps in faster drug dissolution and absorption in the oral cavity resulting in their enhanced bioavailability. A swift onset of action might be beneficial to provide immediate relief from ulcer symptoms, thus making ODFs a fascinating dosage for a wide range of biomedical applications [60]. The disintegration time of OHF (Table 2) was observed as 12.97 ± 0.18 s. Interestingly, a 7.4 % increase in disintegration time is reported for OHVBG compared to OHF, and that for OHVBGEO drastically increased to 24%. The incorporation of NPs and EO in the OHVBGEO film matrix caused a drastic increment in the disintegration time when compared to OHF and OHVBG. An increase in the solid concentration in the film matrix along with the increment in the thickness of OHVBGEO (Table 2) also accounts for the elevation in the disintegration time as previously reported

[23]. Moreover, the films maintain their rapidly disintegrating character with a disintegration duration of <35 s even after the addition of NPs and EO.

3.4 FTIR

The FTIR spectra of OM, HA, and the ODFs are shown in Fig. 5 (a). A shoulder at 1736 cm^{-1} (C=O stretching) and an intense peak at 1638 cm^{-1} (C—OO carboxylate symmetric stretching) were reported for OM. Amide I band (C=O stretching) occurrence in OM corresponding to the 2° structure of the protein is also overlapped at 1638 cm^{-1}. Amide II band mainly due to in-plane N—H bending of OM protein was observed at 1514 cm^{-1}. C—OO carboxylate asymmetric stretching was reported at 1426 cm^{-1}. The shoulders 1322 cm^{-1} and 1372 cm^{-1} correspond to C—OH bond stretching in alcohol and C—O bond stretching in carboxylic groups in the polysaccharide content of OM [61]. A carbohydrate fingerprint area in the wavenumber region from 1200 cm^{-1} to 900 cm^{-1} represents the existence of C—C, C—O (specifically 1054 cm^{-1} and 1246 cm^{-1} respectively), and C—O—C stretching of the polymer backbone including the C—H and C—O—H bending modes are present in OM.

The HA FTIR spectrum depicts a peak at 1654 cm^{-1} which denotes C=O stretching vibration and a small peak at 1612 cm^{-1} represents the amide I (C=O stretching) band. Stretching vibrations of $–COO^-$ was seen at 1410 cm^{-1} Further, shoulders at 1552 cm^{-1} and 1328 cm^{-1} denotes amide II and amide III bands respectively, [36]. C—O—C bond vibrations in the hemiacetalic system of the saccharide unit and C—O—C stretching (O-bridge) at 1036 cm^{-1} and 1154 cm^{-1} respectively were reported for HA. The shoulders at 1328 cm^{-1} and 1374 cm^{-1} again represent the C—O bond stretching in the saccharide unit as reported in OM. A weak peak at 896 cm^{-1} denotes the asymmetric out-of-phase ring vibration in HA.

A doublet of native HA and OM around 1640 cm^{-1} (C=O stretching of carboxylic acid) and 1612 cm^{-1} (amide I) vibrations were reported at 1651 cm^{-1} and 1602 cm^{-1} in the case of OHF. The $C—OO^-$ (1426 cm^{-1}) carboxylate asymmetric stretching of OM was also reported at the wavenumber as 1412 cm^{-1}. An amide II band was noticed at 1319 cm^{-1}. The C—O—C stretching of OM (1054 cm^{-1}) and HA (1036 cm^{-1}) appeared as a strong peak at 1030 cm^{-1} in the OHF with a shoulder at 1155 cm^{-1}. These observations confirm the existence of OM and HA molecules in OHF. The observed slight variations in the peak positions of OHF compared to native OM and HA indicated weak physical interactions between these two molecules.

FTIR spectrum of OHVBG shows 1610 cm^{-1} (C=O stretching of carboxylic acid), 1600 cm^{-1} (amide I), 1401 cm^{-1} (carboxylate asymmetric stretching), and 1319 cm^{-1} (amide II), which

are most characteristic peaks of HA and OM. These peak positions are slightly shifted to lower wavenumbers compared to native HA and OM. The reported sharp peak at 1035 cm^{-1} might be overlapping peaks for C—O—C and Si—O—Si asymmetric stretching. A new peak at 410 cm^{-1} confirmed the existence of Si—O—Si bending from suspended VBG in OHVBG film. The FTIR spectrum of OHVBGEO, which contains EO as a cross-linker was more or less similar to OHVBG film. The presence of signature peaks of BG in OHVBG and OHVBGEO confirms the successful suspension of VBG NPs in the film matrix, this result corroborates with the FE-SEM and EDS study (Fig. 6).

3.5 XRD

The XRD patterns of ODFs are shown in Fig. 5 (b). A broad hump in the 2θ range of 10 ° to 20 ° was reported for OHVBG and OHVBGEO indicating the most amorphous nature of both the films. In addition, emerging broadband near 31° was reported for ODFs, which might account for the proteinaceous content in the OM [62]. The amorphous nature of BG-based formulations is highly preferable as they are thought to be safer and more biodegradable for biomedical applications. Moreover, amorphous biomaterials possess better bioavailability compared to crystalline ones [36].

3.6 Morphology and Elemental Analysis

Fig. 6 (a) represents a micrograph of OHF, which shows a smooth surface morphology indicative of the homogenous distribution of OM and HA during film fabrication and resulted in a smooth structure [63]. When additional additives were added, a more dense and continuous rough morphology was obtained for OHVBG and OHVBGEO (Fig. 6 (b and c) respectively). The observation revealed an even distribution of the NPs in both films. OHVBGEO was observed to have roughest morphology compared to OHF and OHVBG. Hydrophobic essential oils might be separated from the aqueous phase leading to heterogeneity and resulting in the roughest morphology [64] of OHVBGEO. Further, the porous nature of OHVBGEO (depicted in the micrograph Fig. 6 (c)) mimics the structure and properties of natural tissues and thus, is expected to facilitate cell adhesion and proliferation, which supports the growth of new tissues [65]. Fig. 6 (d), (e), & (f) and Fig. 6 (g), & (h) show EDS spectra and mapping respectively for OHF, OHVBG, and OHVBGEO. Interestingly, uniform distribution of Si, Ca, Na, and P elements in OHVBG and OHVBGEO, which are characteristic elements of VBG NPs (shown in Fig. 2 (d)) were reported by EDS spectra and mapping of OHVBG and OHVBGEO films and represent the successful suspension of VBG NPs in the film matrix which corroborates with the FTIR studies (Fig. 5a).

Content uniformity of VBG NPs suspended in the film

FE-SEM micrographs of OHVBGEO from five different sections (supplementary data Fig. S1) depict a uniform dispersion of VBG NPs in the film surface, evidenced by the absence of the same in the OHF micrograph. Further, the UV absorbance spectra (190 nm - 600 nm) of three different diameter films (1 cm, 3 cm, and 5 cm) [45] cut from OHVBGEO shows (supplementary data Fig. S2) that the maximum absorbance peak for VBG NPs (261 nm for the vitamin C loaded on VBG NPs and 191 nm for NPs) was recorded for a 5 cm diameter film. As inferred from Fig. S2, the absorbance value increases with successive increase in the film diameter, suggesting the increased concentration of VBG NPs with the increased diameter. This increment in absorbance with an increase in film size represents the uniform dispersion of VBG NPs across the film samples.

3.7 Thermal stability analysis

Thermal studies were done to evaluate how the films withstand the temperature changes, which provides insights into quality assurance and safety considerations.

3.7.1 TGA

The TGA thermograms (Fig. 7 (a) implies that BG and VBG NPs are thermally stable across a wide range of temperature conditions (400°C-1000°C), and ODFs show continuous thermal degradation up to 800°C. All samples show an initial weight loss of around 100°C due to the loss of bound water [36]. A distinct weight loss of 15.7 w/w % around 100°C-350°C in BG may be attributed to the thermal decomposition of the fructose template, which corroborates with the previous findings [66]. A major weight loss (37.64 w/w %) observed in the VBG thermogram is in the temperature range of 100°C-550°C may correspond to the degradation of fructose and loaded vitamin C [67]. The thermograms of both NPs reveal only 25-30 w/w % of the thermal degradation up to 1000°C. The remnant weight (%) is due to the occurrence of inorganic minerals within the bioglass network, which is in good agreement with FTIR (Fig. 1 (a) and EDS (Fig. 2 (c and d) findings of NPs. The OHF thermogram provides undeniable evidence in which a 99.71% weight loss is recorded up to 800°C due to the presence of only polysaccharide and protein contents. A drastic weight loss near 200°C of all ODFs (OHF, OHVBG, and OHVBGEO) composed of HA and OM may be due to the organic components decomposition [68]. The TGA thermograms for HA and OM are shown in Fig. S3 of the supplementary data. A marginally higher decomposition temperature range for OHVBGEO (300°C-800°C) as compared to OHVBG (280°C-800°C) is assigned to the existence of EO as a crosslinker [48].

3.7.2 DSC

DSC thermograms provide valuable insights into the thermal behaviour of a material, including phase transitions such as glass transition, melting, and crystallization that involve heat exchange associated with enthalpy changes. Fig. 7 (b) represents the DSC thermograms for samples under study. A weak endothermic value reported near 65°C-69°C for BG and VBG might be due to the template/template and loaded vitamin C respectively. The observation evidences the existence of fructose/fructose and vitamin C in BG/VBG and supports findings of FTIR (Fig. 1 (a)) and TGA (Fig. 7 (a)). The glass transition temperature (T_g) for BG and VBG NPs are reported at 298°C and 303.6°C respectively [69]. Endothermic peaks observed near 63°C to 73°C in ODFs may be attributed to the solid-solid transition or some chemical changes involving polymer chain relaxation due to the mobility. This hypothesis corroborates with [70], where the findings suggest a distinguished broad peak at 120°C in the DSC run for okra gum. It is notable that when comparing to the pure polymeric film (OHF T_m = 295.7°C), the addition of VBG NPs did not alter in major, the heat of fusion/ melting temperature (T_m) for OHVBG (292.3°C) indicating weak physical interaction between polymer and VBG NPs. Interestingly, OHVBGEO showed a notable increase in T_m at 305.9 °C. The observation confirms that the addition of EO in OHVBGEO increases crosslinking through physiochemical interactions with the polymer matrix [48]. These findings corroborate with FTIR and TGA results on ODFs (Fig. 5 (a) and 7 (a) respectively).

3.8 Mechanical properties

3.8.1 Rheological measurements

In the present study, understanding the rheological behaviour of ODFs (Fig. 8 (a, b, and c)) helps to assess their adhesive nature along with texture, and its breakdown. The steady shear flow curves in Fig. 8 (a), depict a drop in the apparent viscosity with the increased shear rate, indicating the shear-thinning behaviour of the films. The observation indicates a property of non-Newtonian fluids portraying Bingham's pseudoplastic-like behaviour of films. An increased initial viscosity of the films OHF, OHVBG, and OHVBGEO (33.4%, 72.5%, and 80.4% respectively) as compared to OM were reported at 0.1 s^{-1}. A drastic increase in the viscosity of OHVBG and OHVBGEO is due to crosslinking between the polymeric matrix and VBG in OHVBG and VBG as well as EO in OHVBGEO compared to OHF. The observation is in good agreement with FTIR (Fig. 5 (a)) and thermal analysis (Fig. 7) of ODFs. The maximum shear-thining behaviour of OM (95.34%) as compared to the films OHF (84.5%), OHVBG (86.6%), and OHVBGEO (87.53%) in the range 0.1 s^{-1} – 100 s^{-1}

shear rate is due to the weakest physical interactions between the adjoining OM and HA polymer chains in OM [71].

The oscillatory frequency sweep measurements are shown in Fig.8 (b). Storage modulus (G') is the measure of the elastic behaviour of a material under deformation. In the case of OHVBGEO, G' constantly increases with the increase in the angular frequency (ω) in the range 1-100 rad/s. This suggests that the elastic nature of the film is preserved. Whereas, for OHF and OHVBG, G' increases up to ω value of 25.1 rad/s for OHF and 50.1 rad/s for OHVBG followed by its decrement with the further increase in ω. The observation indicates a change in the behaviour of the film (OHF and OHVBG) from elastic to viscous [46]. Interestingly, Fig.8 (b) shows that a decrease in the G' occurs at a higher ω value for OHVBG (50.1 rad/s) than OHF (25.1 rad/s), possibly due to the interaction of polymeric chains with the BG network. It is noteworthy that the elastic nature of OHVBGEO may be attributed to the infusion of the BG network and EO with the protein and polysaccharide chains causing the chain to unfold and relating to the elasticity of OHVBGEO [46].

Amplitude sweep measurement of ODFs helps in determining the linear viscoelastic region (LVE). LVE helps to analyze the structural integrity of ODFs and also to understand the quick release of nutrients after a breakdown [46]. It also represents the solid-gel-like (G'>G") behaviour of the material before the LVE region followed by a liquid-like (G">G') character [71]. Fig. 8 (c), depicts the amplitude sweep measurement for OHF, OHVBG, and OHVBGEO. OHVBG showed initial higher G' values compared to the G" indicating higher elastic character, whereas OHF and OHVBGEO reveal a higher viscous character due to G" larger than G' in the strain rate range 0.0001 s^{-1} to 0.1 s^{-1}. Deformation in the film structure results after the crossover point when the flow behaviour initiates [71]. G" of OHVBGEO is 57.8% and 26.4% more than OHF and OHVBG respectively at a strain rate value of 0.0001 s^{-1}, indicating the highest viscoelastic-gel-like properties of OHVBGEO. In the case of OHVBGEO, the crossover point becomes apparent at elevated strain values, resulting in the two moduli progressing parallel as strain value increases [46]. The existence of the LVE region at higher strain values for OHVBGEO may be beneficial for the sustained release of nutrients. The gel-like character of OHVBGEO may result in excellent adhesive properties in the oral cavity.

3.8.2 Tensile testing

Tensile testing helps in understanding the durability and handling associated with the ODFs. Fig. 8 (d) represents the stress-strain curves of ODFs. It is evident from Table 3 that suspending VBG NPs enhances the mechanical properties and delays crack propagation of OHVBG and OHVBGEO as compared to OHF [72]. Tensile measurements could not be performed on OHF due to its extremely weak mechanical strength and brittle nature. Fig. 8 (d) depicts that the area under the curve for OHVBGEO is greater, indicating higher toughness than OHVBG. Elongation at break value for OHVBGEO (3.648 ± 0.1 %) was found to be greater than OHVBG (0.68 ± 0.017 %) which results in delayed break and crack propagation of OHVBGEO. The majority of reports reveal the decrement in the tensile strength and increment in elongation at break values upon EO addition, which might be due to the cross-linking with polymer chains creating a stronger network [42]. Enhanced flexibility of OHVBGEO (also shown in supplementary data Multimedia S4) may be attributed to the plasticizing effect of EO, also reported by Evanghelo *et al.* [42] where, the increase in EO concentration results in decreased strength and increased flexibility. The slope of the stress-strain curve indicates elastic or Young's modulus which denotes how the material deforms under a given amount of stress. The higher Young's modulus value of OHVBG (1271.2 ± 3.07 MPa) represents higher stiffness compared to the low value for OHVBGEO (94.03 ± 0.419 MPa) accounting for EO addition. Therefore, the findings suggest the enhanced flexibility of OHVBGEO compared to OHVBG. The findings corroborate with the rheological analysis depicted in Fig. 8 (a, b, and c).

3.9 Antimicrobial Assay

The evaluation of antibacterial activity of the functional film was conducted through a disk diffusion assay as shown in Fig. 9 (A). The observation depicts a pronounced development of a zone of inhibition in the OHVBGEO (Fig. 9 (A) (g) film measuring approximately 10.7 mm ± 0.23 mm for *S.aureus* and 8.7 mm ± 0.20 mm for *E.coli* in diameter. The findings can be attributed to the presence of the EO (clove oil) [48]. This is further validated by the evident zone of inhibition in the presence of only EO as a control (Fig. 9 (A) (f)). The expansion of film's surface area as compared to the EO drop, correspondingly led to an increase in the diameter of the inhibition zone when compared to the control. This might be due to the dispersed droplets of EO in the film. Due to its hydrophobic nature, EO disrupts bacterial phospholipid cell membranes, ultimately resulting in cell death [73].

Journal Pre-proof

3.10 Antioxidant property

The antioxidant assay (Fig. 9 (B) depicts that the OM significantly ($p < 0.05$) results in a decrease in the DPPH free radical ions. The results reveal that, with an increase in OM concentration, the free radical scavenging activity is enhanced. The antioxidant effect is directly related to the disappearance of DPPH˙ after accepting hydrogen from the antioxidant test samples [74]. The highest value of DPPH radical inhibition (91.1 ± 0.88 %) was recorded for standard compound ascorbic acid at its maximum concentration (6 mL/1 mL). Whereas, the lowest radical inhibition value is recorded (29.2 ± 0.56 %) for OM at its lowest concentration (1 mL/1 mL). The results are in good agreement with the previous report [75]. A 10% increase in the radical scavenging activity is reported for the film (OHVBGEO) containing 5 mL of OM with other additives compared to 5 mL of OM alone. This may be due to the vitamin C-loaded on BG NPs and to some extent due to the clove oil (eugenol) contents of the film [76]. Fig. 9 (B) sample image shows that the purple coloration of the DPPH radicals was turned to pale yellow due to the scavenging character of OM and ascorbic acid. The results reflect the antioxidant activity of the OHVBGEO film which might be vital for its biomedical application in preventing ulceration [77].

3.11 Hemocompatibility assay

To evaluate the biosafety, NPs, and ODFs were made to interact with blood RBCs along with appropriate control samples (Fig. 9 (C)). *In vitro,* hemolysis test is an indicator of the injuries to the RBC membrane in which the lysis of erythrocytes leads to the release of hemoglobin (Hb). The released Hb is detected spectrophotometrically at λ_{max} (545 nm) for Hb. Under ISO 10 993-5 1992 standard hemolytic rate (HR%) <5% are classified as hemcompatible. Fig. 9 (C) depicts the HR% for all the samples to lie in this range thereby, under the conditions of this study the tested materials were all safe and had no hemolytic activity. As can be seen from Fig. 9 (C), the HR% for essential oil alone possesses a high value (11.5 ± 0.16%) which may be due to the interaction of eugenol with the cell membrane that reduces some of its characteristics which is related to the previous findings [78]. Further, the OM due to its flavonoid content possesses a considerable HR% value, corroborating with the earlier reports [79]. All the procured samples possess HR% <5% hence, leading to their hemocompatibility.

Conclusion

In this study, the HA and OM blend presents its potential applications in formulating ODFs. Incorporating VBG NPs and EO in the HA and OM film matrix enhanced thermal and mechanical properties compared to the blank. The inclusion of EO further led to the creation

of free spaces as shown by the FE-SEM micrograph which results in the porous nature of the film. This porous nature results in higher WVP values and greater hydrophilic character as revealed by contact angle measurements. The functional ODF is antibacterial and antioxidant in nature. It is a complete source of nutrition for mouth ulcers which also alleviates the pain due to the numbing effect of eugenol in clove oil. The hemocompatibility assay indicates that the obtained ODFs are biosafe. The positive outcomes from all the above studies further motivate future research providing a detailed understanding of the mechanism and *in-vivo* experimentation.

Acknowledgement:

The authors would like to thank the Department of Applied Chemistry, Delhi Technological University-New Delhi for providing funds and resources to carry out the research work.

References

[1] B.B. Yerramathi, B.A. Muniraj, M. Kola, K.K. Konidala, P.K. Arthala, T.S.K. Sharma, Alginate biopolymeric structures: Versatile carriers for bioactive compounds in functional foods and nutraceutical formulations: A review, Int. J. Biol. Macromol. 253 (2023) 127067. https://doi.org/10.1016/j.ijbiomac.2023.127067.

[2] M. Alongi, M. Anese, Re-thinking functional food development through a holistic approach, J. Funct. Foods 81 (2021) 104466. https://doi.org/10.1016/j.jff.2021.104466.

[3] M.M. Essa, M. Bishir, A. Bhat, S.B. Chidambaram, B. Al-Balushi, H. Hamdan, N. Govindarajan, R.P. Freidland, M.W. Qoronfleh, Functional foods and their impact on health, J. Food Sci. Technol. 60 (2023) 820–834. https://doi.org/10.1007/s13197-021-05193-3.

[4] P. Chaiya, V. Patomchaiviwat, T. Phaechamud, Designing and characterization of Yahom buccal mucoadhesive tablets for oral ulcer treatment, Mater. Today Proc. (2023). https://doi.org/10.1016/j.matpr.2023.03.739.

[5] A.K. Gokakin, M. Atabey, A. Koyuncu, O. Topcu, Peptic Ulcer Perforation in Elderly: 10 years' Experience of a Single Institution, Int. J. Gerontol. 10 (2016) 198–201. https://doi.org/10.1016/j.ijge.2013.03.009.

[6] E.A. Bilodeau, R.V. Lalla, Recurrent oral ulceration: Etiology, classification, management, and diagnostic algorithm, Periodontol. 2000 80 (2019) 49–60. https://doi.org/10.1111/prd.12262.

[7] D.U. Owu, A.O. Obembe, C.R. Nwokocha, I.E. Edoho, E.E. Osim, Gastric Ulceration in Diabetes Mellitus: Protective Role of Vitamin C, Int. Sch. Res. Not. 2012 (2012) 362805. https://doi.org/10.5402/2012/362805.

[8] M. Chavan, H. Jain, N. Diwan, S. Khedkar, A. Shete, S. Durkar, Recurrent aphthous stomatitis: A review, J. Oral Pathol. Med. 41 (2012) 577–583. https://doi.org/10.1111/j.1600-0714.2012.01134.x.

[9] K. Xu, C. Zhou, F. Huang, N. Duan, Y. Wang, L. Zheng, X. Wang, W. Wang, Relationship between dietary factors and recurrent aphthous stomatitis in China: a cross-sectional study, J. Int. Med. Res. 49 (2021) 03000605211017724. https://doi.org/10.1177/03000605211017724.

[10] A. Sakly, B. De Wever, B. Jutla, M. Satia, J.P. Bogaert, The safety and efficacy of AphtoFix® mouth ulcer cream in the management of recurrent aphthous stomatitis, BMC Oral Health 16 (2016) 17. https://doi.org/10.1186/s12903-016-0177-0.
[11] A.A. Moghadamnia, M. Motallebnejad, M. Khanian, The efficacy of the bioadhesive patches containing licorice extract in the management of recurrent aphthous stomatitis, Phytother. Res. 23 (2009) 246–250. https://doi.org/10.1002/ptr.2601.
[12] P. Haghpanah, A.A. Moghadamnia, A. Zarghami, M. Motallebnejad, Muco-bioadhesive containing ginger officinale extract in the management of recurrent aphthous stomatitis: A randomized clinical study, Casp. J. Intern. Med. 6 (2015) 3–8.
[13] L.G. Camargo, P. de Freitas Rosa Remiro, G.S. Rezende, S. Di Carla Santos, M. Franz-Montan, Â.M. Moraes, Development of bioadhesive polysaccharide-based films for topical release of the immunomodulatory agent imiquimod on oral mucosa lesions, Eur. Polym. J. 151 (2021) 110422. https://doi.org/10.1016/j.eurpolymj.2021.110422.
[14] S. Choi, J. Jeon, Y. Bae, Y. Hwang, S.-W. Cho, Mucoadhesive Phenolic Pectin Hydrogels for Saliva Substitute and Oral Patch, Adv. Funct. Mater. 33 (2023) 2303043. https://doi.org/10.1002/adfm.202303043.
[15] W.-Z. Jiang, Y. Cai, H.-Y. Li, Chitosan-based spray-dried mucoadhesive microspheres for sustained oromucosal drug delivery, Powder Technol. 312 (2017) 124–132. https://doi.org/10.1016/j.powtec.2017.02.021.
[16] T. Molania, A. Malekzadeh Shafaroudi, M. Saeedi, M. Moosazadeh, F. Valipour, S.S. Rostamkalaei, N. Salehabadi, M. Salehi, Evaluation of cinnamaldehyde mucoadhesive patches on minor recurrent aphthous stomatitis: a randomized, double-blind, placebo-controlled clinical trial, BMC Oral Health 22 (2022) 235. https://doi.org/10.1186/s12903-022-02248-5.
[17] F. Taghizadeh, F. Mehryab, S.A. Mortazavi, S. Rabbani, A. Haeri, Thiolated chitosan hydrogel-embedded niosomes: A promising crocin delivery system toward the management of aphthous stomatitis, Carbohydr. Polym. 318 (2023) 121068. https://doi.org/10.1016/j.carbpol.2023.121068.
[18] L. Wei, S. Wu, W. Shi, A.L. Aldrich, T. Kielian, M.A. Carlson, R. Sun, X. Qin, B. Duan, Large-Scale and Rapid Preparation of Nanofibrous Meshes and Their Application for Drug-Loaded Multilayer Mucoadhesive Patch Fabrication for Mouth Ulcer Treatment, ACS Appl. Mater. Interfaces 11 (2019) 28740–28751. https://doi.org/10.1021/acsami.9b10379.
[19] P. Mahajan, A. Kaur, G. Aggarwal, S.L. Harikumar, Mucoadhesive Drug Delivery System: A Review, (2013).
[20] X. Cao, B. Wu, J. Chen, Z. Liu, Y. Yang, S. Li, H. Zhu, L. Xu, H. Huang, Hydroxypropyl Cellulose-Based Orally Dissolving Film Loaded with Insoluble Dexamethasone for Treatment of Oral Ulcers, Mol. Pharm. (2024). https://doi.org/10.1021/acs.molpharmaceut.4c00391.
[21] M.P. Tedesco, C.A. Monaco-Lourenço, R.A. Carvalho, Characterization of oral disintegrating film of peanut skin extract—Potential route for buccal delivery of phenolic compounds, Int. J. Biol. Macromol. 97 (2017) 418–425. https://doi.org/10.1016/j.ijbiomac.2017.01.044.
[22] M.M. Mehanna, A.T. Mneimneh, S. Domiati, A.N. Allam, Tadalafil-Loaded Limonene-Based Orodispersible Tablets: Formulation, in vitro Characterization and in vivo Appraisal of Gastroprotective Activity, Int. J. Nanomedicine 15 (2020) 10099–10112. https://doi.org/10.2147/IJN.S288552.
[23] R.B. Bodini, S.M.P. Pugine, M.P. de Melo, R.A. de Carvalho, Antioxidant and anti-inflammatory properties of orally disintegrating films based on starch and hydroxypropyl methylcellulose incorporated with *Cordia verbenacea* (erva baleeira)

extract, Int. J. Biol. Macromol. 159 (2020) 714–724. https://doi.org/10.1016/j.ijbiomac.2020.05.075.

[24] D.-Y. Ahn, S.-Y. Kang, J.-A. Han, Impact of calcium addition on the characteristics of hyaluronic acid-based oral films for vitamin D supplementation, Food Hydrocoll. 148 (2024) 109461. https://doi.org/10.1016/j.foodhyd.2023.109461.

[25] P. Dhar, H. Sarma, H.K. Sharma, Fixed Dose Oral Dispersible Tablet of Bitter Drug Using Okra Mucilage: Formulation and Evaluation, J. Drug Deliv. Ther. 10 (2020) 149–158. https://doi.org/10.22270/jddt.v10i5.4393.

[26] A. Garg, N. Panwar, C. Singh, Scholar, A. Garg, EVALUATION OF ANALGESIC ACTIVITY OF SYZYGIUM AROMATICUM W.S.R. TO PAINFUL TOOTH *Corresponding Author, World J. Pharm. Res. 7 (2018) 827–834. https://doi.org/10.20959/wjpr20185-11212.

[27] H. Yasin, F. Tariq, A. Sameen, N. Ahmad, M.F. Manzoor, M. Yasin, T. Tariq, M.W. Iqbal, B. Ishfaq, S. Mahmood, A. Siddeeg, Ethanolic extract of okra has a potential gastroprotective effect on acute gastric lesions in Sprague Dawley rats, Food Sci. Nutr. 8 (2020) 6691–6698. https://doi.org/10.1002/fsn3.1963.

[28] S.N. Coban, I. Polatoglu, E. Eroglu, Methyl cellulose/okra mucilage composite films, functionalized with *Hypericum perforatum* oil and gentamicin, as a potential wound dressing, Int. J. Biol. Macromol. 254 (2024) 127757. https://doi.org/10.1016/j.ijbiomac.2023.127757.

[29] D. Sharma, P. Samanta, P. Guha, S. Mishra, Synthesis and characterization of artificial saliva using Abelmoschus esculentus, J. Biol. Act. Prod. Nat. 13 (2023) 573–582. https://doi.org/10.1080/22311866.2023.2286247.

[30] N. Samiraninezhad, K. Asadi, H. Rezazadeh, A. Gholami, Using chitosan, hyaluronic acid, alginate, and gelatin-based smart biological hydrogels for drug delivery in oral mucosal lesions: A review, Int. J. Biol. Macromol. 252 (2023) 126573. https://doi.org/10.1016/j.ijbiomac.2023.126573.

[31] W. Ge, Y. Gao, L. He, Y. Zeng, J. Liu, Y. Yu, X. Xie, R. Xu, Combination therapy using multifunctional dissolvable hyaluronic acid microneedles for oral ulcers, Int. J. Biol. Macromol. 251 (2023) 126333. https://doi.org/10.1016/j.ijbiomac.2023.126333.

[32] D.-K. Kweon, J.-A. Han, Development of hyaluronic acid-based edible film for alleviating dry mouth, Food Sci. Hum. Wellness 12 (2023) 371–377. https://doi.org/10.1016/j.fshw.2022.07.039.

[33] T. Mehrabi, A.S. Mesgar, Z. Mohammadi, Bioactive Glasses: A Promising Therapeutic Ion Release Strategy for Enhancing Wound Healing, ACS Biomater. Sci. Eng. 6 (2020) 5399–5430. https://doi.org/10.1021/acsbiomaterials.0c00528.

[34] A.C.F. Harrop, K.R. Tupally, P. Pandey, H.S. Parekh, Opportunities for Bioactive Glass in Gastrointestinal Conditions: A Review of Production Methodologies, Morphology, Composition, and Performance, Mol. Pharm. (2023). https://doi.org/10.1021/acs.molpharmaceut.3c00188.

[35] A. Saleem, R. Rehman, S. Hussain, M.A. Salem, F. Ali, S.A.A. Shah, U. Younas, S.M. El-Bahy, Z.M. El-Bahy, M. Iqbal, Biodegradable and hemocompatible alginate/okra hydrogel films with promising stability and biological attributes, Int. J. Biol. Macromol. 245 (2023) 125532. https://doi.org/10.1016/j.ijbiomac.2023.125532.

[36] N. Dey, F.P. Mohny, G. Betsy Reshma, D. Rao, M. Ganguli, D. Santhiya, Bioinspired synthesis of bioactive glass nanocomposites for hyaluronic acid delivery to bone and skin, Int. J. Biol. Macromol. 253 (2023) 127262. https://doi.org/10.1016/j.ijbiomac.2023.127262.

[37] K. Bharti, P. Mittal, B. Mishra, Formulation and characterization of fast dissolving oral films containing buspirone hydrochloride nanoparticles using design of experiment, J.

Drug Deliv. Sci. Technol. 49 (2019) 420–432. https://doi.org/10.1016/j.jddst.2018.12.013.
[38] H. Goel, D. Santhiya, Role of Trigonella foenum-graecum leaf extract in tailoring the synthesis and properties of bioactive glass nanoparticles, Sustain. Mater. Technol. 33 (2022) e00485. https://doi.org/10.1016/j.susmat.2022.e00485.
[39] M. Gautam, D. Santhiya, Pectin/PEG food grade hydrogel blend for the targeted oral co-delivery of nutrients, Colloids Surf. Physicochem. Eng. Asp. 577 (2019) 637–644. https://doi.org/10.1016/j.colsurfa.2019.06.027.
[40] V. Raj, J.-J. Shim, J. Lee, Grafting modification of okra mucilage: Recent findings, applications, and future directions, Carbohydr. Polym. 246 (2020) 116653. https://doi.org/10.1016/j.carbpol.2020.116653.
[41] A.E.O. Elkhalifa, E. Alshammari, M. Adnan, J.C. Alcantara, A.M. Awadelkareem, N.E. Eltoum, K. Mehmood, B.P. Panda, S.A. Ashraf, Okra (Abelmoschus Esculentus) as a Potential Dietary Medicine with Nutraceutical Importance for Sustainable Health Applications, Molecules 26 (2021) 696. https://doi.org/10.3390/molecules26030696.
[42] J.A. do Evangelho, G. da Silva Dannenberg, B. Biduski, S.L.M. el Halal, D.H. Kringel, M.A. Gularte, A.M. Fiorentini, E. da Rosa Zavareze, Antibacterial activity, optical, mechanical, and barrier properties of corn starch films containing orange essential oil, Carbohydr. Polym. 222 (2019) 114981. https://doi.org/10.1016/j.carbpol.2019.114981.
[43] Y. Zhou, X. Wu, J. Chen, J. He, Effects of cinnamon essential oil on the physical, mechanical, structural and thermal properties of cassava starch-based edible films, Int. J. Biol. Macromol. 184 (2021) 574–583. https://doi.org/10.1016/j.ijbiomac.2021.06.067.
[44] V.A. dos S. Garcia, J.G. Borges, V.B.V. Maciel, M.R. Mazalli, J. das G. Lapa-Guimaraes, F.M. Vanin, R.A. de Carvalho, Gelatin/starch orally disintegrating films as a promising system for vitamin C delivery, Food Hydrocoll. 79 (2018) 127–135. https://doi.org/10.1016/j.foodhyd.2017.12.027.
[45] F. Sheikh, M. Naaem Aamir, M.A. Shah, Dr.L. Ali, K. Anwer, M. Zeeshan javaid, Formulation design, characterization and in vitro drug release study of orodispersible film comprising BCS class II drugs, Pak. J. Pharm. Sci. 33 (2020) 343–353. https://doi.org/10.36721/PJPS.2020.33.1.SUP.343-353.1.
[46] M. Kaur, D. Santhiya, Fabrication of soy film with *in-situ* mineralized bioactive glass as a functional food for bone health, Food Biosci. 47 (2022) 101767. https://doi.org/10.1016/j.fbio.2022.101767.
[47] R.G. Bachir, M. Benali, Antibacterial activity of the essential oils from the leaves of Eucalyptus globulus against Escherichia coli and Staphylococcus aureus, Asian Pac. J. Trop. Biomed. 2 (2012) 739–742. https://doi.org/10.1016/S2221-1691(12)60220-2.
[48] M. Kaur, D. Santhiya, UV-shielding antimicrobial zein films blended with essential oils for active food packaging, J. Appl. Polym. Sci. 138 (2021) 1–10. https://doi.org/10.1002/app.49832.
[49] K. Sridhar, A.L. Charles, In vitro antioxidant activity of Kyoho grape extracts in DPPH [rad] and ABTS [rad] assays: Estimation methods for EC 50 using advanced statistical programs, Food Chem. 275 (2019) 41–49. https://doi.org/10.1016/j.foodchem.2018.09.040.
[50] H.Y. Zhou, Y.P. Zhang, W.F. Zhang, X.G. Chen, Biocompatibility and characteristics of injectable chitosan-based thermosensitive hydrogel for drug delivery, Carbohydr. Polym. 83 (2011) 1643–1651. https://doi.org/10.1016/j.carbpol.2010.10.022.
[51] M. Zhang, H. Yang, Y. Liu, X. Sun, D. Zhang, D. Xue, First identification of primary nanoparticles in the aggregation of HMF, Nanoscale Res. Lett. 7 (2012) 38. https://doi.org/10.1186/1556-276X-7-38.

[52] D. Santhiya, F. Anjum, glass-ceramic using CT-DNA as a template, J. Mater. Chem. B 1 (2013) 6329–6338. https://doi.org/10.1039/c3tb21212b.
[53] J. Xiong, Y. Wang, Q. Xue, X. Wu, Synthesis of highly stable dispersions of nanosized copper particles using L-ascorbic acid, Green Chem 13 (2011) 900–904. https://doi.org/10.1039/C0GC00772B.
[54] Tutorial on Powder X-ray Diffraction for Characterizing Nanoscale Materials | ACS Nano, (n.d.). https://pubs.acs.org/doi/full/10.1021/acsnano.9b05157 (accessed March 31, 2024).
[55] A. Doostmohammadi, A. Monshi, R. Salehi, M.H. Fathi, Z. Golniya, Alma.U. Daniels, Bioactive glass nanoparticles with negative zeta potential, Ceram. Int. 37 (2011) 2311–2316. https://doi.org/10.1016/j.ceramint.2011.03.026.
[56] G.E. Fematt-Flores, I. Aguiló-Aguayo, B. Marcos, B.A. Camargo-Olivas, R. Sánchez-Vega, M.C. Soto-Caballero, N.A. Salas-Salazar, M.A. Flores-Córdova, M.J. Rodríguez-Roque, Milk Protein-Based Edible Films: Influence on Mechanical, Hydrodynamic, Optical and Antioxidant Properties, Coatings 12 (2022) 196. https://doi.org/10.3390/coatings12020196.
[57] H. Lu, N. Ji, M. Li, Y. Wang, L. Xiong, L. Zhou, L. Qiu, X. Bian, C. Sun, Q. Sun, Preparation of Borax Cross-Linked Starch Nanoparticles for Improvement of Mechanical Properties of Maize Starch Films, J. Agric. Food Chem. 67 (2019) 2916–2925. https://doi.org/10.1021/acs.jafc.8b06479.
[58] J.-H. Shin, J.-A. Han, Influence of Casting Variables on Release Kinetics of Orally Disintegrating Film, Foods 13 (2024) 1418. https://doi.org/10.3390/foods13091418.
[59] Y. Lee, K. Kim, M. Kim, D.H. Choi, S.H. Jeong, Orally disintegrating films focusing on formulation, manufacturing process, and characterization, J. Pharm. Investig. 47 (2017) 183–201. https://doi.org/10.1007/s40005-017-0311-2.
[60] S. Jacob, S.H.S. Boddu, R. Bhandare, S.S. Ahmad, A.B. Nair, Orodispersible Films: Current Innovations and Emerging Trends, Pharmaceutics 15 (2023) 2753. https://doi.org/10.3390/pharmaceutics15122753.
[61] M.B. Hay, S.C.B. Myneni, Structural environments of carboxyl groups in natural organic molecules from terrestrial systems. Part 1: Infrared spectroscopy, Geochim. Cosmochim. Acta 71 (2007) 3518–3532. https://doi.org/10.1016/j.gca.2007.03.038.
[62] S. Maryam, K. Barkat, I. Khalid, Y. Mehmood, M.A. Syed, N.S. Malik, M. Aslam, Polymeric blends of okra gum/gelatin prepared by aqueous polymerization technique: their characterization and toxicological evaluation, Polym. Bull. 79 (2022) 5339–5363. https://doi.org/10.1007/s00289-021-03561-9.
[63] K. Lewandowska, A. Sionkowska, S. Grabska, B. Kaczmarek, Surface and thermal properties of collagen/hyaluronic acid blends containing chitosan, Int. J. Biol. Macromol. 92 (2016) 371–376. https://doi.org/10.1016/j.ijbiomac.2016.07.055.
[64] Structural, morphological and thermal behaviour characterisations of fish gelatin film incorporated with basil and citronella essential oils as affected by surfactants, Food Hydrocoll. 41 (2014) 33–43. https://doi.org/10.1016/j.foodhyd.2014.03.015.
[65] S. Wu, X. Liu, K.W.K. Yeung, C. Liu, X. Yang, Biomimetic porous scaffolds for bone tissue engineering, Mater. Sci. Eng. R Rep. 80 (2014) 1–36. https://doi.org/10.1016/j.mser.2014.04.001.
[66] K. Samikannu, S. Jesurani, bullet Velmurugan, bullet Sivakumar, bullet Thirupathi, K. Thirunavukarasu, Synthesis and magnetic properties of conventional and microwave calcined barium hexaferrite powder, J. Mater. Sci. Mater. Electron. 23 (2012) 635–639. https://doi.org/10.1007/s10854-011-0460-4.

[67] M. Juhász, Y. Kitahara, S. Takahashi, T. Fujii, Thermal stability of vitamin C: Thermogravimetric analysis and use of total ion monitoring chromatograms, J. Pharm. Biomed. Anal. 59 (2012) 190–193. https://doi.org/10.1016/j.jpba.2011.10.011.
[68] F. Godoy-Alvarez, M. González Torres, D. Giraldo, R. Sanchez, M. Pérez Díaz, M. del Carmen, G. Figueroa-González, O. Reyes-Hernández, J. Sharifi-Rad, H. Cortés C, M. Del Prado, G. Leyva-Gómez, C. Florán, Synthesis by gamma irradiation of hyaluronic acid-polyvinyl alcohol hydrogel for biomedical applications, Cell. Mol. Biol. 67 (2021) 58–63. https://doi.org/10.14715/cmb/2021.67.1.9.
[69] A. Zandi Karimi, E. Rezabeigi, R.A.L. Drew, Crystallization behavior of combeite in 45S5 Bioglass® via controlled heat treatment, J. Non-Cryst. Solids 502 (2018) 176–183. https://doi.org/10.1016/j.jnoncrysol.2018.09.003.
[70] M. de Alvarenga Pinto Cotrim, A.C. Mottin, E. Ayres, Preparation and Characterization of Okra Mucilage (Abelmoschus esculentus) Edible Films, Macromol. Symp. 367 (2016) 90–100. https://doi.org/10.1002/masy.201600019.
[71] S. Mirarab Razi, A. Motamedzadegan, A. Shahidi, A. Rashidinejad, The effect of basil seed gum (BSG) on the rheological and physicochemical properties of heat-induced egg albumin gels, Food Hydrocoll. 82 (2018) 268–277. https://doi.org/10.1016/j.foodhyd.2018.01.013.
[72] N. Gupta, D. Santhiya, *In situ* mineralization of bioactive glass in gelatin matrix, Mater. Lett. 188 (2017) 127–129. https://doi.org/10.1016/j.matlet.2016.11.045.
[73] R. Di Pasqua, G. Betts, N. Hoskins, M. Edwards, D. Ercolini, G. Mauriello, Membrane Toxicity of Antimicrobial Compounds from Essential Oils, J. Agric. Food Chem. 55 (2007) 4863–4870. https://doi.org/10.1021/jf0636465.
[74] J.-K. Moon, T. Shibamoto, Antioxidant Assays for Plant and Food Components, J. Agric. Food Chem. 57 (2009) 1655–1666. https://doi.org/10.1021/jf803537k.
[75] X. Zhu, R. Xu, H. Wang, J. Chen, Z. Tu, Structural Properties, Bioactivities, and Applications of Polysaccharides from Okra [Abelmoschus esculentus (L.) Moench]: A Review, J. Agric. Food Chem. 68 (2020) 14091–14103. https://doi.org/10.1021/acs.jafc.0c04475.
[76] İ. Gülçin, M. Elmastaş, H.Y. Aboul-Enein, Antioxidant activity of clove oil – A powerful antioxidant source, Arab. J. Chem. 5 (2012) 489–499. https://doi.org/10.1016/j.arabjc.2010.09.016.
[77] L.M. da Silva, A. Allemand, D.A.G.B. Mendes, A.C. dos Santos, E. André, L.M. de Souza, T.R. Cipriani, N. Dartora, M.C.A. Marques, C.H. Baggio, M.F. Werner, Ethanolic extract of roots from *Arctium lappa* L. accelerates the healing of acetic acid-induced gastric ulcer in rats: Involvement of the antioxidant system, Food Chem. Toxicol. 51 (2013) 179–187. https://doi.org/10.1016/j.fct.2012.09.026.
[78] C.M. Uchôa Lopes, J.R. Saturnino de Oliveira, V.N. Holanda, A.Y.F. Rodrigues, C.S. Martins da Fonseca, F.F. Galvão Rodrigues, C.J. Camilo, V.L. de M. Lima, H.D.M. Coutinho, R. Kowalski, J.G.M. da Costa, GC-MS Analysis and Hemolytic, Antipyretic and Antidiarrheal Potential of Syzygium aromaticum (Clove) Essential Oil, Separations 7 (2020) 35. https://doi.org/10.3390/separations7020035.
[79] S. Chaudhuri, A. Banerjee, K. Basu, B. Sengupta, P.K. Sengupta, Interaction of flavonoids with red blood cell membrane lipids and proteins: Antioxidant and antihemolytic effects, Int. J. Biol. Macromol. 41 (2007) 42–48. https://doi.org/10.1016/j.ijbiomac.2006.12.003.

Declarations:

Competing Interests:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This is an original study that has never been published in any other journal in any form and will not be published in any other journal.

List of Figures

1. **Scheme: 1** Experimental methodology
2. **Fig. 1 (a)** FTIR spectra **(b)** XRD pattern and **(c)** Zeta potential measurement for BG and VBG NPs.
3. **Fig. 2 (a) & (b)** FE-SEM micrographs at 150K X magnification, **(c) & (d)** represent EDS spectra, and **(e) & (f)** depict EDS mapping for BG and VBG NPs respectively.
4. **Fig. 3 (a) & (b)** TEM micrographs, **(c) & (d)** SAED pattern and **(e) & (f)** Average particle diameter along with descriptive statistics for BG and VBG NPs respectively, using Image-J analysis software.
5. **Fig. 4 (a)** HA structure, **(b) & (c)** OM Polysaccharide and Protein (Lectin) chemical structure respectively.
6. **Fig. 5 (a)** FTIR spectra **(b)** XRD pattern **(c)** Contact angle measurement for ODFs.
7. **Fig. 6 (a), (b) & (c)** FE-SEM micrographs at 50K X magnification, **(d), (e) & (f)** represent EDS spectra, and **(g) & (h)** depict EDS mapping for ODFs.
8. **Fig. 7 (a)** TGA and **(b)** DSC thermograms for BG NPs, VBG NPs, OHF, OHVBG and OHVBGEO films.
9. **Fig. 8** Rheology data showing **(a)** viscosity versus shear rate in the range 0.1 – 100 s^{-1} **(b)** frequency sweep study at constant strain amplitude (0.1% strain) at an angular frequency range 1-100 s^{-1}, **(c)** amplitude sweep studies at a constant frequency (10 rad/s) for OHF, OHVBG, and OHVBGEO **(d)** depicts the stress-strain curve obtained from tensile testing for ODFs.
10. **Fig. 9 (A)** Disk diffusion assay for antimicrobial activity at 37°C against- lane**(1)** (left to right) *Staphylococcus aureus* **(a)** blank, **(b)** BG, **(c)** VBG, **(d)** OHF, **(e)** OHVBG, **(f)** clove oil (+ve control), and **(g)** OHVBGEO. Lane**(2)** (left to right) *Escherichia coli* **(a)** blank, **(b)** BG, **(c)** VBG, **(d)** OHF, **(e)** OHVBG, **(f)** clove oil (+ve control), and **(g)**

Journal Pre-proof

OHVBGEO, **(B)** Antioxidant assay using DPPH˙ method for OM and functional film (OHVBGEO), and **(C)** Hemocompatibility assay representing HR% <5% for synthesized NPs and ODFs.

Table 1. Nutritional analysis of the extracted Okra mucilage.

	Test Parameters	Test Result	Detection Limit	Test Method
1.	Protein g/100g	1.75	-	IS:7219:1973
2.	Magnesium mg/kg	21.45	-	STRC/STP/F129
3.	Potassium mg/100g	Not Detected	0.1	AOAC 965.30
4.	Flavonoid (as Quercetin) mg/100g	1.24	-	HPLC
5.	Carbohydrate g/100g	5.80	-	IS:1656:2007

Table 2. Thickness, Opacity, Surface pH, Contact angle, *In-vitro* disintegration time, and Water vapor permeability measurements of ODFs.

Edible films abbreviation	**Thickness** (mm)	**Opacity** (%)	**Surface pH**	**Contact angle** (°)	***In-vitro* Disintegration time** (s)	**Water vapor permeability** ($\times 10^{-4}$ g/cm ·s·kPa)
OHF	0.09±0.01	11.24±0.56	6.72±0.2	61.1±6.21	12.97±0.18	4.6±0.251
OHVBG	0.14±0.03	13.08±1.02	7.77±0.2	95.03±11.03	13.93±0.28	7.15±0.209
OHVBGEO	0.18±0.02	16.43±1.72	7.06±0.1	59.08±5.16	16.10±0.14	10.81±0.482

Table 3. Mechanical properties of ODFs were obtained using a Universal Testing Machine.

Edible films abbreviation	**Tensile strength** (MPa)	**Elongation at break** (%)	**Elastic modulus / Young's modulus** (MPa)	**Yield strength** (MPa)
OHVBG	7.15 ± 0.044	0.68 ± 0.017	1271.2 ± 3.07	7.24 ± 0.053
OHVBGEO	4.66 ± 0.03	3.648 ± 0.1	94.03 ± 0.419	4.71 ± 0.07

Graphical abstract

Highlights

- Fructose-entrapped BG NPs synthesized using a bio-inspired method and loaded with vitamin C.
- Oral disintegrating film composed of hyaluronic acid, okra mucilage, and suspended VBG NPs as a functional food for mouth ulcers.
- Incorporation of EO in the functional film results in enhanced thermal and mechanical stability and the formation of free spaces in the polymer network.
- The tensile strength decreased and elongation at break increased upon EO addition.
- Excellent antibacterial, antioxidant, and hemocompatible properties of the fabricated ODF.

Journal Pre-proof

Heat and Mass Transfer
https://doi.org/10.1007/s00231-024-03509-5

RESEARCH

PLC and SCADA based temperature control of heat exchanger system through fractional order PID controller using metaheuristic optimization techniques

Basant Tomar[1] · **Narendra Kumar**[1] · **Mini Sreejeth**[1]

Received: 14 April 2024 / Accepted: 29 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
SCADA systems play an important role in tracking the behaviour of critical process variables and connecting geographically dispersed subsystems at the industrial plant level. This article presents a PLC and SCADA-based control framework to automate and supervise the temperature control processes in the heat exchanger plant. The OMRON (NX1P2-9024DT1) PLC is interfaced with the Wonderware InTouch SCADA system to gather data, create a simulated temperature control prototype and carry out the necessary control operations within the heat exchanger plant. The PLC controls the entire process and programming of PLC is done using Sysmac studio automation software using the ladder programming language. The proposed system controls the temperature of the heat exchanger system through PID and Fractional Order PID ($P I^{\lambda}D^{\mu}$) controllers with Integral Anti-windup technique. Various control strategies like Cascade Control, Feedforward Control and Smith Predictor for time delayed process are discussed for controlling the temperature of the process. The performance of both PID and fractional order PID controllers is optimized using adaptive heuristic optimization techniques like Genetic Algorithm (GA), Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO). In control system design and analysis, the calculated performance indices are used as quantitative measures for evaluating the performance of a system. The combined form of temperature controller with Cascade control, Feedforward control and dead-time compensator is modelled and examined for simulation using MATLAB. Simulation and real-time experimentation analysis of the developed controllers are executed with metaheuristic optimization techniques based on different performance indices like ISE, IAE and ITAE.

Keywords Automation · Industry 4.0 · Heat exchangers · Temperature control · Fractional Order PID controllers · Programmable logic controller · Supervisory control and data acquisition · Metaheuristic optimization techniques

1 Introduction

The heat transfer processes account for more than 80% of global energy use today. The primary goal of designing control techniques for industrial processes is to increase energy efficiency [1, 2]. Heat exchangers are used for energy exchange between hot and cold fluids through a solid wall to control the temperature levels of unit operations for a smooth production process and energy-efficient operations [3]. Heat exchanger systems are mostly used in industrial process applications including chemical production processes [4], gas turbines, oil refineries [5], boilers, air conditioning [6], power generation, refrigeration, and petroleum industries [7], etc. Heat exchangers can withstand tremendous pressure by maintaining a specific temperature condition for controlling the outlet fluid temperature with respect to the variations in the operating conditions [8].

High-precision temperature control technology is one of the most critical factors in industrial thermal processing systems. Temperature control is a process of automatically controlling the temperature of large systems to achieve an ideal value in the industrial processes. Temperature stability is required for many biological, chemical production and industrial manufacturing processes [9, 10]. Temperature

✉ Basant Tomar
basanttomar18@gmail.com

Narendra Kumar
narendrakumar@dtu.ac.in

Mini Sreejeth
minisreejeth@dce.ac.in

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

Published online: 06 August 2024

control is a common time delay issue that negatively impacts the performance and stability of controlled systems. Temperature overshooting or undershooting is difficult when there are rapid temperature changes with high precision and stability [11]. Therefore, it is important to design a control scheme to regulate the outlet temperature of a heat exchanger to a certain reference value and enhance the heat transmission efficiency.

Different industries use different forms of heat exchangers on the basis of construction type, transfer process, flow arrangement and phase of liquid as shown in Fig. 1 [12, 13]. Shell and tube heat exchanger (STHEX) is the most prevalent type of heat exchanger in the process industries, mainly due to its flexible and robust design, and capability to operate at high pressure and temperature (cryogenic to high-temperature applications). STHEXs have a larger heat transfer surface-to-volume ratio, and ease of manufacturing in a variety of configurations than double-pipe heat exchangers. STHEX consists of parallel tubes/pipes enclosed in a cylindrical shell in which there is enough space between the tubes and the shell for the fluids to flow through. STHEX allows for easy disassembly for routine maintenance and cleaning and can handle very viscous or corrosive process streams with impurities [14, 15]. STHEX are utilized in a wide range of applications like cooling of water, oil heaters, petrochemical industries, boilers and condensers in power plants [16, 17].

It is very difficult to control a heat exchanger system because of its complex and nonlinear nature due to unknown fluid characteristics i.e. fluid flow patterns and rates, non-causalities, temperature-dependent and time-varying qualities, contact resistance, leakages, ambiguous mechanics and friction [18, 19]. Temperature control is challenging due to the complex thermal interactions in the traditional heater-controlled methods. The conventional control methods are not able to cope up with the difficulties like strong non-linearity, disturbance, uncertainties, large response and delay time of the controlled objects which further results in uncontrollable feedback efficiency for precise control [20, 21]. In the presence of external interference, the current conventional control approaches exhibit high-temperature fluctuation, longer stability correction times, and poor temperature stability. Therefore, an effective control strategy and hardware design are required to maintain a smaller temperature fluctuation amplitude around the desired temperature value for a fluid and heat transfer system.

1.1 Related work

Temperature control algorithms have been widely studied in the process industry [22, 23]. Numerous mathematical modeling approaches and control algorithms have been employed for studying the dynamic behaviour, response characteristics and controlling the temperature of heat exchangers. Proportional Integral Derivative (PID) controllers are most commonly utilized for temperature control in many industrial applications due to their simplicity and applicability [24, 25]. However, it is very difficult to identify the optimal PID parameters in the controller design to obtain the optimal system performance [26, 27]. The evident overshoot in the control cannot be eliminated by the conventional PID control. They can't resolve the time-varying control model parameters because of large lag and long adjustment time.

Advanced control methods such as model predictive control [28], Fractional order PID [29, 30], disturbance-observer-based control [31, 32], fuzzy control [33] and neural network control [34, 35] are used in case of difficulties like large inertia, time delay, and multiple disturbances in the temperature control process. Al-Dhaifallah et al. [36] proposed a fractional-order fuzzy PID controller for the temperature control of a tubular heat exchanger under load variation within the reference environment. The parameters of the fractional-order controller are changed via a rule-based fuzzy logic control method to have improved control performance, resilience and flexibility than conventional PID control schemes. Thepmanee et al. [37] introduced an Allen Bradley PLC using WirelessHART for temperature control and device diagnostic by implementation of SCADA system

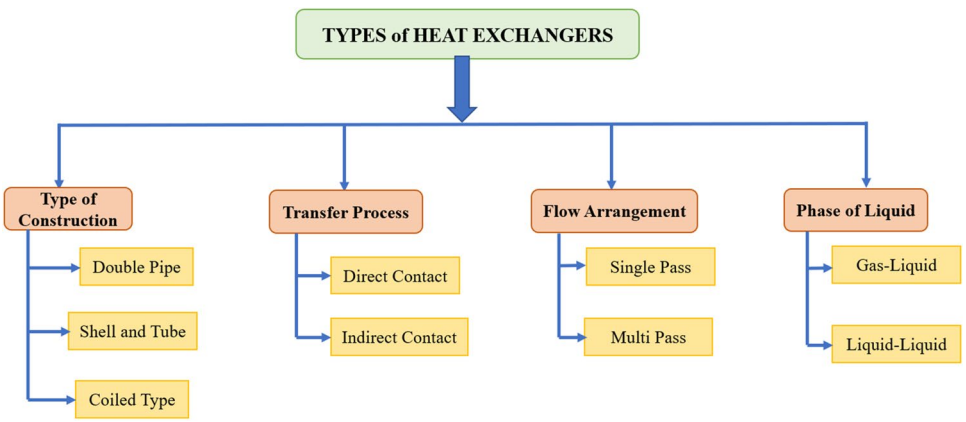


Fig. 1 Types of heat exchangers

along with Prosoft configuration builder and Studio 5000 software.

Lu et al. [38] proposed composition-temperature cascade control strategies for controlling the vapor recompression (VRC) assisted dividing wall column (DWC) with side condenser and reboiler process configurations. These control structures provide high energy efficiency and effective controllability for the industrialization of side heat exchanger processes. Bobič et al. [39] proposed a 1D discretized lumped-parameter system based on the finite volume method to understand the dynamic behaviour such as inlet and outlet temperature disturbances and various fluid flow configurations of a counterflow plate heat exchanger. The predicted transient responses due to temperature disturbances are experimentally verified by using infrared thermography technology tests. Miao et al. [40] introduced a cascade feedforward control technique based on incremental PID and predictive functional control strategy to resolve the delay and anti-disturbance ability problem in the outlet temperature control system of a heat exchanger.

Damasceno et al. [41] presented the modeling and simulation of a Proportional Integral (PI) controller for a STHEX where the controller parameters are tuned and optimized using different techniques like bat algorithms, particle swarm optimization, flower pollination algorithm and cuckoo search algorithm to get the best transient response. Instead of conventional feed-forward control methods, Li et al. [42] introduced an effective temperature-based predictive feedback control algorithm to improve the indoor temperature fluctuation, control flexibility, and anti-interference quality of large-scale district heating substations. Xu et al. [43] proposed an Adam optimization-based recurrent neural network controller with an integral proportional derivative controller for the temperature control of a single-phase hotplate system. The recurrent neural network controller was driven by reference model and feed forward controller to provide better learning reference. Jamil et al. [44] discussed various fractional-order PID controllers for temperature control in different types of heating systems in industries. Hussein et al. [45] introduced an adaptive PID control technique for the temperature control of electric furnaces in industrial applications which used the balloon effect and whale optimization algorithm for tuning the parameters of controller.

Maya-Rodriguez et al. [46] proposed a metaheuristic optimization-based neuro-fuzzy controller for the temperature control of a transesterification biodiesel reactor. This method increased the efficiency of a chemical reactor and minimize the error function and disturbances. Hou et al. [47] introduced a linear active disturbance rejection controller parameter optimization based on a simultaneous heat transfer search algorithm for the cascade control of the main steam temperature control system. The conduction, convection and radiation phases of heat transfer are implemented in parallel during the optimization process to provide significant improvement toward negative effects caused by uncertainties and disturbances. Ouyang et al. [48] introduced a PID approach for controlling the temperature of a continuous stirred tank reactor process which used a sparrow search algorithm for optimizing the PID control parameters. Along with that, it used Gauss Cauchy mutation technique for each iteration in order to enhance the local optimization capability. Zeng et al. [49] proposed a control design for analyzing the outlet temperature performance of temperature control units over the frequency domain to achieve optimal attenuation, high precision and environmental adaptability.

1.2 Aims and objectives

In order to ensure the dynamic control stability of the controller for temperature systems with strong non-linearity, uncertainty, large response and delay time characteristics, it is essential to train the proper hyperparameters and improve the learning and model-free ability of the controller. This paper proposes a novel cascade control strategy using PID and fractional order PID controller ($P I^{\lambda}D^{\mu}$) along with the Integral Anti-windup technique, feedforward control and Smith predictor to control and regulate the outlet temperature of a heat exchanger to a certain reference value and enhance the heat transmission efficiency. The main contributions of this paper are as follows:

- For developing a real-time framework for the temperature control of a heat exchanger plant, this article proposes a PLC and SCADA-based control framework to automate and supervise the temperature control processes in a heat exchanger plant and monitor all the processes using SCADA.
- The OMRON (NX1P2-9024DT1) PLC is interfaced with the Wonderware InTouch SCADA system to gather data and to create a simulated prototype of the temperature control system to monitor the entire process on a real-time basis in the SCADA screen and historize the data to its database. To the best of our knowledge, no other articles have taken into account all of these constraints at the same time.
- The primary objective is to maintain the outflow temperature of the fluid constant by manipulating the steam flow. These are two principle perturbances or disturbances i.e. fluid flow rate and fluid inflow temperature that need to be controlled with feedforward control.
- The proposed system will take the temperature feedback from the sensors installed in the field and automatically control the temperature by adjusting the hot and cold control valves using the logic written in the PLC.
- The SCADA server receives periodic status from sensors, buttons, process state variables, and other PLC-interfaced

devices and stores those details in a database. PLC carries out the necessary control operations within the heat exchanger plant as directed by the SCADA server.

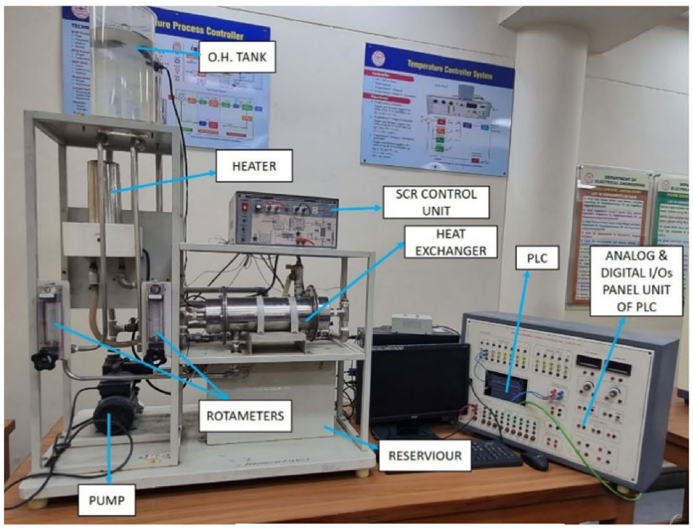
- The designed controller regulates the outlet fluid temperature of a heat exchanger to a certain reference value and enhances the heat transmission efficiency irrespective of the external disturbances, non-linearity and delay characteristics.
- The parameters of the fractional order PID (FPID) controller are tuned by Genetic Algorithm (GA), Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) techniques along with their comparative study to achieve an optimal and desired performance with the fastest steady-state regime and less percentage overshoot.
- The fine-tuning of the controller parameters using GA, ACO and PSO improves the overall performance and generalization capacity of the entire model. These algorithms work with a fitness function for modifying and analyzing a set of potential solutions for evaluating the effectiveness of the model.
- The controller type may have an impact on the stability of the system. Therefore, Integral square error (ISE), Integral absolute error (IAE) and Integral time absolute error (ITAE) cost functions are designed to select optimum parameters of the controller to improve the transient and steady-state response of the controller.
- Most of the studies on PID and Fractional order PID controllers in the literature are primarily supported by simulation results only. On the other hand, this study uses both real-time experimental and simulation results to validate the effective performance of the proposed technique, while guaranteeing both transient and steady-state response of the controller and rejection of external disturbances in comparison to the existing literature.

The remaining article is organized as follows. Section 2 briefs about the experimental architecture of the proposed temperature control system along with the mathematical modeling and SCADA prototype for the temperature control system. Section 3 discusses the developed PID and P $I^{\lambda}D^{\mu}$ cascade control, feedforward control and Smith predictor for time-delayed process controller design. Section 4 discusses the applied GA, PSO and ACO techniques for optimizing the parameters of both PID and P $I^{\lambda}D^{\mu}$ controllers. Finally, the performance analysis of the developed controllers is demonstrated with simulation and real-time experimental results in Sect. 5 followed by the conclusion in Sect. 6.

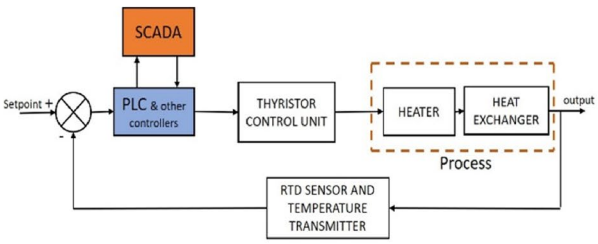
2 System description

The temperature process control setup consists of a combination of heat exchanger and boiler as the main process. Temperature sensors, signal conditioning units, control values and other associated components are also incorporated into the system. The hardware setup is shown in Fig. 2(a) and the block diagram is shown in Fig. 2(b).

The rotameter is one of the types of flow sensors used in process industries. In a rotameter, the obstruction is a float of various shapes (tapered in this process) that rises in a vertical tapered column. The rotameter housed here is mainly used for introducing disturbance and measuring the inlet flow rate. Resistance temperature detectors (RTD) are used as the temperature sensors in the process. RTD exhibits



(a) Hardware Setup



(b) Block diagram of the proposed system

Fig. 2 Experimental architecture of the proposed temperature control system

a change in resistance with temperature. The electric power dissipation must be regulated to avoid errors due to the heating of sensing elements. Typically, 10mW of energy dissipation will raise the temperature to 0.3 °C, which corresponds to low values of current (< 4 mA) and voltage (< 1 V) in industries. For signal conditioning, the temperature is sensed by RTD and it gets converted into (4–20) mA corresponding to the temperature range of (0–100 °C) in the temperature control setup. This current signal is used by the main controller i.e. PLC, which fires the final control element, the SCR control unit.

2.1 Mathematical modelling

The process has to be mathematically modelled for designing the controllers. Almost all industrial processes have non-linear characteristics along with some lag. Generally, these processes are approximated as first-order plus time delay (FOPTD) and second-order plus time delay (SOPTD) models.

Generally, FOPTD can be expressed as

$$G_f(s) = \frac{K_f e^{-t_d s}}{\tau s + 1} \tag{1}$$

SOPTD can be expressed as follows in the general form

$$G_{ff}(s) = \frac{K_{ff} e^{-t_d s}}{(\tau_1 s + 1)(\tau_2 s + 1)} \tag{2}$$

where, K_f is process gain and τ is the time constant of FOPTD, K_{ff} is the process gain and τ_1 and τ_2 are time constants of SOPTD, t_d is the time delay.

Now, the heat exchange transfer function along with process time delay can be expressed as:

$$G_P = \frac{A_P e^{-t_d s}}{\tau_P s + 1} \tag{3}$$

where A_P is process gain, t_d is the process time delay and τ_P is the process time constant.

Disturbance transfer function can be expressed as:

$$G_d = \frac{A_d}{\tau_D s + 1} \tag{4}$$

where A_d is disturbance gain and τ_D is disturbance time constant.

Considering A_d as unity, the disturbance transfer function can be expressed as:

$$G_d = \frac{1}{\tau_D s + 1} \tag{5}$$

Valve and sensing element transfer functions can be expressed as:

$$G_v = \frac{A_v}{\tau_v s + 1},\ G_H = \frac{A_H}{\tau_H s + 1} \tag{6}$$

where A_v is valve constant, τ_v is a valve time constant, A_H is sensing element gain constant and τ_H is sensor time constant.

For reference purpose, the values of the constants which have been considered here are shown in Table 1.

So, the overall transfer function of the whole process is

$$G_{ff}(s) = \frac{5e^{-1s}}{90s^2 + 33s + 1} \tag{7}$$

2.2 Development of SCADA for temperature control system

The OMRON (NX1P2-9024DT1) PLC is interfaced with the Wonderware InTouch SCADA system to gather data and create a simulated prototype of the temperature control system for monitoring the entire process on a real-time basis in the SCADA screen and historize the data to its database. The ladder logic of PLCs is designed using Sysmac studio automation software and Wonderware Intouch SCADA software is used for visualizing the plant independently. A user can enter the temperature set point through SCADA. SCADA screen displays the temperature at various points, the direction of motion, operator caution alarms, auto mode failure, high-temperature alarms, safety alarms, safety shutdown alarms and trends, etc. PLC uses sensors, buttons, process state variables, and other devices to control automated processes and monitor the status of inputs. The SCADA server receives periodic status updates from the PLC-interfaced devices and stores those details in a database under the trends section. PLC carries out the necessary control operations within the heat exchanger plant as directed by the SCADA server. By implementing the proposed system, the operator can easily start the system by setting the

Table 1 Description of the parameters used in mathematical modeling

DESCRIPTION	CONSTANTS	VALUES (units)
Heat exchange transfer function along with process time delay	G_P	50
Process time constant	τ_P	30 s
Disturbance gain	A_d	1
Disturbance time constant	τ_D	10 s
Valve constant	A_v	0.13
Valve time constant	τ_v	3 s
Sensing element gain constant	A_H	0.16
Sensor time constant	τ_H	10 s
Process time delay	t_d	1 s

temperature set point from SCADA and clicking the start button. The system will take the temperature feedback from the sensors installed in the field and automatically control the temperature by adjusting the hot and cold water valves using the logic written in the PLC.

Figure 3 shows the prototype of the developed SCADA system for temperature control of heat exchanger system. The water from the reservoir is pumped to the overhead tank (O.H. Tank), as shown in Fig. 3(a). The O.H. Tank is equipped with a safety sensor for preventing overflowing from the tank. A level transmitter (LT) is installed for level measurement in the O.H. Tank unit. The water flow from the O.H. Tank to the heater is measured by rotameter (ROT_1) and it is controlled with a control valve (CV_1). Initially, the SCR unit is triggered by starting the controller, and then the internal switch to the heater is turned ON. The increasing temperature can be measured by the temperature-sensing element RTD. The change in resistance from the RTD is fed into the Wheatstone's bridge circuit and converted into the appropriate voltage signal (0–10) Volt and its corresponding current signal (4–20)mA using the temperature transmitter. This current signal is coupled with controllers. This control action is used to drive the SCR unit which acts as a final control element unit and is triggered according to the control voltage. The directions of the flow in the plant are shown with the arrows in Fig. 3(b).

3 Controllers design

In this section, various control strategies like cascade control, feedforward control and Smith predictor for time-delayed process are discussed for controlling the temperature of the process.

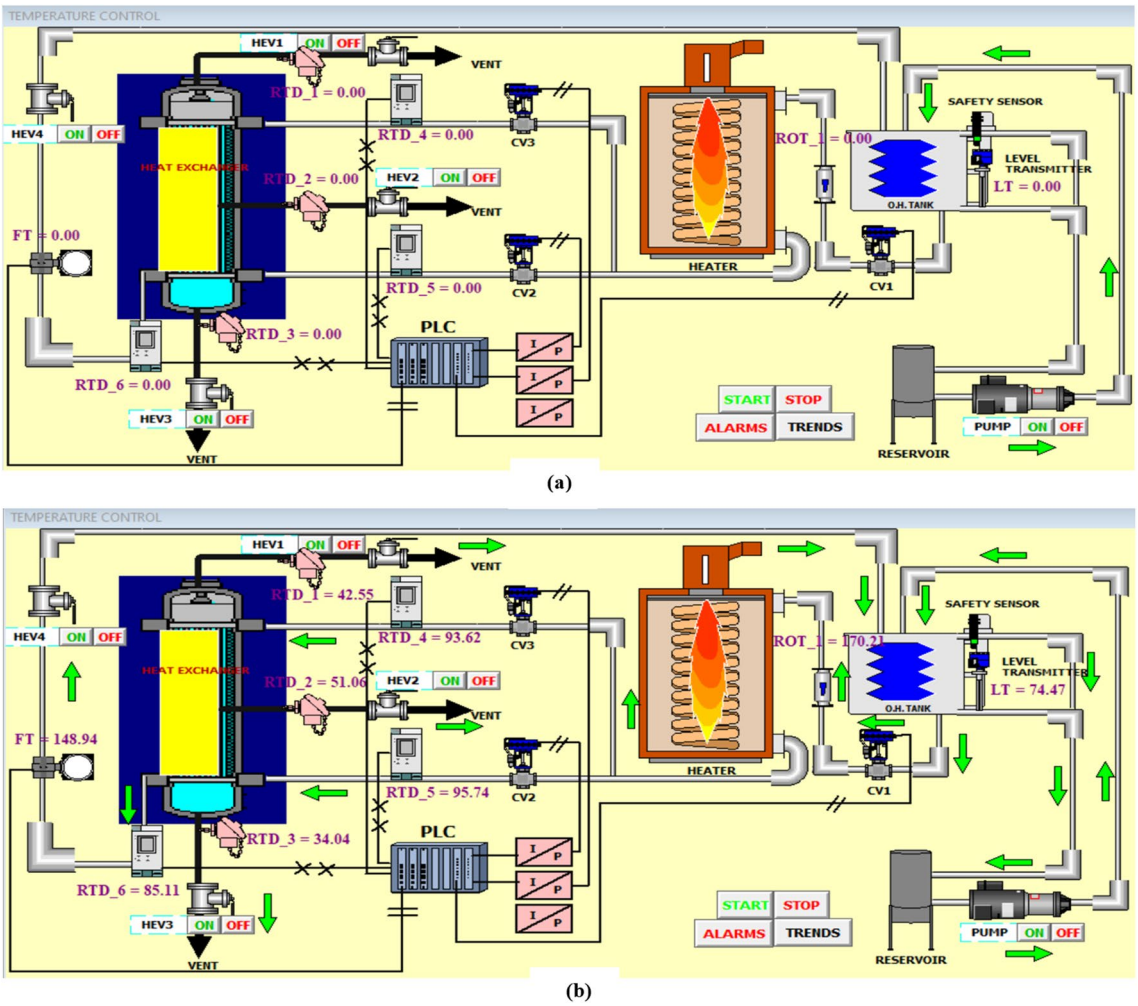


Fig. 3 Prototype of the developed SCADA system for temperature control of heat exchanger system

3.1 Cascade control strategy using PID and FPID controller along with integral anti-windup technique and velocity setpoint

PID and Fractional Order Proportional Integral Derivative (P $I^{\lambda}D^{\mu}$) controller with Integral Anti-windup technique is proposed here. In cascade control configuration, there is more than one measurement for a single manipulated variable. In this configuration, there's a primary or master loop that provides the reference setpoint and the other one is the secondary or slave loop which uses the output of the master loop as its setpoint. The dynamics of the slave loop are much faster than that of the master loop. So, the phase lag of the slave loop is much smaller than that of the master loop. Also, the crossover frequency of the slave loop is much higher than that of the master loop which allows it to accommodate higher gains in the slave loop for regulating the disturbances in the secondary loop without affecting the overall stability of the system. Generally, a proportional control is used in slave loops but Integral control action is also introduced to eliminate the offset error. So, the PI controller is used in the secondary loop. Since flow disturbance is a quick process, therefore, there is no need to increase the proportional gain too much. It will cause too many oscillations. Temperature control process is generally a very slow process as compared to other control processes like flow and pressure. Usually, PID controller action is preferred for the temperature control process. It accommodates high gains for faster response without affecting the stability of the process. The block diagram of the cascade control strategy using PID and FPID controller is shown in Fig. 4.

The PID compensator in the time domain, when λ, μ of the P $I^{\lambda}D^{\mu}$ controller is considered as a unity, is shown in Fig. 4 and the expression is given as

$$K_c(t) = K_p\left\{x_r(t) - r(t)\right\} + K_d\left[V_{sp}\left\{\frac{d}{dt}x_r(t)\right\} - \frac{d}{dt}r(t)\right] + K_i\int\left(x_r(t) - r(t)\right)dt \tag{8}$$

where K_p, K_i, K_d are proportional, integral and derivative constants respectively. V_{sp} is the velocity setpoint and it is considered between 0 and 1.

When $\lambda and \mu$ are considered as non-integer orders, then it is said to be FPID controller. FPID features two additional fractional parameters, integral order (λ) and derivative order (μ) to boost the system's flexibility and enable a more effective implementation than a standard PID controller. The generalized equation of P $I^{\lambda}D^{\mu}$ controller is

$$K_c(t) = K_p\left\{x_r(t) - r(t)\right\} + K_dD^{-\lambda}\left[V_{sp}\left\{\frac{d}{dt}x_r(t)\right\} - \frac{d}{dt}r(t)\right] + K_iD^{\mu}\int\left(x_r(t) - r(t)\right)dt \tag{9}$$

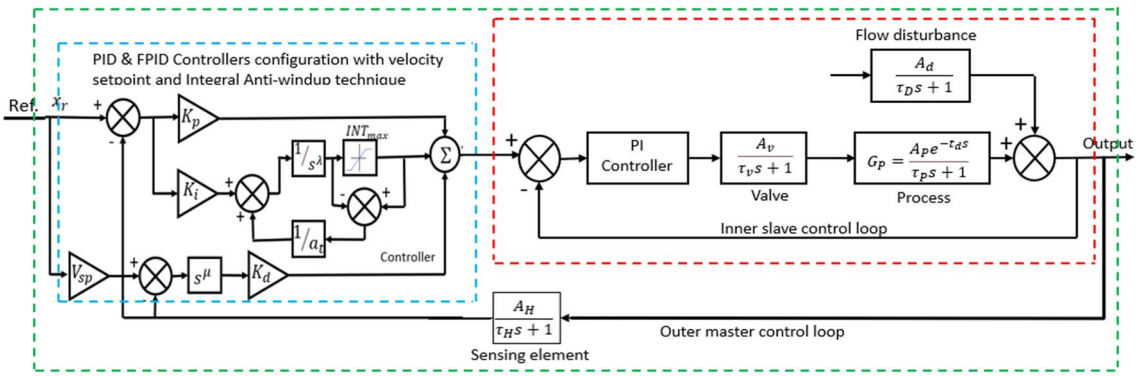


Fig. 4 Block diagram of cascade control strategy using PID and FPID controller

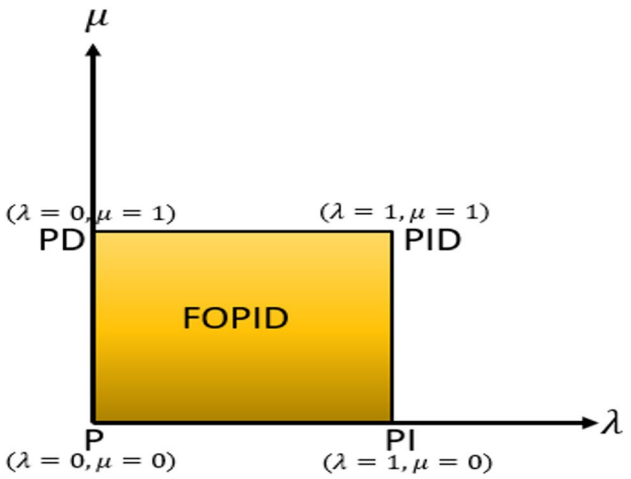


Fig. 5 Point-to-point representation of active regions of P $I^{\lambda}D^{\mu}$ and PID controllers

where both $\lambda and \mu$ are as: $0 < \lambda < 1$ and $0 < \mu < 1$. D is the fractional operator and is defined as

$$D^{-n}f(t) = \frac{1}{\Gamma(n)}\int_0^t f(x)(t-x)^{n-1}dx \tag{10}$$

where, n is non-integer order and Γ is the gamma function. The P $I^{\lambda}D^{\mu}$ controller is the generalization of the PID controller. The active region of P $I^{\lambda}D^{\mu}$ and PID controllers is shown in Fig. 5.

Figure 4 also displays the design of controllers with Integral ANTI-WINDUP. This design incorporates an additional loop around the Integrator, in contrast to the conventional PID controller. By lowering the integrator input, this extra loop inhibits the integrator from reaching saturation and intensifying excessively. At parameter INT_{max}, the Integrator is intended to reach saturation. The integral control action is defined as

$$m_i = \begin{Bmatrix} -\mathrm{INTG}_{max} & n_i < -\mathrm{INTG}_{max} \\ n_i & n_i < -\mathrm{INTG}_{max} < 0 \text{ and } -\mathrm{INTG}_{max} - n_i < 0 \\ \mathrm{INTG}_{max} & \mathrm{INTG}_{max} < n_i \end{Bmatrix} \tag{11}$$

where $INTG_{max}$ is the upper limit of integrator, a_t is the time constant of the anti-windup loop which gets activated only in case of Integrator saturation. The reset time of Integrator depends on a_t. The Integrator is in normal control action when m_i=n_i. But the anti-windup controller makes the saturation error zero in case of Integration saturation and then Integrator output becomes saturation limits, i.e. $INTG_{max}$ and $-INTG_{max}$.

3.2 Feedforward control for slave loop

Feedback slave control loops are not able to achieve perfect control of a chemical industrial process. They react only after detecting any deviation in the output value from the desired value. Unlike feedback control loops, feedforward control loops directly measure the perturbances or disturbances and then anticipate the effect that they will have on the main process output. The control action starts when any change is detected in the disturbance. It changes the manipulated variable precisely to eliminate the disturbance impact on the process-controlled output. The main objective is to maintain the outflow temperature of the fluid constant by manipulating the steam flow. Fluid flow rate and fluid inflow temperature are the two principle perturbances or disturbances that need to be controlled with feedforward control. The block diagram of

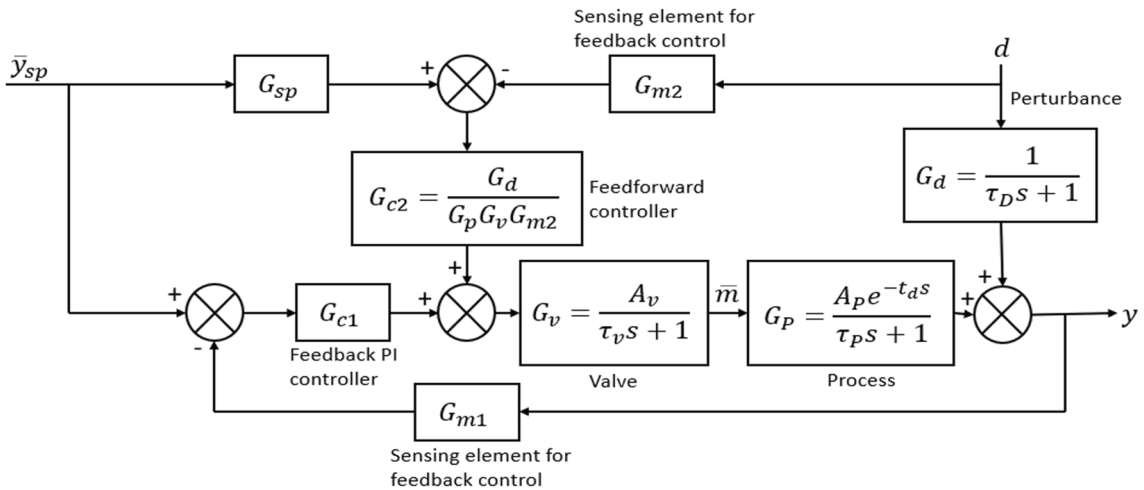


Fig. 6 Block diagram of the feedforward control with feedback control loop

the feedforward control with feedback control loop is shown in Fig. 6. From the block diagram, we have

$$y = G_P\overline{m} + G_d d \tag{12}$$

The manipulated variable is given as

$$\overline{m} = G_v G_{c1}\left(\overline{y}_{sp} - G_{m1}y\right) + G_v G_{c2}\left(\overline{y}_{sp} - G_{m2}d\right) \tag{13}$$

From Eq. (13) and (14) we get

$$y = \frac{G_P G_v(G_{c1} + G_{c2}G_{sp})}{1 + G_P G_v G_{c1} G_{m1}}\overline{y}_{sp} + \frac{G_d - G_p G_v G_{m2} G_{c2}}{1 + G_P G_v G_{c1} G_{m1}}d \tag{14}$$

The transfer functions of the feedforward loop are

$$G_{c2} = \frac{G_d}{G_p G_v G_{m2}} \; and \; G_{sp} = \frac{G_{m2}}{G_d} \tag{15}$$

If $G_p, G_d, G_v and G_{m2}$ are known to be true, then this feedforward loop will compensate for the whole disturbance. While any of the $G_p, G_d, G_v and G_{m2}$ are known by approximation, then

$$G_p G_{c2} G_v G_{sp} \neq 1 and G_d - G_v G_{c2} G_p G_{m2} \neq 0 \tag{16}$$

As a result, the feedforward loop will not compensate for the whole disturbance perfectly but the feedback loop offers the required compensation.

3.3 Smith predictor for time delay process

In the earlier discussions, it was assumed that whenever there is any change in the perturbance and manipulated variables, the effects are observed instantly in the controlled output variable. But in real physical experience, there is always a time interval (duration depends on the complexity of the process) during which no effect is observed on the output-controlled variable. This is termed as dead time or transportation lag. Smith predictor or dead-time compensator is defined as a modification in the classical feedback control system to incorporate the transportation lag as shown in Fig. 7.

$$G_P = \frac{A_P e^{-t_d s}}{\tau_P s + 1} \; or \; G_P = G_s e^{-t_d s} \tag{17}$$

Now, the open loop response for the Fig. 7 will be equal to

$$\overline{y} = \overline{y}_{sp}\left[G_s e^{-t_d s}\right] G_c \tag{18}$$

For undesired effects elimination, it is required to have a feedback signal that does not carry any transportation lag information, such as

$$\overline{y}^* = \overline{y}_{sp}\left[G_s\right] G_c \tag{19}$$

This is only possible if a certain quantity is added in $\overline{y}^*$, such that

$$\overline{y}^{**} = \overline{y}_{sp}\left[G_s\right] G_c\left[1 - e^{-t_d s}\right] \tag{20}$$

which implies

$$\overline{y} + \overline{y}^{**} = \overline{y}^* \tag{21}$$

Usually, in process control, the process and dead time are obtained by approximation. So, let us assume that G_s and t_d are true process characteristics and $\widetilde{G_s} = \frac{A_P}{\tilde{\tau}_P s + 1}$ and $\tilde{t}_d$ are their respective approximation. In this case, the improved Smith predictor is

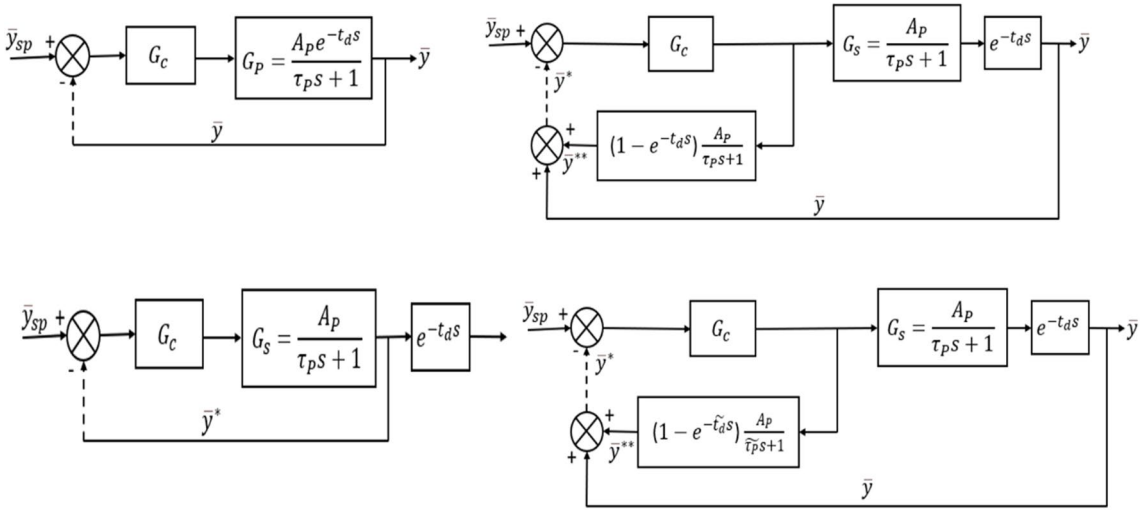


Fig. 7 Block diagram of Smith predictor or dead-time compensator for the time delay process

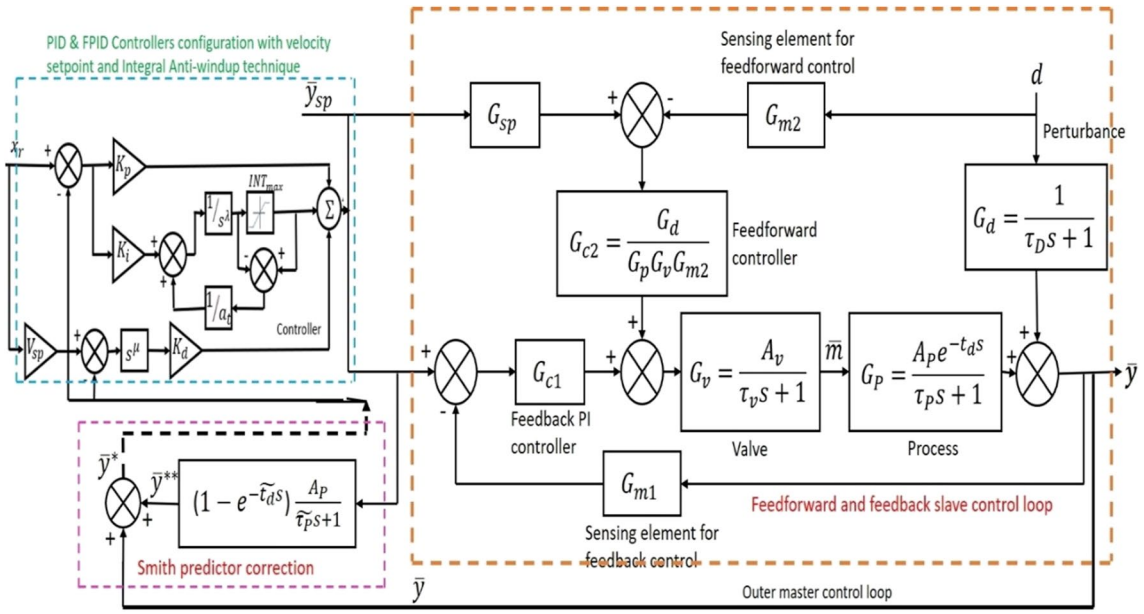


Fig. 8 Temperature controller with cascade control, feedforward control and dead-time compensator

$$\bar{y}^{*} = \bar{y}_{sp}\left[\widetilde{G_s} + \left(G_s e^{-t_d s} - \widetilde{G_s} e^{-\widetilde{t_d} s}\right)\right] G_c \tag{22}$$

Equation (22) implies that $G_c = \widetilde{G_s}$ and $t_d = \widetilde{t_d}$ for the perfect dead-time compensation. Greater the modeling error [$G_c - \widetilde{G_s}$ and $t_d - \widetilde{t_d}$], lower the effectiveness in the compensation. But, the term $(t_d - \widetilde{t_d})$ is more crucial because of the exponential factor.

The combined form of the temperature controller with cascade control, feedforward control and dead-time compensator is shown in Fig. 8.

4 Metaheuristic optimization techniques

The performance of both PID and P $I^{\lambda}D^{\mu}$ controllers are optimized using adaptive heuristic optimization techniques. Genetic algorithm (GA), Ant Colony Optimization (ACO) and particle swarm optimization (PSO) are used for optimizing the parameters of both controllers. In control system design and analysis, the calculated performance indices are utilized as a quantitative matrix for evaluating the performance of a system. The parameters of a control system are adjusted such that the design index reaches its lowest value while respecting the constraints of the controlled system. Integral square error (ISE), Integral absolute error (IAE) and Integral time absolute error (ITAE) are the commonly used indices. These indices are defined as

$$ISE = \int_0^t e^2(t)dt \tag{23}$$

$$IAE = \int_0^t |e(t)|dt \tag{24}$$

$$ITAE = \int_0^t t|e(t)|dt \tag{25}$$

The selection of these criteria depends on the type of characteristics to be controlled. For suppressing large errors, ISE is preferred and IAE is preferred for suppressing small errors. To suppress the errors which persist for a long time, ITAE is preferred. The flow charts of GA, ACO and PSO algorithms are shown in Fig. 9. These algorithms are a little time-consuming but we get better accuracy and precision after a series of experiments and better parameter selection. The values of the set parameters of GA, ACO and PSO are shown in Table 2. The mean computational time was 198.45 s for GA, 176.23 s for PSO and 173.67 s for ACO.

5 Results and discussions

This section discusses the simulation and real-time experimental results of PID and FPID controllers along with optimization techniques for studying temperature control system. Different performance indices like ISE, IAE and

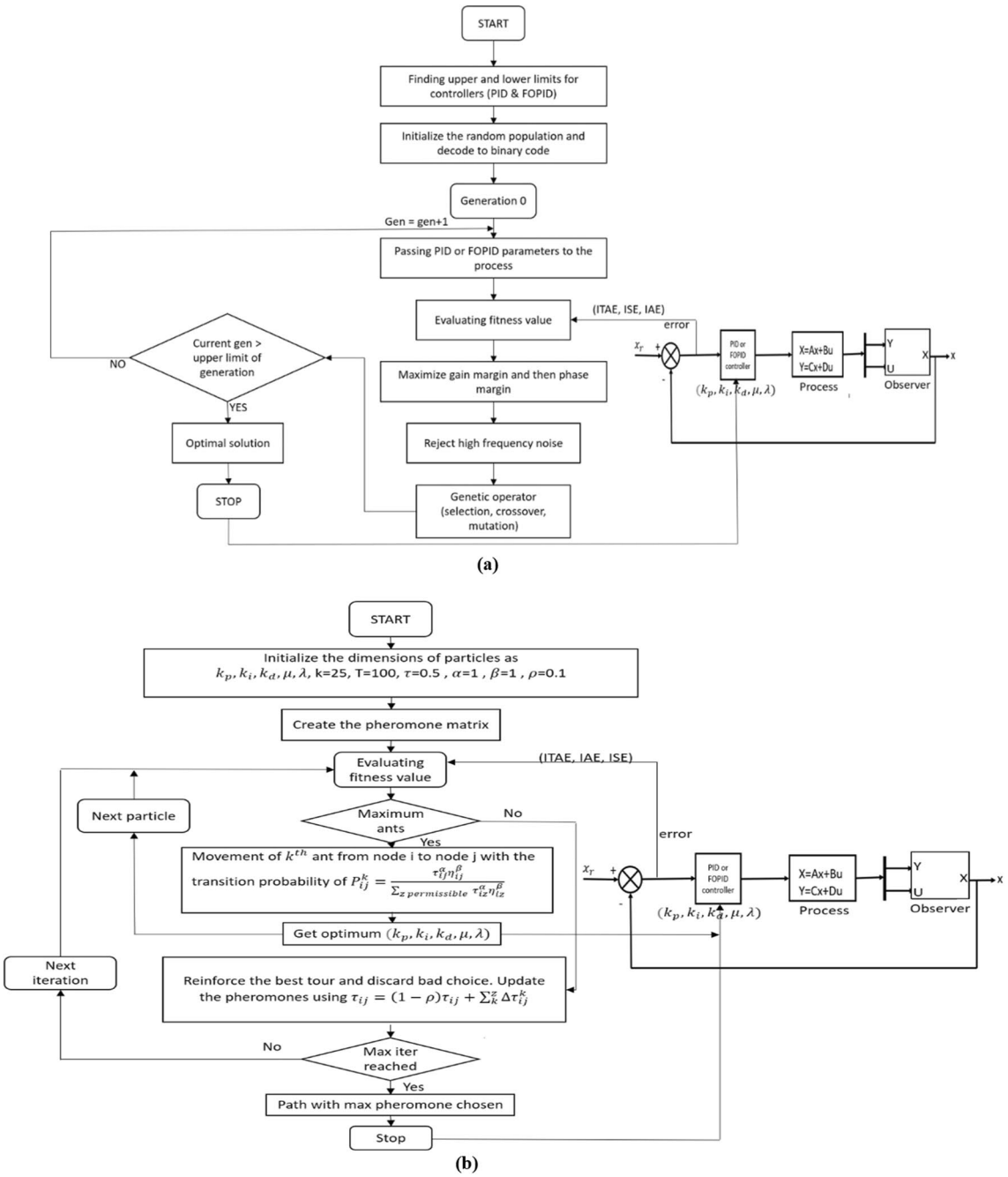


Fig. 9 Flow chart of the proposed **(a)** Genetic Algorithm **(b)** Ant Colony Optimization **(c)** Particle Swarm Optimization

ITAE are implemented for simulation and only ITAE is used for real-time experimentation purpose. The combined form of temperature controller with Cascade control, Feedforward control and dead-time compensator is examined after modeling for simulation using MATLAB 2022a. To keep the computational time measurements consistent, all the experiments were run on a computer with an Intel I7-3770, 3.4 GHz CPU with 32 GB of memory with no GPU acceleration. Verification of results after the implementation of various indices on the simulated model is to be considered on a momentary basis. After that, real-time experimentation is carried out with these controllers over the temperature control setup where $V_{sp} = 0.95$.

1) Comparison of simulated results using PID controller with Metaheuristic Optimization Techniques based on ISE, IAE and ITAE: Fig. 10(a-c) shows the comparison of simulated results using PID controller with metaheuristic

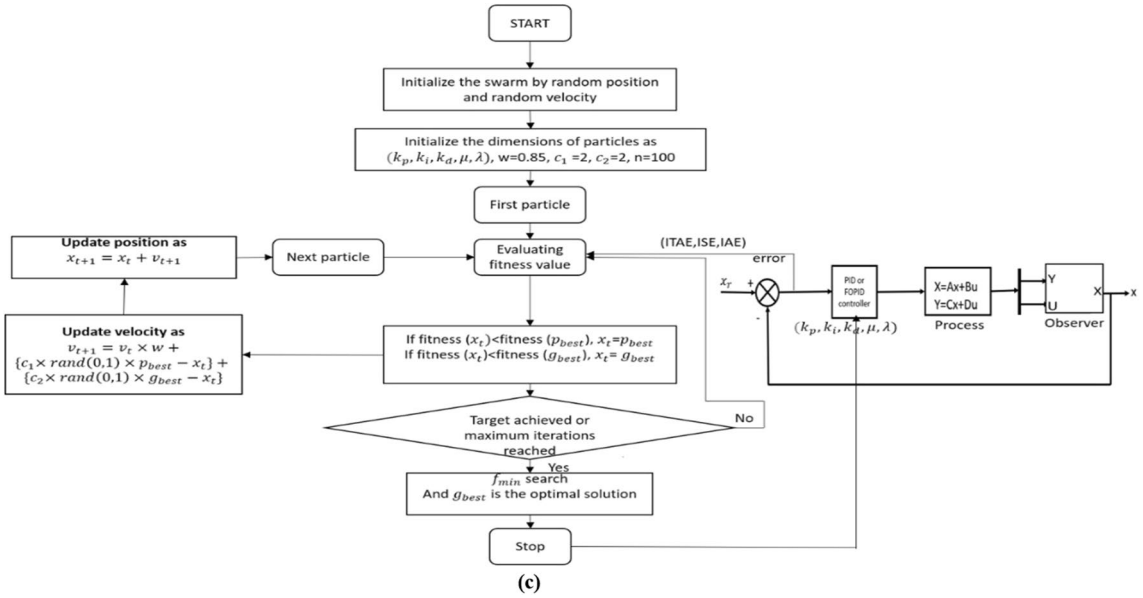


Fig. 9 (continued)

Table 2 Parameters of GA, PSO and ACO algorithm

PARAMETERS of GA		PARAMETERS of ACO		PARAMETERS of PSO	
Parameters	Values	Parameters	Values	Parameters	Values
Number of generations	100	Max iterations (T)	100	Number of iterations	100
Population size	50	Pheromone (τ)	0.5	Number of participles	50
Length of chromosomes	35	Number of ants (k)	50	Inertia	0.85
Probability of crossover	0.85	Pheromone exponential weight (α)	1	Cognitive component (c_1)	2
Probability of mutation	0.3	Pheromone heuristic weight (β)	1	Social component (c_2)	2
		Evaporation parameter (ρ)	0.1		

optimization techniques based on different indices ISE, IAE and ITAE respectively. Values of all the parameters (P, I, D, N) are optimized on the basis of different indices as shown in Table 3. Table 4 depicts the performance analysis of PID and PID with optimization techniques based on different indices. It is observed that all three optimization techniques provide much better results in comparison to the non-optimized PID controller. Among optimization techniques, ACO provides the best settling time and GA provides the least overshoot. Among performance indices, ITAE provides overall balanced results.

2) Comparison of simulated results using FPID controller with Metaheuristic Optimization Techniques based on ISE, IAE and ITAE: Fig. 11(a-c) shows the comparison of simulated results using FPID controller with metaheuristic optimization techniques based on different indices ISE, IAE and ITAE respectively. Values of all the parameters (P, I, D, μ, λ) are optimized based on different indices as shown in Table 5. The performance analysis of FPID and FPID with optimization techniques based on different indices is shown in Table 6. Among optimization techniques, ACO provides the best settling time and the least overshoot. Among performance indices, ITAE provides the best-balanced results. As compared to PID and PID with optimization techniques, FPID and FPID with optimization techniques provide much better performance results.

3) Comparison of real-time experimentation results using PID and FPID controllers with Metaheuristic Optimization Techniques based on ITAE: During simulation, it has been observed that ITAE provides balanced results among all performance indices. So, only ITAE is considered for the real-time analysis of the system. The parametric values of the controllers are decided on the basis of the simulated results discussed above. The results of both PID and FPID controllers with different optimization techniques based on ITAE are shown in Fig. 12 respectively and the performance analysis is shown in Table 7.

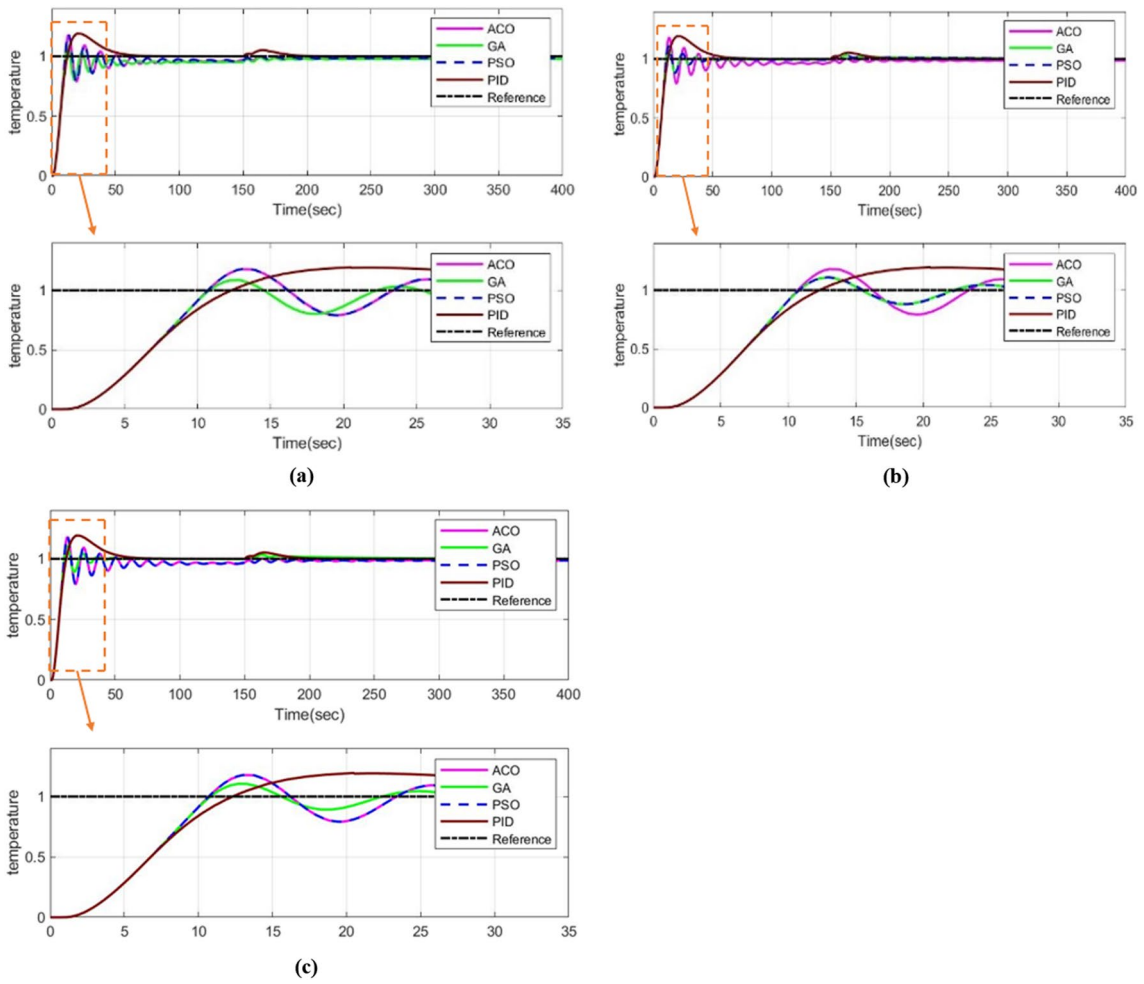


Fig. 10 Comparison of simulated results using PID controller with metaheuristic optimization techniques based on different indices **(a)** ISE **(b)** IAE and **(c)** ITAE

Table 3 Optimized values of all the parameters (P, I, D, N) on the basis of different indices

INDICES →	ISE				IAE				ITAE			
CONTROLLER & OPTIMIZATION TECHNIQUES → PARAMETERS ↓	PID	GA	PSO	ACO	PID	GA	PSO	ACO	PID	GA	PSO	ACO
P	1.694	1.858	1.143	1.116	1.694	1.402	1.012	1.116	1.694	1.658	1.716	1.114
I	0.093	0.002	0.044	0.001	0.093	0.083	0.099	0.003	0.093	0.096	0.099	0.001
D	5.664	4.491	4.089	4.082	5.664	4.193	4.084	4.084	5.664	4.232	4.023	4.083
N	30.42	28.26	28.29	28.12	30.42	28.001	28.18	28.13	30.42	28.51	28.48	28.67

6 Conclusions and future perspective

This paper proposed a novel cascade control strategy using PID and Fractional Order PID ($P\ I^{\lambda}D^{\mu}$) controller along with an integral anti-windup technique and velocity set-point to control and regulate the outlet temperature of a heat exchanger to a certain reference value and enhance the heat transmission efficiency irrespective of the external disturbances, non-linearity and delay characteristics. Various control strategies like Cascade Control, Feedforward Control and Smith Predictor for time delayed process (dead-time compensator) are discussed for controlling the temperature

Table 4 Performance analysis of PID and PID with optimization techniques based on different indices

INDICES →	ISE			IAE			ITAE		
PARAMETERS → CONTROLLER & OPTIMIZATION TECHNIQUES ↓	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT
PID	25.02	276.98	19.48	25.02	276.98	19.48	25.02	276.98	19.48
GA	15.19	269.20	10.82	16.10	267.29	10.53	15.86	263.21	10.74
PSO	13.01	245.65	20.03	12.97	241.34	19.41	15.84	256.56	11.03
ACO	12.97	238.71	19.2	12.39	238.73	19.56	12.97	238.70	19.45

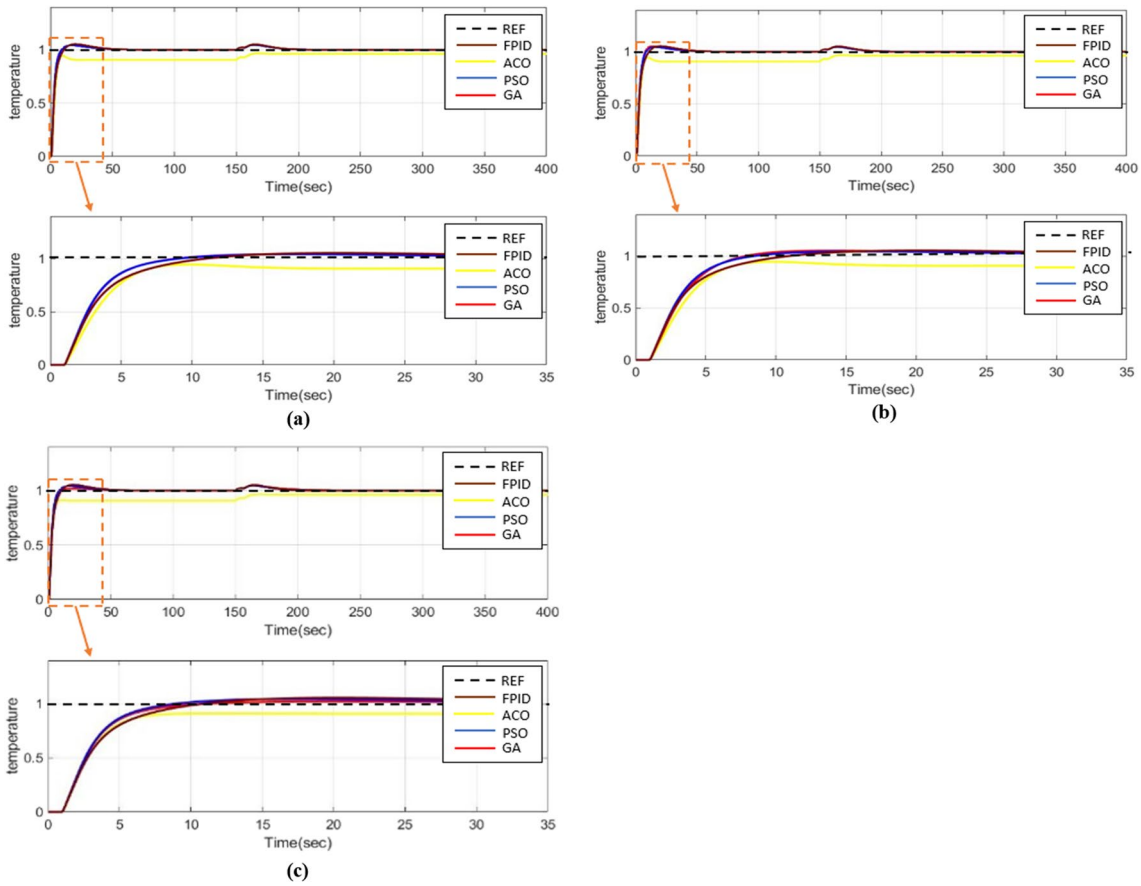


Fig. 11 Comparison of simulated results using FPID controller with metaheuristic optimization techniques based on different indices **(a)** ISE **(b)** IAE and **(c)** ITAE

of the process. The parameters of both PID and P $I^{\lambda}D^{\mu}$ controllers are tuned using adaptive heuristic optimization techniques like GA, ACO and PSO to achieve an optimal and desired performance with the fastest steady-state regime and less percentage overshoot. IAE, ISE and ITAE cost functions are designed to select optimum parameters of the controllers to improve transient and steady-state response of the controllers.

For developing a real-time framework for the temperature control of a heat exchanger plant, this article used a PLC and SCADA-based control framework to automate and supervise the temperature control processes in a heat exchanger plant and monitor all the processes using SCADA. The

Table 5 Optimized values of all the parameters (P, I, D, μ, λ) based on different indices

INDICES →	ISE				IAE				ITAE			
CONTROLLER & OPTIMIZATION TECHNIQUES → PARAMETERS ↓	FPID	GA	PSO	ACO	FPID	GA	PSO	ACO	FPID	GA	PSO	ACO
P	1.694	1.998	1.979	1.955	1.694	1.982	2.011	1.955	1.694	1.989	2.001	1.953
I	0.093	0.099	0.100	0.001	0.093	0.110	0.101	0.001	0.093	0.077	0.110	0.023
D	5.664	5.923	6.021	4.181	5.664	5.183	5.692	4.187	5.664	5.923	5.992	5.582
μ	0.998	0.129	0.001	0.001	0.989	0.183	0.751	0.001	0.999	0.198	0.672	0.001
λ	0.989	0.184	0.011	0.011	0.970	0.839	0.534	0.011	0.998	0.701	0.491	0.010

Table 6 Performance analysis of FPID and FPID with optimization techniques based on different indices

INDICES →	ISE			IAE			ITAE		
PARAMETERS → CONTROLLER & OPTIMIZATION TECHNIQUES ↓	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT
FPID	49.78	1692	5.35	49.78	1691	5.35	49.78	1691	5.35
GA	33.80	1534	4.58	35.05	1536	5.21	34.76	1611	4.77
PSO	32.67	1555	4.56	34.22	1535	4.60	34.22	1565	4.57
ACO	39.34	1366	0.41	39.36	1362	0.40	41.23	1392	0.23

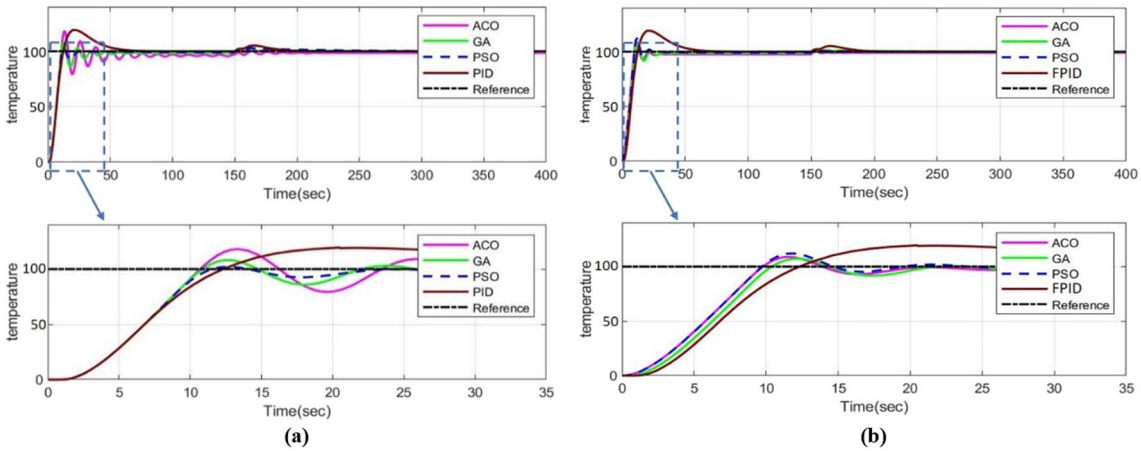


Fig. 12 Real-time analysis of the system using PID and FPID controllers with different optimization techniques based on ITAE

OMRON (NX1P2-9024DT1) is interfaced with the Wonderware InTouch SCADA system to gather data and create a simulated prototype of the temperature control to monitor the entire process on a real-time basis in the SCADA screen and historize the data to its database. PLC carries out the necessary control operations within the heat exchanger plant as directed by the SCADA server. The programming of PLC is done using Sysmac studio automation software using the ladder programming language and Wonderware Intouch SCADA software is used for visualizing the whole plant operations.

In control system design and analysis, the calculated performance indices are used as quantitative measures for evaluating the performance of the system. This study used both real-time experimental and simulation results to demonstrate the effective performance of the developed controllers using

Table 7 Performance analysis of results of both PID and FPID controllers with different optimization techniques based on ITAE

CONTROLLER →	PID			FPID		
PARAMETERS → CONTROLLER & OPTIMIZATION TECHNIQUES↓ (based on ITAE)	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT	RISE TIME (s)	SETTLING TIME (s)	OVERSHOOT
PID/FPID	25.02	276.98	19.48	23.12	276.97	19.48
GA	16.86	245.76	8.14	17.56	137.14	7.27
PSO	20.74	231.96	3.04	14.04	118.18	12.31
ACO	17.03	245.51	19.94	11.09	92.34	9.29

MATLAB with metaheuristic optimization techniques based on different performance indices like ISE, IAE and ITAE while guaranteeing both transient and steady-state response of the controller and rejection of external disturbances in comparison to the existing literature. The outcomes of simulation and experimental assessment implies that fractional order PID controller provides overall better control performances as compared to PID compensator for the temperature control of a heat exchanger plant. During simulation, the optimized PID controller with ACO algorithm provides the best setting time and the optimized PID controller with GA algorithm provides the minimum peak overshoot. However, the optimized FPID controller with the ACO algorithm provides the least settling time and least peak overshoot. ITAE performance index provides overall balanced results both in the case of optimized PID and P $I^{\lambda}D^{\mu}$ controllers respectively. As compared to the PID controller, P $I^{\lambda}D^{\mu}$ controller provides better performance results. During real-time experimentation, PID optimized with PSO algorithm provides balanced results and FPID optimized with ACO provides the most balanced results and FPID controller provides better performance results as compared to PID controller. Since FPID controller provides much-improved results than the PID controller and it is observed that FPID optimized with ACO provides overall balanced results during the simulation and in real-time experimentation. So, among all the optimization techniques (GA, PSO & ACO) in this study, ACO provides overall better and balanced results. ACO attains the desired performance with the fastest steady-state regime and less percentage overshoot by tuning the parameters of the proposed Fractional Order P $I^{\lambda}D^{\mu}$ controller.

The proposed system uses EtherCAT protocol which offers high speed, low hardware cost, wide range of products, low latency/delay, flexible topology support and incredible synchronization than other real-time protocols in industrial control systems. The proposed SCADA system achieves automatic control and management of industrial processes, improves worker safety, and reduces manpower and physical resources instead of the distributed control system. The significance of investigating and implementing automated systems across multiple sectors is increased by this work.

The control scheme proposed in this paper also has some limitations that need to be studied further. Future recommendations of the proposed work are as follows:

1. Since it is a component and cost-heavy setup and all the components are interlinked. So, if one of the components start behaving abnormally, the whole process will be compromised. So, timely maintenance and checking is required.
2. Through careful selection of parameter values, we are just exploiting GA, PSO and ACO algorithms to get the best possible results in terms of precision and accuracy. These algorithms have a good convergence rate and result in lower error rates. But there is always a scope for improvement in the obtained results from these optimization techniques based on appropriate choice of representations, parametrization and techniques. In future work, some recently developed algorithms such as the Grey Wolf Optimization algorithm (GWO), and artificial ecosystem optimization algorithm (AEO) can be used and the comparison of the results obtained with all the algorithms can be drawn to achieve an optimal and desired performance.
3. There is always some latency between the actual process parameters in real-time and shown parametric values on SCADA. This latency can be reduced in future work by improving the hardware and software linked to the system.

Author's contributions B.T.: Conceptualization and methodology, writing-original draft preparation, investigation, software, validation, editing; N.K. and M.S.: Research design, formal analysis, guidance, and reviewing. All authors have read and approved the revised final manuscript.

Funding The authors received no specific funding for this study.

Data availability No datasets were generated or analysed during the current study.

Declarations

Code availability Code is available.

Competing of interests The authors declare no competing intere.

Conflict of interest The authors declare no competing interests.

References

1. Wang B, Klemeš JJ, Li N, Zeng M, Varbanov PS, Liang Y (2021) Heat exchanger network retrofit with heat exchanger and material type selection: A review and a novel method. Renew Sustain Energy Rev 138:110479
2. Klemeš JJ, Kravanja Z (2013) Forty years of heat integration: pinch analysis (PA) and mathematical programming (MP). Curr Opin Chem Eng 2(4):461–474
3. Dulău M, Karoly M, Dulău TM (2018) Fluid temperature control using heat exchanger. Procedia Manuf 22:498–505
4. Kaur I, Singh P (2021) State-of-the-art in heat exchanger additive manufacturing. Int J Heat Mass Transf 178:121600
5. Safaei A, Nezhad AH, Rashidi A (2020) High temperature nanofluids based on therminol 66 for improving the heat exchangers power in gas refineries. Appl Therm Eng 170:114991
6. Afram A, Janabi-Sharifi F (2014) Theory and applications of HVAC control systems–A review of model predictive control (MPC). Build Environ 72:343–355
7. Laszczyk P (2017) Simplified modeling of liquid-liquid heat exchangers for use in control systems. Appl Therm Eng 119:140–155
8. Tugashova LG, Zatonskiy AV (2021) Comparison of Methods for Heat Exchanger Control. Theor Found Chem Eng 55(1):53–61
9. Feliu-Batlle V, Rivas-Perez R (2021) Control of the temperature in a petroleum refinery heating furnace based on a robust modified Smith predictor. ISA Trans 112:251–270
10. Kumar M, Prasad D, Giri BS, Singh RS (2019) Temperature control of fermentation bioreactor for ethanol production using IMC-PID controller. Biotechnology Reports 22:e00319
11. Gao Z, Trautzsch TA, Dawson JG (2002) A stable self-tuning fuzzy logic control system for industrial temperature regulation. IEEE Trans Ind Appl 38(2):414–424
12. Taborek J, Hewitt GF, Afgan NH (1983) Heat exchangers–theory and practice International Center for Heat and Mass Transfer (Eds: J. Taborek, G.F. Hewitt, N.H. Afgan), Hemisphere, Washington, D.C
13. Venegas T, Qu M, Nawaz K, Wang L (2021) Critical review and future prospects for desiccant coated heat exchangers: Materials, design, and manufacturing. Renew Sustain Energy Rev 151:111531
14. Rößler F, Krumova V, Gewald S, Bauernfeind A, Freko P, Thomas I, Zander HJ, Rehfeldt S, Klein H (2022) Hazard Analysis of Fixed-Tube-Sheet Shell-and-Tube Heat Exchangers. Chem Ing Tec 94(5):727–737
15. Jin HZ, Gu Y, Ou GF (2021) Corrosion risk analysis of tube-and-shell heat exchangers and design of outlet temperature control system. Pet Sci 18(4):1219–1229
16. Jafari SM, Saramnejad F, Dehnad D (2018) Designing and application of a shell and tube heat exchanger for nanofluid thermal processing of liquid food products. J Food Process Eng 41(3):e12658
17. Bichkar P, Dandgaval O, Dalvi P, Godase R, Dey T (2018) Study of shell and tube heat exchanger with the effect of types of baffles. Procedia Manuf 20:195–200
18. Pekař L (2020) Modeling and identification of a time-delay heat exchanger plant. In: Advanced analytic and control techniques for thermal systems with heat exchangers. Academic Press, pp 23–48
19. Zhang T, Wen JT, Catano J, Zhou R (2009) Stability analysis of heat exchanger dynamics. In: 2009 American control conference. IEEE, pp 3656–3661
20. Al-Dhaifallah M (2019) Heat exchanger control using fuzzy fractional-order PID. In: 2019 16th International multi-conference on systems, signals & devices (SSD). IEEE, pp 73–77
21. Li R, Wu F, Hou P, Zou H (2020) Performance assessment of FO-PID temperature control system using a fractional order LQG benchmark. IEEE Access 8:116653–116662
22. Palaniyappan TK, Yadav V, Tayal VK, Choudekar P (2018) PID control design for a temperature control system. In: 2018 International conference on power energy, environment and intelligent control (PEEIC), pp 632–637
23. Cervantes-Escorcia N, Santos-Sánchez OJ, Rodríguez-Guerrero L, Romero-Trejo H, González-Facundo A (2019) Optimal PI and PID temperature controls for a dehydration process. Arab J Sci Eng 44:2519–2534
24. Pamela D, Godwin Premi MS (2017) Wireless control and automation of hot air temperature in oven for sterilization using fuzzy PID controller and adaptive smith predictor. Wireless Pers Commun 94:2055–2064
25. Zou H, Li H (2017) Improved PI-PD control design using predictive functional optimization for temperature model of a fluidized catalytic cracking unit. ISA Trans 67:215–221
26. Aldemir A, Hapoğlu H (2016) Comparison of PID tuning methods for wireless temperature control. J Polytech 19(1):9–19
27. Aldemir A (2018) PID controller tuning based on phase margin (PM) for wireless temperature control. Wireless Pers Commun 103:2621–2632
28. Oravec J, Bakošová M, Trafczynski M, Vasičkaninová A, Mészáros A, Markowski M (2018) Robust model predictive control and PID control of shell-and-tube heat exchangers. Energy 159:1–10
29. Dong G, Deng M (2022) Operator & Fractional Order Based Nonlinear Robust Control for a Spiral Counter-flow Heat Exchanger with Uncertainties and Disturbances. Machines 10(5):335
30. Tiwari D, Pachauri N, Rani A, Singh V (2016) Fractional order PID (FOPID) controller based temperature control of bioreactor. In: 2016 International conference on electrical, electronics, and optimization techniques (ICEEOT). IEEE, pp 2968–2973
31. Jin Q, Yu Y, Xia Y (2024) Active control for system with multiple microchannel heat exchangers using disturbance observer. Numer Heat Tran Part A: Appl 85(6):958–973
32. Utama YAK, Hari Y (2017) Design of PID disturbance observer for temperature control on room heating system. In: 2017 4th International conference on electrical engineering, computer science and informatics (EECSI). IEEE, pp 1–6
33. Mohanty MD, Mohanty MN (2020) Design of fuzzy logic-based PID controller for heat exchanger used in chemical industry. In: New paradigm in decision science and management: proceedings of ICDSM 2018. Springer Singapore, pp 371–379
34. Vasičkaninová A, Bakošová M (2015) Control of a heat exchanger using neural network predictive controller combined with auxiliary fuzzy controller. Appl Therm Eng 89:1046–1053
35. Carvalho CB, Carvalho EP, Ravagnani MA (2020) Implementation of a neural network MPC for heat exchanger network temperature control. Braz J Chem Eng 37(4):729–744
36. Al-Dhaifallah M (2023) Fuzzy fractional-order PID control for heat exchanger. Alex Eng J 63:11–16

37. Thepmanee T, Pongswatd S, Asadi F, Ukakimaparn P (2022) Implementation of control and SCADA system: Case study of Allen Bradley PLC by using WirelessHART to temperature control and device diagnostic. Energy Rep 8:934–941
38. Lu K, Xu L, Fu Q, Yin X, Yuan X (2022) Composition-temperature cascade control of vapor recompression assisted dividing wall column with side heat exchanger. Chem Eng Res Des 177:24–35
39. Bobič M, Gjerek B, Golobič I, Bajsić I (2020) Dynamic behaviour of a plate heat exchanger: Influence of temperature disturbances and flow configurations. Int J Heat Mass Transf 163:120439
40. Miao W, Xu B (2023) Application of feedforward cascade compound control based on improved predictive functional control in heat exchanger outlet temperature system. Appl Sci 13(12):7132
41. Damasceno NC (2017) PI controller optimization for a heat exchanger through metaheuristic bat algorithm, particle swarm optimization, flower pollination algorithm and cuckoo search algorithm. IEEE Lat Am Trans 15(9):1801–1807
42. Li Z, Liu J, Yang X, Huang X, Jia L, Li M (2023) Novel effective room temperature-based predictive feedback control method for large-scale district heating substation. Appl Therm Eng 218:119241
43. Xu S, Shi S, Jiang W, Hashimoto S (2023) Cooperative Control of Recurrent Neural Network for PID-based Single Phase Hotplate Temperature Control Systems. IEEE Access 11:105557–105569
44. Jamil AA, Tu WF, Ali SW, Terriche Y, Guerrero JM (2022) Fractional-order PID controllers for temperature control: A review. Energies 15(10):3800
45. Hussein MM, Alkhalaf S, Mohamed TH, Osheba DS, Ahmed M, Hemeida A, Hassan AM (2022) Modern temperature control of electric furnace in industrial applications based on modified optimization technique. Energies 15(22):8474
46. Maya-Rodriguez MC, Carvajal-Mariscal I, López-Muñoz R, Lopez-Pacheco MA, Tolentino-Eslava R (2023) Temperature control of a chemical reactor based on neuro-fuzzy tuned with a metaheuristic technique to improve biodiesel production. Energies 16(17):6187
47. Hou G, Gong L, Wang M, Yu X, Yang Z, Mou X (2022) A novel linear active disturbance rejection controller for main steam temperature control based on the simultaneous heat transfer search. ISA Trans 122:357–370
48. Ouyang M, Wang Y, Wu F, Lin Y (2023) Continuous reactor temperature control with optimized PID parameters based on improved sparrow algorithm. Processes 11(5):1302
49. Zeng Z, Zhang W, Yu H, Cao D, Li X (2023) Dynamic thermal model and control performance analysis over frequency domain of temperature control unit. Case Stud Therm Eng 45:102975

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.

Prediction of Air Quality Index Using Random Forest and Prophet Tool

1st Abhishek Walia
Software Engineering Department
Delhi Technological University
Delhi, India
abhishekw836@gmail.com

2nd Ajay Paliwal
Software Engineering Department
Delhi Technological University
Delhi, India
ajay1182001@gmail.com

3rd Sanjay Patidar
Software Engineering Department
Delhi Technological University
Delhi, India
sanjaypatidar@dtu.ac.in

4th Rakeshkumar Mahto
Department of Electrical and Computer Engineering
California State University, Fullerton
California, USA
ramahto@fullerton.edu

2024 19th Annual System of Systems Engineering Conference (SoSE) | 979-8-3503-6591-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/SOSE62659.2024.10620972

***Abstract*—The Air Quality Index (AQI) serves as a critical tool for informing the public about the negative impact of air pollution on human health. In Indian cities, where air pollution levels have been rising sharply, accurate assessment and forecasting of AQI are more important than ever. This study evaluates urban air quality and forecasts future AQI by analyzing a wide array of pollutants, including nitrogen oxides (NO_x), ammonia, carbon monoxide, sulfur dioxide, ozone, benzene, toluene, PM2.5, PM10, nitric oxide, nitrogen dioxide, and xylene. We employ a Random Forest classifier for detection and use the Prophet library for forecasting. The performance of our model is demonstrated by precision, accuracy, and recall metrics, with the model achieving an accuracy score of 99.57%. This high level of accuracy indicates the model's robust capability in predicting AQI and underscores its potential utility in urban air quality management.**

***Index Terms*—Random forest classifier, Air Quality Index (AQI), precision, accuracy, forecasting**

I. Introduction

To sustain human existence and development, atmospheric conditions must be conducive to growth. Among all the atmospheric conditions, air quality is one of the key elements to maintaining human life. Therefore, it is essential to have a comprehensive understanding of air quality and mechanisms to monitor it precisely. Due to rapid population growth and swift industrialization resulted in a steady deterioration in air quality. This is due to the increase in the emission of toxic gases and the simultaneous release of hazardous compounds into the atmosphere. Consequently, the rise in air pollution has resulted in millions worldwide suffering from physiological ailments [1] and respiratory disorders. As a result, scientific data in [2] identifies air pollution as a primary environmental threat. Therefore, it is necessary to develop a new benchmark to quantify the air quality and accordingly identify its consequences on humans.

Air Quality Index (AQI) is a numerical indicator utilized to quantify an area's air pollution. Generally, AQI computations require twelve different parameters that include contaminates such as nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), Ozone (O_3) [3], [4], particulate matter 10 micrometers or less in diameter (PM10), particulate matter 2.5 micrometers or less in diameter (PM2.5), benzene [5], [6], [7], [8] and Ammonia. In some applications, six pollutants are specifically used to calculate the AQI: NO_2, CO, PM10, PM2.5, SO_2, and O_3 [5], [9]. The selection of a particular pollutant depends on the intended use and a number of other variables, such as measuring techniques, data accessibility, and monitoring frequency. Many meteorological stations provide hourly and daily AQI in real time. Due to the severity of poor AQI on respiratory health, it becomes critical to develop a mechanism to predict its value in the future. That way, resources can be appropriately deployed during severe AQI.

In a study conducted by [1], statistical models such as automatic moving averages (ARMA) and multiple linear regression are applied to historical data of AQI to predict the future behavior of variables in AQI. The proposed models demonstrated a high accuracy. However, the statistical model could not capture the dynamic nature of meteorological data. This can lead to a less precise estimate of exposure levels. Therefore, there is a need for advanced techniques that can process a large dataset and capture patterns in them. The techniques in machine learning can meet these requirements, and they can be utilized to predict AQI.

This capability of machine learning in AQI was extensively demonstrated in recent research that leverages advanced methods to improve accuracy and reliability in environmental monitoring. For example, Madan et al. in [5] conducted a detailed review of using machine learning methods, including decision trees and random forests, to predict the AQI. According to their research, the random forest method provides superior prediction capabilities for the AQI. Building on this advancement, Rao et al. utilized a profound learning design called BILSTM [7]. They achieved exceptional mean absolute error and root mean square error (RMSE), demonstrating

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 09:33:23 UTC from IEEE Xplore. Restrictions apply.

significant improvements in performance over previous models for PM2.5 prediction. Continuing the research work in the application of machine learning in AQI, Chavi Srivastava et al. highlighted machine learning algorithms as a promising avenue for forecasting air pollution levels [9]. In this work [9], various machine learning techniques were tested to predict key pollutants such as CO, NO_2, O_3, SO_2, PM10, and PM2.5, leveraging meteorological data and environmental variables. However, all the machine learning based AQI prediction models referenced in this paper have utilized robust datasets with minimal null or missing data points. An AQI dataset with many null values due to sensor errors or technical issues could result in inaccurate prediction. Hence, despite the promise shown by the machine learning technique, there is still a crucial need to develop techniques robust enough to accurately predict AQI, even with missing data points, while maintaining high accuracy.

This study employs advanced machine-learning techniques to enhance the prediction and forecasting of the AQI. By extracting features from the dataset and employing innovative data cleaning that includes replacing null values with column means, we ensure the integrity and usability of data for analysis. This work utilizes the random forest classifier [5] for AQI index prediction, providing a reliable and accurate model for real-time air quality assessment. Furthermore, this paper leverages the Prophet Library [10], [11], an advanced forecasting tool, to project AQI trends over the upcoming year based on historical data. Hence, this work will provide actionable insights to make informed, actionable decisions and build strategies to combat air pollution effectively.

The rest of the paper is structured as follows: Section II briefly describes the dataset used for this work. This is followed by Section III, which presents the methodology employed in this paper. The methodology section offers an overview and systematically describes this study's data cleansing, forecasting, and prediction techniques. Subsequent to this, Section IV, the results and analysis section, presents the performance of the random forest technique based on various performance evaluation metrics. We then conclude in the conclusion section, highlighting the work done in this paper.

II. Dataset Description

In this study, we utilized the dataset sourced from Kaggle [12]. This comprehensive dataset provides detailed records of air quality measurements from multiple monitoring stations across India in 26 cities from 2015 - 2020, as shown in Fig. 1. The date attribute in the dataset is in YYYY-MM-DD format. Moreover, the dataset consists of a wide range of pollutants that includes particulate matter such as PM2.5 and PM10, NO_2, SO_2, CO, O_3, and others. Hence, this dataset reflects the varied aspects of air quality relevant to public health and environmental policy.

Each entry in the dataset corresponds to daily averages per city, thereby facilitating the tracking of pollution patterns and assessment of the efficacy of environmental regulations over

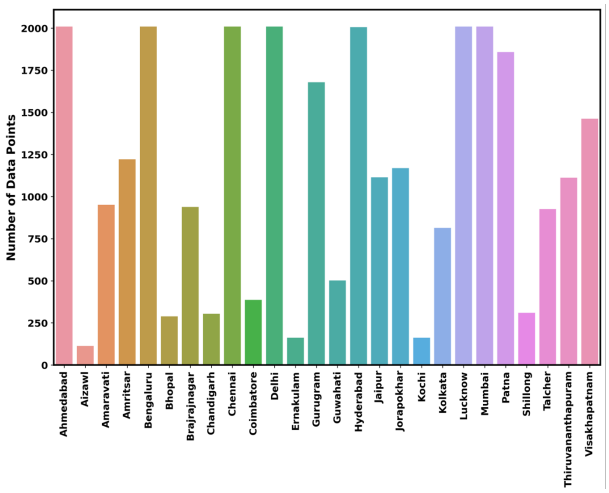


Fig. 1: Distribution of data points across various cities

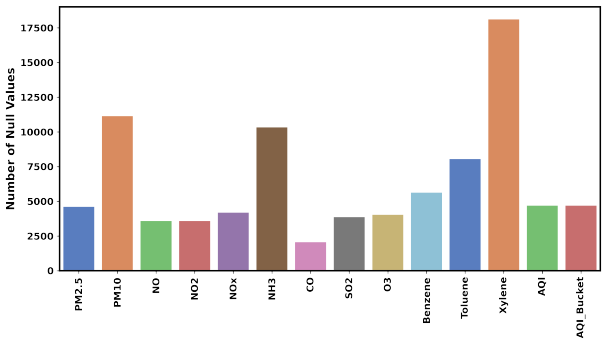


Fig. 2: Distribution of data points across various cities

time. Moreover, the dataset consists of extensive coverage over consecutive years, allowing for an in-depth study of the progression and seasonal variations in air quality. Hence, such an extensive dataset is essential for forecasting and planning purposes.

The dataset consists of many null values in different pollutant categories, as shown in Fig. 2. These null values in various categories pose a challenge for data analysis and may compromise the reliability of subsequent findings. In the subsequent section, we will present a methodology to address and mitigate these null values.

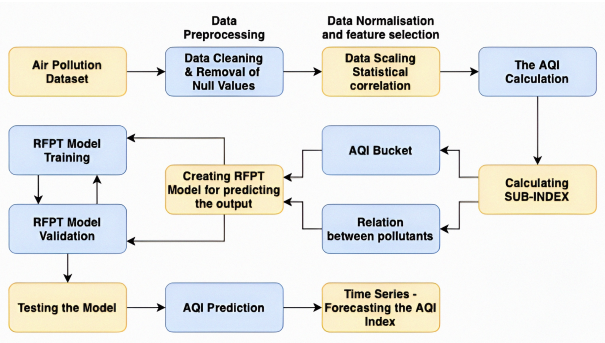


Fig. 3: Flow Diagram for prediction and forecasting of AQI

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 09:33:23 UTC from IEEE Xplore. Restrictions apply.

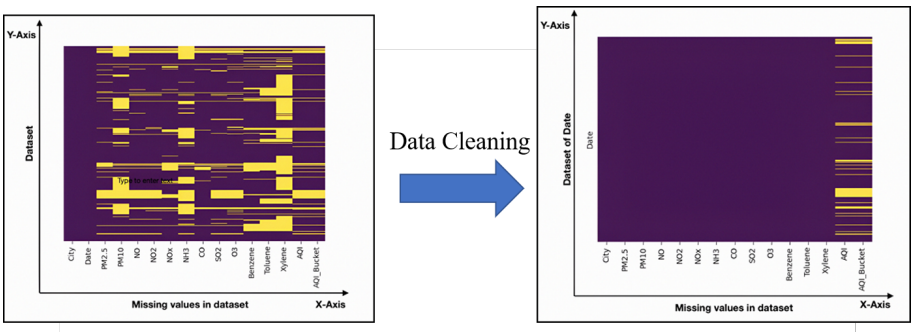


Fig. 4: Comparative heatmaps of data before and after cleaning.

III. Methodology

The flow diagram shown in Fig. 3 presented how the data from the dataset is processed in various stages that lead to predicting the AQI of different cities in India. This flow diagram shows how the dataset is cleaned to eliminate null values from it. Later, the pollutants that are correlated with each other are selected for train and testing the random forest classifier to predict the AQI. Subsequently, we use the prophet tool to forecast the AQI, which enables us to reduce the impact of air pollution on the environment. Prior to modeling and predicting the AQI, data preprocessing is necessary to address the numerous null values that could impact model's performance.

A. Data Cleaning

Data preprocessing is essential to ensure the accuracy of the environmental models used in this study. The initial phase involved a data cleaning process to rectify the issue of missing values within the dataset as shown in Fig. 4. Given the heterogeneous nature of air quality across urban areas, the replacement of missing pollutant data was performed on a city-wide basis.

The pollutants under consideration for the AQI include Nitrogen oxides (NOx), Ammonia (NH_3), CO, PM2.5, PM10, SO_2, and O_3. These constituents play a critical role in the computation of the AQI, which is a determinant factor in assessing air quality standards.

An analysis of the dataset's missing values revealed a significant percentage of missing entries across various pollutants, as shown in Fig. 4. To combat this, a strategy was employed where missing values were supplanted with the mean of the corresponding pollutant, calculated for each city individually. This approach ensures that the substituted values reflect the local environmental conditions, thus preserving the integrity of the dataset. Consequently, the dataset's consistency was reinforced, rendering it suitable for subsequent stages of modeling and analysis.

B. Calculating the Sub-Index

A sub-index is calculated for each contaminant utilizing suitable measures, encompassing concentrations and their likely impacts on well-being. The highest sub-index indicates the overall AQI, removing the maximum values from the sub-indices in order to fill the NULL values in the AQI column. The calculation uses specific guidelines that relate concentration levels to health impact scores, as shown in Fig. 5. In the process of determining the overall AQI, the highest sub-index among all pollutants serves as the AQI. This method ensures that the AQI reflects the most severe pollution level among the monitored contaminants.

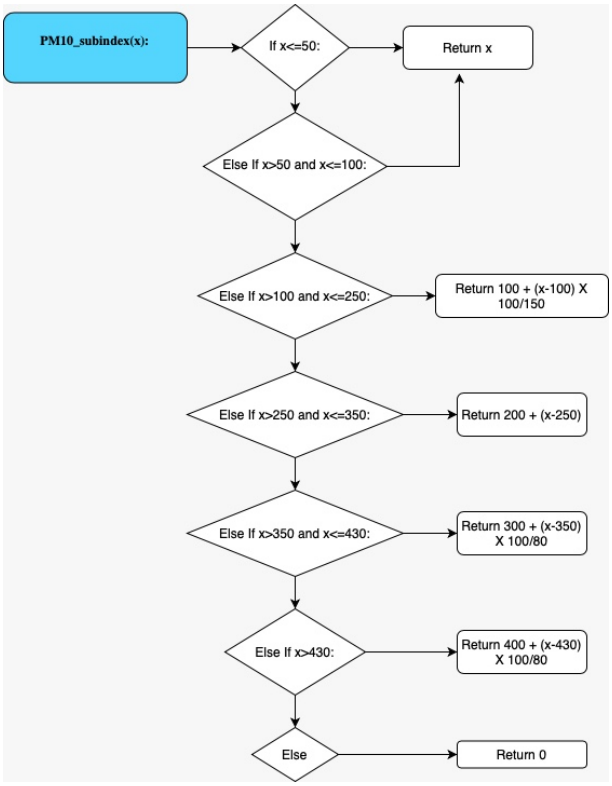


Fig. 5: Flowchart illustrating the step-by-step computation of the PM10 sub-index based on concentration levels and their corresponding health impact scores.

C. Categorize AQI Bucket

After calculating the AQI, it is essential to categorize each AQI value into a descriptive bucket representing the air quality level, such as 'Good,' 'Satisfactory,' or 'Poor.' This classification is done according to predefined ranges found in Table. I. The process involves filling any null values in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 09:33:23 UTC from IEEE Xplore. Restrictions apply.

TABLE I: World Health Organization's general guidelines for AQI categories and associated health impacts

AQI Category	Associated Health Impact
Good 0 to 50	Negligible effect.
Satisfactory 51 to 100	Could result in slight respiratory discomfort for individuals with sensitivities.
Moderately 101 to 200	May induce respiratory discomfort in individuals with lung conditions like asthma, as well as discomfort in those with heart disease, children, and older adults.
Poor 201 to 300	Long-term exposure might cause respiratory distress, especially in heart disease patients.
Very Poor 301 to 400	Long-term exposure can cause respiratory sickness, and people with pre-existing disorders may be more susceptible to its effects.
Severe 401 to 500	While people with lung or heart issues may incur major health consequences, even healthy persons may encounter respiratory side effects.

the AQI bucket column based on the calculated AQI. This categorization helps simplify the interpretation of AQI values, making it easier for policymakers and the public to understand the air quality levels and associated health implications.

D. Analyzing the Relationship Between Pollutants

To explore the interrelationships between different air pollutants, we utilized various open-source libraries in Python. The primary library employed was the Seaborn Library [13], in conjunction with pandas and Matplotlib [14]–[16], to create a correlation heatmap. The correlation heatmap for this study is depicted in Fig. 6. This visualization aids in identifying pollutants that have similar sources or exhibit similar behaviors under specific atmospheric conditions, which can be crucial for targeted air quality management strategies. Having established the correlations between various pollutants and AQI, it is now appropriate to select a suitable modeling technique for accurate prediction.

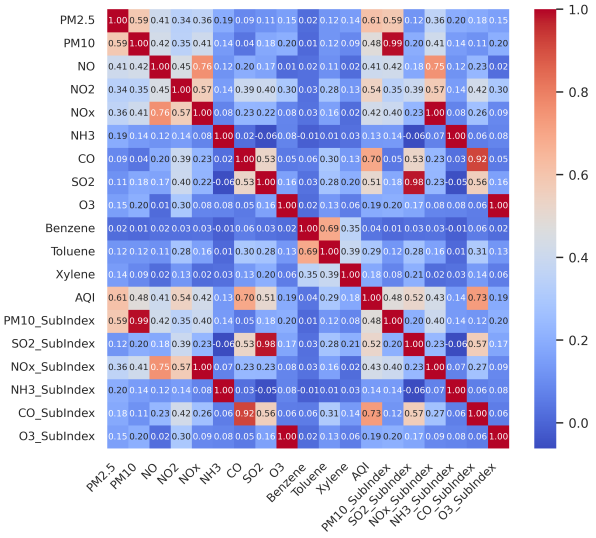


Fig. 6: Correlation Heat-Map matrix (provides dependency of pollutant and others)

E. Random Forest Model for AQI Prediction

Random forest [5], [17], [18] is a powerful ensemble-learning algorithm renowned for its ability to construct numerous decision trees during training and amalgamate their outputs for robust predictions. Employing bagging techniques, each tree is trained on a random subset of the data, mitigating overfitting and enhancing generalization. Furthermore, the algorithm introduces feature randomization by considering only a subset of features at each split, thereby augmenting the diversity among individual trees.

In classification tasks, the final prediction is determined through a majority vote; in regression, it relies on the average prediction. Notably, random forest excels in versatility, adeptly handling both classification and regression tasks with remarkable predictive accuracy. It demonstrates resilience to noisy data, shows less sensitivity to hyper-parameter tuning, and offers insights into feature importance, rendering it widely applicable across diverse domains. We opted for random forest over support vector machines and decision trees to predict the AQI index. The support vector machine was dismissed due to its higher complexity, computational expense, challenges with imbalanced data, and increased memory requirements. The decision tree was also ruled out because of its susceptibility to overfitting, tendency for high variance and bias, and perceived lack of reliability compared to random forest. Given the strengths of the random forest in handling AQI prediction, we now turn our attention to the Prophet library to explore its capabilities in time series forecasting, offering a comparative perspective on modeling efficiency and accuracy.

F. Prophet Library

Prophet, an open-source forecasting tool developed by Facebook, specializes in analyzing and predicting time series data characterized by patterns such as seasonality, holidays, and recurring events. Renowned for its simplicity and accuracy, Prophet requires minimal user input while delivering effective forecasts [19].

Key features of the Prophet library [10] include automatic detection of seasonality, the ability to incorporate custom holidays and special events, tunable parameters for adjusting the strength of seasonality and trend flexibility, and the capability to handle missing data and outliers. Prophet supports both additive and multiplicative components, providing flexibility in capturing different patterns in the data. Prophet requires input data in the 'ds' and 'y' columns of a DataFrame. The 'ds' column should follow the Pandas format, which is ideally HH:MM:SS for timestamps and YYYY-MM-DD for dates. The 'y' column should contain numeric values representing the measurement intended for forecasting.

The Prophet tool involves the following steps:

1) To determine the city with the highest level of pollution, the process begins with data cleaning, followed by populating the AQI and AQI_BUCKET columns and subsequently creating a graphical representation. As a result, Ahmedabad was identified as the most polluted city.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 09:33:23 UTC from IEEE Xplore. Restrictions apply.

2) Develop a model for forecasting air quality. The model was created using the random forest classifier.
3) Forecast future AQI by analyzing past data trends, forecasting for 365 days was performed using this method.

G. Forecasting the AQI index

We used Prophet to estimate future values of the AQI. Prophet requires input data to be organized into a DataFrame with the columns 'ds' and 'y.' The 'ds' column indicates date stamps, which should follow the Pandas format—ideally HH:MM:SS for timestamps and YYYY-MM-DD for dates. The 'y' column must be numeric and represents the measurement targeted for forecasting.

IV. Results

This section presents the outcomes of our predictive modeling efforts, specifically measuring the random forest model's performance using established performance metrics. Each metric, selected for its relevance to AQI prediction goals, offers a unique perspective on the model's predictive accuracy and reliability.

A. Evaluation Metrics

The random forest model's performance in predicting air pollutant concentrations was evaluated using statistical metrics such as accuracy, precision, and recall. As Table II shows, the model achieved perfect precision, recall, and F1-scores of 1.00 across most classes, demonstrating its exceptional ability to accurately classify different categories.

1) Accuracy Score: The accuracy score measures the proportion of total predictions that the model classified correctly. It is computed by:

$$\text{Accuracy Score} = \frac{\text{NS}}{\text{TN}} \tag{1}$$

Where:

NS : Number of correctly classified samples

TN : Total number of samples

The accuracy score of the samples was 99.57%.

2) Precision: It is characterized as the proportion of genuine positive predictions to all instances (true positives and false positives) that are anticipated to be positive. It provides information about the model's ability to correctly identify positive examples while reducing the number of negative instances that are incorrectly classified.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

In the precision formula shown in (2), TP represents the number of True Positives, which are instances correctly identified as positive by the model. FP stands for False Positives, referring to instances that the model incorrectly classified as positive when they are actually negative.

3) Recall: It is a performance metric in machine learning that evaluates a classification model's ability to recognize all significant instances of a given class accurately. Also known as the true positive rate or sensitivity, it measures the proportion of relevant cases that are correctly identified.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

In (3), TP represents the number of True Positives, and FN stands for False Negatives. FN represent the number of instances where the model incorrectly classifies an actual positive case as negative, thus failing to detect a condition that is truly present.

TABLE II: Classification Metrics

Class	Precision	Recall	F1-Score	Support
0	1.00	0.99	1.00	245
1	1.00	1.00	1.00	1639
2	1.00	1.00	1.00	1998
3	1.00	1.00	1.00	904
4	1.00	1.00	1.00	780
5	1.00	1.00	1.00	341

B. Summary of Performance Metrics

Overall, the random forest algorithm achieved an accuracy score of 99.57%, underscoring the model's robust performance in predicting air quality indices.

C. Time Series Decomposition

Forecasts are produced for uni variate time series datasets using an open-source tool called Prophet. It is straightforward to use and designed to automatically determine a reasonable set of model hyper-parameters, with the goal of generating reliable forecasts for data that by default has trends and seasonal structure.

Fig. 7 is created with the Prophet tool. It provides the capacity to forecast future patterns. Using data from the preceding seven years, the program forecasted Delhi's trend for the ensuing 365 days.

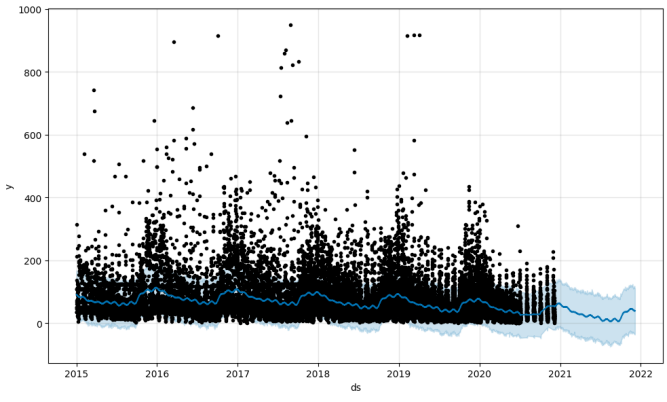


Fig. 7: Observed vs. predicted AQI values using the Random Forest model, with a trend line and confidence intervals for forecasted data points.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 09:33:23 UTC from IEEE Xplore. Restrictions apply.

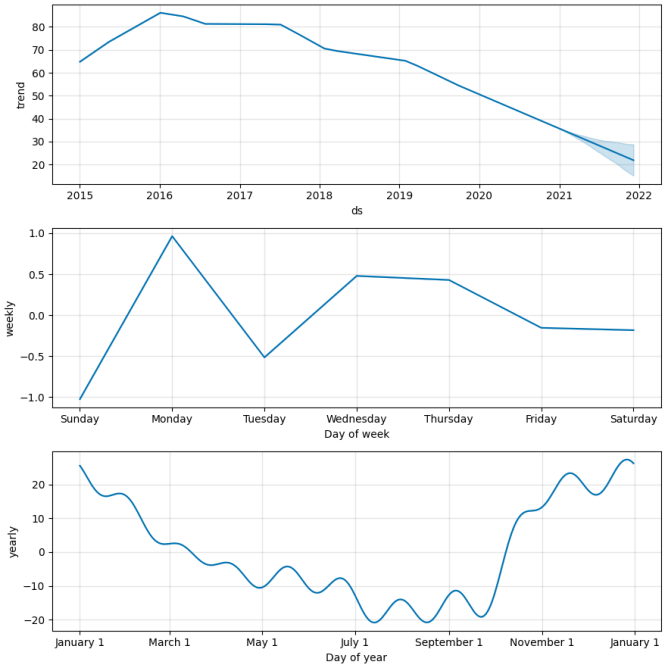


Fig. 8: Trend and seasonality analysis from the random forest model, showing the yearly trend, weekly patterns, and yearly seasonality in the AQI predictions.

As Fig. 8 shows, using plot-components, the yearly and weekly trend of the AQI with respect to time. The initial plot indicates that the annual fluctuation in the AQI is decreasing. The second plot makes clear that there is a noticeable increase in pollution at the beginning of the week and a decrease at the conclusion of the week. The final one emphasizes that, in comparison to other months, the AQI is rather low from July through September.

To ensure our model's robustness, additional validation steps are essential. Investigating different imputation strategies, including multiple imputation or k-nearest neighbors (KNN), could improve our AQI forecasts [20]. Comparing our model's output with results from advanced methods like SVM or Neural Networks will further validate its efficacy. These measures are critical for confirming the model's practical utility in air quality management.

Conclusion

The composition of the atmosphere, comprising gasses and particulate matter, significantly influences its quality. These contaminants adversely affect air quality and can lead to serious health complications with prolonged exposure. By examining the mean data over the previous five years, we may assess which urban areas require assistance in reducing their pollution levels by referring to the study's results. Furthermore, the predictive model can also be used to estimate future AQI for cities other than Delhi. This combined understanding empowers individuals to actively contribute to addressing pollution, a pressing modern day concern. The research in air quality monitoring holds substantial potential, as it involves predicting the current AQI and forecasting its future values. This approach allows for proactive measures to improve the AQI immediately, enhancing the overall quality of air from the present onward.

References

[1] A. S. Nagpure, B. R. Gurjar, and J. C. Martel, "Human health risks in national capital territory of delhi due to air pollution," *Atmospheric Pollution Research*, vol. 5, no. 3, pp. 371–380, 2014.

[2] P. K. Hopke, "Contemporary threats and air pollution," *Atmospheric Environment*, vol. 43, no. 1, pp. 87–93, 2009.

[3] M. Sarmadi, S. Rahimi, M. Rezaei, D. Sanaei, and M. Dianatinasab, "Air quality index variation before and after the onset of covid-19 pandemic: a comprehensive study on 87 capital, industrial and polluted cities of the world," *Environmental Sciences Europe*, vol. 33, no. 1, p. 134, 2021.

[4] M. C. S. Gadekar, "Air quality index (aqi) basics," *International Journal of Research Publication and Reviews*, vol. 2582, p. 7421, 2022.

[5] T. Madan, S. Sagar, and D. Virmani, "Air quality prediction using machine learning algorithms – a review," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 140–145, IEEE, 2020.

[6] M. D. et al., "Machine learning algorithms for air pollutants forecasting," in *2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, pp. 109–113, IEEE, 2020.

[7] V. R. Pasupuleti, P. Kalyan, and H. K. Reddy, "Air quality prediction of data log by machine learning," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 1395–1399, IEEE, 2020.

[8] S. Yarragunta and M. A. Nabi, "Prediction of air pollutants using supervised machine learning," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1633–1640, IEEE, 2021.

[9] C. Srivastava, S. Singh, and A. P. Singh, "Estimation of air pollution in delhi using machine learning techniques," in *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 304–309, IEEE, 2018.

[10] E. Zunic, K. Korjenic, K. Hodzic, and D. Donko, "Application of facebook's prophet algorithm for successful sales forecasting based on real-world data," *International Journal of Computer Science and Information Technology*, vol. 12, pp. 23–36, Apr. 2020.

[11] Z. Z. Oo and P. Sabai, "Time series prediction based on facebook prophet: a case study, temperature forecasting in myintkyina," *International Journal of Applied Mathematics Electronics and Computers*, vol. 8, no. 4, pp. 263–267, 2020.

[12] , "Air quality data in india (2015 - 2020)." https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india, 2024. Accessed: Apr. 22, 2024.

[13] M. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, p. 3021, Apr 2021.

[14] P. Lemenkova, "Python libraries matplotlib, seaborn and pandas for visualization geo-spatial datasets generated by qgis," *Analele Stiintifice ale Universitatii "Alexandru Ioan Cuza" din Iasi-seria Geografie*, vol. 64, no. 1, pp. 13–32, 2020.

[15] F. Nelli, *Python Data Analytics: Data Analysis and Science Using PANDAs, Matplotlib and the Python Programming Language*. Apress, 2015. Accessed: Apr. 22, 2024.

[16] W. McKinney, "Pandas, python data analysis library." URL http://pandas.pydata.org, 2015.

[17] V. M. Madhuri, G. G. Samyama, and S. Kamalapurkar, "Air pollution prediction using machine learning supervised learning approach," *International Journal of Science and Technology Research*, vol. 9, no. 4, pp. 118–123, 2020.

[18] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, and G. Arulkumaran, "Prediction of air quality index using machine learning techniques: a comparative analysis," *Journal of Environmental and Public Health*, vol. 2023, pp. 1–26, 2023.

[19] P. Aggarwal and S. Jain, "Impact of air pollutants from surface transport sources on human health: A modeling and epidemiological approach," *Environment International*, vol. 83, pp. 146–157, 2015.

[20] E.-L. Silva-Ramírez, R. Pino-Mejías, and M. López-Coello, "Single imputation with multilayer perceptron and multiple imputation combining multilayer perceptron and k-nearest neighbours for monotone patterns," *Applied Soft Computing*, vol. 29, pp. 65–74, 2015.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 09:33:23 UTC from IEEE Xplore. Restrictions apply.

Uncorrected Proof

Journal of
Water & Climate Change

© 2024 The Authors

Journal of Water and Climate Change Vol 00 No 0, 1 doi: 10.2166/wcc.2024.077

Prediction of inlet-to-outlet width ratio of type-A piano key weir using fuzzy neural network (FNN)

Nipun Bansal[a], Keshav Bhardwaj[a], Deepak Singh [b,*] and Munendra Kumar[b]

[a] Department of Computer Engineering, Delhi Technological University, New Delhi 110042, India
[b] Department of Civil Engineering, Delhi Technological University, New Delhi 110042, India
*Corresponding author. E-mail: kumaindeepaksingh@gmail.com

DS, 0000-0001-9397-3220

ABSTRACT

A piano key weir (PKW) is a nonlinear (labyrinth-type) weir with a small spillway footprint and a large discharge carrying capacity. It enables water bodies to continue functioning at elevated supply levels while causing no damage to dam structures, resulting in increased storage. PKW's geometrical structure is extremely complex, and geometrical aspects have a significant impact on its efficiency as well as on energy dissipation. Among them, relative width ratio (W_i/W_o) (i.e. inlet-to-outlet key width ratio) is a critical parameter that affects the PKW's discharge efficiency and energy dissipation across the weir significantly. In this article, a fuzzy neural network (FNN) is used to predict the width of the PKW's inlet-to-outlet key and understand the resulting hydraulic behaviours. The dataset used in this study was collected experimentally, which adds to the study's authenticity because it is not a conventional dataset. The model's performance is evaluated by the root mean square error (RMSE) and mean absolute error (MAE); both values are 0.0305 and 0.0222, respectively. According to the dataset, these scores tell the model's reliability as it is in the ideal range.

Key words: fuzzy neural network (FNN), hydraulic behaviours, inlet-to-outlet key width ratio, mean absolute error (MAE), root mean square error (RMSE)

HIGHLIGHTS

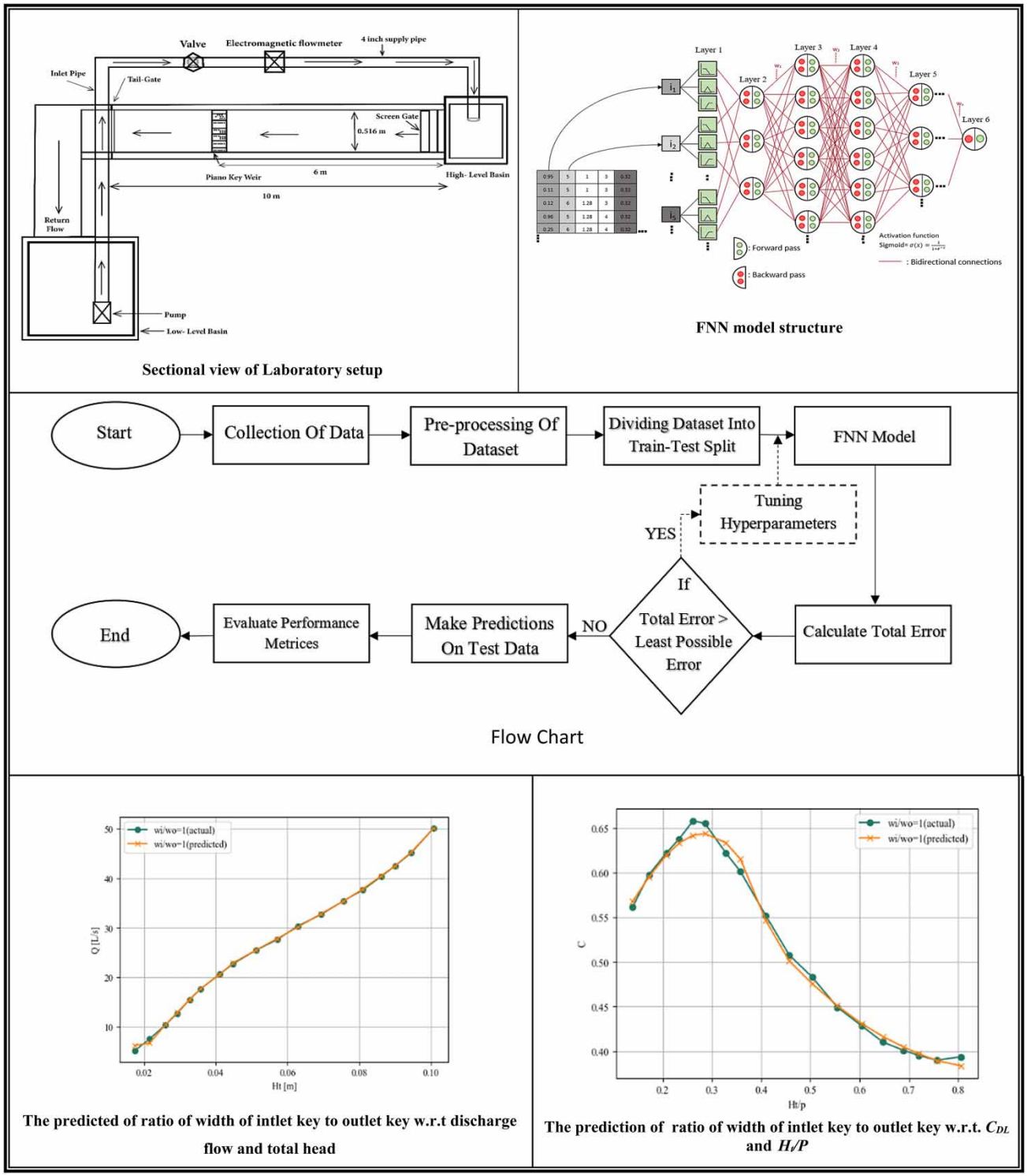
- A new fuzzy neural network (FNN) model is proposed to predict the inlet-to-outlet key width ratios of piano key weir.
- It is helpful to predict the other hydraulic behaviour of the weir such as energy dissipation.
- A thorough derivation of the tuning of the hyperparameters suggested in FNN is presented in this article.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

GRAPHICAL ABSTRACT



LIST OF SYMBOLS AND ABBREVIATIONS

P	PKW's height
L	Overall development length
W	Weir's width

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

H_t	Cumulative head over the weir
W_i	Width of inlet key
W_o	Width of outlet key
N	Cycle number
S_i	Slope of inlet key
S_o	Slope of outlet key
B_i	Length of inlet key
B_o	Length of outlet key
C_{DL}	Coefficient of discharge along developed crest length
Q	Discharge flow of PKW
E_L	Relative energy dissipation
E_1 or E_2	Energy dissipation in sections 1 and 2
MLPNN	Multi-level perceptron neural network
ANFIS	Adaptive Neuro-fuzzy Inference System
PSO	Particle swarm optimization
GA	Genetic algorithm
MFO	Moth flame optimization
FA	Factor analysis
SVR	Support vector machine
ANN	Artificial Neural Network
x	Input for the FNN model
w_i	Weights associated with ith layer in FNN model
e	Epsilon
x_i^o	Outputs associated with ith layer in FNN model
∂	Partial differentiation

1. INTRODUCTION

Piano key weir (PKW) is a spillway that is used in water management at dams. In dam engineering, there are height and space restrictions; at that given condition, PKW is a very effective management solution and is nonlinear in design. As the name suggests, it resembles piano keys, a structure where keys are grouped in a stepped manner to serve the purpose. When PKW is compared with traditional linear spillways, it has much better discharge capacity, energy dissipation, and lower material utilization. This is a very cutting-edge structure used to manage and regulate water flows in dams. PKW is an upgraded form of labyrinth weirs, which were first researched by and gained notoriety as a more effective and economical weir (Lempérière & Ouamane 2003). A recent project in France that uses PKW to boost the water release capacity of the country's current suction dams (Laugier *et al.* 2017) and as a promising solution for dam weir on derailed structures in Vietnam (Khanh 2017) is rapidly increasing. Since 2011, significant knowledge exchange during geographical conferences has guided the compilation of reference material summarizing the current state of PKW technology (Erpicum *et al.* 2011, 2013, 2017). All around the world, more than 35 PKWs have been successfully built, including in the UK, Algeria, South Africa, France, Switzerland, Vietnam, India, Australia, and Sri Lanka (Crookston *et al.* 2019).

The first PKWs were built in 2006 at the Goulours Dam in France and more recently at the Hazelmere Dam in South Africa. They significantly increased the possibility of reservoir and spillway release. An upcoming PKW implementation at Tzaneen Dam is in progress (van Deventer *et al.* 2015; Chemaly 2017). Notably, recent interest in PKW implementations extends to interior weir applications, complementing the primary focus on utilizing PKWs as lateral weirs for runoff release (Karimi *et al.* 2018). Over the last two decades, detailed summaries of PKW geometry and hydraulic behaviour have been presented by Abhash & Pandey (2020) and Singh & Kumar (2021). Weir altitude (P), relative advancement L/W ratio (crest length), (H_t/P) ratio, alveolar width (W_i/W_o), and cycle number (N) for a constant channel width are geometric parameters that have a significant impact on head-discharge efficiency and energy dissipation (where L is the overall development length, W is weir's width, H_t is cumulative head over the weir, and P is weir altitude). PKW behaviour is further influenced by the type of crest, lengths of overhangs, upstream apex walls, and raising the crest via a parapet wall (Anderson & Tullis 2013). Numerous investigations have explored scientific hypotheses, case studies, and original research on PKW energy loss and labyrinths (Bieri *et al.* 2011; Ribeiro *et al.* 2011; Khanh 2013). Silvestri *et al.* (2013) observed that in PKWs with stepped chutes, residual energy was low at the spillway toe and increased with release and spillway length. Al-Shukur & Al-Khafaji (2018) investigated how the PKW slope affected energy loss and showed that the proportion of energy dissipation decreased as the slope decreased. Computational techniques employing AI models such as MLPNN, ANFIS, PSO, GA, MFO, and FA have been used to define PKW

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

hydraulic behaviour based on geometrical parameters (Zounemat-Kermani & Mahdavi-Meymand 2019). Karbasi & Azamathulla (2016) used GEP for hydraulic jump aspects over a rough bed, comparing its performance with conventional AI approaches SVR and ANN. Many studies have tried to estimate how deep scour can occur along dikes, weirs, and piers that are partially submerged in water (Rashki *et al.* 2019; Pandey *et al.* 2020; Birbal *et al.* 2021; Ghasempour *et al.* 2021; Emadi *et al.* 2022; Singh *et al.* 2022). Therefore, more new ideas and rivals are needed to improve PKW technology. Bhukya *et al.* (2022) have published comprehensive reviews on PKWs for discharge measurement and power dissipation, highlighting the importance of conducting thorough experimental and computational studies to comprehend PKW hydraulic characteristics.

Amin *et al.* (2019) have provided valuable information on energy loss in free streams below rectangular sharp-crested weirs, emphasizing the importance of careful hydraulic system design to enhance efficiency and reduce negative effects on the environment and structure. This study aims to examine how the difference between the width of the air sacs at the entrance and the exit (W_i/W_o) influences the flow and energy loss of PKWs (PKWs). Previous studies have only tried to explore how the W_i/W_o ratio affects the discharge capacity, coefficient, and energy dissipation in PKWs. However, there is still a big lack of knowledge about this topic. Consequently, this experimental investigation was conducted to provide comprehensive insights into how the W_i/W_o ratio impacts hydraulic efficiency and downstream energy dissipation, mainly focusing on type-A PKWs. Pralong *et al.* (2011) indicated various geometrical parameters associated with PKWs (shown in Figure 1), which are frequently used in design analysis and provide sufficient information about the PKWs. The study utilizes a fuzzy neural network (FNN) to enhance our understanding of this critical PKW design and performance parameter. Poonia *et al.* (2021a, 2021b, 2021c) focus on the problem of resource distribution and ineffective response to disasters, which makes natural disasters worse. The aim is to resolve issues like poor coordination, late responses, and corruption among responsible agencies. The author proposes a step-by-step approach to tackle disaster management using blockchain technology; the first step is data management: a blockchain framework that ensures the timely sharing of trustable and traceable records related to various aspects of disasters. The second step is automated aid distribution, built on smart contracts to automate the fast transfer of emergency relief, prevent corruption in its use, and allocate it wisely.

Kumar *et al.* (2021) reviewed and understood the change in extreme circumstances to develop effective, durable strategies. The document identifies regions in India prone to droughts and floods to prepare in advance and reduce risk. The author offers a solid approach for assessing risks and boosting the resilience of essential infrastructure. The proposed approach first assesses the potential impact of extreme weather on critical infrastructure and then develops strategies that can withstand climate-related challenges. Tiwari *et al.* (2024) explore the relationship between climate change and water quality by focusing on the electrical conductivity in the Narmada basin. AI techniques can be used to predict water quality. The author develops 10 AI models to predict the EC levels at the Sandia station in the Narmada basin. This information can be used to manage the region's drinking water, irrigation, and other uses. Model 8 performs exceptionally well among all models, with an R^2 value of 0.889. Poonia *et al.* (2021a, 2021b, 2021c) addressed the lack of comprehensive analysis of the joint dependence of drought duration and its effect across Indian river basins. There is a research gap that hinders the development of effective drought strategies. The author suggests modelling the joint dependence structure of drought features using a bivariate copula-based technique. The author aims to identify patterns and hotspots of severe and prolonged drought events across India. Poonia *et al.* (2021a, 2021b, 2021c) aim to address the research gap related to the unpredictability of climate model projections

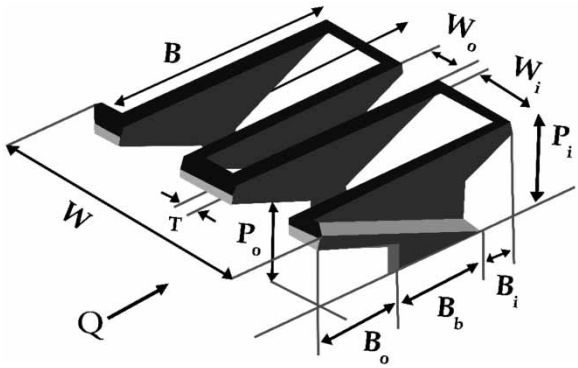


Figure 1 | Cross-sectional view of piano key weir (adapted from Pralong *et al.* (2011)).

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

and scenarios in assessing CWR and CIR in the Eastern region of Himalayan, specifically in Sikkim. Das *et al.* (2020) provide a thorough analysis of how climate change affects a Sikkim crop field, particularly rice, wheat, and maize. They use calibrated AquaCrop simulations based on historical data to project future crop yields under different emission scenarios and GCMs.

Additionally, they add possibility theory to examine uncertainty with GCMs. Jamal *et al.* (2024) aim to address the challenge and land loss in states like West Bengal and India. Existing systems for protecting riverbanks include different materials and methods to determine stone size and effectiveness under different temporal variations. The author identifies a lack of research finding the most effective stone for safeguarding against bank failure. The author suggests a way to test how a riverbank reacts to different sizes of stones using a physical model. Poonia & Tiwari (2020) aim to satisfy the demand for rainfall-runoff modelling in Madhya Pradesh's Hoshangabad Narmada River watershed. This area often experiences floods, so it is important to have accurate runoff estimation for good flood management. The author suggests using ANN-based models to simulate water flow in the Hoshangabad area. Different ways of measuring how well the models work is used, such as how close they are to the real data, how much error they have, and how much they improve the results. Kantharia *et al.* (2024) want to find out how well they can predict the amount of water that flows out of the Damanganga River daily by using information about the rain and the soil. The study's objective is to fill the gap by creating a model using ANFIS that is customized for the Damanganga basin, taking into account soil moisture at various depths as input variables. We measure how well the model fits the data by using two numbers: the R^2 and the NSE.

Although there is growing awareness of the urban heat island (UHI) phenomenon and its consequences, Sharma *et al.* (2023) discovered a scarcity of in-depth investigations on this topic, especially in regions with hot semi-arid climates, such as India. The author's model intends to fill the current knowledge gaps on UHI effects, and it employs remote sensing and geospatial analysis methods to examine how surface temperatures and UHI effects vary throughout the year, especially considering the influence of land use and water body cooling. Baudhanwala *et al.* (2024) discuss the necessity for accurate rainfall prediction models in regions like South Gujarat, India. Severe rainstorms are happening more often in the area, and this causes problems for water, farming, buildings, and people's lives. The writer utilizes four distinct machine learning (ML) techniques: support vector regression (SVR), multiple linear regression (MLR), decision tree (DT), and random forest (RF). The research suggests that a width ratio between 1.25 and 1.30 provides the best results in terms of hydraulic efficiency, as indicated by the highest discharge coefficient. Beyond this range, energy dissipation increases, yielding diminishing returns. This research offers valuable insights for PKW design and highlights the applicability of FNNs in hydraulic engineering, ultimately enhancing civil engineering and water resource management practices.

A real-world dataset and the suggested strategy were used to demonstrate the contribution. An overview of the contribution would be as follows:

- The dataset utilized in this research is experimentally obtained and used to evaluate proposed methodologies; it is not predefined. The dataset utilized in this work was obtained through experimentation. Unlike traditional methods, which rely on pre-defined datasets, this study offers a unique dataset, enhancing the data's validity.
- A new model named FNN is proposed; this method has never been used before. The research void in hydraulic behaviour pattern prediction for PKWs will be filled by proposing FNN.
- A thorough derivation of the tuning of the hyperparameters suggested in FNN is presented in this article. A crucial part of building a model is tuning; the contribution offers a systematic and precise way to adjust the parameters to the specifications. The derivation is novel and contributes to improving the model's performance.

The organization of the article is as follows: Section 2.1 talks about the experimental set-up and dataset used in the process, Section 2.2 shows the methodology used in the article and also the tuning of hyperparameters, then further in Section 3, experimental results are discussed and in Section 4, conclusion is given.

2. MATERIAL AND METHODS

2.1. Experimental set-up and model fabrication

In this study, Singh & Kumar (2023a, 2023b) publication of the PKW's width ratio data is used. They performed a series of laboratory experiments using a horizontal flume of a rectangular cross-section of width 0.516 m, height 0.6 m, and length 10 m. The discharge flow rate was measured with an electromagnetic flowmeter within an accuracy limit of $\pm 0.2\%$. A metal monitor gate was incorporated into the flume headbox to enhance upstream approach flow uniformity. The average flow velocity was determined using an acoustic Doppler velocimeter (ADV). The models were constructed from 8-mm-

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

thick transparent acrylic sheets, assembled using chloroform, shown in Figure 2. Detailed specifications of geometrical parameters and collected data during the research are outlined in Table 1. Where, W_i is the width of the input key, W_o is the width of the outlet key, L is the developed crest length, W states the width of the channel of PKW, H_t is the total head, Q is the discharge flow of PKW, P is the height of PKW, S_i is the slope of the inlet key, S_o is the slope of the outlet key, B_i is the length of the inlet key, B_o is the length of the outlet key, E_L is relative energy dissipation, and E_1 or E_2 is energy dissipation at a particular section.

2.2. Methodology

This article proposes a novel FNN. The biological neurons inspire neurons in ML. Pattern recognition and prediction are required in many situations where automation is challenging. Neural networks (NNs) were developed to address these issues, as they can automate the process and provide accurate data predictions. NN processes the input along with the

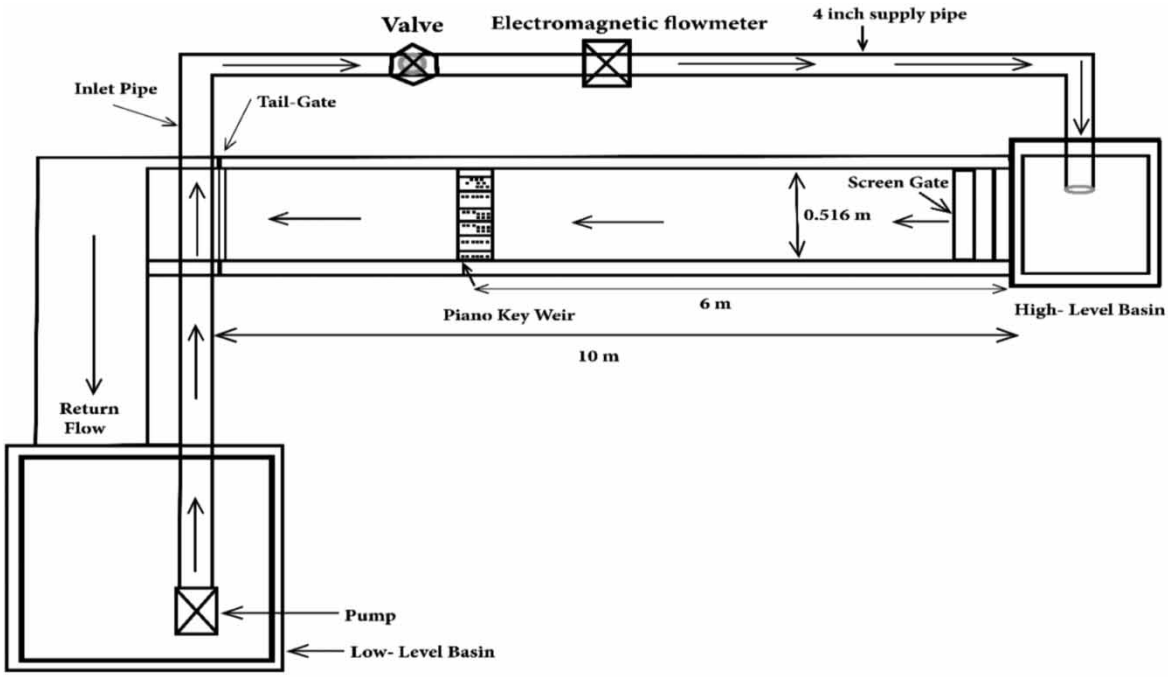


Figure 2 | Schematic plan view of experimental set-up.

Table 1 | Range of data collected in the present study

S. No.	W_i/W_o	L/W	$S_i = S_o$	H_t (m)	Q (L/s)	$B_i/P = B_o/P$	Range of E_L/E_1	Range of E_2/E_1	No. of readings
1	1.00	5	1.08	0.0300–0.0971	10.17–50.26	0.69	0.8093–0.1930	0.1907–0.8096	18
2	1.10	5	1.08	0.0304–0.0986	10.14–50.26	0.69	0.7860–0.1785	0.2140–0.8214	18
3	1.20	5	1.08	0.0307–0.0989	10.19–50.07	0.69	0.7700–0.1731	0.2300–0.8268	18
4	1.25	5	1.08	0.0317–0.1011	10.28–50.18	0.69	0.7533–0.1734	0.2467–0.8265	18
5	1.30	5	1.08	0.0322–0.1004	10.09–50.00	0.69	0.7356–0.1655	0.2644–0.8344	18
6	1.35	5	1.08	0.0310–0.0891	10.16–50.07	0.69	0.7297–0.1501	0.2703–0.8498	18
7	1.40	5	1.08	0.0303–0.0985	10.19–50.13	0.69	0.7031–0.1435	0.2969–0.8564	18
8	1.50	5	1.08	0.0310–0.0992	10.15–50.45	0.69	0.6822–0.1411	0.3118–0.8588	18
9	2.00	5	1.08	0.0313–0.0995	10.29–49.82	0.69	0.6518–0.1342	0.3482–0.8657	18

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

hyperparameters. A fuzzy system is a logic that deals with binary values, divides the data based on membership function, and gives values as a degree of truth. Equations are given a more human touch through the use of fuzzy logic.

The FNN model combines the benefits of fuzzy logic and NNs. Figure 3 illustrates the FNN model employed in the process. The model consists of six layers, including the first layer. Once the data has been gathered, it is prepared for the FNN model by going through a pre-processing stage. Before we can use ML, we need to prepare the data, which improves its quality and makes it suitable for our purposes. This research involved the elimination of redundant data, the correction of inconsistencies, the normalization of data, and the management of missing values. Initially, the dataset was examined for missing values, which could impact the analysis outcome if not appropriately handled. Data loss has different causes, such as mistakes in data input or faulty devices. The presence of repeated redundant data, such as rows or columns, was resolved earlier. The same data can influence outcomes and create prejudices. Data are normalized and scaled to ensure features are equally available for analysis.

Data standardization is essential when evaluating features with different units or scales. There is an input layer where we retrieve data in the form of $x(t) = \{i_1(t), i_2(t), i_3(t), i_4(t), \ldots, i_n(t)\}$ from the dataset, where n is the total number of samples and t shows the iteration number. The fuzzification layer is the first layer of the model after the input layer. The membership function used in the fuzzification layer is the trapezoidal membership function, and a graphical and mathematical representation of this function is shown in Figure 4. Where a_1, b_1, a_2, b_2, a_3, and b_3 are the inputs of randomly generated parameters, and x is provided to the fuzzification layer.

Layer 1 of Figure 3 illustrates how each input is divided into three components following the fuzzification layer: $A_1, A_2,$ and A_3. The production layer, the second layer in the model, takes the output of the first layer as input and outputs as x_1, combining the fuzzy components from the membership function's output. $x_1^1 = A_1^1 \cdot A_1^2 \cdots A_1^n$, where n is the number of inputs in one iteration, subscript shows the number of the layers, and superscript shows the number of nodes in the present layer; as the membership function has three components, the production layer will have only three nodes combining the inputs. Similarly, we can calculate x_1^2 and x_1^3. Subsequently, we will employ a membership function, namely the sigmoid function, on the output of the second layer and all subsequent layers. The sigmoid function is a membership function that helps produce output from 0 to 1, making it easy to handle outputs and then further process them. The sigmoid membership

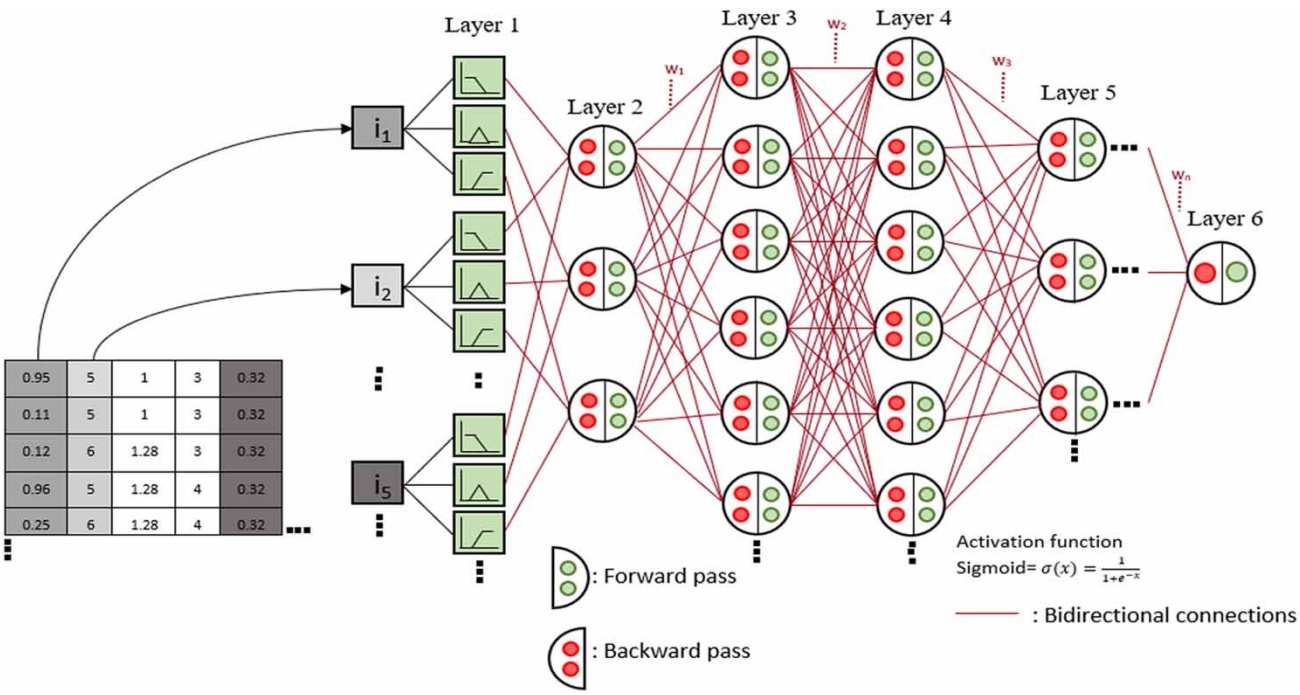


Figure 3 | Fuzzy neural network model.

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

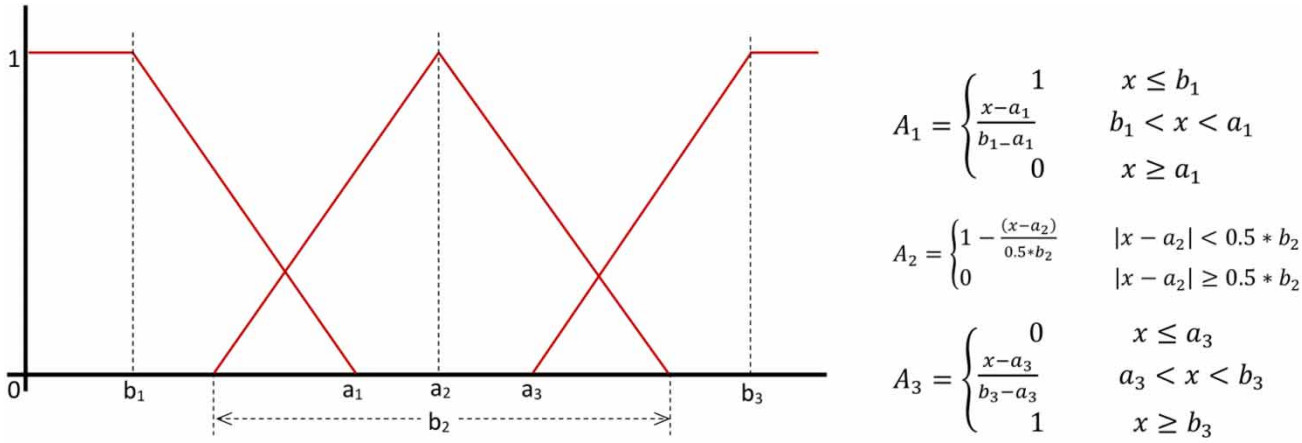


Figure 4 | Trapezoidal membership function.

function is shown in Equation (1):

$$x_1^o = \frac{1}{1+e^{-x_1}} \tag{1}$$

where x_1 denotes the output of the first layer, which is fed to the sigmoid function and x_1^o is the layer's final output will be sent as input to the following layer. Further, this output of the second layer is passed to the third layer, a neural layer in FNN. w_1 is the weight for this neural layer, and it will first be produced at random and will give output as x_2, in generic form, it will be given as follows:

$$x_2 = f(x_1^o, w_1) \tag{2}$$

Then, the output is passed through the sigmoid membership function in the second layer, and it can be presented as a similar notation shown in Equation (3):

$$x_2^o = \frac{1}{1+e^{-x_2}} \tag{3}$$

Comparably, the function for the fourth layer is displayed in Equation (4). It accepts input as x_2^o and handles weights as w_2 to produce an output of x_3.

The output of the sigmoid function is then passed to the final layer, the network's last layer, x_3^o is the final output of the layer.

$$x_3 = f(x_2^o, w_2) \tag{4}$$

Equation (5) shows the function for the fifth layer. It can handle weights as w_3 and take in input x_3^o to produce an output x_4. The output of the sigmoid function is then passed through the layer, and the final result is x_4^o.

$$x_4 = f(x_3^o, w_3) \tag{5}$$

Equation (6) shows how the fifth layer works. The input x_4^o is processed and the weights w_4 are used to generate an output x_4. The layer's last result is x_5^o. The result of this process is then passed through the sigmoid function.

$$x_5 = f(x_4^o, w_4) \tag{6}$$

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

The last result of the FNN will be written as x_5^o, shown in Equation (7):

$$x_5^o = \frac{1}{1 + e^{-x_5}} \tag{7}$$

Because the initial parameters were randomly generated, it is reasonable to expect that there will be mistakes in our output after the feed-forward process. The following section explains adjusting the settings to make the error as small as possible. The flow of data is presented in the flowchart, as shown in Figure 5.

2.2.1. Tuning of hyperparameters

Let's denote the outcome of FNN with 'z' which equals x_5^o in Equation (7). The output expected after the prediction is 't'. The mean square error (MSE) for FNN is shown in Equation (8):

$$E = \frac{1}{2n}\sum_{i=1}^{n}(z - t)^2 \tag{8}$$

where 'n' is the total number of samples of input. The backtracking process is depicted in generalized form; for example, the value of n is taken as 1 for simplifying the derivations. The error function in Equation (8) will give the total error of the model, and it will be minimized using the process of gradient descent. Gradient descent is an algorithm where, with each iteration, we try to update the parameters, which are weights in this case, to minimize the error.

2.2.2. Layer six

According to the gradient descent algorithm, to minimize the error value in output, the partial differentiation of the error function is required w.r.t. parameters, i.e. weights of the respective layer. The deduction of the partial differentiation for the sixth layer is shown in Equation (9):

$$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial x_5^o} \cdot \frac{\partial x_5^o}{\partial x_5} \cdot \frac{\partial x_5}{\partial w_4} \tag{9}$$

where $\partial E/\partial x_5^o$ is the partial differentiation of error function (E) w.r.t. x_5^o, which is the output of the last layer shown in Equation (10):

$$\frac{\partial E}{\partial x_5^o} = (x_5^o - t) \tag{10}$$

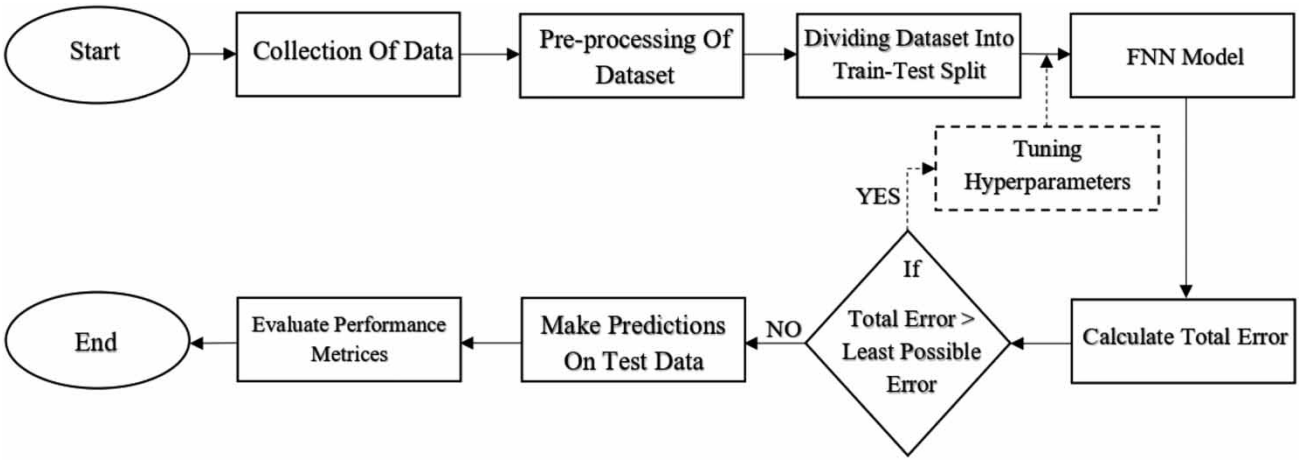


Figure 5 | Flowchart of the FNN model.

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

where $\partial x_5^o/\partial x_5$ is the partial differentiation of the output of a layer after the sigmoid membership function w.r.t. output of the layer without the membership function, shown in Equation (11) and also given a generalized notation, i.e. σ_5. Every layer will repeat this step, making it easy to calculate and donate it with a generalized term.

$$\frac{\partial x_5^o}{\partial x_5} = \frac{e^{-x_5}}{(1+e^{-x_5})^2} = \sigma_5 \tag{11}$$

where $\partial x_5/\partial w_4$ is the partial differentiation of x_5 which is the output without membership function w.r.t. w_4. As shown in Equation (12), the weights between the present and previous layers must be updated to achieve a lower error value.

$$\frac{\partial x_5}{\partial w_4} = f'(x_4^o, w_4) \cdot x_4 \tag{12}$$

In Equation (12), $f'(x_n^o, w_n)$ is the denotation for differentiated function which has final outputs of the previous layers as inputs and respective weights. Now combine the partial differentiations from Equations (10) to (12), and put together in Equation (9):

$$\frac{\partial E}{\partial w_4} = (x_5^o - t) \cdot \sigma_5 \cdot f'(x_4^o, w_4) \cdot x_4^o \tag{13}$$

Some general values can be clubbed together to simplify the terms, repeating further in derivation. Δ_5 is assumed notation for the ease of calculations as shown in Equation (14):

$$\Delta_5 = (x_5^o - t) \cdot \sigma_5 \cdot f'(x_4^o, w_4) \tag{14}$$

Now, putting the simplified assumed notation from Equation (14) to the expanded Equation (13):

$$\frac{\partial E}{\partial w_4} = \Delta_5 \cdot x_4^o \tag{15}$$

After getting the partial differentiation of error function w.r.t. respective weight, the weight for the respective layer can be updated using Equation (16):

$$\text{updated_}w_4 = w_4 - \alpha \frac{\partial E}{\partial w_4} \tag{16}$$

where α is the learning rate of the model, and the learning rate is the tuning parameter in the optimization process, which regulates the model's parameter changes in steps that lead to the lowest possible error value.

2.2.3. Layer five

To update the weights of this layer, i.e. w_{3-}, the partial differentiation of E w.r.t. w_3 is needed, deduced in Equation (17):

$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial x_5^o} \cdot \frac{\partial x_5^o}{\partial x_5} \cdot \frac{\partial x_5}{\partial x_4^o} \cdot \frac{\partial x_4^o}{\partial x_4} \cdot \frac{\partial x_4}{\partial w_3} \tag{17}$$

Here are values for $\partial E/\partial x_5^o$ and $\partial x_5^o/\partial x_5$ are available in previous equations, i.e. Equations (10) and (11). $\partial x_5/\partial x_4^o$ is the partial differentiation of the sixth layer w.r.t. output to the final output of the fifth layer in FNN, shown in Equation (18):

$$\frac{\partial x_5}{\partial x_4^o} = f'(x_4^o, w_4) \cdot w_4 \tag{18}$$

$\partial x_4^o/\partial x_4$ is the partial differentiation of the final output of the fifth layer after the membership function w.r.t. the output of the layer without membership function. It is very similar to Equation (11). As it is the differentiation of the sigmoid function only,

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

the generic notation continues, as shown in Equation (19):

$$\frac{\partial x_4^o}{\partial x_4} = \sigma_4 = \frac{e^{-x_4}}{(1+e^{-x_4})^2} \tag{19}$$

$\partial x_4/\partial w_3$ is the partial differentiation of the output of the respective layer w.r.t. weights between the respective layer and the previous layer, shown in Equation (20):

$$\frac{\partial x_4}{\partial w_3} = f'(x_3^o, w_3) \cdot x_3^o \tag{20}$$

Putting deduced Equations (18)–(20) and previously calculated Equations (10) and (11) in Equation (17) for an expanded form, which is shown in Equation (21):

$$\frac{\partial E}{\partial w_3} = (x_5^o - t) \cdot \sigma_5 \cdot f'(x_4^o, w_4) \cdot w_4 \cdot \sigma_4 \cdot f'(x_3^o, w_3) \cdot x_3^o \tag{21}$$

There are some repetitive terms that generic assumed notations can replace. So, using Equation (14) in Equation (21).

$$\frac{\partial E}{\partial w_3} = \Delta_5 \cdot w_4 \cdot \sigma_4 \cdot f'(x_3^o, w_3) \cdot x_3^o \tag{22}$$

Some general values can be clubbed together to simplify the terms, repeating further in derivation. Δ_4 is assumed notation for the ease of calculations as shown in Equation (23).

$$\Delta_4 = \Delta_5 \cdot w_4 \cdot \sigma_4 \cdot f'(x_3^o, w_3) \tag{23}$$

After all the deductions and using all generalized notations, the final equation for the partial differentiation of E w.r.t. w_3 is shown in Equation (24):

$$\frac{\partial E}{\partial w_3} = \Delta_4 \cdot x_3^o \tag{24}$$

According to the gradient descent algorithm, the weights for this layer will be updated, as shown in Equation (25):

$$\text{updated_}w_3 = w_3 - \alpha \frac{\partial E}{\partial w_3} \tag{25}$$

2.2.4. Layer four

To update the respective parameters if this layer is w_2, partial differentiation of E w.r.t. w_2 is needed and deduced in Equation (26):

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial x_5^o} \cdot \frac{\partial x_5^o}{\partial x_5} \cdot \frac{\partial x_5}{\partial x_4^o} \cdot \frac{\partial x_4^o}{\partial x_4} \cdot \frac{\partial x_4}{\partial x_3^o} \cdot \frac{\partial x_3^o}{\partial x_3} \cdot \frac{\partial x_3}{\partial w_2} \tag{26}$$

Here, we can directly take values from Equation (10), (11), (18), and (19). The rest of the values are creating a pattern and repeating themselves. So, the expanded form of Equation (26) is shown in Equation (27):

$$\frac{\partial E}{\partial w_2} = (x_5^o - t) \cdot \sigma_5 \cdot f'(x_4^o, w_4) \cdot w_4 \cdot \sigma_4 \cdot f'(x_3^o, w_3) \cdot w_3 \cdot \sigma_3 \cdot f'(x_2^o, w_2) \cdot x_2^o \tag{27}$$

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Using Equation (14) in Equation (27) for simplification of the equation:

$$\frac{\partial E}{\partial w_2} = \Delta_5 \cdot w_4 \cdot \sigma_4 \cdot f'(x_3^o, w_3) \cdot w_3 \cdot \sigma_3 \cdot f'(x_2^o, w_2) \cdot x_2^o \tag{28}$$

Using Equation (23) in Equation (28),

$$\frac{\partial E}{\partial w_2} = \Delta_4 \cdot w_3 \cdot \sigma_3 \cdot f'(x_2^o, w_2) \cdot x_2^o \tag{29}$$

From the above calculations, Δ_3 can be computed in similar pattern and is shown in Equation (30):

$$\Delta_3 = \Delta_4 \cdot w_3 \cdot \sigma_3 \cdot f'(x_2^o, w_2) \tag{30}$$

Using Equation (30) in Equation (29) to simplify and find the final equation of the partial differentiation of E w.r.t. w_2:

$$\frac{\partial E}{\partial w_2} = \Delta_3 \cdot x_2^o \tag{31}$$

After getting the final equation in Equation (31), updated weight can be calculated as shown in Equation (32):

$$\text{updated_}w_2 = w_2 - \alpha \frac{\partial E}{\partial w_2} \tag{32}$$

2.2.5. Layer three

From previous findings and generic notations, the partial differentiation of E w.r.t w_1 is deduced in Equation (34) from Equation (33) easily.

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial x_3} \cdot \frac{\partial x_3}{\partial x_2^o} \cdot \frac{\partial x_2^o}{\partial x_2} \cdot \frac{\partial x_2}{\partial w_1} \tag{33}$$

$$\frac{\partial E}{\partial w_1} = \Delta_3 \cdot w_2 \cdot \sigma_2 \cdot f'(x_1^o, w_1) \cdot x_1^o \tag{34}$$

From Equation (34) and previous patterns, Δ_2 can be computed easily and is shown in Equation (35):

$$\Delta_2 = \Delta_3 \cdot w_2 \cdot \sigma_2 \cdot f'(x_1^o, w_1) \tag{35}$$

Putting Equation (35) in Equation (34),

$$\frac{\partial E}{\partial w_1} = \Delta_2 \cdot x_1^o \tag{36}$$

Weights can be updated easily after getting the final equation for the partial differentiation, as shown in Equation (37):

$$\text{updated_}w_1 = w_1 - \alpha \frac{\partial E}{\partial w_1} \tag{37}$$

2.2.6. Layer two

This layer is the production layer, which combines all the components of the fuzzification layer. This particular layer has no variable weights; hence, updating the parameters is impossible.

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

2.2.7. Layer one

In the first layer, while forward pass, the trapezoidal function is used for the fuzzification shown in Figure 4. Firstly, the partial differentiation of the A_1, A_2, and A_3 will be taken on the bases of the respective variables a_1, b_1, a_2, b_2, a_3, and b_3 which is shown in equations from Equations (38) to (43):

$$\frac{\partial A_1}{\partial a_1} = \frac{(x - b_1)}{(b_1 - a_1)^2} \tag{38}$$

where a_1 and b_1 are the hyperparameters for the first component of fuzzy logic, i.e. A_1. And Equation (38) shows the partial differentiation of component A_1 w.r.t. variable a_1.

$$\frac{\partial A_1}{\partial b_1} = \frac{(a_1 - x)}{(b_1 - a_1)^2} \tag{39}$$

where the hyperparameters for the first fuzzy logic component or A_1 are a_1 and b_1. Additionally, Equation (39) illustrates component A_1's partial differentiation with respect to the variable b_1.

$$\frac{\partial A_2}{\partial a_2} = \frac{1}{0.5 \cdot b_2} \tag{40}$$

where a_2 and b_2 are the hyperparameters for the second fuzzy logic component or A_2. Furthermore, the partial differentiation of component A_2 with respect to the variable a_2 is shown in Equation (40).

$$\frac{\partial A_2}{\partial b_2} = \frac{(x - a_2)}{0.5 \cdot b_2^2} \tag{41}$$

where the hyperparameters for the second fuzzy logic component or A_2 are a_2 and b_2. Moreover, Equation (41) illustrates the partial differentiation of component A_2 with respect to the variable b_2.

$$\frac{\partial A_3}{\partial a_3} = \frac{(x - b_3)}{(b_3 - a_3)^2} \tag{42}$$

where a_3 and b_3 are the hyperparameters for the third fuzzy logic component or A_3. Additionally, Equation (42) shows how a component A_3 is partially differentiated with respect to the variable a_3:

$$\frac{\partial A_3}{\partial b_3} = \frac{(a_3 - x)}{(b_3 - a_3)^2} \tag{43}$$

where the hyperparameters for the third fuzzy logic component or A_3 are a_3 and b_3. Furthermore, Equation (43) illustrates the partial differentiation of component A_3 with respect to the variable b_3.

Further, to update all six variables, differentiation of the respective E w.r.t variables is needed, as shown in the equations from Equations (44) to (52).

$$\frac{\partial E}{\partial a_1} = \frac{\partial E}{\partial x_2} \cdot \frac{\partial x_2}{\partial x_1^o} \cdot \frac{\partial x_1^o}{\partial x_1} \cdot \frac{\partial x_1}{\partial A_1} \cdot \frac{\partial A_1}{\partial a_1} \tag{44}$$

Equation (44) illustrates the partial differentiation of the error function with respect to the fuzzy parameter a_1.

$$\frac{\partial E}{\partial a_1} = \Delta_2 \cdot w_1 \cdot \sigma_1 \cdot 1 \cdot \frac{(x - b_1)}{(b_1 - a_1)^2} \tag{45}$$

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

In Equation (45), the parameter Δ_2 is from Equation (35). w_1 is the parameter related to layers 2 and 3 of the FNN as seen in Figure 3, σ_1 is the partial differentiation of the sigmoid function utilized in layer 3, a_1 and b_1 are the parameters related to fuzzy logic, and x is the input provided to the FNN model.

Similarly, the remaining values will also be calculated as

$$\frac{\partial E}{\partial b_1} = \Delta_2 \cdot w_1 \cdot \sigma_1 \cdot 1 \cdot \frac{(a_1 - x)}{(b_1 - a_1)^2} \tag{46}$$

The partial differentiation of the error function with respect to the fuzzy parameter b_1 is shown in Equation (46).

$$\frac{\partial E}{\partial a_2} = \Delta_2 \cdot w_1 \cdot \ \sigma_1 \cdot 1 \cdot \frac{1}{0.5 \cdot b_2} \tag{47}$$

Equation (47) displays the partial differentiation of the error function with respect to the fuzzy parameter a_2.

$$\frac{\partial E}{\partial b_2} = \Delta_2 \cdot w_1 \cdot \ \sigma_1 \cdot 1 \cdot \frac{(x - a_2)}{0.5 \cdot b_2^2} \tag{48}$$

The partial differentiation of the error function with respect to the fuzzy parameter b_2 is shown in Equation (48).

$$\frac{\partial E}{\partial a_3} = \Delta_2 \cdot w_1 \cdot \ \sigma_1 \cdot 1 \cdot \frac{(x - b_3)}{(b_3 - a_3)^2} \tag{49}$$

Equation (49) displays the partial differentiation of the error function with respect to the fuzzy parameter a_3.

$$\frac{\partial E}{\partial b_3} = \Delta_2 \cdot w_1 \cdot \ \sigma_1 \cdot 1 \cdot \frac{(a_3 - x)}{(b_3 - a_3)^2} \tag{50}$$

The partial differentiation of the error function with respect to the fuzzy parameter b_3 is shown in Equation (50). These values can be updated easily with the help of general terms in Equations (51) and (52).

$$\text{updated_}a_i = a_i \ - \ \alpha \frac{\partial E}{\partial a_i} \tag{51}$$

$$\text{updated_}b_i = b_i \ - \ \alpha \frac{\partial E}{\partial b_i} \tag{52}$$

From this complete hyper-tuning process, some generalized equations can be deduced as the '*i*th' term and shown in equations from Equations (53) to (57).

$$\frac{\partial E}{\partial w_i} = \Delta_{i+1} \cdot x_i^{\text{o}} \tag{53}$$

where $\partial E/\partial w_i$ is the partial differentiation of E w.r.t. weights associated with the respective layer and x_i^{o} is the final output of

Table 2 | Performance evaluation of predicted parameters by the FNN model for training and testing datasets

Dataset	H_t vs Q	H_t/P vs C_{DL}	H_t/P vs E_L	H_t/P vs E_2/E_1	Average
RMSE	0.0582	0.0206	0.0216	0.0216	0.0305
MAE	0.0383	0.0150	0.0183	0.0174	0.0222

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

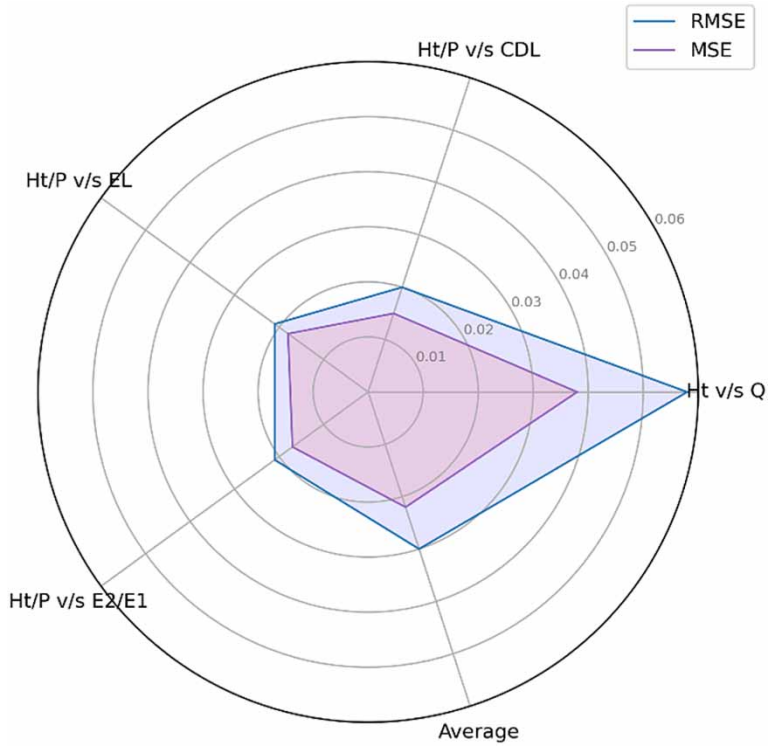


Figure 6 | RMSE and MAE values.

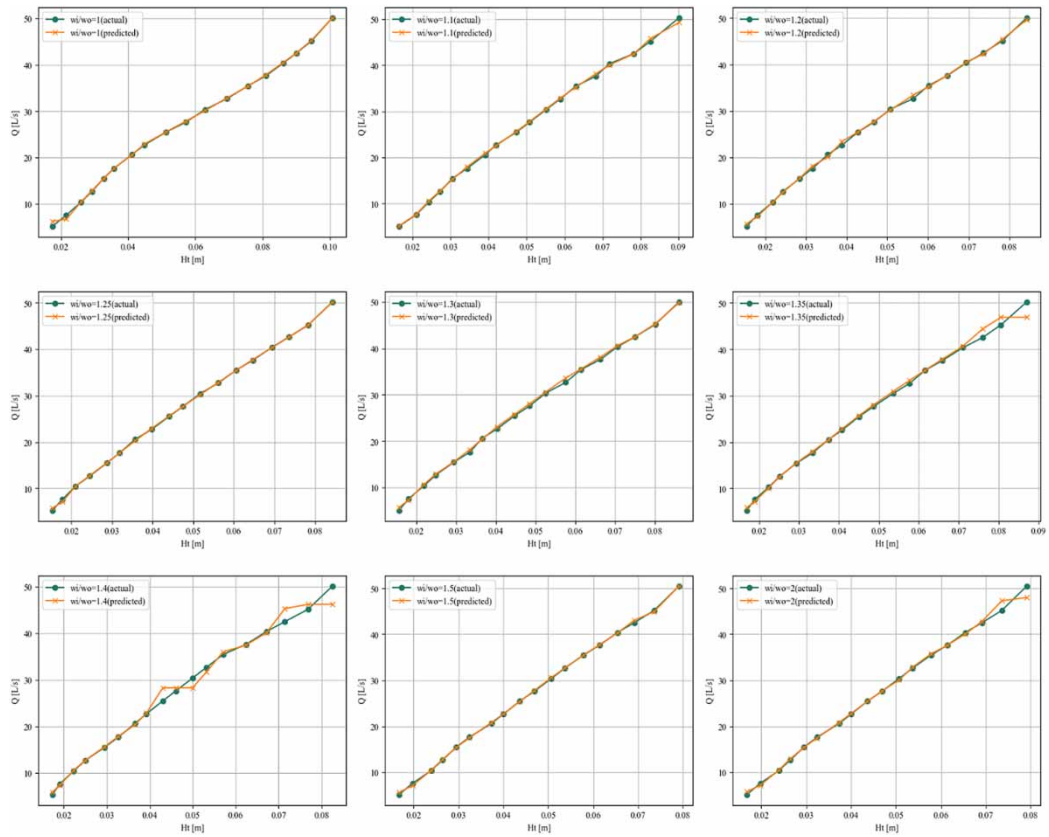


Figure 7 | Stage–discharge curve Q (L/s) vs H_t (m).

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

the previous is associated with the weights for the ith layer.

$$\Delta_i = \Delta_{i+1} \cdot w_i \cdot \sigma_i \cdot f'(x_{i-1}^{o}, w_{i-1}) \tag{54}$$

$$\sigma_i = \frac{e^{-x_i}}{(1+e^{-x_i})^2} \tag{55}$$

where σ_i is the differentiation of the sigmoid function for the ith layer.

$$w_i = w_i - \alpha \cdot \frac{\partial E}{\partial w_i} \tag{56}$$

Equation (56) is the generic formula to update the weight for the layer.

$$\Delta_L = (x_L^{o} - t) \cdot \sigma_i \cdot f'(x_{i-1}^{o}, w_{i-1}) \tag{57}$$

Equation (57) is similar to Equation (54) only L in Equation (57) is used for the variables of the last layers.

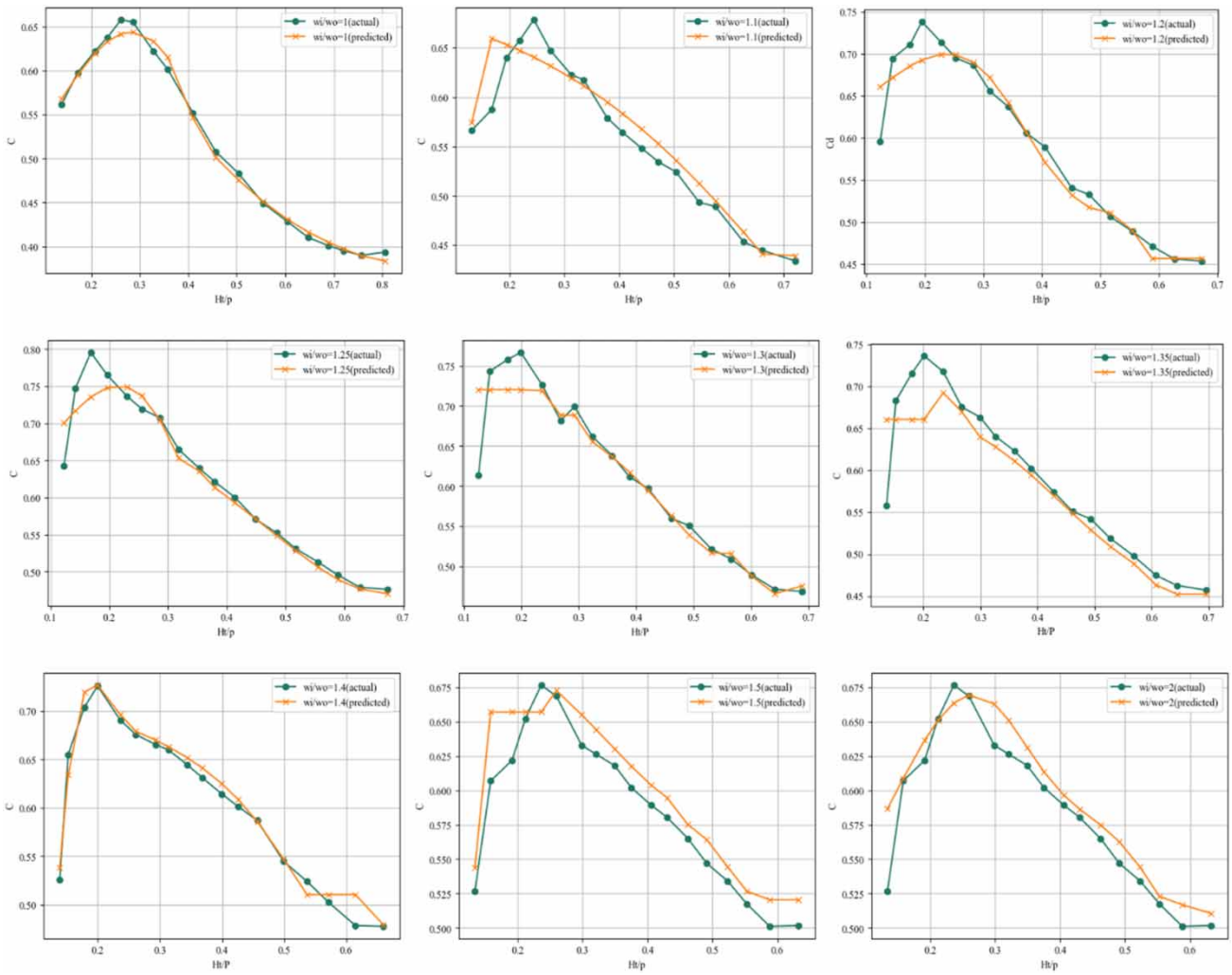


Figure 8 | Discharge coefficient variation curve (C_{DL} vs H_t/P).

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

3. RESULTS AND DISCUSSION

The dataset is collected experimentally, pre-processed, and passed to the FNN model. The method used to validate data is splitting the data into two categories, i.e. training and testing data. Both datasets are chosen randomly and divided in a ratio of 4:1. The training data is used to train the model by tuning the hyperparameters. Then, the model is validated using the testing data. This method allows the judgement of the model's performance to be balanced.

Measures used for the evaluation of the model are root mean square error (RMSE) and mean absolute error (MAE). MAE tells us about the average error, which can offer a straightforward idea about the model's performance. RMSE gives more importance to significant errors, which is a more sensitive measure to determine the accuracy of the predictions. Values for RMSE and MAE are shown in Table 2 and graphically in Figure 6.

The PKW discharge coefficient is determined by assessing the developed length within the $0.24 < H_t/P < 0.79$ range. The stage–discharge relationship serves as a fundamental characteristic of each flow measurement structure. It is illustrated as a curve, plotting discharge against the head for all W_i/W_o values ranging from 1.0 to 2.0, as depicted in Figure 7. The FNN model's performance metrics indicate an RMSE value of 0.0582 and an MAE value of 0.0383, as visualized in Figure 5. These metrics are vital indicators of the model's accuracy in predicting the given data. An RMSE of 0.0582 signifies the average magnitude of prediction errors, while an MAE of 0.0383 denotes the mean absolute difference between expected and actual values (see Table 2). Such metrics are pivotal in assessing the model's reliability and precision in research, contributing to a comprehensive understanding of its predictive capabilities.

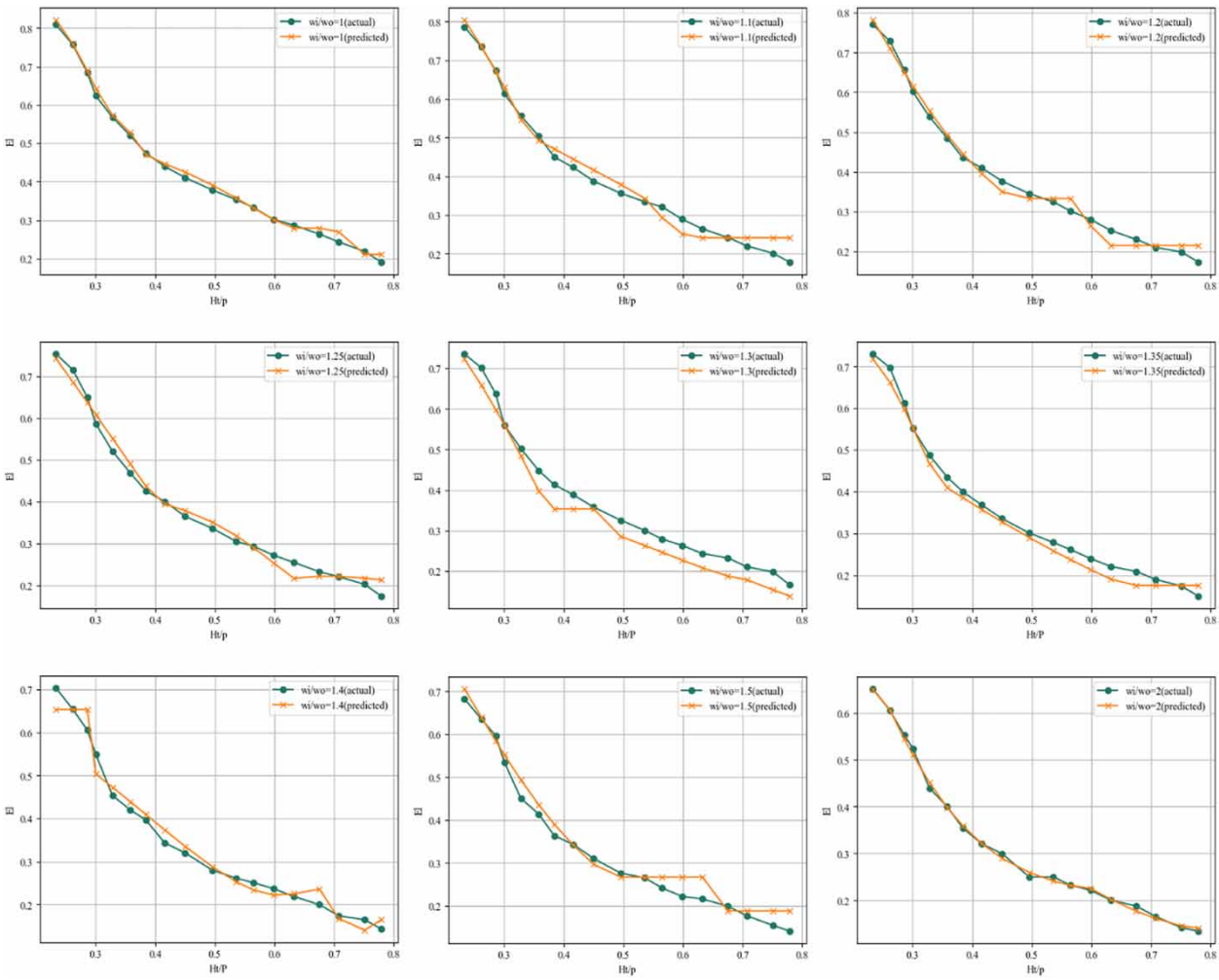


Figure 9 | Energy dissipation curve (E_L vs H_t/P).

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

To delineate the optimal W_i/W_o proportion range, considering the highest C_{DL} values representing peak discharge efficiency, we present the results in Figure 8, depicting C_{DL} as a function of H_t/P. Figure 8 clearly illustrates that W_i/W_o ratios of 1.25 and 1.3 yield the most substantial improvements in discharge efficiency. They are closely followed by W_i/W_o values of 1, 1.1, 1.2, 1.25, 1.3, 1.35, 1.4, 1.5, and 2. This observation underscores that the most advantageous discharge performance is concentrated between the W_i/W_o range of 1.25 and 1.3. The predictions made by the model in all the cases of the ratio of widths and the parameters are almost coinciding, as is visible in the graphs. Thus, the model proves to help predict the hydraulic behaviours from the change of width of the inlet key and the outlet key.

Furthermore, the data in Figure 8 reveals that $W_i/W_o = 1.25$ achieves notably higher discharge efficiency than $W_i/W_o = 1.3$ for $H_t/P < 0.35$. However, for the range $0.35 < H_t/P < 0.44$, $W_i/W_o = 1.3$ surpasses $W_i/W_o = 1.25$ regarding discharge efficiency. Notably, $W_i/W_o = 1.4$ stands out as the configuration with the highest discharge capacity within the $0.44 < H_t/P < 0.81$ range. In particular, PKW setups with W_i/W_o values of 1.25 and 1.3 demonstrate a remarkable 7–17% increase in efficiency compared with $W_i/W_o = 1.0$ and an approximately 8–13% gain over $W_i/W_o = 2.0$. It's worth noting that the FNN model excels in precision, as evidenced by its RMSE of 0.0206 and MAE of 0.0150, as shown in Figure 6.

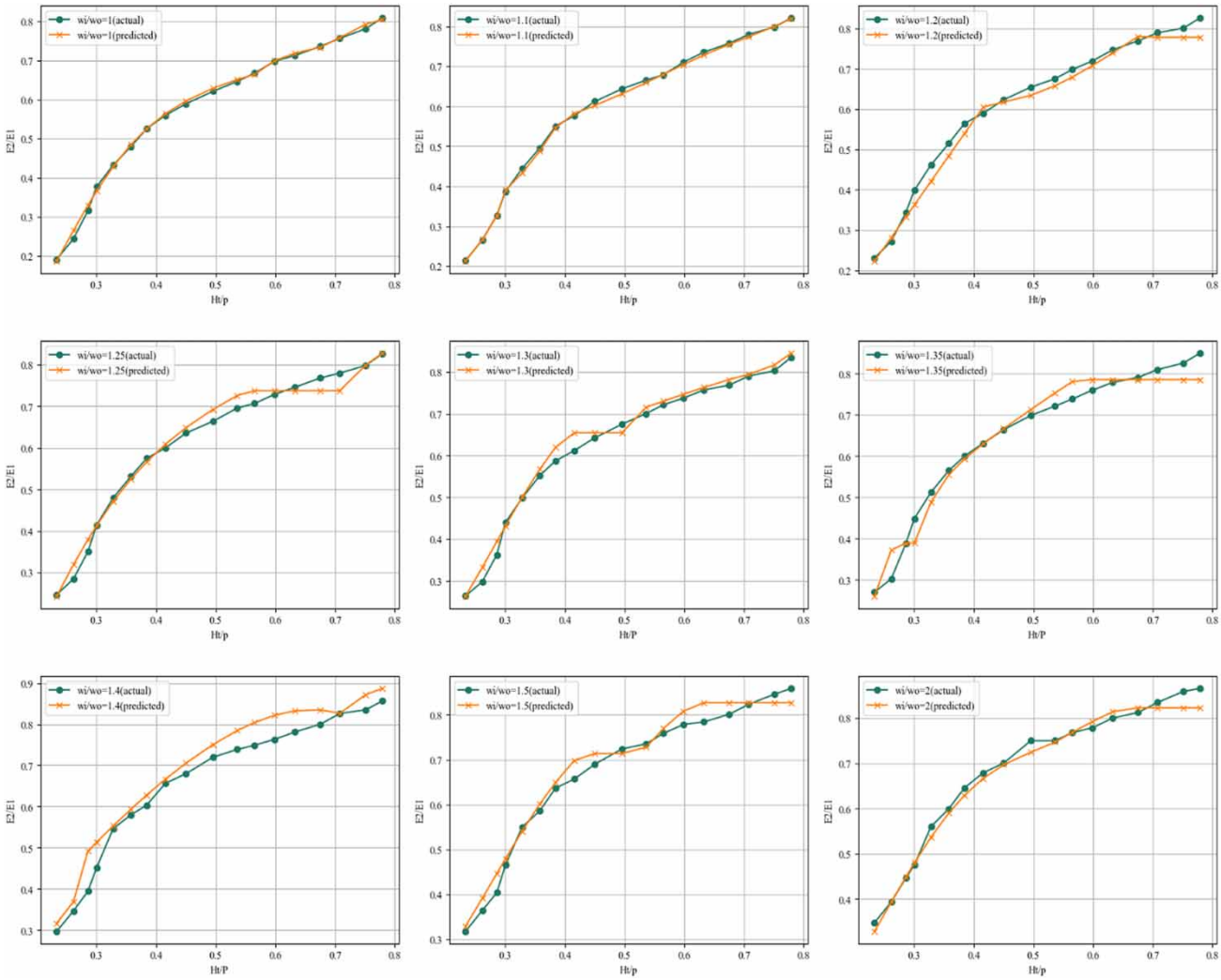


Figure 10 | Residual energy curve (E_2/E_1 vs H_t/P).

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

The FNN models exhibited consistent trends in relative energy dissipation, and their findings were validated against previous studies with a high degree of accuracy in measurement. In this research, we noted that energy dissipation rates tend to be higher when H_t/P is less than 0.42, a trend across all PKW models. This observation deviates from previous research, which reported lower energy dissipation at these values. Conversely, for H_t/P values greater than 0.55 and within the range of $0.42 < H_t/P < 0.55$, the relative energy dissipation rate $[E_L = (E_1 - E_2)/E_1]$ displayed a more complex, intermingled behaviour. Figures 9 and 10 illustrate the variation of relative energy dissipation $[E_L = (E_1 - E_2)/E_1]$ at the base of PKWs as a function of the upstream head ratio (H_t/P).

4. CONCLUSION

To precisely compute the optimal inlet-to-outlet key width ratio of PKW and the influence of the different width ratios on energy dissipation, the experimental data from a previous study by Singh & Kumar (2023a, 2023b) were used to create the FNN algorithm-based model in this research. According to the proposed model for predicting the key's width ratio, the approaches resulted in a highly nonlinear relationship between the width ratio and input parameters, with promising prediction results. The hydraulic behaviours measured in the process are discharge flow over the PKW, coefficient of discharge along developed crest length (C_{DL}) and relative energy dissipation (E_L). The peak efficiency occurs at a width ratio (W_i/W_o) of approximately 1.2755–1.28, with a 7–17% efficiency advantage over $W_i/W_o = 1$ and 8–13% over $W_i/W_o = 2.0$. The energy losses over the weir decreases as the W_i/W_o ratio increases, with the highest relative energy dissipation corresponding to the lowest width ratio (i.e. $E_L = 0.8093$ or 80.93%, the corresponding $W_i/W_o = 1$) and the lowest energy loss for the highest width ratio (i.e. $E_L = 0.5818$ or 58.18% the related $W_i/W_o = 2.0$). This means the energy dissipation across the weir for $W_i/W_o = 2.0$ indicates 12–23% less energy dissipation than $W_i/W_o = 1$. The performance of the algorithm is measured based on the RMSE value and MAE value. More studies are needed on prediction and various soft computing techniques. Furthermore, doing an experimental investigation or CFD simulation while accounting for scaling effects is possible.

ACKNOWLEDGEMENTS

The faculty, staff, and technical team of the Hydraulic Laboratory of the Civil Engineering Department are much appreciated by the authors for their support, and they are also beseeched with sympathy.

FUNDING INFORMATION

This study has not received any funding.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abhash, A. & Pandey, K. K. 2020 A review of piano key weir as a superior alternative for dam rehabilitation. *ISH Journal of Hydraulic Engineering* **28** (supp1), 541–551.

Al-Shukur, A. H. K. & Al-Khafaji, G. H. 2018 Experimental study of the hydraulic performance of piano key weir. *International Journal of Energy and Environment* **9** (1), 63–70.

Anderson, R. M. & Tullis, B. P. 2013 Piano key weir hydraulics and labyrinth weir comparison. *Journal of Irrigation and Drainage Engineering* **139** (3), 246–253.

Baudhanwala, D., Mehta, D. & Kumar, V. 2024 Machine learning approaches for improving precipitation forecasting in the Ambica River basin of Navsari District, Gujarat. *Water Practice & Technology* **19** (4), 1315–1329.

Bhukya, R. K., Pandey, M., Valyrakis, M. & Michalis, P. 2022 Discharge estimation over piano key weirs: A review of recent developments. *Water* **14** (19), 3029.

Birbal, P., Azamathulla, H., Leon, L., Kumar, V. & Hosein, J. 2021 Predictive modelling of the stage–discharge relationship using gene-expression programming. *Water Supply* **21** (7), 3503–3514.

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Crookston, B. M., Erpicum, S., Tullis, B. P. & Laugier, F. 2019 Hydraulics of labyrinth and piano key weirs: 100 years of prototype structures, advancements, and future research needs. *Journal of Hydraulic Engineering* **145** (12), 02519004.

Das, J., Poonia, V., Jha, S. & Goyal, M. K. 2020 Understanding the climate change impact on crop yield over Eastern Himalayan Region: Ascertaining GCM and scenario uncertainty. *Theoretical and Applied Climatology* **142**, 467–482.

Emadi, A., Sobhani, R., Ahmadi, H., Boroomandnia, A., Zamanzad-Ghavidel, S. & Azamathulla, H. M. 2022 Multivariate modeling of river water withdrawal using a hybrid evolutionary data-driven method. *Water Supply* **22** (1), 957–980.

Erpicum, S., Nagel, V. & Laugier, F. 2011 Piano key weir design study at Raviege dam. In: *Labyrinth and Piano Key Weirs–PKW 2011*, pp. 43–50.

Erpicum, S., Laugier, F., Pfister, M., Pirotton, M., Cic, M. & Schleiss, J. 2013 *Labyrinth and PKWs II – PKW-2013*.

Erpicum, S., Laugier, F., Ho Ta Khanh, M. & Pfister, M. 2017 *Labyrinth and PKWs III – PKW-2017*. CRC Press, London, UK.

Eslinger, R., & Crookston, K. & M, B. 2020 Energy dissipation of type A piano key weirs. *Water* **12** (5), 1253.

Jamal, M., Khatun, S., Jha, S., Mehta, D., Dixit, S., Kumar, V. & Tiwari, D. K. 2024 Experimental investigation of the efficiency of stone revetment for different temporal variations with the static water condition. *Journal of Water and Climate Change* **15** (4), 1969–1980.

Kantharia, V., Mehta, D., Kumar, V., Shaikh, M. P. & Jha, S. 2024 Rainfall–runoff modeling using an adaptive neuro-fuzzy inference system considering soil moisture for the Damanganga basin. *Journal of Water and Climate Change*. jwc2024143.

Karimi, M., Attari, J., Saneie, M. & Jalili Ghazizadeh, M. R. 2018 Side weir flow characteristics: Comparison of piano key, labyrinth, and linear types. *Journal of Hydraulic Engineering* **144** (12), 04018075.

Khanh, M. H. T. 2013 The piano key weirs: 15 years of research & development–prospect. In: *Labyrinth and Piano Key Weirs II*, p. 3.

Khanh, M. H. T. 2017 History and development of piano key weirs in Vietnam from 2004 to 2016. In: *Labyrinth and Piano Key Weirs III*. CRC Press, pp. 3–16.

Kumar, N., Poonia, V., Gupta, B. B. & Goyal, M. K. 2021 A novel framework for risk assessment and resilience of critical infrastructure towards climate change. *Technological Forecasting and Social Change* **165**, 120532.

Laugier, F., Vermeulen, J. & Blancher, B. 2017 Overview of design and construction of 11 piano key weirs spillways developed in France by EDF from 2003 to 2016. In: *Labyrinth and Piano Key Weirs III – PKW*, pp. 37–51.

Lempérière, F. & Ouamane, A. 2003 The piano keys weir: A new cost-effective solution for spillways. *International Journal on Hydropower & Dams* **10** (5), 144–149.

Pandey, M. & Md Azamathulla, H. 2021 Discussion of 'gene-expression programming, evolutionary polynomial regression, and model tree to evaluate local scour depth at culvert outlets' by Mohammad Najafzadeh and Ali Reza Kargar. *Journal of Pipeline Systems Engineering and Practice* **12** (2), 07021001.

Pandey, M., Zakwan, M., Sharma, P. K. & Ahmad, Z. 2020 Multiple linear regression and genetic algorithm approaches to predict temporal scour depth near circular pier in non-cohesive sediment. *ISH Journal of Hydraulic Engineering* **26** (1), 96–103.

Poonia, V. & Tiwari, H. L. 2020 Rainfall-runoff modeling for the Hoshangabad Basin of Narmada River using artificial neural network. *Arabian Journal of Geosciences* **13** (18), 944.

Poonia, V., Das, J. & Goyal, M. K. 2021a Impact of climate change on crop water and irrigation requirements over eastern Himalayan region. *Stochastic Environmental Research and Risk Assessment* **35**, 1175–1188.

Poonia, V., Goyal, M. K., Gupta, B. B., Gupta, A. K., Jha, S. & Das, J. 2021b Drought occurrence in different river basins of India and blockchain technology-based framework for disaster management. *Journal of Cleaner Production* **312**, 127737.

Poonia, V., Jha, S. & Goyal, M. K. 2021c Copula-based analysis of meteorological, hydrological and agricultural drought characteristics across Indian river basins. *International Journal of Climatology* **41** (9), 4637–4652.

Pralong, J., Vermeulen, J., Blancher, B., Laugier, F., Erpicum, S., Machiels, O., Pirotton, M., Boillat, J. L., Leite Ribeiro, M. & Schleiss, A. J. 2011 A naming convention for the piano key weirs geometrical parameters. In *Proceedings of the International Conference on Labyrinth and Piano Key Weirs (PKW 2011)*. CRC Press, Taylor & Francis Group, Leiden, Netherlands, pp. 271–278.

Ribeiro, M. L., Boillat, J. L., Schleiss, A. J., Le Doucen, O. & Laugier, F. 2011 Experimental parametric study for hydraulic design of PKWs. In *Proc. Int. Conf. Labyrinth and Piano Key Weirs, Liege B*. CRC Press, pp. 183–190.

Sharma, K. V., Kumar, V., Gautam, L., Choudhary, S. & Mathew, A. 2023 Geo-physical seasonal deviations of land use, terrain analysis, and water cooling effect on the surface temperature of Pune city. *Journal of Water and Climate Change* **14** (12), 4802–4820.

Silvestri, A., Archambeau, P., Pirotton, M., Dewals, B. & Erpicum, S. 2013 Comparative analysis of the energy dissipation on a stepped spillway downstream of a piano key weir. In: *Labyrinth and Piano Key Weirs II*, pp. 111–120.

Singh, D. & Kumar, M. 2021 Hydraulic design and analysis of piano key weirs: A review. *Arabian Journal for Science and Engineering* 1–15.

Singh, D. & Kumar, M. 2022a Computation of energy dissipation across the type-A PKW by using gene expression programming technique. *Water Supply* **22** (8), 6715–6727.

Singh, D. & Kumar, M. 2022b Energy dissipation of flow over the type-B piano key weir. *Flow Measurement and Instrumentation* **83**, 102109.

Singh, D. & Kumar, M. 2022c Gene expression programming for computing energy dissipation over type-B piano key weir. *Renewable Energy Focus* **41**, 230–235.

Singh, D. & Kumar, M. 2023a Effect of the inlet-to-outlet key width ratio of piano key weir on its hydraulic behaviour. *Flow Measurement and Instrumentation* **91**, 102342.

Singh, D. & Kumar, M. 2023b Study of the energy dissipation over the type-A piano key weir. *KSCE Journal of Civil Engineering* **27** (4), 1568–1584.

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Singh, U. K., Jamei, M., Karbasi, M., Malik, A. & Pandey, M. 2022 Application of a modern multi-level ensemble approach for the estimation of critical shear stress in cohesive sediment mixture. *Journal of Hydrology* **607**, 127549.

Tiwari, D. K., Singh, K. R. & Kumar, V. 2024 Forecasting of water quality parameters of Sandia station on Narmada basin using AI techniques, Central India. *Journal of Water and Climate Change*. jwc2024520.

van Deventer, N. J., Matlala, J., Mulder, E. & Hoosen, S. 2015 Tzaneen dam raising: Hydraulic model study. In *Proceedings of the SANCOLD Conference 2015*, pp. 223–230.

Zounemat-Kermani, M. & Mahdavi-Meymand, A. 2019 Hybrid meta-heuristics artificial intelligence models in simulating discharge passing the piano key weirs. *Journal of Hydrology* **569**, 12–21.

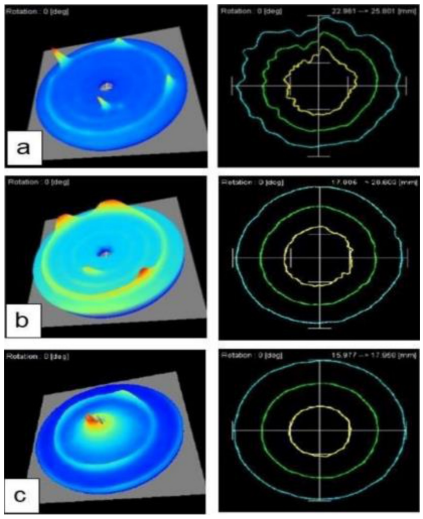
First received 17 January 2024; accepted in revised form 30 July 2024. Available online 13 August 2024

Downloaded from http://iwaponline.com/jwcc/article-pdf/doi/10.2166/wcc.2024.077/1464558/jwc2024077.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Check for updates

Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only.
It will not be published as part of main.



A functionally graded leaf spring plate with five layered gradations in the central section has been fabricated. The quantity of molten material needed for each gradation has been determined using a novel experimental technique.

Received: 25 December 2023 | Revised: 11 March 2024 | Accepted: 8 April 2024

DOI: 10.1002/mawe.202300276

ARTICLE

WILEY VCH

Property investigation of functionally graded materials leaf spring plate fabricated through stir casting process by using new gradient evaluation method

Untersuchung der Eigenschaften von funktionell abgestuften Werkstoffen von im Rührgussverfahren hergestellten Blattfederplatten durch eine neue Gradientenbewertungsmethode

R. K. Singh | **R. C. Singh**

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

Correspondence
R. K. Singh, Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India.
Email: ranjeetsing@gmail.com

Abstract

Automobile and aviation industries require lightweight solutions in the suspension system to improve overall performance and payload capacity. This can be achieved by materials having high strength to weight ratio. Fiber composites, known for their lightweight and high strength, face a major issue of delamination, which shortens the lifespan of fiber-reinforced polymer leaf springs. Current researchers are emphasizing the use of functionally graded materials as a promising alternative to tackle this problem. In the present work, a functionally graded leaf spring plate was fabricated using stir casting technique with five layered gradations in the mid portion. The amount of molten material required for each gradation was ensured using a novel experimental technique. On the other hand, the matrix and reinforcement have been adopted from the output of the analysis conducted using CES EduPack 2019 software. Low density and high fatigue strength were primary focal point for the consideration of matrix and reinforcement materials. Aluminium 7075 and boron carbide were considered as matrix and reinforcement materials respectively. Stir casting technique, used in the present work, is simple, suitable and cost effective technique which creates a gradation at central part of the leaf.

KEYWORDS
graded material, leaf spring plate, stir casting

SCHLÜSSELWÖRTER
Blattfederplatten, gradierter Werkstoff, Rührgießen

© 2024 Wiley-VCH GmbH.

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1 | INTRODUCTION

To make comfortable ride and to minimize various types of vibrations, bumps and impacts due to bad road conditions, springs play a crucial role in automobile industries. Leaf spring is the oldest suspension, used in automobiles. Revolution in design and manufacturing was introduced at the beginning of 19th century [1]. Leaf springs are fabricated as a stack of prismatic beams with longer lengths placed on top of each other. This design ensures that the spring is thicker at the middle to resist bending and thinner at the ends where it is attached to the vehicle body [2]. To improve load carrying capacity, these beams are initially pre-stressed [3]. The steels used for such leaf springs are primarily, SAE 1095, SAE 5160, SAE 6150 and, SAE 9254 [3,4].

During last few decades, automobile industries are attempting to use materials having high strength to weight ratio. This attempt will increase fuel economy of automobiles as well as meet the demands of natural resources conservation. The suspension leaf spring is one of the potential things for weight decrease in vehicles because it has 10% to 20% of the un-sprung weight [5]. Fortunately, fiber composites have excellent strength-to-weight ratio, higher natural frequency, superior fatigue strength which make them excellent for leaf spring [2–5]. However, demerits of fiber composite materials coming from discontinuous variation of properties at transition zone between two solids and delamination over a period of time. Demerits of fiber composite can be overcome by the use of low weight and high strength material i.e. metal matrix composite.

Metal matrix composites, particularly aluminum matrix composites, offer advantages like high strength, stiffness, thermal stability, low density, and improved wear resistance [6–9]. The metal matrix composites also provide better toughness and higher transverse properties compared to fiber composites. The group of metal matrix composites comprises different metals utilized as matrices like aluminum, magnesium, copper, and titanium. Different reinforcements have been utilized to accomplish desired properties. Each metal has remarkable applications best suit its properties however out of all, aluminium-metal matrix composites are the most broadly utilized composite materials [10,11]. Aluminium matrix composites have striking properties, for example, lightweight, excellent corrosion resistance, high strength, low thermal expansion coefficient, great wear resistance, and simplicity of machining [12]. Aluminum 7075, among various aluminium alloys, possesses the highest specific strength and toughness [13]. Similarly, in the reinforcement category boron carbide is a charming reinforcement material because of its good chemical and thermal stability. Boron carbide has higher hardness and lower density compared to silicon carbide and aluminium oxide [14–16]. However, directional strength to weight ratio of aluminium matrix composites is lower than the fiber composites. It means that it is not a perfect replacement of fiber composite. Thus above discussions reveal that both aluminium matrix composites and fiber composites will not be the perfect material for leaf springs. Although, aluminium matrix composites have one more advantage over fiber composite that its strength to weight ratio can be improved by tailoring in required direction this lead towards evolution of a modern material known as functionally graded material (functionally graded materials) using aluminium matrix composites.

Functionally graded materials are a class of microscopically heterogeneous materials in which the material compositions are designed to change material properties fluently over volume [17]. Generally solid based, liquid based and gas based methods are used to fabricate functionally graded materials [18]. There are various types of fabrication processes for producing functionally graded materials. Further, functionally graded materials can be divided majorly into two broad groups i.e. thin and bulk functionally graded materials. Coatings are considered into thin category; while bulk functionally graded materials have thick gradation, used for extreme environment. Bulk functionally graded materials are typically made out of two distinct materials, for example, metal and ceramic and they exploit the superior property from every constituent [17,19]. Bulk functionally graded materials are produced using powder metallurgy technique, slip casting, centrifugal casting method, gel casting, solid freeform technology etc. [20]. In addition to the diverse methods employed for the synthesis of functionally graded materials, various techniques have emerged, each possessing unique advantages and limitations. The versatility of functionally graded materials stems from their tailored composition, which changes gradually across the material, providing enhanced performance in specific applications. Beyond conventional methods like powder metallurgy and chemical vapor deposition, several alternative approaches have gained prominence. A concise comparison to highlight the distinctive features of functionally graded materials fabrication techniques are presented in Table 1.

In the present work, one leaf (second leaf from the top) of the rear suspension system of the Tata long platform truck 3518 model was chosen for fabrication. The selection of this specific model is based on its structural characteristics and the critical role played by the second leaf, which is designed to have the same length as the top leaf, providing crucial support to the eye and

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Comparative study of various methods for fabrication of functionally graded materials.

Method	Advantages	Limitations
Powder metallurgy	Widely established, cost-effective	Limited control over gradient profiles, fine-scale challenges
Chemical vapor deposition	High purity, precise composition control	Limited material compatibility, complex equipment
Additive manufacturing	Customizable structures, intricate geometries	Material constraints, build size limitations
Electrochemical deposition	Uniform coatings, controllable deposition rates	Limited to conductive materials, challenges in thickness
Spark plasma sintering	Rapid processing, high densification	Llimited material compatibility, complexity in shapes

resistance to torsional loads acting at the eyes [2]. Selecting the Tata long platform truck 3518 model for second leaf spring study presents numerous advantages. This model is widespread used in the automotive and logistics industries in India and ensures that the current research aligns with real-world applications and industry standards. The Tata 3518′s robust design, substantial payload capacity, and common usage in commercial transportation allow us to assess the leaf spring‘s performance under typical, heavy-load conditions. The truck‘s versatility across diverse terrains provides insights into the leaf spring‘s adaptability, while the ready availability of technical data enhances the precision of our study. Literature reviews have revealed that maximum bending moment appears at the central portion of semi elliptical leaf spring, due to which most of the failures occurs in this zone. Thus, the motive of this work is to strengthen the central portion of the leaf spring by providing gradation in this zone. This will improve strength to weight ratio of plate as well as improve effective use of material. Decision is made on the selection of suitable method, for the fabrication of such plate, which is based on the critical review of bulk manufacturing techniques of functionally graded materials. Selection criteria is based on some prerequisites (methods must have) such as feasibility of achieving required gradation, low fabrication cost, simplicity, mass productivity with good quality. Each method has its own advantages and demerits like, powder metallurgy technique is not suitable for fabrication of this large plate because, size of compaction die will be large and bulky. Further, heavy and costly machine will be required to produce such compaction force and extra setup is required to provide uniformly distribution of this force over the punch. Similarly, centrifugal casting is also not suitable for the fabrication of such plate because radius of curvature of leaf spring plates are very high thus drum size will be very large as well as very powerful motor will be required to fulfill this need. On the other hand, other bulk functionally graded materials fabrication techniques are also not suitable to provide required gradation except stir casting technique. It was found that quality of stir casted products is highly influenced by pouring temperature, pouring rate and the angle of stirring blade. The stir casting process is simplest, economical and most commercially used technique in liquid state processing [21, 22]. Since stir casting technique is capable to fulfill prerequisites in the direction of fabrication of the plate. Therefore, stir casting method is selected for the fabrication of the plate.

The change in properties of straight beam of functionally graded materials is commonly controlled in one direction, either along beam centerline or across the beam thickness [23–26]. Bidirectional gradation of properties is provided in case of functionally graded curved beam [27]. In this work, bidirectional gradation across the plate thickness was furnished at the middle portion.

The mechanical characterization of materials plays a pivotal role in understanding their structural behavior and performance under applied loads. In leaf springs, residual stress plays very important role because it directly influences the fatigue life. Residual stress in tensile nature decreases the fatigue life while the compressive residual stresses have enhanced the fatigue life [28, 29]. In metal matrix composites, residual stresses are incited during fabrication because of differences in coefficients of thermal expansion of matrix and reinforcement materials [30]. Several methods such as sin α, cos α based on x-rays and, four-point bend test can be used to measure residual stress. In x-ray method, the strain is measured by bragg diffraction crystal. Other two methods i. e. sin α and cos α can also be used to measure the residual stress. In this research work, cos α method was used to measure residual stress. Residual stress measured by cos α method is a fast technique and it provides more accurate result for fine grain materials [31–33]. The cos α method is useful in the analysis of in plane biaxial stresses. In this method 2-dimensional detector was used to record whole debye scattered ring data in an exposure of x-rays [31].

In the context of the present study on functionally graded leaf spring plates, the three-point bend test serves

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

as a crucial tool for evaluating the flexural strength of compositions featuring aluminum 7075 and boron carbide. The specific flexural strength of a composite material depends on factors like the percentage of boron carbides, the quality of the composite manufacturing process, and other considerations [34]. The results obtained from this test contribute to a comprehensive understanding of how these materials perform under bending loads, guiding the development of lightweight and durable solutions for automotive suspension systems.

The novelty of this work lies in the fact that no such gradation scheme is adopted earlier for leaf spring plates. Total amount of material required in each gradation has evaluated experimentally and computationally and their results are compared. Decision of gradation at the mid portion of leaf is based on variation of bending and shear stresses under the bending moment, Figure 1a. As per basic theory of bending, the magnitude of maximum shear stress acts at the mid, but its magnitude is much lesser than the maximum bending stress, Figure 1c. Further maximum bending stress is acting at top and bottom layers of the plate, Figure 1b [35]. Thus, the intensity of stresses decreases either towards top to centre or towards bottom to the centre of thickness.

2 | MATERIALS AND PROCESSING

2.1 | Selection of matrix and reinforcement materials

In the present work, the matrix material was selected as commercial grade aluminium 7075 alloy. Micro particles of boron carbide were used as reinforcement. The chemical composition of matrix material has 5.23 percent – zinc, 2.1 percent – magnesium, 1.45 percent – copper, 0.30 percent – manganese, 0.23 percent – chromium, 0.22 percent – silicon, 0.22 percent – iron, 0.16 percent – titanium, and balance percent – aluminium. The average size of reinforcement particles was 3.9019 μm. The matrix and reinforcement material have been selected on the basis of fatigue strength and density of the material. To fabricate low weight and high fatigue leaf spring plate, selection was performed by using software (CES EduPack 2019). Software has shown fatigue strength vs. density diagram of different materials, Figure 2. Density of both selected reinforcement and matrix are approximately same whereas, the fatigue strength of reinforcement is much higher than the matrix material, Figure 2. Thus, the literature review and the software confirmed that aluminium 7075 and reinforcement boron carbide are suitable combination of material for leaf spring.

2.2 | Selection of functionally graded materials gradation in leaf spring plate

Stepwise gradation was adopted in the fabrication of functionally graded materials leaf spring plate. Five-layer gradation was provided in the middle portion of spring plate and the remaining portion of the leaf (both ends of the plate) have the same composition, Figure 3. The decision on composition used in different layers is based on intensity of stress at different positions along the thickness of the plate and effect of reinforcement on the properties of composition. To resist maximum stress intensities at the top and bottom layers, the weight percentage of boron carbide should be maximum as well as

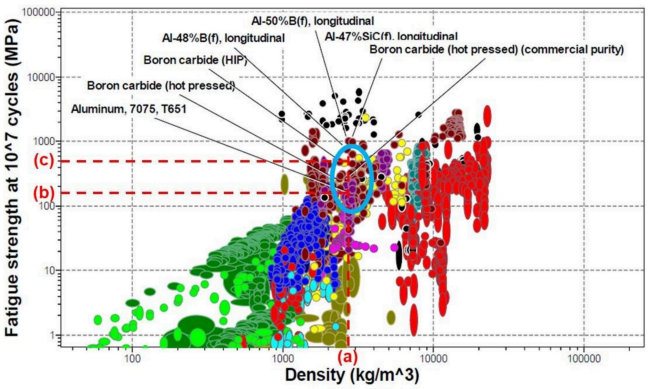


FIGURE 2 Fatigue strength at 10^7 cycles (MPa) vs. density (kg/m^3).

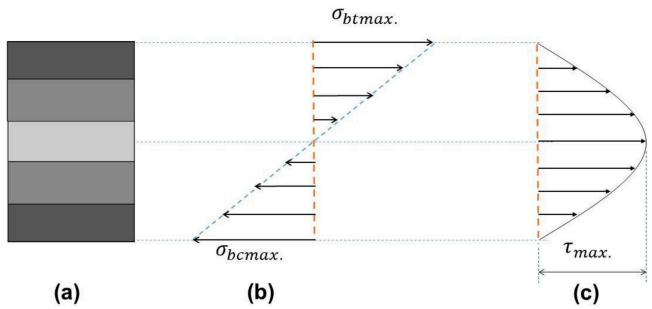


FIGURE 1 (a) Gradation at middle of leaf (b) theoretical bending stress distribution (c) shear stress distribution.

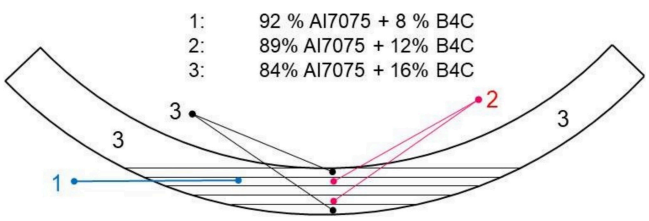


FIGURE 3 Compositions in different portion of leaf spring plate.

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to resist minimum stress intensity acted at the middle layer of thickness, the percentage of reinforcement (boron carbide) should be low. Thus, to achieve this 8 weight percent, 12 weight percent, and 16 weight percent of reinforcement have been used to get uniform variation from mid of the thickness to top and bottom. Further, to resist twisting moment which will act near the both ends of the plate, and to maintain uniformity of properties and microstructure between plate ends and top layer, the composition of material used for both ends will be same as used in top layer [2, 36].

2.3 | Experimental and computational investigation of required amount of material for different gradation

In order to get the actual amount of material required for uniform thickness of each gradation at the center of the leaf spring plate, an experiment has been performed. Initially 6 kg white portland cement, water and a pattern of leaf spring plate were used to fabricate mold. Appropriate quantity of water and cement were mixed together to get slurry. Further, mould of leaf spring plate with one side opening was fabricated with the help of pattern and slurry. Open side of the mould was covered with glass. To get seamless joint between these and to obstruct liquid absorption of the mold, the surface of the mould was covered by a micro film of polymer. At the centre of the pattern, five equal partitions were marked on the thick side of wooden leaf pattern. In order to get these markings on glass, the pattern was put inside the mould by taking care that marking side of the pattern should appear on glass side. A scaled cup was used to pour colored water into the mold to reach each mark on the glass, Figure 4. Through this process, the amount required for each gradation was noted. This process was repeated five times to get accurate result.

2.3.1 | Validation of gradation material requirement

A computational analysis was also performed to evaluate the amount of molten material required in each gradation. In the computational study, the validation process involved rigorous testing against different datasets and benchmarking against established metrics, ensuring model accuracy and robust performance under multiple scenarios. Important dimension of reference spring plate was measured on a coordinate measuring machine. After that, the dimensions of fabricated pattern was measured on same coordinate measuring machine for the validation of dimensional accuracy of the pattern. In the next step, a three dimensional model was generated by using software abaqus 2018. Further, central thickness of this model was divided horizontally into five equal parts, Figure 5. Now the volume of each partition was evaluated through this software.

The experiment was performed continuously five times to get five readings of volume of each gradation. The average volume of liquid from bottom to top layers was 63.6 ml, 130.2 ml, 168.4 ml, 250 ml, and 449 ml respectively, Figure 6. The average volume of each gradation was compared with computational results. It is noticed that both results were slightly deviated (within 2%) from each other. The experimental results were slightly greater than the computational results. The deviation in both results suggests that there could be some errors during experimentation. Since computational software provided to approximate results thus the average value of computational and experimental was considered as optimum amount of material for each gradation. Further, final amount of material in each gradation was decided by considering 2 percent losses in gating such as runner, sprue etc., Table 2.

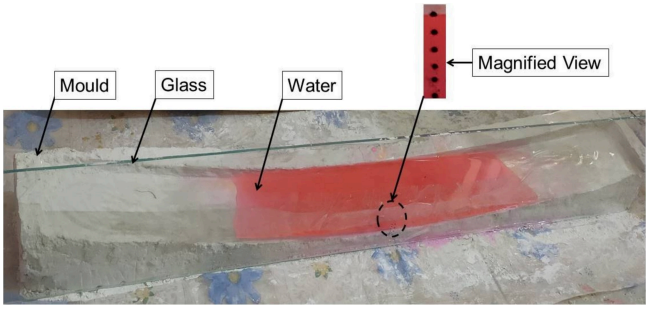


FIGURE 4 Experimental setup for evolution of the material required in each gradation.

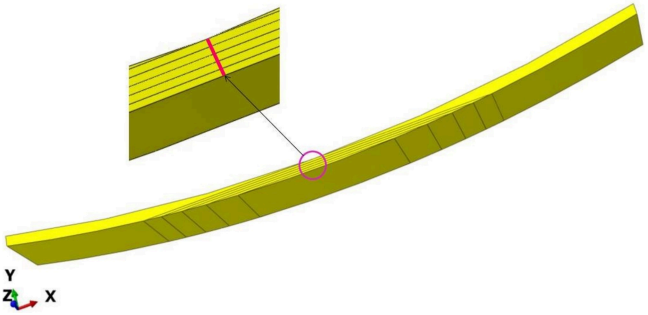


FIGURE 5 3-D model for computational investigation of required material in each gradation.

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

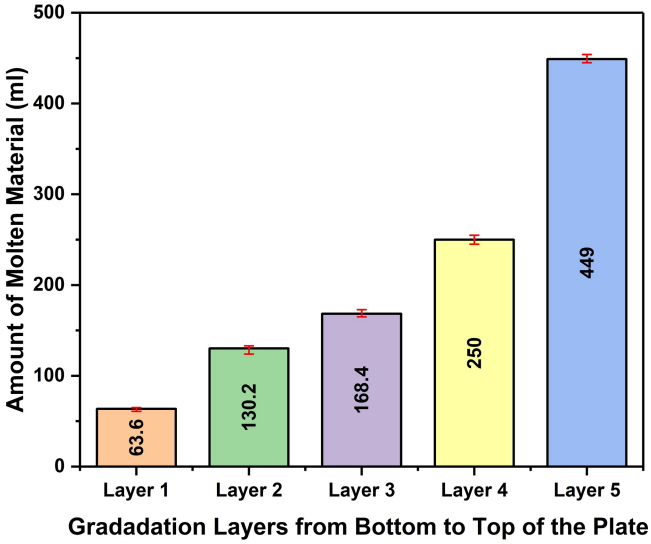


FIGURE 6 Average amount of molten material required in each gradation.

2.4 | Functionally graded materials leaf fabrication

The functionally graded materials leaf was fabricated using a combination of mechanical stirrer and gravity die casting approaches. A mould is the primary need of gravity die casting. Appropriate moulds were achieved with the use of proper patterns and suitable sand. The dimensions of the wooden pattern have been determined by considering different allowances such as draft allowance, shrinkage allowance, machining allowance, and shaking allowance. Among all the allowances, only shaking allowance is negative. In this work, the dimensions of the wooden pattern have been determined by considering a machining allowance of 1 mm on each surface of the pattern, a shaking allowance of 1 mm on length and width, a shrinkage allowance of 2.5 percent in the volume of the pattern, and a taper allowance of 1 degree on vertical sides [37].

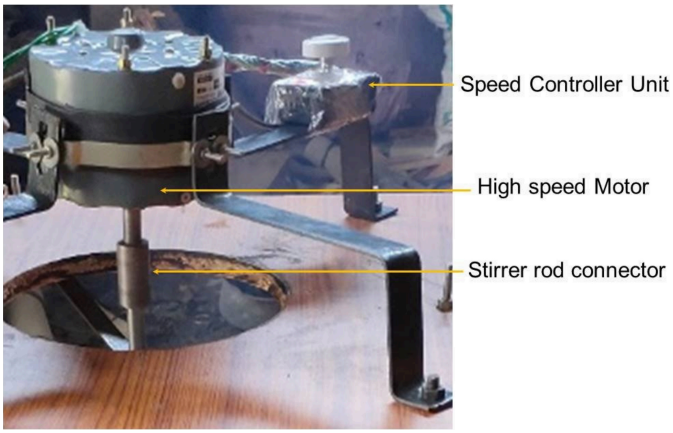


FIGURE 7 Stirrer with a speed controller.

The mould was prepared in green sand containing ash of coconut shell and camphor powder about 1.5 weight percent. Coconut shell ash and camphor powder between 1 weight percent to 2 weight percent of green sand exhibited better mould permeability together with sufficient green strength [38]. In the stir casting process, three stir casting setups are used to prepare three compositions simultaneously. Speed controllers have been used in each stirrer, Figure 7. In each stir casting setup, the matrix material is melted and then blended with reinforcement by stirring, Figure 8. Each setup contains a furnace in which each aluminum 7075 matrix of weight 2 kg is melted at a temperature of 800 °C. To prevent rapid oxidation at high temperatures, an inert gas environment was created by feeding argon gas into the furnace. Three compositions of boron carbide powders of weight percentages of 8 percent, 12 percent, and 16 percent have been preheated in a muffle furnace. Laboratory reagent-grade potassium titanium fluoride flux with a purity of 99.9 percent was used as a wetting agent. The mixture of boron carbide particles and the same amount of wetting agent were prepared. The preheated boron carbide and potassium titanium fluoride powders were then added to their respective melted matrix and stirred

TABLE 2 Amount of composites required in different gradation at center part of the leaf spring plate.

Gradation number from bottom to top of the leaf	Amount of liquid required (ml)								
	Experimental readings								
	1	2	3	4	5	Average value	Computational result	Average value (ml)	Amount selected with 2% additional allowance (ml)
1	64	64	61	65	64	63.6	61.3	62.5	63.7
2	132	132	124	130	133	130.2	133.0	131.6	134.2
3	169	173	170	165	165	168.4	167.6	168.0	171.4
4	247	253	245	250	255	250.0	248.7	249.4	254.3
5	450	454	445	446	450	449.0	450.1	449.6	458.5

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

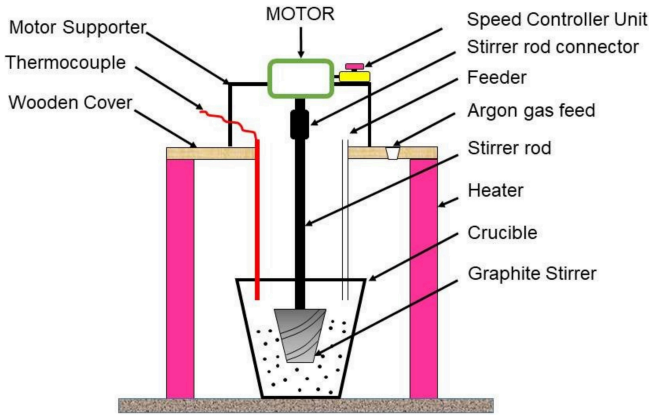


FIGURE 8 Schematic diagram of stir casting set-up.

for homogeneous mixing. A three-blade graphite stirrer, rotated at a speed of 550 min^{-1}, was used in all three set-ups to create a vortex. Further, five cups of varying volumes were used to pour melts through the central runner of the mold, and the runners at both ends of the mold were filled with melts containing 16 weight percent reinforcement simultaneously. The entire process of pouring melts was executed within a minute. The remaining melts from all three crucibles were used to cast specimens for different testing. The specimens were prepared to study the mechanical properties of the three composites. The cast samples and leaf spring plate were heat-treated to the temper six conditions.

3 | CHARACTERIZATION

3.1 | Experimental setup for residual stress:

The surface of each samples used for the measurement of residual stress was chemically etched at 100 °C for 7 minutes. Well treated samples were tested with the help of x-ray residual stress measuring machine, Figure 9. This machine uses chromium K-alpha radiation which uses Debye ring cos α method with the help of fitting lorentz approximation method. At the time of investigation, the x-ray device having tube voltage 30 kV, tube current 1 A, wave length (λ) of chromium K-alpha 2.909 A0 and x-ray incident angle (ψ_0) 35° was ensured. This device uses imaging plates as 2-dimensional detectors.

Residual stress measured by cos α method is a very simple and quick technique. In this method, the stress is measured within 3 minutes in a single exposure and it provides more accurate results. The imaging plates were used to collect strain data around the circumference of Debye-Scherrer ring. The imaging plates were placed



FIGURE 9 X-ray residual stress measuring machine.

normal to the x-ray and at azimuthal angle α, Figure 10. The value of strain (ε_α)for a known diffraction angle θ, is calculated as per equation 1[31].

$$\varepsilon_\alpha = \frac{\cos^2 2\theta}{2\ \mathrm{L}\ \cos\theta}(\mathrm{r}_\alpha - \mathrm{r}_0) \tag{1}$$

where L is the distance between the specimen and the imaging plate, $r_\alpha$is the radius of the debye ring at the azimuthal angle α and r_0 is the radius of stress-free crystals [31–33]. The relation between strain and cos α was used to determine the magnitude and the nature of stress. When the rotation angle is 0°, the stress is determined from the slope of the strain and cos α.

3.2 | Microstructure

The phase composition and components of composites used to create gradation in plate were analyzed by x-ray diffraction patterns. Origin 2017 software was used to plot x-ray diffraction pattern of each composites. Each peak of x-ray diffraction patterns were then fitted by using Gaussian fitting method. Further, Gaussian fitting method has been used to calculate broadening of each peak or full width at half maximum in degrees. In order to check the effect of reinforcement on crystallites size, the crystallites size was then evaluated by using Scherrer equation 2, as given below:

$$\mathrm{D} = \frac{\mathrm{K}\lambda}{\beta\ \cos\theta} \tag{2}$$

In equation 2 , D is crystallites size in (nm), K is Scherrer constant (0.94), λ is wave length of copper K-alpha radiation (0.15406 nm), β is full width at half

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

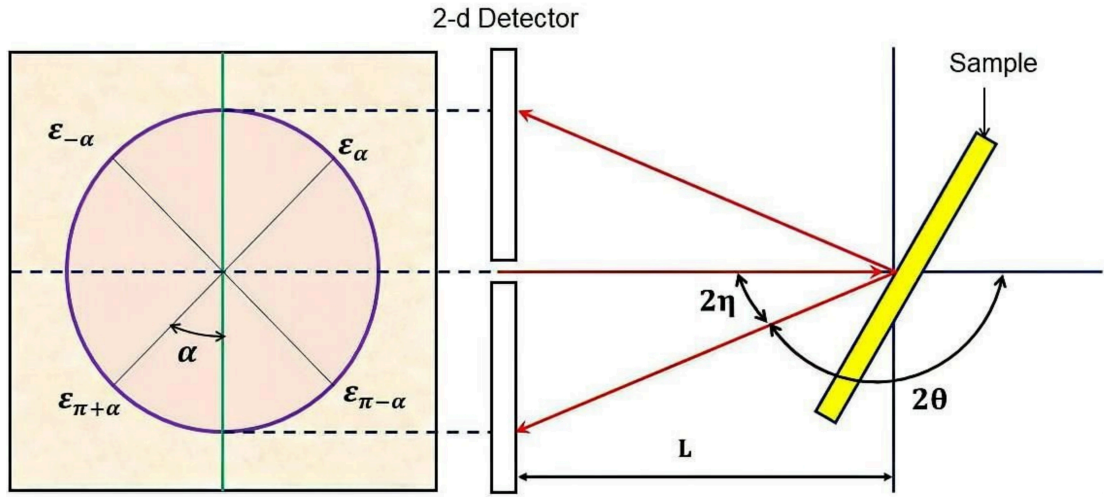


FIGURE 10 Strain measurement in Debye-Scherrer ring in cos α method.

maximum and θ is the peak position (Bragg angle) in radians.

Microstructural morphology of composite, reinforcement and wettability element was investigated by field-emission scanning electron microscopy (Zeiss, Sigma vp).

3.3 | Experimental setup for hardness measurement

Brinell hardness tester was used to measure the hardness of all three compositions of aluminium-boron carbide composites used in the fabrication of functionally graded materials plate. The test was performed according to ASTM E10 standards. The readings were taken at five places on each layer in order to obtain average value of hardness.

At the top and bottom layer, the bending strength will be maximum, so high hardness will resist more bending stress. Moreover, at harder surface, wear rate will be minimum with the mating leafs. Further, this type of variation of material will not lose its toughness, which is a desirable property for the leaf spring.

3.4 | Experimental setup for flexural strength measurement

Three-point bend test was performed on electronic universal testing machine of capacity 10 kN. The samples were fabricated as per ASTM E855. The test was performed on five samples of each composition which was used in the fabrication of FGM leaf plate. To get more accurate results, two types of feeds were applied. In each composition of functionally graded material sample, first 3 samples were tested under feed of 0.5 mm/min. and other two samples were tested under the feed of 1 mm/min.

4 | RESULTS AND DISCUSSION

The experimental readings presented the actual measurements of liquid required for each gradation. There is observable variability in the experimental data, as seen in the ranges for each gradation. The computational results provide an average value for each gradation, aiming to offer a more consolidated understanding of the liquid requirements. The close alignment between experimental readings and computational results suggests the reliability of the computational model in predicting the average liquid requirement. A qualitative trend is observed as we move from the bottom to top of the leaf spring. The liquid requirement generally increases across gradations, indicating a systematic variation in composite needs along the vertical axis of the leaf spring. The inclusion of a 2 percent additional allowance in the analysis reflects a practical consideration. Comparing the average value with the amount selected with additional allowance, we find a consistent pattern. This additional allowance, ranging from 2.5 ml to 9.2 ml, ensures a margin of safety or accounts for potential uncertainties in real-world applications.

The dimensions of wooden pattern were prepared by providing allowances in a sequence i.e. machining allowance, shaking allowance, shrinkage allowance and draft allowance. The pattern of leaf spring plate is fabricated by wood, Figure 11a. With the use of pattern and green sand, the mould was fabricated successively.

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

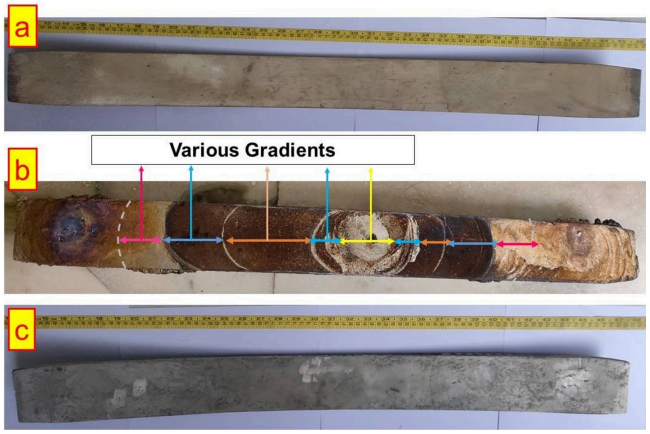


FIGURE 11 (a) Pattern of spring plate (b) Stir casted functionally graded materials leaf (c) Finished functionally graded materials leaf spring plate.

Further, molten composite, which was homogeneously mixed in three crucibles, were poured into the mould through pouring basin. Visual inspection of rear surface of casted leaf showed the colour of this surface is changing from center to the both ends in steps, Figure 11b. This deviation revealed that the gradation will form successively. After machining, the visual inspection on the finished surface, such gradations are not visible, Figure 11c. Thus, microscopic investigation was performed for the verification of the gradation at the central portion of the leaf.

Debye-Scherrer ring profiles shows with distortion of all temper 6 heat treated samples under normal incidence of x-rays on imaging plate, Figure 12. Figure caption also shows full width at half maximum average around the circumference of Debye ring. Comparison between the Debye rings of three compositions shows that the smoothness and uniformity of Debye ring and distortion profiles increases towards higher weight percentage of reinforcement, Figure 12. With the increase in weight percentage of reinforcement, distorted ring as well as Debye-Scherrer ring becomes more uniform which reveals that grain refinement increases with the increase of reinforcement [32, 33].

In coarse grain composite, having tensile residual stress, an increase in weight percent of reinforcement the grains will be refined which represents that the residual stresses increases in compressive nature. This reveals that fatigue property increases with the increase in weight percentage of reinforcement, Table 3.

X-ray diffraction patterns of all of the aluminium 7075 composites having 8 weight percent, 12 weight percent and 16 weight percent of boron carbide reinforcement revealed that the intensity of peaks increases with the increase in percentage of reinforcement, Figure 13. The consistency of peaks and components have verified

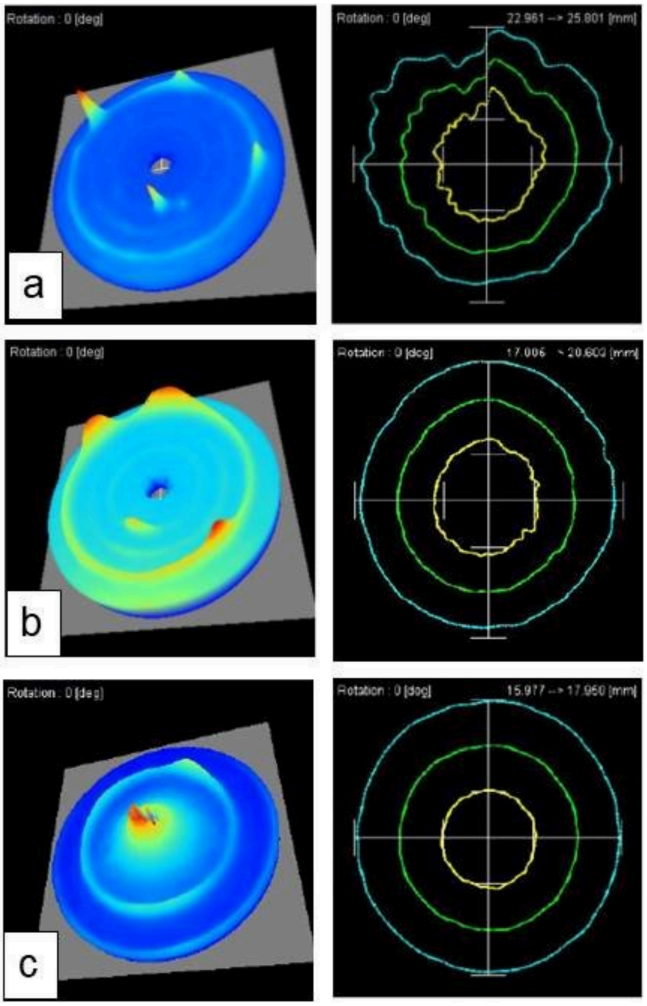


FIGURE 12 3-D Debye-Scherrer ring and distortion for (a) Al 7075 8 wt.–% B4 C, (b) Al 7075 12 wt.–% B4 C, and (c) Al 7075 16 wt.–% B4 C.

TABLE 3 Residual stress for different weight percent of reinforcement.

Material composition	Residual stress (MPa)
Al 7075 + 8 wt.–% B4 C	−24
Al 7075 + 12 wt.–% B4 C	−73
Al 7075 + 16 wt.–% B4 C	−144

by international centre for diffraction data files. The peaks x-ray diffraction patterns were consistent with the international centre for diffraction data files number 89–2837 and 07–0275 for aluminium 7075, and boron carbide respectively.

The pattern of x-ray diffraction depicted the presence of boron carbide and aluminium for composites. The intensity of boron carbide peak is higher in 101 plane, Figure 13. With the increase of weight percentage of

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

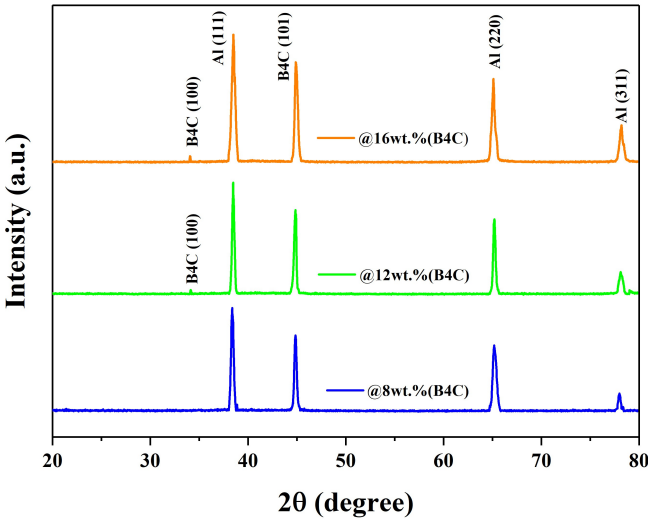


FIGURE 13 Cascade model of XRD pattern of compositions used in functionally graded materials plate.

reinforcement, the peaks are slightly sifted towards higher angles.

The crystallite sizes of all three composites, used for the gradation, are determined by Scherrer's equation (1). The results obtained by using Scherrer's equation revealed that with the increase in weight percent of reinforcement, the crystallite size slightly increases [39]. Average crystallite size for the aluminium- boron carbide composites having 8 weight percent, 12 weight percent, and 16 weight percent are 29.70947285 nm, 28.14782544 nm, and 26.54855502 nm respectively. The observation of x-ray diffraction plot of all composites are shown at each peak, the value of full-width at half-maximum increases with increase in 2-theta. Thus for each composite, the crystallite size decreases with the increase in 2-theta angle.

The particle size of reinforcement (boron carbide) and wetting element (potassium titanium fluoride) were investigated with the help of their micrographs and image J analyzer. Further, origin 2017 was used to obtain average particle size and to generate histogram with normal distribution curve. Each particle of boron carbide was measured from its field-emission scanning electron microscopy image, Figure 14a. The average particle size of boron carbide was 3.9019 μm, Figure 14b. Similarly, image J analyzer was used on field-emission scanning electron microscopy image of potassium titanium fluoride particles to evaluate size of each particle, Figure 14c. The average particle size of potassium titanium fluoride was 1.00655 μm, Figure 14d.

An image of functionally graded materials plate thickness was depicted in Figure 15a. The field-emission scanning electron microscopy micrograph of all three compositions of functionally graded materials was used

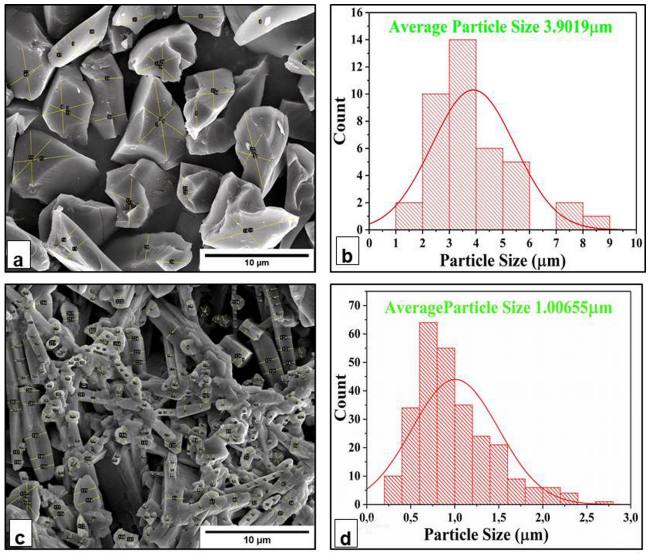


FIGURE 14 (a) Field-emission scanning electron microscopy image of ultrafine boron carbide particles (b) Crystallite size distribution of ultrafine boron carbide particles (c) Field-emission scanning electron microscopy image of ultrafine K2TiF6 particles (b) Crystallite size distribution of ultrafine K2TiF6 particles.

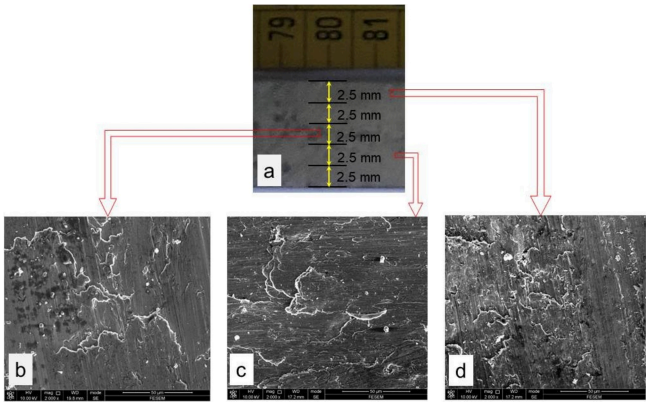


FIGURE 15 (a) Gradation thickness at the center of plate (b) Field-emission scanning electron microscopy image of composition having 8 percent boron carbide (c) Field-emission scanning electron microscopy image of composition having 12 percent boron carbide (d) Field-emission scanning electron microscopy image of composition having 16 percent boron carbide.

in order to verify the dispersion of micro particles of boron carbide. Inspection has revealed that the reinforcement is uniformly distributed at low percentage of reinforcement whereas with the increase in percentage of reinforcement, the dispersion was slightly random, Figure 15b, c. At higher percentage of reinforcement the distribution of boron carbide was non-uniform and micro pores are present, Figure 15d.

Since the composition of aluminium- boron carbide composites varies from centre to top surface and centre to bottom surface of the plate in same manner, thus the

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

measurement of hardness was executed from bottom to central layer. It has revealed that the hardness varies from the central layer (8 weight percent boron carbide) to the bottom and top layers (16 weight percent boron carbide) in ascending order. Further, the average brinell hardness number values of 8 wt.–%, 12 wt.–% and 16 wt.–% boron carbide were 141.2, 166.6 and 202.4 respectively which are remarkably higher than that of base alloy, Figure 16.

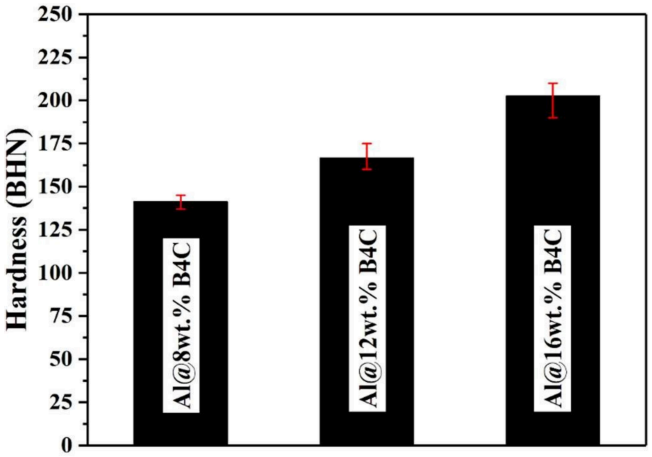


FIGURE 16 Hardness values of bulk functionally graded materials.

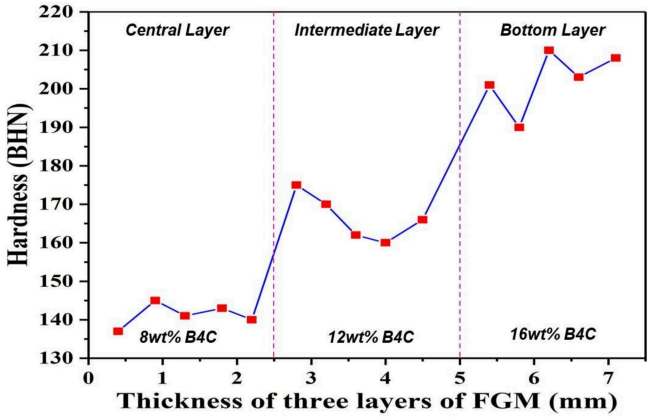


FIGURE 17 Variation of hardness values along half of the functionally graded materials plate thickness.

The variation of hardness along the transverse axis of the plate was measured from middle of the plate to the bottom of plate, Figure 17. The observed hardness variation along the transverse axis of the plate has a significant effect, suggesting a possible material anisotropy that can affect mechanical properties and structural integrity, requiring tailored engineering considerations to achieve optimal performance in specific positions or applications. The result showed that the bottom layers have the highest hardness in comparison to the middle. Since top and bottom layers have same composition, therefore, both have maximum hardness value, Table 4. The observed variation in hardness from the central layer to the top and bottom layers is purposeful. This gradation in hardness is strategically designed to maximize bending strength at the extremities. Specifically, the ascending order of hardness ensures that the material exhibits optimal resistance to bending stress at both the top and bottom layers [40]. Consequently, this tailored hardness distribution minimizes wear rates and preserves the material's toughness, making it an ideal choice for leaf springs. A comparative study of hardness was carried out, Table 5 [40–42]. Results depicted that marginal improvement was found when weight percentage of reinforcement was increased from 15% to 16%.

The results of the three-point bending test show the bending strength of the functionally graded leaf spring plate with different compositions of aluminium 7075 and

TABLE 4 Brinell hardness values for different weight percent of reinforcement.

	Brinell hardness values at different locations (BHN)					
Material composition	**1**	**2**	**3**	**4**	**5**	**Mean (BHN)**
Al 7075 + 8 wt.–% B4 C	137	145	141	143	140	141.2
Al 7075 + 12 wt.–% B4 C	175	170	162	160	166	166.6
Al 7075 + 16 wt.–% B4 C	201	190	210	203	208	202.4

TABLE 5 Comparative study of hardness of aluminium 7075 matrix composites.

Reference	Matrix material	Reinforcement material	Reinforcement percentage	Hardness (BHN)
[40]	Al7075	B4 C	15	198.4
[41]	Al7075	SiC + egg shell	15 + 5	223
[42]	Al7075	B4 C	6	84.2
Present work	Al7075	B4 C	16	202.4

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

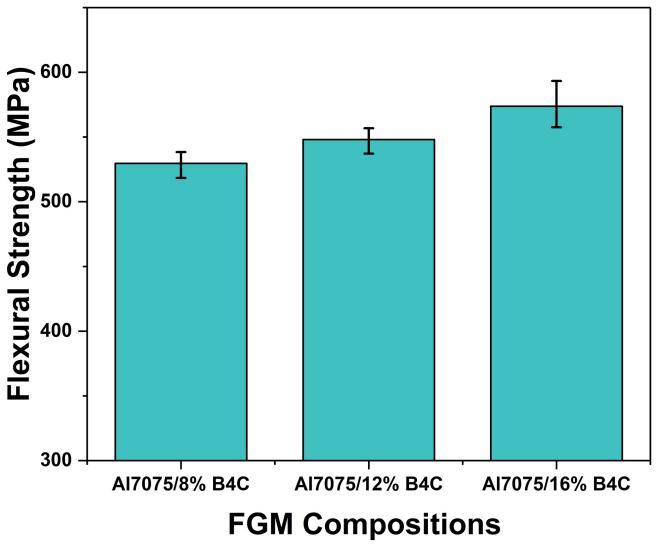


FIGURE 18 Flexural strength of different composition of functionally graded materials.

boron carbide. There is a remarkable trend where the increasing weight percentage of B4C corresponds to a continuous increase in flexural strength. Sample 3 with 16 wt.–% boron carbide shows the highest average flexural strength of the samples, Figure 18. The included error values provide valuable information about the reliability and consistency of the results and explain the robustness of the observed mechanical behavior.

5 | CONCLUSIONS

In the present work, a functionally graded leaf spring plate was fabricated using stir casting technique with five layered gradations in the mid portion. The matrix and reinforcement have been adopted from the output of the analysis conducted using CES EduPack 2019 software. The key findings are summarized as follows:

- There is a systematic increase in the amount of liquid required for the fabrication of the leaf spring from the bottom face to the top face. . The computational results align with the experimental readings, and the inclusion of an additional 2 percent allowance offers a practical weighting for potential variations.
- Stir casting technique is simple, suitable and cost effective technique which creates a gradation at central part of the leaf. This work is novel in its aspect, as similar findings are seldom reported in the literature.
- The hardness value increase with the increase of percentage of reinforcement. Since the top and bottom surfaces of the plate have the highest percentage of reinforcement, thus due to high hardness, both surfaces will have minimum wear.
- The increase in flexural strength with higher boron carbide weight percentages indicates that the reinforcement is effective in enhancing the composite material's mechanical properties.
- The addition of boron carbide as reinforcement refines the grain structure. The intensity of compressive residual stress increases with the increment of weight percentage of reinforcement. Consequently, it can confine crack propagation. Likewise, better fatigue properties are acquired by adding the boron carbide particles as the reinforcement into the aluminium matrix composite.

FUTURE SCOPE

Subsequent studies may focus more intently on gradation profile optimization to attain improved mechanical properties which in turn improves the flexural strength at specified locations as reported in the present work. Such initiatives would support the industry's goal of increased economy and sustainability by encouraging the development of robust and lightweight vehicle suspension system solutions.

ACKNOWLEDGEMENTS

The authors like to acknowledge Shree Vaishno foundry works at Ghaziabad for providing facilities related to casting of functionally graded materials leaf plate and samples. The authors also acknowledge the classic engineers, Faridabad, Haryana for providing temper 6 heat treatment facilities.

ORCID

R. K. Singh http://orcid.org/0000-0003-3776-1658

REFERENCES

1. D. Landau, P. H. Parr, *J. Franklin Inst.* **1918**, *185*, 481.
2. M. M. Shokrieh, D. Rezaei, *Compos. Struct.* **2003**, *60*, 317.
3. H. A. Al-Qureshi, *J. Mater. Process. Technol.* **2001**, *118*, 58.
4. Y. Yamada, T. Kuwabara, *Materials for springs*, Springer Science & Business Media, **2007**.
5. K. Tanabe, T. Seino, Y. Kajio, *SAE Trans.* **1982**, *91*,1628.
6. J. Wang, X. Deng, S. Du, F. Cheng, F. Li, L. Lu, H. Zhang, *InterCeram: Int.Ceram. Rev.* **2014** ,*63*, 286.
7. M. K. Surappa, *Sadhana* **2003**, *319*.
8. V. Khanna, V. Kumar, S. A. Bansal, *IOP Conf. Ser.: Mater. Sci. Eng.* **2021**, *1033*, 012057.
9. P. S. Bains, S. S. Sidhu, H. S. Payal, *Mater. Manuf. Processes* **2016**, *553*.

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10. J. M. Torralba, C. E. da Costa, F. Velasco, *J. Mater. Process. Technol.* **2003**, *203*.
11. A. Kumar, M. Y. Arafath, P. Gupta, D. Kumar, C. M. Hussain, A. Jamwal, *Mater. Today: Proc.* **2020**, *1417*.
12. M. C. Santos, A. R. Machado, W. F. Sales, M. A. Barrozo, E. O. Ezugwu, *Int. J.Adv.Manuf.* **2016**, *3067*.
13. G. B. Veeresh Kumar, C. S. P. Rao, N. Selvaraj, M. S. Bhagyashekar, *J. Miner. Mater. Charact. Eng.* **2010**, *43*.
14. J. Jung, S. Kang, *J. Am. Ceram. Soc.* **2004**, *47*.
15. M. Aizenshtein, N. Froumin, E. Shapiro-Tsoref, M. P. Dariel, N. Frage, *Scr. Mater.* **2005**, *1231*.
16. N. K. Shrestha, M. Kawai, T. Saji, *Surf. Coat. Technol.* **2005**, *2414*.
17. L. M. Kachanov, *Fundamentals of the Theory of Plasticity*, Dover Publications, INC. Mineola, New York, **2004**.
18. R. K. Singh, V. Rastogi, *Mater. Today: Proc.* **2021**, *3930*.
19. G. E. Knoppers, J. W. Gunnink, J. Van Den Hout, W.Van Vliet, Presented at *In Intelligent Production Machines and Systems: First I* PROMS Virtual Conference*, Elsevier, Amsterdam, December, **2005**, pp. 467–474.
20. A. G. Arsha, E. Jayakumar, T. P. D. Rajan, V. Antony, B. C. Pai, *Mater. Des.* **2015**, *1201*.
21. W. S. Ebhota, A. S. Karun, F. L. Inambao, *Int. J. Eng. Res. Afr.* **2016**, *47*.
22. A. O. Owoputi, F. L. Inambao, W. S. Ebhota, *Int. J. Appl. Eng. Res.* **2018**, *16141*.
23. K. Sarkar, R. Ganguli, *Composites, Part B* **2014**, *361*.
24. A. Majumdar, D. Das, *Proc. Inst. Mech. Eng., Part L.* **2018**, *769*.
25. H. Niknam, A. Fallah, M. M. Aghdam, *Int. J. Non-Linear Mech.* **2014**, *141*.
26. U. Eroglu, *Compos. Struct.* **2016**, *204*.
27. A. Pydah, A. Sabale, *Compos. Struct.* **2017**, *867*.
28. B. G. Scuracchio, N. B. de Lima, C. G. Schön, *Mater. Des.* **2013**, *672*.
29. N. K. Mukhopadhyay, S. K. Das, B. Ravikumar, V. R. Ranganath, S. Ghosh Chowdhury, *Eng. Failure Anal.* **1997**, *161*.
30. Z. Wang, T. K. Chen, D. J. Lloyd, *Metall. Trans. A* **1993**, *197*.
31. S. Taira, K. Tanaka, T. Yamasaki, *Japan Society of Materials Science* **1978**.
32. K. Tanaka, *Mech. Eng. Rev.* **2019**, *18*.
33. A. Sharma, V. Rastogi, A. K. Agrawal, *Trans. Indian Inst. Met.* **2020**, *2995*.
34. Lall, Pradeep, A. R. R. Pandurangan, K. Blecker, *J. Electron. Packag.ASME* **2024**, *1*.
35. S.Timoshenko, *Strength of Materials, 3E, Part 1*, CBS Publishers and Distributors Pvt. Ltd. **2021**.
36. Y. Kong, S. Abdullah, M. Omar, S. Haris, *Eng. Failure Anal.* **2016**, *146*.
37. S. Daronde, A. Kuthe, S. Keerti, R. Khatirkar, A. Bagde, M. Kamble, S. Dahake, *J. Mater. Eng. Perform.* **2022**, *262*.
38. P. K. Sahoo, S. Pattnaik, M. K. Sutar, *Mater. Sci. Forum* **2020**, *29*.
39. Q. Yang, L. Zhao, H. Xiao, N. Zhao, *Adv. X-Ray Anal.* **2000**, *151*.
40. A. Baradeswaran, A. E. Perumal, *Composites, Part B* **2013**, *146*.
41. S. Kumar, A. K. Srivastava, R. K. Singh, *Ann. Chim. Sci. Mat* **2020**, *295*.
42. T. H. Manjunatha, Y. Basavaraj, M. Nagaral, V. Venkataramana, J. I. Harti *IOP Conf. Ser.: Mater. Sci. Eng.* **2018**, *376*.

How to cite this article: R. K. Singh, R. C. Singh, *Materialwiss. Werkstofftech.* **2024**, *55*, e202300276. https://doi.org/10.1002/mawe.202300276

15214052, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mawe.202300276 by Delhi Technological University, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Check for updates

Received: 6 February 2024 | Revised: 14 June 2024 | Accepted: 31 July 2024

DOI: 10.1111/exsy.13697

REVIEW

Expert Systems WILEY

Prospect of large language models and natural language processing for lung cancer diagnosis: A systematic review

Arushi Garg[1] | Smridhi Gupta[1] | Soumya Vats[1] | Palak Handa[1,2] | Nidhi Goel[1]

[1]Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India

[2]Research Center for Medical Image Analysis and Artificial Intelligence, Department of Medicine, Danube Private University, Krems, Austria

Correspondence
Nidhi Goel, Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India.
Email: nidhi.iitr1@gmail.com

Palak Handa, Research Center for Medical Image Analysis and Artificial Intelligence, Department of Medicine, Danube Private University, Krems, Austria.
Email: palakhanda97@gmail.com

Abstract

Lung cancer, a leading cause of global mortality, demands a combat for its effective prevention, early diagnosis, and advanced treatment methods. Traditional diagnostic methods face limitations in accuracy and efficiency, necessitating innovative solutions. Large Language Models (LLMs) and Natural Language Processing (NLP) offer promising avenues for overcoming these challenges by providing comprehensive insights into medical data and personalizing treatment plans. This systematic review explores the transformative potential of LLMs and NLP in automating lung cancer diagnosis. It evaluates their applications, particularly in medical imaging and the interpretation of complex medical data, and assesses achievements and associated challenges. Emphasizing the critical role of Artificial Intelligence (AI) in medical imaging, the review highlights advancements in lung cancer screening and deep learning approaches. Furthermore, it underscores the importance of on-going advancements in diagnostic methods and encourages further exploration in this field.

KEYWORDS
comparative analysis, large language models, lung cancer, natural language processing

1 | INTRODUCTION

Lung cancer, a pervasive and formidable disease, poses a significant global health challenge with profound implications for public well-being (Islami et al., 2015). Characterized by the uncontrolled growth of abnormal cells in the lungs, it stands as a leading cause of cancer-related deaths worldwide. The World Health Organization's 2020 report highlighted approximately 2.2 million new cases and 1.8 million deaths from lung cancer, emphasizing the urgent need for enhanced diagnostic techniques and treatment modalities to combat its alarming mortality rates (Islami et al., 2015). Addressing this pressing need requires not only a paradigm shift in our approach but also the integration of cutting-edge technologies to improve early detection and treatment strategies.

Traditional diagnostic methods for lung cancer often face limitations in accuracy and efficiency, considering the low doctor-to-patient ratio and expert technicians across the globe, further underscoring the need for innovative solutions (Handa, Gupta, et al., 2023; Howlader et al., 2020; Islami et al., 2015; Kumari et al., 2023; Tang et al., 2023). Pioneering early detection methods are essential for improving patient outcomes and mitigating the profound socio-economic consequences associated with this disease. The comprehensive evaluation of lung health through diverse imaging modalities, such as high-resolution Computed Tomography (CT) scans and Positron Emission Tomography (PET) scans, plays a crucial role in detecting and characterizing lung nodules, particularly among high-risk populations (Collins et al., 2007; MacRedmond et al., 2004). Additionally, bronchoscopy and various biopsy methods provide valuable insights into specific types of lung cancer, guiding subsequent treatment decisions (Bunyaviroch & Coleman, 2006). Magnetic Resonance Imaging (MRI) offers intricate soft-tissue contrast and contributes to a comprehensive understanding of lung health when integrated with other imaging techniques (Arroliga & Matthay, 1993). High-resolution CT and PET scans offer valuable insights into lung health but are costly, difficult to interpret, and inaccessible to underdeveloped countries. Additionally, while informative, biopsy methods and bronchoscopy may be invasive and subject to sampling errors.

 © 2024 John Wiley & Sons Ltd.

Machine Learning (ML) algorithms have significantly augmented the diagnostic capabilities of radiologists, enabling more accurate identification of abnormalities in X-rays and CT scans (Mascaux et al., 2018). Deep learning (DL) techniques, particularly through Convolutional Neural Networks (CNNs), have demonstrated unprecedented proficiency in detecting lung tumours by discerning intricate patterns within medical images (Koyama et al., 2013). Large Language Models (LLMs), exemplified by models like Generative Pre-trained Transformer (GPT) 3.5, extend beyond imaging technologies, showcasing expertise in processing vast textual data, including medical literature and research papers (Travis et al., 2020). They provide comprehensive insights that enhance diagnostic accuracy by analysing diverse medical reports, including radiology and pathology findings and clinical notes. This holistic approach reduces errors and misinterpretations, thereby facilitating more precise diagnoses (He et al., 2023). Additionally, LLMs and NLP leverage patient records and genetic data to personalize treatment plans, ensuring optimal efficacy and patient satisfaction. Furthermore, their ability to swiftly retrieve pertinent information from extensive datasets enables healthcare professionals to make timely decisions, particularly in emergency situations. The integration of innovative technologies offers insights into the current state of the field. It sets the stage for future developments that promise to redefine the landscape of lung cancer research and improve patient outcomes (Naik et al., 2023).

Figure 1 illustrates the organizational structure of the review paper, providing a visual representation of its sections and sub-sections. It serves as a guide for navigating the review paper's contents and understanding its overall framework for the readers. Section 1 details the introduction, motivation, and contribution of the review paper. In the Section 2, the review meticulously explores global incidence and mortality rates, regional disparities, age distribution, and the diagnostic journey from initiation to after-care and treatment options. It delves into challenges faced during the diagnostic process and investigates the integration of AI, particularly Computer aided Design (CAD) systems, emphasizing their significance in analysing CT scans and detecting abnormalities using ML and DL techniques. Section 3 offers a detailed analysis of the evolution of LLMs, comparing models, and assessing their strengths and limitations. It examines the role of LLMs in the medical domain, with a focus on their potential applications in supporting healthcare professionals and augmenting medical decision-making processes. Furthermore, it is expanded into Section 4 which discusses the intersection of LLMs and lung cancer diagnosis, highlighting the emerging opportunities and challenges in leveraging these models for oncological research and clinical practice. It is followed by future research directions and conclusion in Section 5 and 6 respectively.

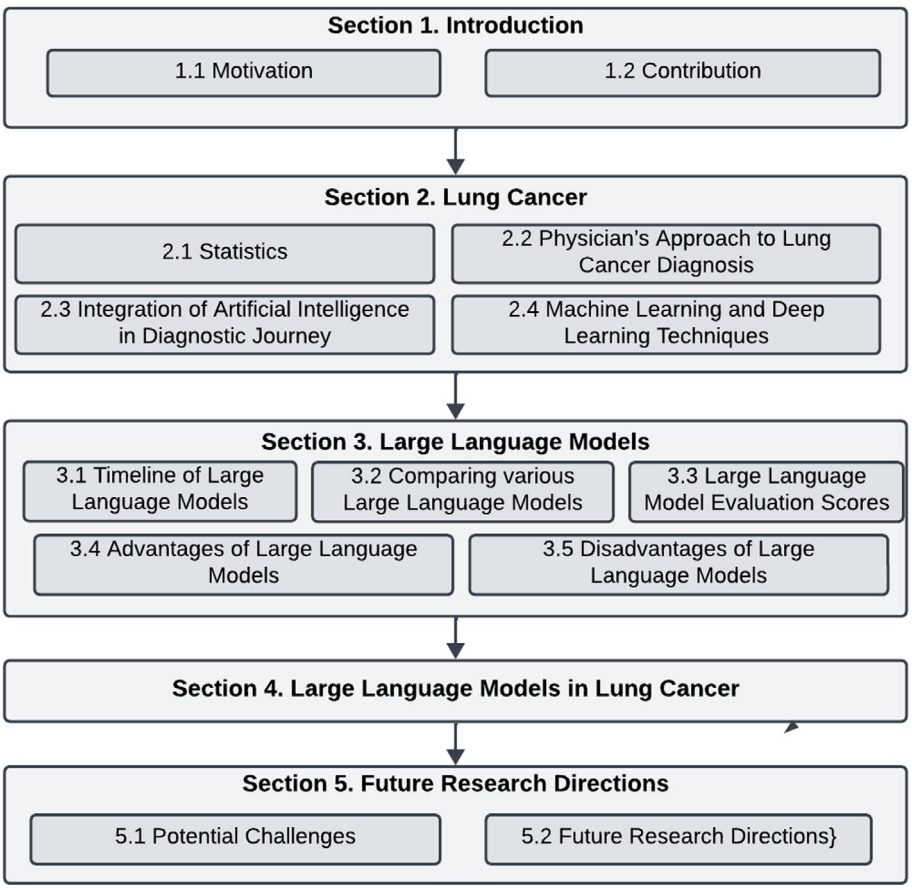


FIGURE 1 Illustration of the organizational structure of the systematic review paper, providing a visual guide for navigating its sections.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1.1 | Motivation

In the realm of lung cancer research, substantial efforts have been invested (Ali et al., 2023; Gayap & Akhloufi, 2024; Jassim & Jaber, 2022; Montazeri et al., 2021; Mridha et al., 2022; Mumtaz et al., 2024); however, the increasing prevalence of the disease highlights the need for a more thorough exploration of diagnostic and treatment techniques, as indicated in Gupta et al. (2023). To the best of author's knowledge, there is no systematic review on the prospect of LLMs and NLP for lung cancer diagnosis. Only one review paper has been published for NLP systems to diagnose and determine level of lung cancer (Montazeri et al., 2021). The authors have compared seven studies which worked on text-based data and reported in-sufficient results and the need of more studies in this field.

Now we discuss the existing reviews in this field to establish the motivation behind this systematic review. Interesting reviews like Ali et al. (2023), Gayap and Akhloufi (2024), Jassim and Jaber (2022), and Mridha et al. (2022) have been generated from the year 2021–2024 to highlight the importance of AI in the automatic detection of lung cancer. Mridha et al. (2022) focused on providing a comprehensive survey on the existing datasets, image preprocessing, segmentation methods, optimal feature extraction and selection methods, evaluation measurement matrices in classification, detection, and segmentation of lung cancer imaging up to the year 2021. Jassim and Jaber (2022) reported the existing works in lung nodule classification from 2016 to 2021. Gayap and Akhloufi (2024) demonstrated the advent of DL in lung cancer detection through 2D and 3D imaging. Another scoping review done by Ali et al. (2023), focused on the application of vision transformers for the diagnosis and prognosis of lung cancer patients. The authors have provided an interactive dashboard called the 'lung cancer dashboards' which consists of important papers, datasets, and statistics in this field for reading purposes (link).

A few scoping reviews have been published to showcase the importance of LLMs in the healthcare such as Mumtaz et al. (2024) and Rydzewski et al. (2024) but have not included the advent of lung cancer diagnosis. Rydzewski et al. (2024) focused on conducting comprehensive evaluation of LLMs in the field of clinical oncology with the help of a detailed questionnaire. Similarly, Mumtaz et al. (2024) has reported the current application of LLMs in various medical specialities like dentistry, dermatology, oncology, and mental health. In response to this imperative, our comprehensive investigation is set on the objectives to:

1. Emphasize the significance of early diagnosis: Recognizing the importance of early detection in managing lung cancer, we underscore the imperative for swift and accurate diagnosis. Detecting lung cancer early on significantly enhances treatment options and patient outcomes. Early diagnosis enables timely intervention, potentially reducing the severity of the disease and improving the effectiveness of treatment modalities. Moreover, it can mitigate the risk of complications and enhance the overall prognosis for individuals affected by lung cancer. Through our research endeavours, we aim to advance strategies for early detection, thereby bolstering efforts to combat this disease and improve the well-being of patients worldwide.

2. Systematic review: Our endeavour entails a meticulous and systematic review of existing literature and research findings about lung cancer. In addition to scrutinizing the breadth of available knowledge, we aim to delve into the various modalities utilized in diagnosing and treating lung cancer. This encompasses exploring the utilization of LLMs and NLP techniques within the field. Our review will focus on elucidating the potential of LLMs and NLP in enhancing the interpretation of complex medical data, thereby assisting healthcare professionals in making informed and timely diagnostic decisions.

3. Utilization of NLP and LLMs in early diagnosis: We recognize the potential of NLP techniques and LLMs in facilitating early lung cancer diagnosis. By leveraging NLP algorithms to analyse diverse sources of textual data such as electronic health records, medical literature, and patient reports, we aim to identify patterns, risk factors, and diagnostic indicators associated with the onset and progression of lung cancer. Comprehensively reporting the evaluation metrics needed to evaluate such AI pipelines. Furthermore, integrating LLMs, with their capacity for understanding, interpretation, and generating human-like text, can enhance the interpretation of complex medical data and assist healthcare professionals in making timely and accurate diagnostic decisions. Our research will explore the feasibility and efficacy of employing NLP and LLMs as complementary tools in early lung cancer detection and characterization, aiming to improve patient outcomes through proactive intervention strategies.

Through this multifaceted approach, we aspire to generate valuable insights that can inform and contribute to advancements in the understanding and management of lung cancer, ultimately aiming to enhance diagnostic accuracy, treatment efficacy, and overall research outcomes.

1.2 | Contribution

This review focuses on utilizing LLMs to detect, treat, and prevent lung cancer. The contributions of the review paper are as follows:

1. Role of AI-driven computer-aided diagnosis systems for advanced lung cancer diagnosis: The present review comprehensively delved into various aspects of lung cancer, including global statistics, the physician's diagnostic approach, the integration of AI in the diagnostic journey, and ML and DL techniques. Detailed insights into the prevalence, mortality rates, regional discrepancies, and age distribution of lung cancer cases worldwide have been provided. Furthermore, it elucidated the systematic process followed by physicians in diagnosing lung cancer, emphasizing the role of medical history review, physical examination, screening tests, diagnostic imaging, biopsies, and multidisciplinary consultations.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The review also explored how AI, mainly through CAD systems, enhances the interpretation of medical images like CT scans for early detection and how ML and DL techniques improve diagnostic precision through automated medical data analysis. Overall, it underscored the pivotal role of AI and ML/ DL technologies in advancing lung cancer detection, diagnosis, and treatment strategies.

2. Comparative analysis and critical evaluation of LLMs: The systematic review provides a comprehensive comparative analysis and critical evaluation of various LLMs in lung cancer applications with the help of various evaluation metrics. It examines their strengths and limitations and their detection, treatment, and prevention performance. By synthesizing insights on different LLMs, the review offers valuable guidance to researchers and practitioners, facilitating informed decisions in selecting the most suitable model for specific needs. This comparative analysis and critical evaluation provide a nuanced understanding of LLM effectiveness and potential challenges in lung cancer management, contributing to advancements in the field and improving diagnostic accuracy, treatment efficacy, and patient outcomes.
3. Comprehensive exploration of existing literature: The review meticulously examines and consolidates the extensive body of research concerning the intricate interplay between lung cancer and LLMs with the help of chord diagram and comparative analysis, enriching our understanding of their evolving role in the domain. Additionally, it goes beyond mere synthesis by elucidating potential avenues for future investigation, serving as a guiding beacon for researchers. By offering insights into untapped research directions and innovative applications of LLMs, the review propels the advancement of lung cancer detection methods and fosters continuous progress in the field, ultimately contributing to the collective efforts to combat this widespread disease.

Overall, the systematic review paper aims to provide valuable insights that can inform and contribute to advancements in the understanding and management of lung cancer. Its comprehensive analysis strives to enhance diagnostic accuracy, treatment efficacy, and overall research outcomes in the fight against this pervasive disease.

2 | LUNG CANCER

Lung cancer stands as a formidable challenge in the realm of global health, a relentless force claiming lives at an alarming rate and standing tall as a leading cause of mortality worldwide. Its prevalence extends beyond geographical boundaries, affecting diverse populations and transcending socio-economic disparities, making it a critical concern in the contemporary medical landscape. It ranks as the second most commonly diagnosed cancer and occupies the leading position in terms of global cancer mortality (Sung et al., 2021). The escalation of lung cancer cases has primarily been linked to the increase in cigarette smoking, a trend projected to have reached its apex in developed nations, such as the United States, during the latter half of the twentieth century (Doll & Hill, 1954). While lung cancer demonstrates some of the poorest survival rates among prevalent cancer types, significant variations exist in the 5-year survival rates among different countries. These rates range from below 10% in developing nations like India to a comparatively higher rate of 32.9% observed in Japan (Allemani et al., 2018).

Lung cancer comprises various histologically diverse sub-types that are crucial for a comprehensive understanding and effective management of the disease. Primarily classified as Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC), these delineations underscore the complexity of lung malignancies. 81% of lung cancer cases fall under non-small cell lung cancer, categorized into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Gridelli et al., 2015). SCLC constitutes around 14% of lung cancer cases, while the remaining diagnoses consist of lung carcinoid tumours and other less prevalent lung cancer varieties (Sharma, 2022).

While cigarette smoking is responsible for approximately 80% of lung cancers (Islami et al., 2018), a significant toll is incurred among individuals who have never smoked, with their mortality being ranked among the top 10 causes of cancer death when categorized separately (Siegel et al., 2023). The difficulty in early diagnosis is compounded by the absence of distinctive symptoms in the initial stages of the disease. Patients may remain asymptomatic or exhibit subtle signs easily mistaken for common respiratory ailments, thereby delaying diagnosis until the cancer has reached an advanced and less treatable phase. Coupled with this diagnostic challenge is the unfortunate stigma attached to lung cancer, often overshadowing efforts to raise awareness and secure early intervention.

One of the striking challenges posed by lung cancer lies in its elusive nature, often evading detection until advanced stages. Unlike some other forms of cancer with identifiable early symptoms, lung cancer frequently progresses silently, its presence concealed until it reaches an advanced and often incurable phase. This progression significantly hampers treatment efficacy, leading to limited survival rates and emphasizing the imperative of early detection strategies. Moreover, the absence of widely accessible and highly sensitive screening methods adds to the complexity of timely diagnosis. While advancements in imaging technologies and diagnostic tools have emerged, their widespread availability and cost-effectiveness remain obstacles in many regions, hindering the implementation of large-scale screening programs that could potentially detect lung cancer in its early, more manageable stages.

Furthermore, the intricate landscape of lung cancer adds layers of complexity, encompassing various histological subtypes and molecular alterations. This heterogeneity presents a formidable challenge in developing targeted therapies that address the diverse genetic mutations and cellular characteristics driving the disease. The pursuit of precision medicine tailored to individual patient's unique molecular profiles stands as a promising frontier, yet its full realization remains a work in progress.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Despite these formidable challenges, strides have been made in understanding the underlying mechanisms and risk factors associated with lung cancer. Research endeavours continue to unravel the intricate interplay of genetic predisposition, environmental exposures—such as tobacco smoke, radon, and air pollution—and other contributing factors, shedding light on potential avenues for prevention and intervention. In conclusion, lung cancer as a primary contributor to global mortality, underscores the pressing need for concerted efforts in early detection, innovative treatments, and widespread awareness campaigns. Overcoming the barriers posed by its elusive nature demands collaborative initiatives among healthcare professionals, researchers, policymakers, and communities worldwide. Only through a unified approach can strides be made toward mitigating the burden of lung cancer and enhancing outcomes for those affected by this formidable disease.

2.1 | Statistics

2.1.1 | Global incidence and mortality

The global burden of lung cancer persisted with an estimated 2.21 million new cases reported worldwide in 2020. This statistic underscores the disease's prominence as one of the most frequently diagnosed cancers on a global scale. Tragically, despite advances in healthcare, lung cancer retained its grim status as the leading cause of cancer-related mortality, claiming approximately 1.8 million lives in 2020 (Islami et al., 2018). This figure represents an alarming proportion, accounting for nearly 1 in 5 cancer-related deaths globally. With an estimated 1,369,000 cases, lung cancer stands as the second most prevalent cancer in men, following prostate cancer, and in women, following breast cancer, with 725,000 cases (Bray et al., 2018). The age-standardized cumulative lifetime risk of being diagnosed with lung cancer is 3.8% among men and 1.77% among women, as depicted in Figure 2 (Islami et al., 2018).

2.1.2 | Regional disparities

The prevalence of lung cancer demonstrates notable regional discrepancies, with East Asia emerging as the forefront region, recording an estimated 1.01 million incident cases and subsequent deaths, signifying a substantial burden. Following closely, northern America reported 253,537 cases. Within this context, absolute numbers accentuate the impact on a country-by-country basis. In 2020, China, the United States, and Japan emerged as the primary contributors to the global burden, representing a significant portion of incident cases: 815,563, 227,875, and 138,532, respectively (Sharma, 2022). Similarly, in terms of fatalities, these nations reported the highest counts: 714,699 in China, 138,225 in the United States, and 82,369 in Japan. These statistics underscore the intertwining of lung cancer prevalence with the economic and developmental status of nations (Sung et al., 2021) as represented in Figure 3.

2.1.3 | Age distribution

While lung cancer can affect individuals across all age groups, the risk escalates substantially after the age of 45 (Sung et al., 2021). Peak occurrences in both incidence and mortality were observed within the 65–69 years age bracket, with a subsequent decline in prevalence beyond this

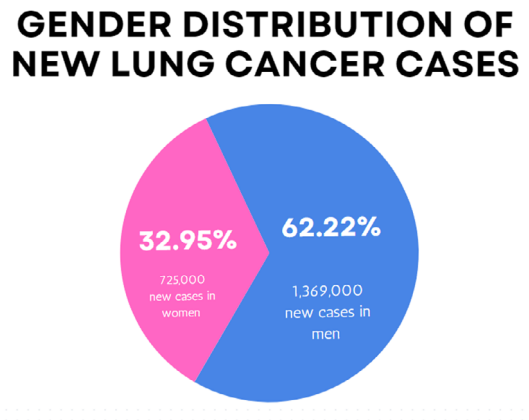


FIGURE 2 The gender distribution in newly reported cancer cases is illustrated by a pie chart, based on data from Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide. (Bray et al., 2018; Islami et al., 2018).

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

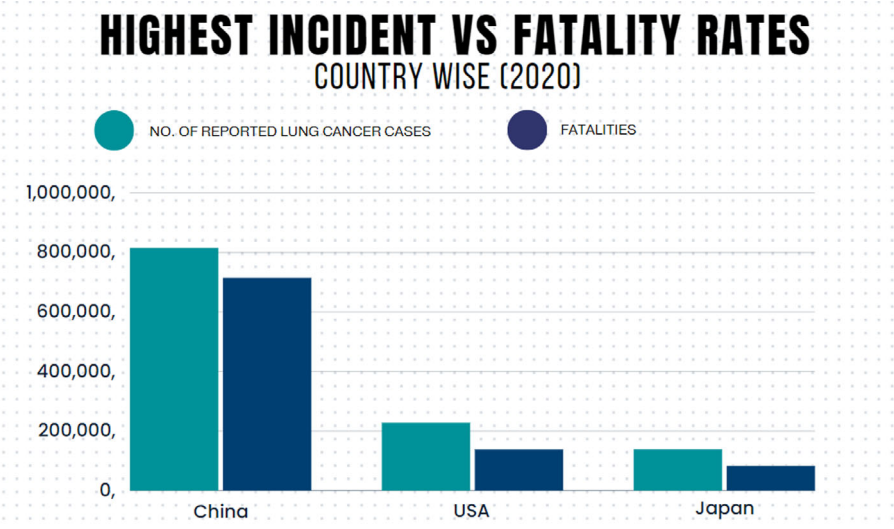


FIGURE 3 Bar graph representing the top 3 countries in the world with the highest reported number of incident cases of Lung Cancer and fatalities (Sharma, 2022; Sung et al., 2021).

demographic. Age categories below 45 years comprise only 2.4% of incident cases and 1.9% of deaths. Moreover, while historically more prevalent among men, The reduction in gender disparity concerning lung cancer incidence is attributed to shifting smoking patterns and exposure to other risk factors (Sung et al., 2021). These statistics, sourced from reputable global health organizations and epidemiological studies, elucidate the staggering magnitude of lung cancer's impact on a global scale. They serve as a sobering reminder of the urgent necessity for intensified research efforts, heightened awareness campaigns, and more effective interventions to combat this pervasive and deadly disease.

2.2 | Imaging modalities

The complexity of lung cancer diagnosis involves the utilization of various diagnostic modalities. Among these, a crucial role is played by CT scans, which are renowned for their high-resolution cross-sectional imaging capabilities. They are considered a cornerstone in the diagnostic process. They offer valuable insights into lung lesions, including the precise delineation of nodules, tumours, and their exact positions within the lung parenchyma (Gindi et al., 2014). Beyond mere identification, CT scans enable precise evaluation of lymph node involvement and the presence of metastases, pivotal in determining the extent of disease spread and guiding subsequent therapeutic decisions (De Bondt et al., 2007). In conjunction with CT scans, PET scans play a pivotal role in the diagnostic armamentarium. PET scans, when combined with CT imaging (PET/CT), provide a comprehensive assessment by evaluating metabolic activity. This amalgamation of metabolic data with anatomical details not only aids in distinguishing between benign and malignant lesions but also serves as a vital tool for accurate staging and the detection of distant metastases, crucial in devising a targeted treatment strategy (Silvestri et al., 2007).

Biopsy procedures, ranging from bronchoscopy to Fine Needle Aspiration (FNA), core needle biopsy, and surgical biopsy, hold paramount significance in confirming the diagnosis of lung cancer. These procedures, while invasive, are indispensable in obtaining tissue samples for histopathological scrutiny. Beyond confirmation, these biopsies provide critical insights into the histological characteristics of the tumour and its genetic profile, paving the way for personalized and targeted therapeutic interventions (Islam & Walker, 2013). While offering less detailed images than CT scans, chest X-rays retain their place as an initial screening modality for lung abnormalities. Despite their limitations in detecting smaller lesions, X-rays provide a rapid and cost-effective means of visualizing larger masses or nodules, initiating the diagnostic cascade for further investigation and confirmation (Sharma et al., 2009). The five major modalities used in lung cancer diagnostics are represented in Figure 4. The utilization of MRI as a diagnostic modality for lung cancer presents a promising avenue, particularly in contrast to PET/CT. The emerging PET/MRI technology offers enhanced soft-tissue contrast while substantially reducing radiation exposure. This improvement in imaging capabilities suggests the potential for acquiring additional pertinent information in lung cancer diagnosis (Henzler et al., 2012). From a technical standpoint, lung MRI exhibits promise as a viable screening tool. Evaluations have outlined a benchmark sensitivity of greater than 75% as an imperative criterion for an effective early detection program for lung cancer, underscoring the potential of MRI in meeting these critical diagnostic requirements (Sommer et al., 2014).

Collectively, these diagnostic modalities synergize to provide a comprehensive evaluation of lung cancer, aiding clinicians in accurate diagnosis, staging, and the formulation of personalized treatment strategies for patients affected by this complex disease.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

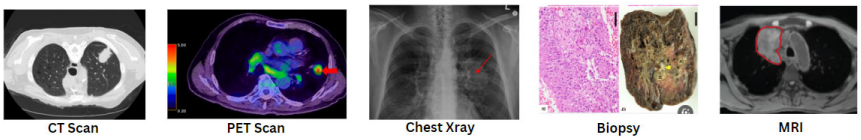


FIGURE 4 Pulmonary nodule on chest CT. Thin-section CT (Left) revealing a ground-glass dominant nodule highlighted within the red box. Thin-section CT (Right) displaying a lobulated contour of the pure ground-glass nodule.

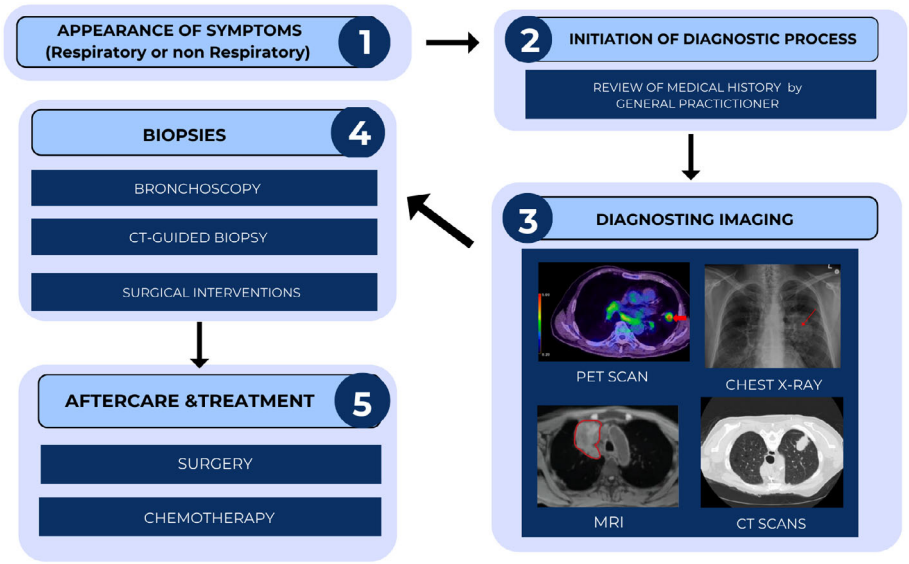


FIGURE 5 A flowchart showcasing the physician's approach to lung cancer diagnosis. Depicting the systematic steps and essential considerations, providing valuable insights into the precision and care integral to the diagnostic process.

2.3 | Physician's approach to lung cancer diagnosis

In this sub-section, we discuss the initial stages of lung cancer screening, the role of practitioners, modalities utilized, and after-care and treatment options. Figure 5 depicts a flowchart for the physician's approach to lung cancer diagnosis.

2.3.1 | Initiation of the diagnostic process

Upon the suspicion of the presence of lung cancer, a methodical exploration to confirm or dismiss the diagnosis is undertaken by healthcare professionals. This systematic journey begins with a comprehensive review of the patient's medical history. A meticulous examination delves into symptom analysis, past medical records, lifestyle habits (with a particular emphasis on smoking history), and familial predispositions to cancer. The process unfolds with a thorough physical examination, including the auscultation of lung sounds and a meticulous assessment for any signs of respiratory distress or anomalies.

2.3.2 | Role of the general practitioner

In the early stages of the diagnostic trajectory, the pivotal figure in this investigative process is the General Practitioner (GP). Acting as the primary point of contact, the GP orchestrates the initial consultation, where a profound examination takes place (Guldbrandt et al., 2017). The GP's discerning role involves recognizing potential indicators of lung cancer, which prompts further steps in the diagnostic journey. This collaborative effort with the patient lays the foundation for subsequent screenings and diagnostic procedures.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3.3 | Screening for potential abnormalities

In response to symptoms or identified risk factors, the GP may recommend screening tests, such as chest X-rays, as part of the preliminary investigative phase. This facilitates the identification of potential abnormalities within the lung tissue. In certain cases, expedited access to secondary care is arranged, involving diagnostic radiology and biopsy procedures while the patient is still hospitalized. These screenings serve as the initial step in confirming or ruling out the presence of lung cancer (Walton et al., 2013).

2.3.4 | Diagnostic imaging

Initiated with chest X-rays, diagnostic imaging serves as the cornerstone for unveiling visible abnormalities within the lung tissue. While chest X-rays provide an initial overview, the diagnostic journey advances with the integration of more sophisticated imaging modalities, notably CT scans and PET scans. CT scans, with their unparalleled resolution, delve into intricate details of lung structures, detecting lesions as small as 2–3 mm (Folio, 2012). The addition of contrast during CT scans enhances the differentiation of tissues, spotlighting potential cancerous lesions and their surrounding environment. PET scans, on the other hand, reveal not just structural abnormalities but also cellular activity, offering a functional perspective crucial for comprehensive diagnostics. This amalgamation of advanced imaging modalities ensures a thorough evaluation, enabling healthcare professionals to precisely identify and characterize suspicious nodules or masses, laying the groundwork for targeted investigations.

2.3.5 | Biopsies

When imaging unveils potential lesions or nodules requiring further scrutiny, the diagnostic trajectory advances to tissue sampling through diverse biopsy techniques. Minimally invasive approaches, such as bronchoscopy, enable direct visualization and sampling from the airways, minimizing patient discomfort (Shen et al., 2021). Alternatively, CT-guided biopsies offer pinpoint accuracy by guiding the biopsy needle directly to the targeted lesion (Arnolli et al., 2015). Surgical interventions, including thoracotomies and Video-Assisted Thoracoscopic Surgeries (VATS), become imperative for inaccessible or larger lesions, ensuring comprehensive tissue procurement. This biopsy phase not only provides histo-pathological insights but also integrates molecular and genetic testing. These advanced analyses unravel specific genetic mutations and biomarkers within cancerous tissue, steering the diagnostic process toward a personalized understanding of the disease. This intricate interplay of imaging and biopsy procedures enhances diagnostic precision, guiding subsequent therapeutic strategies tailored to the individual patient (Tam et al., 2016). After tissue sampling, staging procedures utilizing CT scans, PET scans, or MRIs are implemented to ascertain the extent and potential spread of the cancer within the body. This comprehensive evaluation culminates in a multidisciplinary consultation involving oncologists, pulmonologists, radiologists, and pathologists. The collective expertise converges to decipher the most appropriate treatment trajectory, meticulously tailored to the cancer's stage, sub-type, molecular characteristics, and the patient's circumstances (Tod et al., 2008).

2.3.6 | After-care and treatment options

If the diagnosis reveals lung cancer, the subsequent steps depend on the staging and characteristics of the cancer. After-care often involves determining the most suitable treatment plan, and the primary options include surgery, chemotherapy, and other targeted therapies. If the tumour is localized and surgery is a viable option, it may involve the removal of the tumour or even a portion of the lung. Surgical procedures can vary, from lobectomy (removing a lobe of the lung) to pneumonectomy (removing the entire lung) (Alexander, 1935). The decision for surgery depends on factors such as the size and location of the tumour and the overall health of the patient. For cancers that have spread beyond the lungs or are inoperable, chemotherapy becomes a common treatment option (Lemjabbar-Alaoui et al., 2015). Chemotherapy uses drugs to kill cancer cells or slow their growth. In recent years, targeted therapies that focus on specific molecules involved in cancer growth have shown promise. These therapies are often discussed with an oncologist who specializes in cancer treatment.

2.3.7 | Challenges in diagnostic journey

The diagnostic journey for lung cancer involves coordinating multiple specialists, leading to potential delays. Efficient communication and collaboration among specialists are crucial for optimizing treatment plans and ensuring a streamlined patient experience. This orchestrated process, initiated by the GP's assessment, sets the stage for a comprehensive and methodical diagnostic journey aimed at achieving an accurate and timely lung cancer diagnosis.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.4 | Integration of artificial intelligence in diagnostic journey

In the dynamic landscape of medical diagnostics, the advent of AI has ushered in trans-formative changes, notably in the realm of lung cancer diagnosis. A pivotal player in this paradigm shift is the implementation of CAD systems, particularly for the interpretation of CT scans (Lee et al., 2009). These advanced systems leverage intricate ML algorithms to scrutinize vast datasets, effectively replicating and enhancing the diagnostic processes traditionally performed by healthcare professionals.

2.4.1 | Why computer aided diagnosis?

CAD is a groundbreaking procedure in the realm of medical information, playing a pivotal role in aiding healthcare professionals in the interpretation of medical images. The advent of CAD for lung CT images represents a revolutionary stride toward the early and timely detection of pulmonary abnormalities (Parveen & Kavitha, 2012). It functions as a valuable supplementary perspective when medical practitioners conduct examinations as part of lung cancer screening (El-Baz et al., 2013). Employing sophisticated image processing algorithms and techniques, these systems enhance the clarity and distinctiveness of medical images. The subsequent isolation of ROI within the images enables a focused analysis for the detection of nodules, thereby facilitating disease diagnosis (Khawaja et al., 2004; see Figure 6). An eminent advantage of CAD systems lies in their efficiency in image analysis. Through rapid scanning and interpretation of CT scans, these systems enable the swift identification of potential abnormalities in lung tissues (Shariaty & Mousavi, 2019).

The incorporation of AI algorithms significantly elevates sensitivity and specificity, allowing for nuanced detection of suspicious nodules or masses (El-Baz et al., 2013). Importantly, CAD systems contribute to the standardization of the diagnostic process, ensuring uniform criteria and reducing variability across different healthcare settings. This not only enhances the reliability of diagnoses but also establishes a consistent framework for medical professionals. Furthermore, CAD systems exhibit a remarkable ability to seamlessly integrate data from various imaging modalities, including CT scans. This integration enhances their capacity to discern intricate details and patterns that might elude conventional diagnostic methods. The synergy between advanced image processing techniques and AI algorithms equips CAD systems with the potential to revolutionize the landscape of medical imaging, particularly in the realm of early detection and diagnosis of lung abnormalities (El-Baz et al., 2013).

2.4.2 | CAD system for CT scans

Within medical practice, CT stands as a widely recognized and responsive imaging technique utilized for the detection and diagnosis of pulmonary nodules. Its distinct advantages encompass superior spatial resolution, cost-effectiveness, widespread availability, and non-invasiveness. Moreover, CT scans reign as the predominant modality employed for the detection of lung cancer due to their unparalleled efficacy and reliability in this specific diagnostic domain (Mhaske et al., 2019). CT scanning proves highly effective in identifying pulmonary nodules and micro-calcifications due to its capacity to generate three-dimensional images of the chest cavity. These images offer high-resolution views of nodules, micro-calcifications, and pathological tumours. The utilization of CT scans in hospitals for diagnosing pulmonary nodules through computer-processed imaging has become prevalent in medical practice. Radiologists must examine diverse attributes such as length, width, location, texture, and shape, among others, in the nodules and micro-calcifications depicted in CT scan images to effectively perform the diagnostic process. The conventional

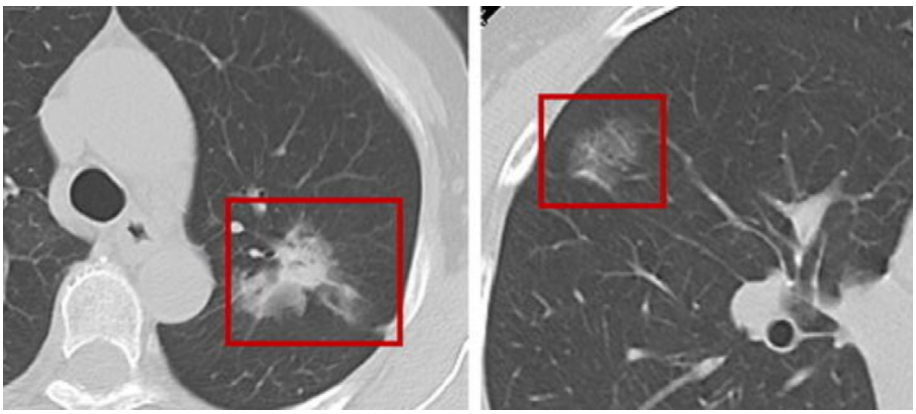


FIGURE 6 Pulmonary nodule annotated in a chest CT scan (Tsou et al., 2015).

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

approach for detecting and classifying micro-calcifications typically involves a sequence of steps, starting with enhancement, followed by segmentation, subsequent feature extraction, and concluding with classification (Maurer, 2021).

Numerous image processing techniques prove advantageous in identifying micro-calcifications and categorizing cancer into benign and malignant types, utilizing features related to texture and shape. For detection, the focus is specifically on utilizing lung CT images to detect the existence of micro-calcifications. Micro-calcification refers to minuscule calcium deposits that develop in a manner indicative of potential cancer (Lanyi, 2000). These calcifications are categorized into benign or malignant based on characteristics such as shape, size, boundaries, distribution pattern, and the presence of calcium compounds. For instance, benign cancer often arises from calcium oxalate, whereas malignant form is associated with calcium phosphate. Through a review of existing literature, it's evident that micro-calcification stands out as a significant factor contributing to certain cancerous conditions (Chaudhury et al., 2021). The CAD system relies on training from extensive datasets to learn patterns and features associated with both normal and abnormal lung tissues. This training phase is crucial for enabling the algorithm to discern subtle differences signifying the presence of lesions. Subsequently, the algorithm classifies lesions based on predefined criteria, distinguishing between benign and potentially cancerous abnormalities.

2.4.3 | Picking regions of interest and image processing

The extraction of ROIs from masked CT scan images is a pivotal step in lesion classification. This involves the meticulous application of segmentation techniques to identify and isolate areas of interest within the medical images (Hu et al., 2020). Specifically, in the context of lung cancer diagnosis, the algorithm is trained to recognize features such as size, shape, and density variations associated with potential lesions. After identifying the ROI, the CT scan undergoes the process of segmentation, separating the objects from their surroundings. Subsequently, features are extracted from these segmented objects, encompassing various characteristics such as shape, texture, and statistical measures, laying the foundation for further analysis. The higher intensity of micro-calcification compared to other lung zones prompts the need to identify it as a specific ROI. This involves utilizing image processing methods to differentiate these calcifications from surrounding tissues. Understanding the distinct pattern of these calcifications becomes imperative for in-depth investigation and the classification of anomalies observed in the lungs.

2.4.4 | Role of neural networks in nodule detection

A neural network is designed to replicate the actions of a radiologist when processing an image. The detection of lung nodules emulates the radiologist's initial task of scanning the chest radiograph to identify potential areas of concern. Subsequently, the network meticulously scrutinizes each suspected area by evaluating the contrast information of the bright spot in its surrounding background (Husain & Shigeru, 2000). Initially, the computer scans the perimeter of the specified area. Employing both boundary detection and sphere template matching methods aids in identifying potential nodules. The former enhances the detection of circular entities, while the latter focuses on identifying objects that are circular or near-circular. This approach termed the double-matching technique, aims to mimic the initial search conducted by radiologists for circular objects. However, these techniques solely replicate the radiologists' primary search for round shapes.

After the ROI is carved, the CT scan undergoes the process of segmentation. This segmentation procedure separates the objects from their surroundings. Subsequently, features are extracted from these segmented objects, encompassing various characteristics such as shape, texture, and statistical measures, laying the foundation for further analysis. The extracted features are utilized as inputs for diverse ML models, including linear discriminant analysis (Fukunaga, 1993), quadratic discriminant analysis (Fukunaga, 1993), multilayer perceptron (Rumelhart et al., 1986), and Support Vector Machines (SVM) (Vapnik, 1999). These models, in turn, are trained by providing them with sets of input features alongside their corresponding class labels.

2.4.5 | Machine learning models and their training

During the training phase, the ML models absorb knowledge from the provided input features and class labels. Notably, when a training sample belongs to a particular class, the relevant output unit is assigned a label of 1, while other output units receive a label of 0. Upon the completion of the training phase, the system is poised to classify unknown samples. Input features of an unknown sample are introduced to the trained models. The classification of the unknown sample is determined by identifying the output unit exhibiting the highest value, indicating the predicted class to which the sample is likely to belong Suzuki (2012). Figure 7 depicts a step-wise method for automatic classification using ML and DL algorithms. The amalgamation of advanced imaging technologies, image processing methodologies, and ML algorithms represents a promising paradigm shift in pulmonary nodule detection and classification. As these technologies continue to evolve and integrate, their collective impact promises enhanced diagnostic accuracy, streamlined workflows, and improved patient outcomes within the realm of lung cancer diagnostics and care.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

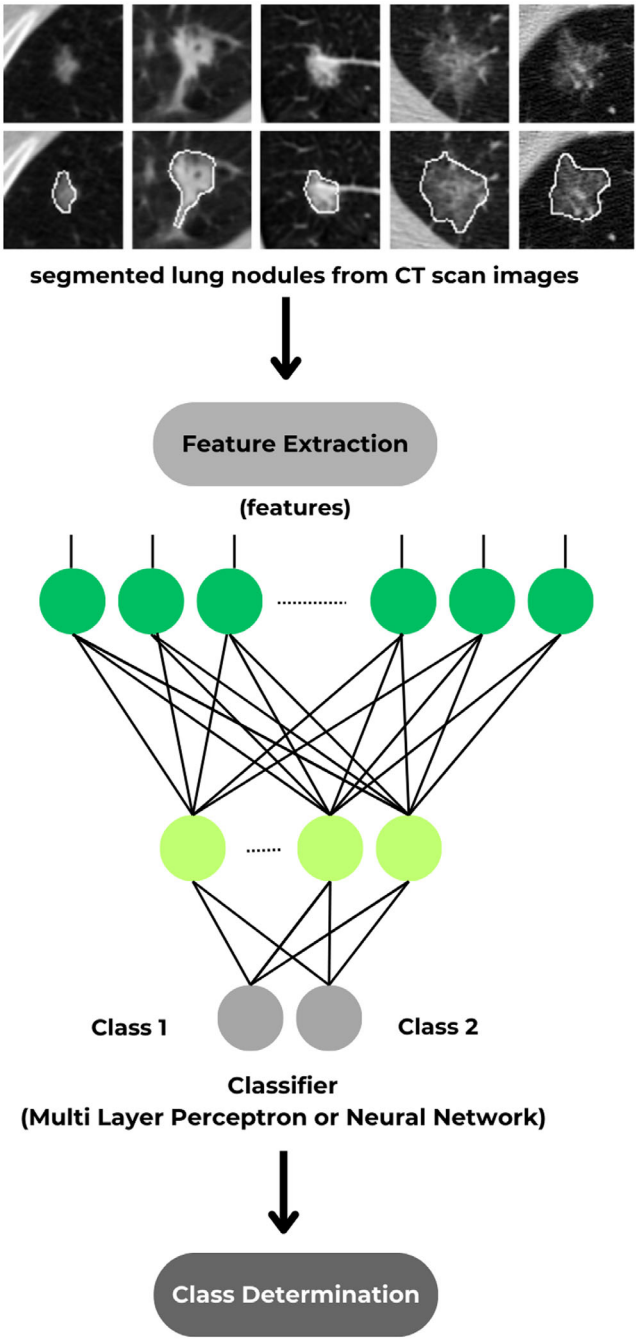


FIGURE 7 The conventional classification method involves extracting features such as texture, diameter, and size from the segmented image. These extracted features are then fed into the classifier, utilizing suitable machine learning algorithms to determine the class.

2.4.6 | Advent of deep learning techniques

In recent years, the progression of AI, particularly in DL, has exerted a substantial influence on both business and academic domains, notably reshaping radio-logical research. This evolution has significantly impacted the approach to clinical imaging challenges, with discernible instances where AI models have demonstrated superiority over human clinicians (Ardila et al., 2019). Within this evolving landscape, CAD has emerged as an indispensable tool aiding radiologists in the evaluation of lung cancer tumours (Zheng et al., 2021). Despite the increased utilization of CAD systems for nodule detection, their initial limitations in accuracy prompted a shift toward ML algorithms. Researchers aimed to address the distinction between benign and malignant pulmonary nodules in CT images, employing diverse ML methodologies such as K-Nearest Neighbour (KNN) (Ajitha et al., 2022), Decision trees (Preethi & Ganapathy, 2022), and SVM (Kareem et al., 2021) as classifiers. However, given the accelerated advancements in AI, there has been a substantial reorientation toward the adoption and refinement of DL techniques. Early endeavours in

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ML for lung cancer detection, as demonstrated by Ardila et al. (2019) laid the foundation by employing feature-based ML algorithms on chest CT scans. The pivotal shift toward DL was observed in the work by Nam et al. (2019), which highlighted the potential of CNNs in detecting pulmonary nodules on radio-graphs, marking a significant milestone in DL-based approaches for lung cancer detection.

Dataset creation and annotation played a crucial role in the progress of DL models. Setio et al. (2017) significantly contributed to this area by introducing the Lung Image Database Consortium (LIDC) (link) and the National Cancer Institute (NCI) dataset (link), providing essential resources for training and evaluating DL models for nodule detection. Today the LIDC dataset stands as the most renowned and esteemed repository of annotated lung cancer images globally. Its status as the leading dataset is attributed to its comprehensive collection of meticulously annotated CT scans, fostering invaluable advancements in lung cancer research and diagnostic methodologies. There are a few other datasets like the Cancer Genome Atlas – Lung Adenocarcinoma and Lung Squamous Cell Carcinoma (TCGA-LUAD and TCGA-LUSC) dataset (link), Early Lung Cancer Action Project (ELCAP) dataset (link) (Henschke et al., 1999), SPIE-AAPM Lung CT Challenge dataset (link), Tianchi Alibaba competition dataset (link), National Lung Screening Trial (NLST) dataset (link), RIDER Lung CT dataset (link), and the OncoSG NSCLC Cohort dataset (Prazanowska & Lim, 2023). Interested readers may refer link for more dataset specifications. Tool kits like DICOM, pylidc, CaPTk (Cancer Imaging Phenomics Toolkit), 3D Slicer, and ITK-SNAP help in processing 2D, 3D images and develop their corresponding masks, perform registration, and segmentation.

Ali et al. (2018) explored reinforcement learning for nodule detection, showcasing the potential of adaptive DL models for improved accuracy. Tang et al. (2019) presented NoduleNet, an integrated system crafted to adeptly manage pulmonary nodule screening, minimize false positives, and execute segmentation comprehensively. In parallel, Gruetzemacher and Gupta (2016) employed Deep Neural Network (DNN) designs encompassing all four types of convolutional layers. In contrast, Dou et al. (2017) introduced an automated detection system for pulmonary nodules utilizing CNNs. Their approach involved three specific systems: a conventional deep network, a residual network, and a hybrid-loss residual network. Moreover, Zhang et al. (2018) focused on leveraging spatial pyramid pooling in DL architectures to improve the classification accuracy of lung nodules. Additionally, Xia and Zhang (2023) introduced a novel deep learning-based graph model network, leveraging PET/CT imaging modalities to automatically fuse dual-modality information for tumour segmentation, demonstrating effective integration and improved segmentation accuracy. Furthermore, Winkels and Cohen (2018) demonstrate the efficacy of 3D group convolutions (G-Convs) in enhancing CNNs for pulmonary nodule detection, showcasing improved performance with less annotated data. Additionally, Zhao et al. (2023) introduced an enhanced Faster R-CNN model with multi-scale training strategies, path augmentation, and de-formable convolutions for improved pulmonary nodule detection.

Furthermore, after the rapid advancements in ML and DL techniques in lung cancer diagnosis, the exploration of new technologies, such as LLMs, has become prominent. The following section transitions to an in-depth examination of LLMs, to elucidate their significance and potential applications. This will involve a comprehensive overview of LLMs to provide insights into their functionality and utility, followed by a specific focus on their relevance in lung cancer diagnosis.

3 | LARGE LANGUAGE MODELS

An LLM is a sophisticated manifestation of ML, intricately designed to comprehend and generate human language through intricate neural networks (Tang et al., 2023). Its core functionality lies in assessing the probability of a specific word following a given input within a sentence, achieved through extensive training on vast textual datasets known as corpora. This training regimen results in a pre-trained language model, equipped to understand and generate coherent human-like language. These models exhibit versatility through fine-tuning, enabling applications in diverse practical domains like translation, law, and medicine (Handa, Chhabra, et al., 2023; Song et al., 2023), showcasing the concept of transfer learning (Ge et al., 2023). The depth of an LLM is closely tied to its architectural complexity. Built upon a foundational architecture resembling interconnected neural networks observed in the human brain, LLMs boast a substantial number of parameters, exemplified by models like GPT with over 100 billion parameters (Ji et al., 2023). The training process involves exposing these models to copious amounts of unlabelled text data, primarily through self-supervised or semi-supervised learning paradigms. Self-supervised learning, particularly advantageous, empowers models to extract insights from un-annotated text, circumventing the resource-intensive reliance on manually labelled data (Hu et al., 2023).

Empirical evidence supports the superior performance of larger networks with more parameters, exhibiting enhanced capacity to retain information and discern intricate patterns compared to smaller counterparts (Tu et al., 2023). Despite the computational challenges and substantial energy requirements associated with training LLMs, ongoing efforts aim to introduce more compact yet optimally trained models, ensuring sustainable and efficient deployment (Wei et al., 2023). While LLMs excel in tasks like predicting the next word in a sentence, their remarkable ability to capture language structure and meaning is noteworthy, marking a trans-formative epoch in NLP and AI (Han et al., 2023). The impact of LLMs extends into the broader landscape of AI and human-machine interaction, prompting researchers and developers to enhance their capabilities and address associated challenges (Eloundou et al., 2023). Notably, the interpretability and explainability of LLMs are significant areas of exploration. Despite unparalleled performance, these models often operate as black boxes, prompting active research to unravel decision-making processes, promoting transparency, and aiding users in understanding predictions (Hosseini & Horbach, 2023). Bias in LLMs is a focal point of discussion, as

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

these models may inherit and propagate biases present in training data. Efforts are directed toward developing mechanisms to mitigate bias, promote fairness, and ensure ethical deployment (Joublin et al., 2023).

Computational demands of training large models are under refinement, with researchers exploring novel architectures and optimization techniques to balance model size, performance, and environmental sustainability (Li, Su, et al., 2023). Efficient training approaches, such as sparse neural networks or knowledge distillation, are investigated to reduce computational resources while preserving efficacy (Makridakis et al., 2023). In application, LLMs contribute to fields like healthcare for medical diagnosis and research, as well as the development of intelligent virtual assistants, chatbots, and language interfaces. Collaborative efforts foster the development of benchmark datasets and evaluation metrics for assessing LLM performance, promoting healthy competition, and innovation (Zhang, Zheng, et al., 2023). The trajectory of LLMs involves pushing linguistic capabilities and addressing broader societal implications, with discussions on regulation, responsible AI practices, and societal impacts gaining momentum (Singha et al., 2023). Striking a balance between innovation and ethical considerations is crucial to ensure positive contributions to society without perpetuating harm (Singha et al., 2023). In essence, LLMs represent a pivotal chapter in the ongoing narrative of AI, with continued refinement and responsible deployment holding the potential to transform technology integration into our daily lives.

3.1 | Timeline of large language models

The evolutionary trajectory of LLMs within the domains of AI and NLP has witnessed significant milestones. Commencing in the 1960s, the journey traces its origins to the rudimentary chatbot Eliza, developed by MIT researcher Weizenbaum (1966). Despite Eliza's simplicity, it laid foundational concepts for subsequent language processing advancements (Shum et al., 2018; Weizenbaum, 1966). Notable progress occurred in the late 1990s with the introduction of Long Short-Term Memory (LSTM) networks, enhancing neural network capabilities for handling extensive datasets and contextual understanding (Staudemeyer & Morris, 2019; Yu et al., 2019). The Stanford CoreNLP suite in the early 2010s contributed to sentiment analysis and named entity recognition, augmenting NLP sophistication. A pivotal moment in LLM evolution transpired in 2017 with the introduction of transformer models (Han et al., 2021). Google Brain's Transformer model showcased the efficacy of attention mechanisms, enabling models to focus on specific input sequence segments during output generation (Gillioz et al., 2020; Tetko et al., 2020). This marked the foundation for subsequent influential models such as GPT and Bidirectional Encoder Representations from Transformers (BERT). The paradigm shifted in 2018 with a focus on developing larger models (Ravichandiran, 2021). Google's BERT, unveiled in 2019 with 340 million parameters, became versatile for diverse NLP tasks (Sur, 2020). However, the trans-formative GPT-3 in 2020, with a colossal 175 billion parameters, showcased unprecedented capabilities across tasks like language translation and code generation (Devlin et al., 2018).

Capitalizing on GPT-3's success, GPT-4 emerged in 2023 with a staggering one trillion parameters, approximately five times larger than its predecessor and vastly surpassing BERT's scale (Nori et al., 2023). GPT-4 incorporated computer vision capabilities, demonstrating the industry's commitment to pushing neural network architecture boundaries (Peng et al., 2023). A novel feature called ‘steerability’ introduced in recent LLMs, allows users to customize model behaviour, enhancing adaptability for diverse applications (Mao et al., 2023). Despite the trend toward larger models, alternatives like DeepMind's Retrieval-Enhanced Transformer (RETRO) demonstrated superior performance in certain tasks, showcasing the potential of smaller, more efficient models (Norlund et al., 2023). Post-GPT-4, recognition of practical constraints in data centres has led to a shift in focus toward exploring the effectiveness of smaller LLMs (Thoppilan et al., 2022).

The LLM landscape extends beyond GPT, with contributors like Google's LaMDA, Microsoft's Copilot, NVIDIA's Megatron, and Meta's initiatives (Bird et al., 2022; Shoeybi et al., 2019), each offering unique contributions to the field. The timeline of LLM development highlights critical milestones, including self-attention mechanisms, multi-headed attention, and the transition from LSTM to transformer architectures. Research focus has expanded to encompass human-preference alignment, external information retrieval, mixture-of-expert models, and multi-modality (Rajbhandari et al., 2020). LLM applications are gaining prominence across sectors such as content creation, conversational AI, healthcare, and finance. In conclusion, the progression from GPT-1 to GPT-4 signifies the continual evolution of LLMs from nascent stages to formidable problem-solving tools (Yang et al., 2022). The integration of computer vision, exploration of steer-ability, and emphasis on customization represent the forefront of current research. Researchers strive to unlock the full potential of large-scale language models while navigating challenges and ethical considerations inherent in their development and deployment. Table 1 offers a comparative overview of various LLMs. This overview offers insights into the landscape of LLM development, highlighting the diverse entities involved and the evolution of these models over time. Additionally, the table examines critical parameters such as model size, number of parameters, and training data size, providing a comprehensive overview of each LLM's architecture and capabilities. Furthermore, it evaluates the open-sourcing status and fine-tuneability of each model, shedding light on their accessibility and adaptability for different applications. Figure 8 provides a comprehensive timeline of LLM evolution. The evolution of LLMs traces back to the inception of NLP in 1966, progressing through decades of research and innovation. Breakthroughs such as GPT-3, introduced in 2023, epitomize the transformative journey of language understanding and generation. These models have revolutionized NLP by leveraging vast amounts of data and sophisticated neural architectures, enabling remarkable capabilities in tasks like translation, summarization, and question-answering. With each iteration, LLMs continue to push the boundaries of linguistic comprehension and interaction, marking significant milestones in the evolution of artificial intelligence.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Comparative Study on Various LLMs exploring the organization and date of launching along with the parameters. Open sourcing and fine-tuneability are also compared (Kumari et al., 2023).

Model	Organization	Date	Parameters	Open source	Fine-tuneability
ELMo	AI2	February 2018	94,000,000	Yes	Yes
GPT	OpenAI	June 2018	110,000,000	Yes	No
BERT	Google	October 2018	340,000,000	Yes	Yes
XLM	Facebook	January 2019	655,000,000	Yes	Yes
GPT-2	OpenAI	March 2019	1,500,000,000	Yes	Yes
RoBERTa	Facebook	July 2019	355,000,000	Yes	Yes
Megatron-LM	NVIDIA	September 2019	8,300,000,000	Yes	Yes
T5	Google	October 2019	11,000,000,000	Yes	Yes
Turing- NLG	Microsoft	February 2020	17,000,000,000	Yes	No
GPT-3	OpenAI	June 2020	175,000,000,000	No	No
PaLM	Google	April 2022	540,000,000,000	Yes	Yes
GPT-4	OpenAI	March 2023	—	No	No

3.2 | Comparing various large language models

3.2.1 | ChatGPT

The deployment of ChatGPT by OpenAI signifies a notable progression in conversational AI, employing advanced algorithms for seamless human-AI interactions. Utilizing Reinforcement Learning from Human Feedback (RLHF), ChatGPT, introduced on November 30, 2022, incorporates adaptive learning facilitated by a GPT-based architecture with 175 billion parameters (Ray, 2023). Its pricing model includes a complimentary usage tier and a $20 per month subscription, providing enhanced access and response times (Wu, He, et al., 2023). This tiered approach aligns with OpenAI's strategy of offering additional features to subscribers. ChatGPT's reliability stems from continuous evolution via RLHF, enabling iterative learning and adaptation based on user input (Kalla & Smith, 2023). Despite not achieving perfect accuracy, ChatGPT's learning mechanisms drive ongoing improvement of responses. Rapid adoption is evident, with over one million registered users within 4 days of launch (Haleem et al., 2022). Users demonstrate diverse applications, from coding assistance to creative endeavours like poetry and songwriting, showcasing the platform's versatility (Chatterjee et al., 2023). OpenAI introduces GPT-4, building upon ChatGPT's success, boasting enhanced capabilities in creativity, collaboration, and technical writing, with improved safety measures (Lo, 2023; Roumeliotis & Tselikas, 2023). GPT-4 also introduces image recognition, supporting 26 languages and offering customizable AI behaviour at $20 per month (Lancaster, 2023). This represents a significant advancement in conversational AI.

3.2.2 | Bidirectional encoder representations from transformers

Building upon the success of BERT (Bidirectional Encoder Representations from Transformers) introduced by Google Research in 2018, BARD and GEMINI represent the next generation of LLMs from Google AI, pushing the boundaries of NLP (Ravichandiran, 2021). While BERT relied on a transformer architecture with 12 or 24 encoder layers, BARD and GEMINI leverage the power of PaLM 2 architecture, boasting 137 billion parameters – significantly exceeding BERT's capabilities (Sur, 2020). This substantial increase in parameters allows BARD and GEMINI to capture more intricate language patterns and relationships. Similar to BERT, BARD, and GEMINI benefit from contextualized word embeddings, considering the entire sentence for nuanced understanding (Sur, 2020). This grants them superior performance in various NLP tasks, including question answering and general language understanding. However, unlike BERT, BARD and GEMINI incorporate real-time data into their training process, enabling them to stay up-to-date with the evolving nature of language (Devlin et al., 2018). Furthermore, BARD and GEMINI excel in code generation, a capability not present in BERT. This empowers them to assist programmers with various tasks, such as writing different parts of code or translating between programming languages (Poerner et al., 2019). Their massive scale, incorporation of real-time data, and code-generation capabilities position them at the forefront of language understanding and interaction.

3.2.3 | Large language model meta AI

Meta's LLaMA (Large Language Model Meta AI) encompasses a series of foundational language models ranging from 7 billion to 70 billion parameters. These models leverage a transformer-based DNN architecture (Deep Neural Network) and achieve strong performance on various NLP tasks

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

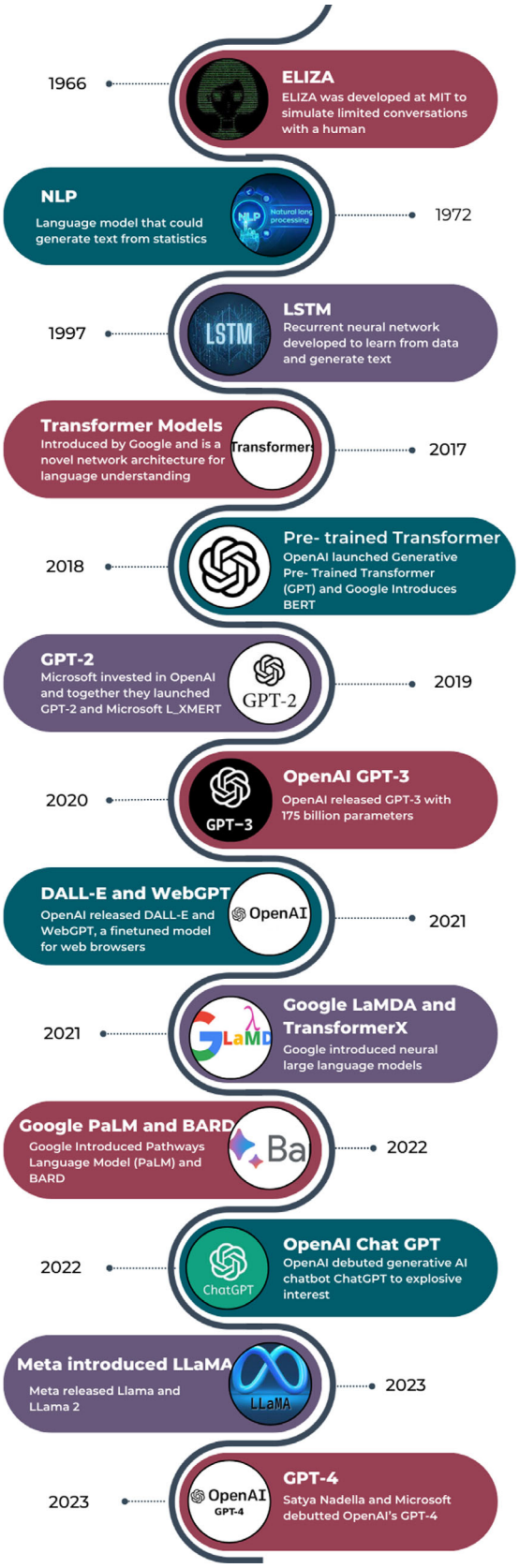


FIGURE 8 Evolution of LLMs: From the inception of natural language processing in 1966 to the emergence of groundbreaking models like GPT-3 in 2023, representation of the transformative journey of language understanding and generation.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

despite their reduced parameter footprint compared to some prior models (Gao, Han, et al., 2023). LLaMA is trained on a massive dataset encompassing Common Crawl, C4, GitHub, Wikipedia, Books, ArXiv, and StackExchange, enabling robust fine-tuning for diverse NLP applications (Inan et al., 2023). Notably, the architecture incorporates a multilingual corpus during training, exposing the model to languages like Bulgarian, Catalan, Czech, and Dutch (Ma et al., 2023). Pre-trained LLaMA models are available through the Hugging Face platform, specifically under the repository Facebook research/llama (Wu, Zhang, et al., 2023). Beyond text generation, these models demonstrate significant adaptability through fine-tuning various NLP tasks and datasets (Zhang, Li, & Bing, 2023). Benchmarking against established models like GPT-3, GPT-NeoX, Gopher, Chinchilla, and PaLM reveals LLaMA's strengths in common sense reasoning, trivia, reading comprehension, closed-book question answering, and code generation (Zhang, Li, & Bing, 2023). While potentially outperformed by specialized models like Minerva in mathematical reasoning, LLaMA's superior code generation capabilities position it as a valuable tool for diverse NLP research and applications (Hu et al., 2023).

3.2.4 | Falcon

The Falcon LLM 40B, developed by the Technology Innovation Institute (UAE), stands out as a 40-billion parameter auto-regressive decoder-only LLM (Penedo et al., 2023). Trained for 2 months on the AWS Cloud using 384 GPUs, it leverages a curated 1 trillion token dataset derived from the Refined Web corpus (Sreevaatsav et al., 2023). Notably, Falcon prioritizes transparency by being open-source and achieving top performance on the Open LLM Leaderboard, signifying its competitiveness with other state-of-the-art models (Zou et al., 2023). Furthermore, Falcon offers fine-tuning capabilities through versions like Falcon-7B-Instruct and Falcon-40B-Instruct specifically designed for chat-oriented tasks. Additionally, its compatibility with the transformers library enhances its usability for various NLP applications (Rebedea et al., 2023). In essence, Falcon establishes itself as a powerful and transparent LLM, achieving top ranking on the Open LLM Leaderboard (Umapathi et al., 2023). Its open-source nature and optimized architecture position it as a valuable tool for developers and researchers exploring advanced language models across diverse applications (Jayakumar et al., 2023).

A comparative analysis of LLMs, including ChatGPT 3.5, BARD, LLaMa 2, and Falcon, revealing differences in architecture, parameter availability, input data, and generation capabilities is represented in Table 2. ChatGPT 3.5 employs a transformer-based architecture, while BARD and LLaMa 2 leverage PaLM 2 architecture. Parameter availability varies, with Falcon presumed to have a substantial parameter count. Input data sources differ, and all models generate text-based outputs but lack image or file-generation capabilities.

3.3 | Large language model evaluation scores

Mosaic ML's Mosaic Eval Gauntlet employs a robust evaluation matrix for assessing LLMs, incorporating 34 benchmarks across six competency domains (Menghani, 2023). Derived from an extensive analysis of LLM literature and established evaluation frameworks like Eleuthera AI Eval Harness and Stanford CRFM's Holistic Evaluation of Language Models (HELM), these competency categories encapsulate the anticipated strengths of robust foundation models (Shayegani et al., 2023).

TABLE 2 Comparative Analysis of LLMs, namely, ChatGPT 3.5, BARD, LLaMa 2, and Falcon based on various parameters such as architecture, parameter availability, input data, and generation (Roumeliotis et al., 2024).

Feature	ChatGPT 3.5	BARD	LLaMa 2	Falcon
Organization	Open AI	Google	Meta	TII, UAE
Architecture	Generative Pre-trained transformer (GPT)	Pathways Language Model 2 (PaLM 2)	Deep Neural Network architecture	Decoder-only framework
Parameters Availability	175 Billion Available	137 Billion Available in Beta	70 Billion Available as pre-trained model in Hugging face	40 Billion Available as open source
Up to date information	Until September 2021	Upto date information	Until August 2020	Upto date information
Real-Time Data?	No	Yes	No	No
Generate images?	No	No	No	No
Accept File Input?	No	No	No	Yes
Can generate code?	Yes	No	Yes	Yes

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The evaluative methodology is designed to capture the diverse capabilities of LLMs tailored for proficiency in a range of natural language processing tasks. The decision to aggregate evaluation scores, rather than relying on singular metrics, is driven by two primary considerations (Cao et al., 2024). First, LLMs are conceptualized as genuine generalist models adept at handling an extensive array of tasks with minimal prompting. The diversity of potential tasks renders representation through a confined set of academic benchmarks insufficient. Aggregating scores across numerous benchmarks allows for a nuanced understanding of the model's multifaceted capabilities (Wu et al., 2024). Second, reliance on a limited set of metrics may introduce misleadingly high variance, potentially amplifying perceived differences in models' proficiencys. Aggregating scores across a spectrum of benchmarks mitigates the signal-to-noise ratio, offering more robust assessments of relative model performance (Ge et al., 2023).

Table 3 represents the evaluation metrics for GPT, Falcon, Llama 1 and Llama 2 with various sizes for different parameters such as code, common sense reasoning, word knowledge, reading comprehension, math, Massive Multitask Language Understanding (MMLU), BIG-Bench Hard evaluation (BBH), and Artificial General Intelligence (AGI) (Westreich, 1990). Code evaluates the model's proficiency in understanding and generating programming code. It assesses the ability to interpret, generate, and manipulate code snippets in various programming languages (Dery et al., 2024). Common sense reasoning evaluates the model's capability to comprehend and respond to situations or queries that require inherent human-like understanding and reasoning, often involving everyday knowledge and context (Teng et al., 2023).

World knowledge assesses the model's grasp of general world knowledge across diverse domains. It evaluates how well the model can answer questions or generate content based on factual information about the world (Hong et al., 2023). Reading comprehension evaluates the model's ability to understand and answer questions related to given passages or texts. It assesses how effectively the model can extract information and derive meaningful insights from written content (Laskaridis et al., 2024). The math parameter assesses the model's mathematical reasoning abilities. It evaluates how well the model can solve mathematical problems, understand mathematical concepts, and generate accurate responses to math-related queries (Luo et al., 2023). MMLU evaluates the model's performance across a diverse set of language understanding tasks, showcasing its versatility in handling multiple NLP tasks simultaneously. It reflects the model's multitasking capabilities (Laskaridis et al., 2024). The BBH evaluation assesses the model's performance on challenging benchmarks that demand advanced language understanding and reasoning skills. It represents a more rigorous evaluation to test the model's robustness (Jin et al., 2023). The AGI parameter evaluates the model's performance on tasks that are indicative of artificial general intelligence. It assesses the model's ability to generalize knowledge and apply reasoning skills across a broad spectrum of challenges, aiming for human-like intelligence (Bai et al., 2023).

The strategic decision to present multiple separate aggregate scores, categorized by competency, instead of a singular metric, caters to the varied requirements of end users deploying models for specific use cases (Zhang, Chen, et al., 2023). This breakdown of model performance into essential competencies facilitates informed decision-making during the model training process. The evaluation procedure utilizes the Mosaic ML LLM Foundry's In-Context-Learning (ICL) evaluation framework, characterized by multi-GPU capabilities and native FSDP compatibility. It ensures equitable model comparisons by avoiding prompt tuning or reliance on prompt strings (Antaki et al., 2023). The benchmarks cover a spectrum of evaluation tasks, and resultant scores are aggregated within each competency category, providing composite scores and their averages. To address challenges posed by benchmarks with random guessing baselines, scores undergo rescaling to ensure they are constrained to less than or equal to 1 (Awasthi et al., 2023). This involves subtracting the random baseline accuracy and rescaling the remainder by 1 minus the baseline

TABLE 3 LLM evaluation metrics for GPT, Falcon, Llama 1, and Llama 2 with various sizes for different parameters such as code, common sense Reasoning, world knowledge, reading comprehension, math, massive multitask language understanding, big-bench hard evaluation, and artificial general intelligence (Li, Tang, et al., 2023).

Model	Size	Code	Common sense reason	World knowledge	Reading comprehension	Math	MMLU	BBH Eval	AGI
GPT	7B	20.5	57.4	41.0	57.5	4.9	26.8	31.0	23.5
	30B	28.9	64.9	50.0	64.7	9.1	46.9	38.0	33.8
Falcon	7B	5.6	56.1	42.8	36.0	4.6	26.2	28.0	21.2
	40B	15.2	69.2	56.7	65.7	12.6	55.4	37.1	37.0
Llama 1	7B	14.1	60.8	46.2	58.5	6.95	35.1	30.3	23.9
	13B	18.9	66.1	52.6	62.3	10.9	46.9	37.0	33.9
	33B	26.0	70.0	58.4	67.6	21.4	57.8	39.8	41.7
	65B	30.7	70.7	60.5	68.6	30.8	63.4	43.5	47.6
Llama 2	7B	16.8	63.9	48.9	61.3	14.6	45.3	32.6	29.3
	13B	24.5	66.9	55.4	65.8	28.7	54.8	39.4	39.1
	34B	27.8	69.9	58.7	68.0	24.2	62.6	44.1	43.4
	70B	37.5	71.9	63.6	69.4	35.2	68.9	51.2	54.2

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

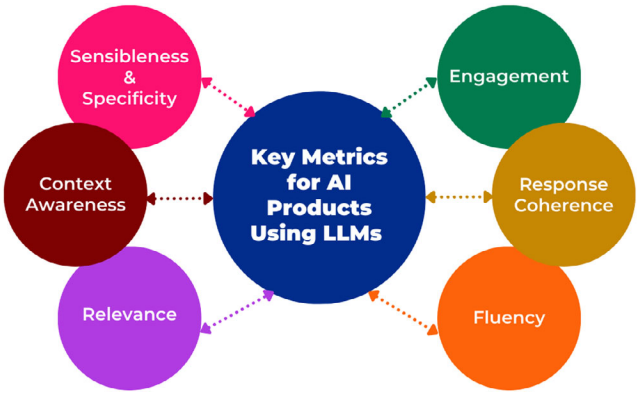


FIGURE 9 Representation of various key metrics for artificial intelligence products using large language models.

accuracy, ensuring a consistent portrayal of accuracy above chance across diverse benchmarks (Bang et al., 2023). Figure 9 represents various key metrics for AI products using LLMs.

3.4 | Advantages of large language models

LLMs have ushered in a paradigm shift in generative AI applications, with their prominence beginning in 2022. Noteworthy instances, including ChatGPT, Bing Chat, and Bard, leverage LLMs to automate the intricacies of content and data generation (Sainz et al., 2023). This extends to diverse applications such as image generators exemplified by Stable Diffusion and DALL-E, thereby showcasing the profound versatility of LLMs. Their aptitude for nuanced contextual comprehension renders them highly suitable for an extensive array of natural language processing tasks, spanning from the intricacies of content creation to the nuanced analysis of datasets and the adept addressing of complex questions (Wu, Fei, et al., 2023). The scalability and versatility inherent in LLMs are direct outcomes of their training on expansive datasets. Larger model sizes have been empirically correlated with heightened performance in the realm of natural language processing tasks (Wang et al., 2014). Beyond static capabilities, some LLMs exhibit an enviable degree of adaptability through continuous refinement via updates and fine-tuning processes, thereby ensuring a sustained relevance amidst the ever-evolving nuances of language patterns (Chen et al., 2023). This adaptability extends its benefits across a spectrum of applications, from speech recognition and language translation to writing assistance and digital virtual support (Zhang, Krishna, et al., 2023).

The developmental landscape of LLMs is characterized by a diverse array of modelling approaches, fostering an environment of perpetual innovation. Autoregressive modelling, the transformative advent of the transformer architecture, and a plethora of other modelling methodologies contribute synergistically to the dynamic evolution of LLMs (Lai et al., 2023). Furthermore, the emergence of open-source LLMs aligns seamlessly with the overarching ethos of democratizing artificial intelligence, accelerating its developmental trajectory, and underscoring the pivotal role of open-source practices in the field (Wu, Bansal, et al., 2023).

Parallel to these advantages, LLMs hold the potential to democratize access to knowledge and information, akin to the seminal roles played by the internet and conventional search engines (Westreich, 1990). The prowess of LLMs shines through in facilitating comprehension of intricate textual content, seamlessly translating languages, and fostering interactive experiences (Akinci D'Antonoli et al., 2024). General-purpose chatbots, positioned as formidable alternatives to conventional search engines, find application niches in personalized knowledge extraction, health information dissemination, and the nuances of business communications. The innate ability of LLMs to facilitate human-like interactions renders them apt for an expansive spectrum of applications (Zhang, Yu, et al., 2023).

3.5 | Disadvantages of large language models

LLMs face several critical challenges. One significant hurdle is the immense requirement for vast datasets. This inevitably restricts access primarily to well-resourced organizations (Banerjee et al., 2023). Ethical considerations arise concerning data utilization and privacy, especially when large tech companies scrape supposedly public information (Li, Zhang, & Chen, 2023). Furthermore, continuous retraining or fine-tuning is essential to maintain LLM relevance, posing a challenge for entities lacking access to high-quality data (Spark & Luckner, 2022). The computational resources required for LLM training and deployment present another substantial barrier, particularly for smaller organizations. Employing advanced architectures like transformers necessitates powerful and expensive hardware, raising environmental concerns (Ke et al., 2023). This high computational

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

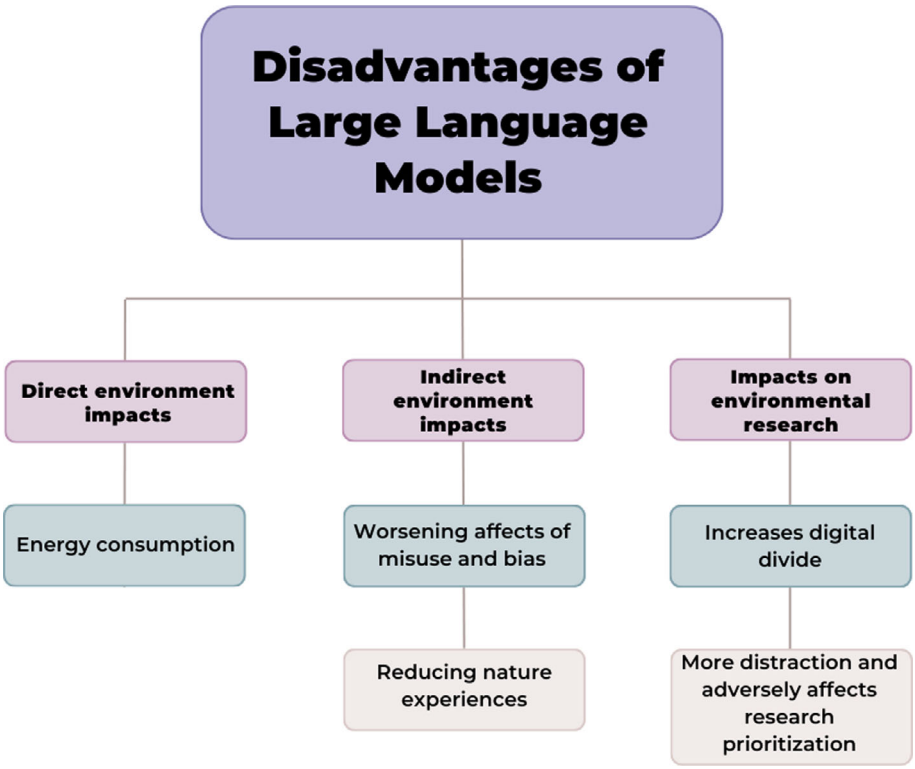


FIGURE 10 Disadvantages of LLMs representing direct and indirect environmental impacts and adverse impacts on environmental research.

cost not only limits accessibility but also contributes to environmental impact through energy consumption and ecological footprint (Phillips, 2023). LLMs are also susceptible to bias and hallucination. Without rigorous data curation, these models can perpetuate biases inherent in their training data, leading to ethical issues (Chang et al., 2024). Hallucination, where LLMs generate seemingly credible outputs lacking factual basis, is a significant concern observed in advanced chatbots. Mitigating misinformation and deceptive outputs requires proactive measures from developers.

The widespread adoption of LLMs presents concerns regarding unforeseen consequences, including potential job displacement and excessive reliance on generative applications. Automation of tasks like writing, programming, and customer service introduces dynamics that might negatively impact employment in specific sectors (Gao, Wang, et al., 2023). Additionally, LLMs create opportunities for malicious activities, such as deepfakes, misinformation propagation, targeted surveillance, and phishing scams (Chen et al., n.d.). These practices pose tangible societal risks and necessitate a comprehensive approach to mitigating potential harms as represented in Figure 10. While LLMs represent a significant advancement in AI due to their exceptional generative capabilities, a balanced understanding of their advantages and disadvantages is crucial during deployment and development (Ribar et al., 2023). The complexities of bias mitigation, ethical considerations, and environmental impact highlight the need for a responsible and conscientious approach to harnessing the potential of these transformative models for societal benefit (Bai et al., 2024).

4 | LARGE LANGUAGE MODELS IN LUNG CANCER

LLMs like GPT-3.5 may play a pivotal role in the realm of lung cancer research and healthcare. These sophisticated models adeptly process extensive textual data, aiding in literature reviews, knowledge extraction, and staying abreast of the latest research, diagnostic methods, and treatment advancements in lung cancer. LLMs provide valuable insights for clinical decision support by analysing patient data and medical reports, contributing to well-informed decision-making (Haver et al., 2023). Their natural language processing capabilities prove beneficial in extracting meaningful information from Electronic Health Records (EHR), facilitating a comprehensive grasp of patient histories. Additionally, LLMs assist in generating patient-friendly educational materials and contribute to text generation for reports and documentation, thereby enhancing communication and efficiency in the healthcare sector. While not a substitute for healthcare professionals, LLMs significantly enhance information processing and knowledge dissemination in lung cancer research and healthcare (Panagoulias et al., 2023).

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

LLMs and NLP in diagnosing lung cancer over ML and DL techniques include their ability to process unstructured medical data such as clinical notes, pathology reports, and radiology images, enabling comprehensive patient assessment (Iannantuono et al., 2023). LLMs excel in understanding complex medical language and context, aiding in accurate diagnosis and personalized treatment planning (Naik et al., 2023). Additionally, NLP facilitates efficient information extraction and summarization from vast medical datasets, enhancing decision-making processes. However, challenges arise in the interpretability of their outputs, potential biases in training data, and the need for robust validation in clinical settings to ensure reliability (Hub, Handa, Nautiyal, et al., 2024; Hub, Handa, Saini, et al., 2024; Panagoulias et al., 2023). While powerful in image analysis and pattern recognition, ML and DL techniques may require large labelled datasets and computational resources for training, limiting their scalability and accessibility in clinical practice.

The examination of the use of LLMs in identifying lung cancer adheres to the PRISMA guidelines (Agarwal & Kumar, 2022; Page et al., 2021). The decision to adopt PRISMA is driven by its effectiveness in improving the identification of Critical Research (CR), providing direction for the selection, identification, and assessment of pertinent studies. The research made use of four distinct databases, which include Google Scholar, Scopus, PubMed, and several preprint platforms such as ArXiv, TechRxiv, MedRxiv, and ChemRxiv. The search strategy involved employing different strings such as "Lung Cancer," "((LLM) AND (Lung Cancer))," "((Large Language Model) AND (lung cancer))," and "((Oncology) AND (LLM))" or "((Oncology) AND (Large Language Model))." Figure 11 illustrates the search string used during the search process.

Apart from the electronic search, a manual search was carried out to identify pertinent research papers. In the identification stage, an initial selection of 900 research publications was made. Subsequently, during the screening phase, 500 research articles were chosen after excluding unsuitable, duplicate, and irrelevant papers. Following a review of titles, abstracts, and introductions, 420 research articles were excluded. The remaining 80 research articles underwent further analysis based on removal criteria, resulting in the exclusion of 65 articles. After this, 15 research articles progressed to the eligibility phase. In the final eligibility phase, a comprehensive assessment of the complete papers led to the elimination of nine research articles. The criteria governing the selection and removal of research articles are detailed in Table 4. Figure 12 visually illustrates the distinct phases of the PRISMA process for this review.

Niu and Wang (2023) illustrate the feasibility of developing a CT Large Image-Text (LIT) model that serves multiple tasks for diagnosing lung cancer. This model integrates an LLM and a Large Image Model (LIM). The LIT model exhibits proficiency in various medical tasks, including lung segmentation, lung nodule detection, and lung cancer classification. To improve model performance and applicability, the authors suggest the necessity of larger datasets and the resolution of obstacles in sharing patient text reports and genomic data under Institutional Review Board (IRB) approvals. Highlighting the importance of open-source practices, the Cerebras-GPT family stands out as the exclusive industry-leading family of open-source models currently accessible. Rahsepar et al. (2023) conducted a comparison of question-answering capabilities related to lung cancer prevention, screening, and radiology terminology. ChatGPT-3.5 demonstrated higher accuracy than Google Bard, Bing, and Google search engines. However, neither ChatGPT nor the search engines consistently provided flawless answers. Notably, ChatGPT-3.5 and the Google search engine exhibited greater consistency in their responses compared to Google Bard. The accuracy of ChatGPT-3.5 in delivering correct or partially correct answers surpassed that of Google Bard by approximately 1.5 times. This study underscores the potential benefits and challenges associated with the use of large language models, such as ChatGPT and Google Bard, in addressing non-expert questions related to lung cancer.

Next-generation sequencing (NGS) is a routine clinical practice in advanced NSCLC. NGS reports are information-dense, and clinical interpretation remains a challenge. Hamilton et al. (2023) presented that the AI-generated NGS reports had an average length of 117 words, ranging from 44 to 232 words. The median number of recommended treatments in the reports was 5, varying between 3 and 8 treatments. Preferred first-line treatments for specific oncogenic drivers were included in 68.8% of the reports, while recommendations to explore clinical trials were listed in 53.8%. The Relevance Score (RS) for the entire sample was 0.59, indicating moderate relevance of the recommendations. However, the Accuracy Score (AS) was 46.0, suggesting limited accuracy in the treatment options provided in the NGS reports. The study concludes that ChatGPT can efficiently generate concise NGS reports for NSCLC patients with driver oncogenes, but there is room for improvement in terms of both relevance and accuracy. However, certain transformer-based models (T523 and Clinical BERT) were excluded from the primary analysis due to insufficient phenotype results, emphasizing the need for continued exploration and refinement in the field. Table 5 illustrates the efforts undertaken in lung cancer research through the utilization of LLMs. Figure 13 illustrates the chord diagram of the considered articles.

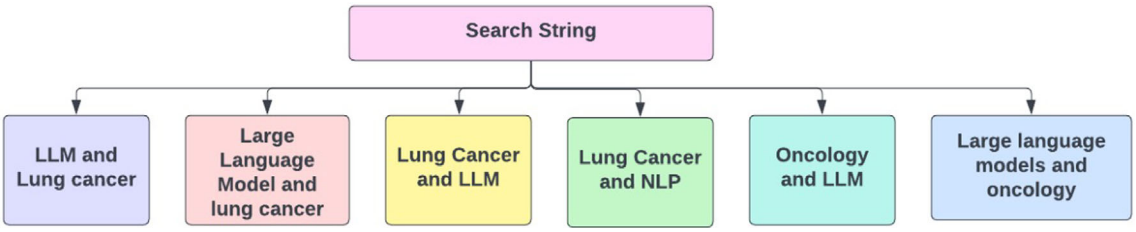


FIGURE 11 Illustration depicting the search query employed to acquire relevant literature on lung cancer and LLMs from Google Scholar, Scopus, PubMed, and preprint platforms (ArXiv, TechRxiv, MedRxiv, ChemRxiv).

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Selection and removal criteria for selection of research articles.

S. no.	Parameter	Selection criteria	Removal criteria
1	Time duration	Research articles published between 2019 and 2023	Research articles published before 2019
2	Analysis	Research articles including lung cancer and its detection	Research articles including other cancer and its detection
3	Comparison	Research article focus on LLM used for lung cancer analysis	Research article focus on other techniques used for lung cancer analysis
4	Study	Research involving prompt analysis and experimental results	Research involving case study and articles in different language other than English

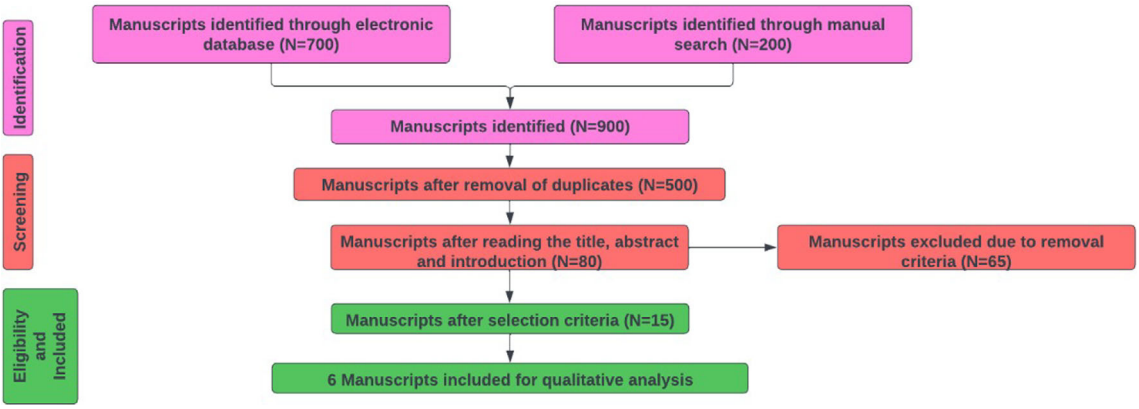


FIGURE 12 PRISMA review process conducted for the comprehensive analysis of large language models in the context of lung cancer research.

The study showcases the potential of the CT Large Image-Text (LIT) model in lung cancer diagnostics, employing a fusion of LLM and LIM. Emphasizing the importance of larger datasets and overcoming data-sharing challenges, the authors highlight the Cerebras-GPT family's leading role in open-source models. In question-answering comparisons, ChatGPT-3.5 demonstrates superior accuracy over search engines, underlining its efficacy for non-expert inquiries. The AI-generated NGS reports for advanced NSCLC patients, while concise, reveal room for improvement in both relevance and accuracy, emphasizing the ongoing need for refinement in the field. Overall, these efforts underscore the transformative potential of LLMs in advancing lung cancer research and diagnostics.

Bhattarai et al. (2023) evaluated Open AI's GPT-4 model for identifying clinical phenotype in NSCLC patients' EHRs. The study aimed to assess GPT-4's ability to extract crucial information on disease stages, treatments, and recurrence patterns from complex medical records. Results showed GPT-4 outperforming rule-based models and other transformers like GPT-3.5-turbo, scispaCy, and medspaCy, highlighting its advanced capabilities. Similarly, Fink et al. (2023) compared ChatGPT and GPT-4 in extracting oncologic phenotypes from free-text CT reports on lung cancer. GPT-4 demonstrated superior accuracy (98.6%) in extracting lesion parameters, surpassing ChatGPT's 84.0%. It also showed higher accuracy (98.1%) in identifying metastatic disease, outperforming ChatGPT's 90.3%. GPT-4 received higher Likert scale scores for factual correctness and accuracy, with a significantly lower confabulation rate (1.7% v/s 13.7%) compared to ChatGPT. These findings emphasize GPT-4's potential in efficiently extracting oncologic information from free-text CT reports for more accurate and reliable outputs.

Naik et al. (2023) demonstrated that LLMs exhibit heightened accuracy in predicting causal graph edges, surpassing current state-of-the-art methods. The pivotal role of LLMs in automating the creation of Causal Directed Acyclic Graphs (DAGs) provides nuanced insights into patient responses to therapy. Incorporating additional data modalities, such as genomic variations, enriches the understanding of patient responses. The specific focus on oncology, particularly Non-Small Cell Lung Cancer (NSCLC), emphasizes the targeted application of LLMs in Causal Structure Learning (CSL). Validation strategies utilizing Bayesian Dirichlet estimators and observational data contribute to enhancing the credibility of LLM-generated graphs.

The recent strides in LLMs have ignited interest in their application for groundbreaking biomedical research, particularly in Causal Discovery or Causal Structure Learning (CSL) (Zhang, Bauer, et al., 2023). Demonstrating this potential, Naik et al. (2023) have showcased improved accuracy in generating Causal Directed Acyclic Graphs (Causal DAGs) using LLMs with data extracted from EHR and Molecular Genomic Reports. The Bayesian Dirichlet equivalent uniform (Bdeu) score is employed for model evaluation. Causal Modelling is recognized as pivotal in biomarker discovery (Papaleo et al., 2011), and the methodology incorporates prompt engineering, including zero-shot prompts (Suzgun et al., 2022). Each prompt is meticulously formatted to query potential cause-effect relationships and the resulting DAGs are optimized through Bayesian estimation.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Comparative analysis of the efforts undertaken in lung cancer research through the utilization of large language models.

Ref.	Methodology	Factors of reproducibility	Inference	Accuracy	Limitations	Research	Chatbot used
Niu and Wang (2023)	LLM + LIM	The methodology builds multi-task CT models for lung cancer diagnosis by combining LLMs and large image models (LIMs).	The LIT model adeptly handles diverse medical tasks, including lung segmentation, nodule detection, and cancer classification.	Proposed model LIT-BioGPT outperformed LIT-CLIP with a classification accuracy of 85.1% compared to 79.5%.	Limited dataset size and diversity may impact model generalizability, and challenges in sharing patient data under IRB approvals restrict dataset access.	Feasibility assessment for a multi-task CT Large Image-Text model in lung cancer. Integrating LLM and LIM as encoders that capture task-specific information.	—
Rahsepar et al. (2023)	Prompt Engineering on LLMs	Findings highlight LLMs' reliability in information retrieval and decision-making for lung cancer.	ChatGPT-3.5 was more likely to provide correct or partially correct answers than Google Bard, by approximately 1.5-fold for lung cancer.	ChatGPT-3.5 gave 70.8% correct answers, Google Bard did not answer 19.2% questions, and gave 51.7% correct answers.	The study lacks exploration of biases in data and models, with increasing LLM performance affecting findings' relevance over time.	Assess and compare ChatGPT and Google Bard responses to non-expert queries about lung cancer.	ChatGPT, Bard
Hamilton et al. (2023)	Prompt Engineering on LLMs	Dataset prompts offer insights into chatbot strengths and limitations.	The average length of the AI-generated NGS reports was 117 words (range: 44–232. ChatGPT recommendations were promising, given the complexity of the task, with no prompting or training provided to the AI.	The Relevance Score was 0.59, and the Accuracy Score was 46.0%.	ChatGPT-generated NGS reports show limited accuracy and inconsistent alignment with NCCN guidelines, indicating discrepancies in clinical standards.	Evaluate the clinical relevance and accuracy of ChatGPT-generated Next-Generation Sequencing (NGS) reports for stage IV non-small cell lung cancer patients.	Chatgpt
Bhattarai et al. (2023)	Prompt Engineering on LLMs + NER	The method utilizes Relevance Score (RS) and Accuracy Score (AS) to assess AI-generated reports in comparable healthcare contexts.	GPT-4 achieves higher F1 score, precision, and recall compared to medspaCy and scispaCy's models. GPT-3.5-turbo performs similar to that of GPT-4.	Average F1 scores for clinical phenotype identification: GPT-4 (0.62), GPT-3.5-turbo (0.8475), scispaCy (0.605), medspaCy (0.5975).	Traditional metrics complicate comparing LLM-generated text to human reference, introducing discrepancies across LLMs.	Evaluate GPT-4's capability in identifying clinical phenotypes from EHR text in NSCLC. Compare its performance with GPT-3.5-turbo, scispaCy, and medspaCy in identifying disease stages, treatments, and progression.	GPT-4, GPT-3.5, scispaCy, medspaCy
Fink et al. (2023)	Prompt Engineering on LLMs	GPT-4 excels in oncologic phenotype extraction from free-text CT reports, showing superior reasoning.	GPT-4 excels in extracting oncologic phenotypes from CT reports, outperforming ChatGPT with superior reasoning and fewer confabulations.	GPT-4 excelled in lesion parameter extraction (98.6% vs. ChatGPT's 84.0%) and identified metastatic disease more accurately (98.1% vs. ChatGPT's 90.3%).	Biases in prompt engineering and reliance on user-defined prompts may affect generalizability. Unaddressed ethical considerations in LLM decision-making.	Evaluate ChatGPT and GPT-4's performance in mining oncologic phenotypes from free-text CT reports on lung cancer. Emphasizes accuracy, correctness, and oncologic reasoning using user-defined prompts.	Chatgpt, GPT-4
Naik et al. (2023)	Prompt Engineerings on LLMs + Bayesian estimation	LLMs used to generate Causal DAGs, revealing linkages between diverse biomarkers in biomedical research.	Findings reveal varying degrees of treatment modalities' influence on specific biomarkers, enhancing understanding of healthcare dynamics.	Findings reveal varying degrees of treatment modalities' influence on specific biomarkers, advancing healthcare dynamics understanding.	Data extraction in 2018 and reliance on general-purpose LLMs pose limitations in gene set comprehensiveness and causal relationship discernment in the medical domain.	Examines how LLMs automate Causal DAGs for Non-Small Cell Lung Cancer, comparing their performance with existing methods.	—

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

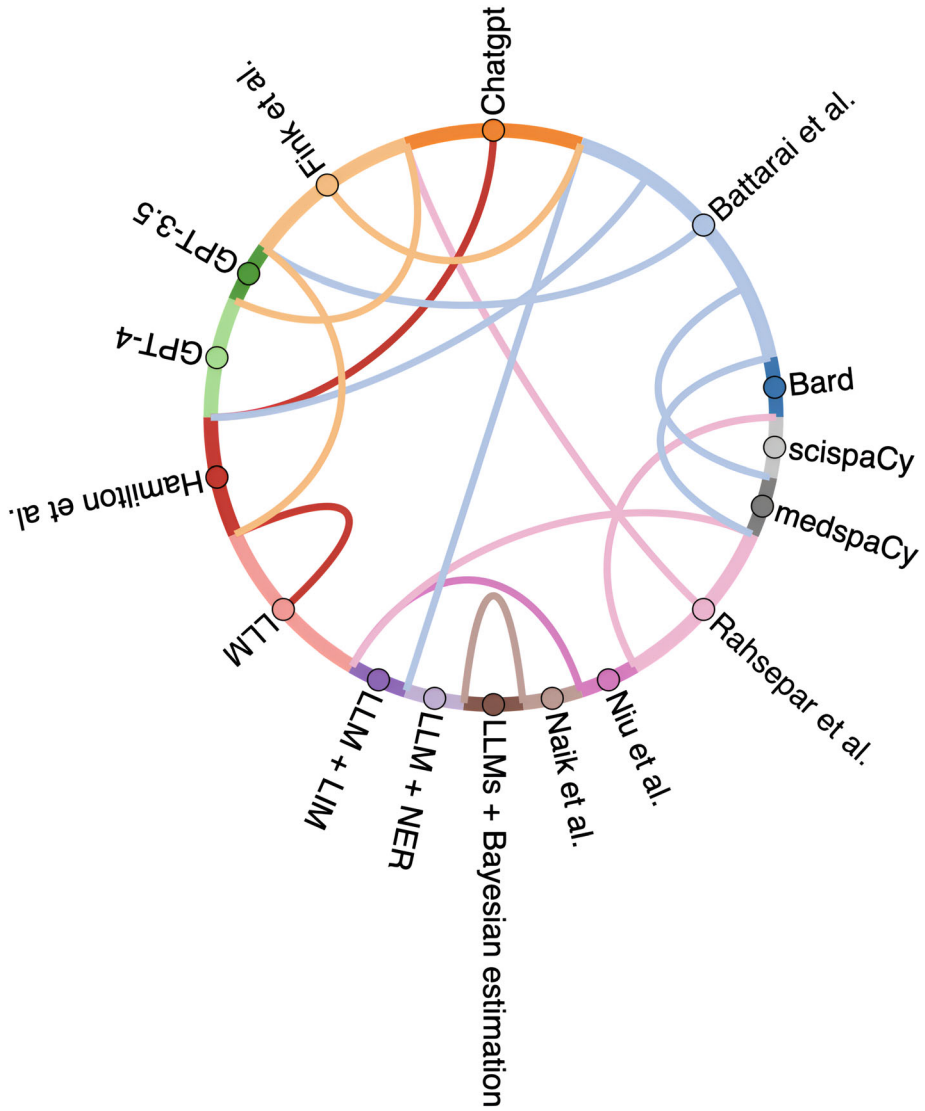


FIGURE 13 The chord diagram visually represents the relationship between authors, various methodologies, and the associated chatbots utilized in their research on large language models in the context of lung cancer. Each chord represents the collaborative interplay between methodology and specific chatbots (if any) employed by various authors, providing a comprehensive overview of the research landscape in this critical field.

the study highlights a potential drawback; while achieving better results is feasible, utilizing general-purpose LLMs for biomedical research poses a disadvantage due to their lack of specialized knowledge essential for discerning causal relationships in the medical domain. The intricate relationship between authors, various methodologies, and the associated chatbots (if any) utilized in their research on LLMs in the context of lung cancer is depicted in Figure 13.

5 | FUTURE RESEARCH DIRECTIONS

5.1 | Potential challenges

While LLMs hold immense promise for lung cancer research, significant challenges need to be addressed to fully realize their potential. This section outlines key future research directions, acknowledging current limitations and proposing how LLMs can contribute, while also highlighting potential barriers to their implementation.

1. Limited dataset size and diversity: The current landscape of lung cancer research is hampered by constrained datasets, limiting the generalizability of models. Efforts in the future should prioritize the creation of expansive and diverse datasets encapsulating various patient

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

demographics and cancer subtypes. LLMs can play a pivotal role in facilitating data aggregation from disparate sources and ensuring compliance with privacy regulations through anonymization techniques.

2. Understanding LLM outputs: Traditional evaluation metrics may fall short in assessing the quality of LLM-generated outputs in the medical domain. Future research should focus on developing comprehensive evaluation methodologies that gauge clinical relevance, factual accuracy, and adherence to clinical guidelines. LLMs themselves can contribute by analysing large datasets annotated by humans to devise improved evaluation metrics.
3. Ethical considerations and explainability: The ethical implications of LLM-driven decision-making in lung cancer diagnosis and treatment warrant careful consideration. Implementing explainable AI techniques is imperative to ensure transparency and instill trust among healthcare professionals. LLMs can contribute by generating explanations for their reasoning processes, thereby enhancing understanding and trust.
4. Computational resources: The computational demands associated with training and deploying LLMs pose a significant barrier. Continued advancements in hardware and software optimization are indispensable for broader adoption in lung cancer research.
5. Explainability and trust: Establishing trust in LLM-generated outputs among clinicians remains a formidable challenge. Addressing concerns related to interpretability and potential biases is crucial for successful integration into clinical practice.

5.2 | Future research prospects

Building upon the identified challenges, future research should harness the transformative potential of LLMs to propel lung cancer research and diagnosis forward. Here's a detailed exploration of key areas for investigation:

1. Moving beyond question-answering: Current research predominantly focuses on LLMs' question-answering abilities. Future endeavours should expand into tasks such as real-time analysis of clinical text data, personalized treatment plan generation, and identification of high-risk patients specific to lung cancer. By diversifying LLM applications, researchers can enhance their utility in addressing complex clinical challenges unique to lung cancer diagnosis and treatment.
2. Domain-specific pre-training: Pre-training LLMs on extensive datasets of lung cancer research literature and clinical data can enrich their understanding of lung cancer nuances. This domain-specific knowledge empowers LLMs to produce more contextually relevant outputs tailored to the intricacies of lung cancer diagnosis and treatment, thus improving their efficacy in clinical settings.
3. Developing robust evaluation metrics: Beyond conventional accuracy measures, novel evaluation metrics are essential for assessing LLM performance in lung cancer research. These metrics should evaluate clinical relevance, adherence to best practices, and the ability to generate novel insights specific to lung cancer diagnosis and treatment. Leveraging LLMs to analyse large datasets of human-annotated clinical data can aid in the development of these new metrics, ensuring a comprehensive evaluation of LLM capabilities in lung cancer research.
4. Integration of multimodal data: Beyond genomic data, integrating clinical imaging data, pathology reports, and patient medical history can provide a holistic view of lung cancer. LLMs excel at processing diverse data sources, uncovering hidden patterns and relationships critical for accurate diagnosis and treatment planning in lung cancer. Advanced data fusion techniques, facilitated by LLMs, can effectively combine information from various modalities, enhancing diagnostic accuracy and personalized treatment options for lung cancer patients. Biomedical challenges with multi-modal lung cancer data could help in collaborations and innovation in this field.
5. Bias mitigation and fairness: Identifying and mitigating hidden biases in LLMs is paramount for ensuring fairness and reliability in lung cancer research. LLMs can analyse training data to detect biases and recommend balanced prompts to address them. Implementing fairness-aware training algorithms, guided by LLM analyses, can mitigate biases related to patient demographics, socioeconomic status, or disease subtype, promoting equitable healthcare outcomes in lung cancer diagnosis and treatment.
6. Clinical validation and adoption: Rigorous clinical trials are necessary to evaluate the efficacy and safety of LLM-powered tools in lung cancer diagnosis and treatment. These trials should assess LLM impact on patient outcomes, clinician workflow, and healthcare delivery costs. Real-world implementation studies following successful clinical trials are crucial to assess LLM integration into existing clinical workflows and gauge perceptions among healthcare professionals and patients.
7. Ethical and regulatory compliance: Ensuring transparency, explainability, and adherence to data privacy regulations are essential for ethical LLM deployment in lung cancer research. LLMs can generate explanations for their outputs, fostering trust among clinicians and patients. Strict data governance protocols and clear regulatory frameworks are imperative to safeguard patient privacy and ensure responsible LLM implementation in lung cancer diagnosis and treatment.

6 | CONCLUSION

In summary, the present review of advanced technologies, particularly LLMs and NLP, illuminates their promising role in revolutionizing lung cancer research and diagnostic methodologies. The integration of LLMs and LIMs in a multitude of medical tasks, coupled with the exploration of

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

NLP techniques for extracting invaluable information from clinical narratives, showcases their versatility and transformative potential. The robustness of these technologies, exemplified by the adept performance of LLMs in question-answering tasks and the efficient generation of AI-driven NGS reports for NSCLC patients, underscores their capacity to elevate the efficiency and efficacy of healthcare processes. While acknowledging the significant strides made by these technologies, it is essential to recognize the intricacies and nuances inherent in their application. The comparative analysis of question-answering capabilities, while highlighting the strengths of LLMs, reveals challenges related to achieving consistent and flawless responses across diverse platforms. The evaluation of AI-generated NGS reports underscores the efficiency of models like ChatGPT, yet it underscores the on-going need for continuous refinement to enhance both relevance and accuracy for optimal clinical use.

Moreover, this review emphasized the prevalence of NLP studies in lung cancer diagnosis, signalling a growing interest in leveraging technology for improved healthcare outcomes. However, it simultaneously underscores the need for a more balanced exploration of NLP applications, especially in critical areas such as cancer staging, urging researchers to navigate potential pitfalls for more comprehensive and accurate results. Looking toward the future, the proposed avenues for further research present an exciting opportunity to address current limitations and propel the field forward. The exploration of diverse data modalities, the incorporation of higher-level genomic markers, and the integration of specialized medical knowledge represent a forward-looking perspective that promises to enhance the capabilities of LLMs and NLP in lung cancer research. This comprehensive approach not only aims to fill existing gaps but also ensures a nuanced understanding of the complexities associated with lung cancer, fostering a continuous cycle of improvement and innovation in diagnostic methodologies that ultimately translates into improved patient outcomes. In conclusion, this systematic review provides valuable insights into the evolving landscape of lung cancer research, setting the stage for a future where advanced technologies play a pivotal role in shaping the trajectory of healthcare in the battle against this formidable disease.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Palak Handa https://orcid.org/0000-0003-4447-8172
Nidhi Goel https://orcid.org/0000-0001-6337-9052

REFERENCES

Agarwal, T., & Kumar, V. (2022). A systematic review on bat algorithm: Theoretical foundation, variants, and applications. *Archives of Computational Methods in Engineering*, *29*, 2707–2736.

Ajitha, E., Diwan, B., & Roshini, M. (2022). Lung cancer prediction using extended knn algorithm. In *2022 6th international conference on computing methodologies and communication (ICCMC)* (pp. 1665–1670). IEEE.

Akinci D'Antonoli, T., Stanzione, A., Bluethgen, C., Vernuccio, F., Ugga, L., Klontzas, M. E., Cuocolo, R., Cannella, R., & Koçak, B. (2024). Large language models in radiology: Fundamentals, applications, ethical considerations, risks, and future directions. *Diagnostic and Interventional Radiology*, *30*(2), 80–90.

Alexander, J. (1935). Observations on total pulmonary lobectomy and pneumonectomy. *Annals of Surgery*, *101*(1), 393.

Ali, H., Mohsen, F., & Shah, Z. (2023). Improving diagnosis and prognosis of lung cancer using vision transformers: A scoping review. *BMC Medical Imaging*, *23*(1), 129.

Ali, I., Hart, G. R., Gunabushanam, G., Liang, Y., Muhammad, W., Nartowt, B., Kane, M., Ma, X., & Deng, J. (2018). Lung nodule detection via deep reinforcement learning. *Frontiers in Oncology*, *8*, 108.

Allemani, C., Matsuda, T., di Carlo, V., Harewood, R., Matz, M., Nikšić, M., Bonaventure, A., Valkov, M., Johnson, C. J., Estève, J., Ogunbiyi, O. J., Azevedo e Silva, G., Chen, W. Q., Eser, S., Engholm, G., Stiller, C. A., Monnereau, A., Woods, R. R., Visser, O., … Lewis, C. (2018). Global surveillance of trends in cancer survival 2000–14 (concord-3): Analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *The Lancet*, *391*(10125), 1023–1075.

Antaki, F., Touma, S., Milad, D., El-Khoury, J., & Duval, R. (2023). Evaluating the performance of ChatGPT in ophthalmology: An analysis of its successes and shortcomings. *Ophthalmology Science*, *3*(4), 100324.

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, *25*(6), 954–961.

Arnolli, M. M., Hanumara, N. C., Franken, M., Brouwer, D. M., & Broeders, I. A. (2015). An overview of systems for CT-and MRI-guided percutaneous needle placement in the thorax and abdomen. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *11*(4), 458–475.

Arroliga, A. C., & Matthay, R. A. (1993). The role of bronchoscopy in lung cancer. *Clinics in Chest Medicine*, *14*(1), 87–98.

Awasthi, R., Mishra, S., Mahapatra, D., Khanna, A., Maheshwari, K., Cywinski, J., Papay, F., & Mathur, P. (2023). Humanely: Human evaluation of LLM yield, using a novel web-based evaluation tool. *medRxiv*, 2023–12.

Bai, G., Chai, Z., Ling, C., Wang, S., Lu, J., Zhang, N., Shi, T., Yu, Z., Zhu, M., Zhang, Y., & Yang, C. (2024). Beyond efficiency: A systematic survey of resource-efficient large language models. *arXiv preprint arXiv:2401.00625*.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., & Hui, B. (2023). Qwen technical report. *arXiv preprint arXiv: 2309.16609*.

Banerjee, D., Singh, P., Avadhanam, A., & Srivastava, S. (2023). Benchmarking LLM powered chatbots: Methods and metrics. *arXiv preprint arXiv: 2308.04624*.

Bang, J., Choi, Y., Kim, M., Kim, Y., & Rhu, M. (2023). Vtrain: A simulation framework for evaluating cost-effective and compute-optimal large language model training. *arXiv preprint arXiv:2312.12391*.

Bhattarai, K., Oh, I. Y., Sierra, J. M., Payne, P. R., Abrams, Z. B., & Lai, A. M. (2023). Leveraging Gpt-4 for identifying clinical phenotypes in electronic health records: A performance comparison between GPT-4, GPT-3.5-turbo and spacy's rule-based & machine learning-based methods. *bioRxiv*, 2023–09.

Bird, C., Ford, D., Zimmermann, T., Forsgren, N., Kalliamvakou, E., Lowdermilk, T., & Gazit, I. (2022). Taking flight with copilot: Early insights and opportunities of AI-powered pair-programming tools. *Queue*, *20*(6), 35–57.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424.

Bunyaviroch, T., & Coleman, R. E. (2006). Pet evaluation of lung cancer. *Journal of Nuclear Medicine*, *47*(3), 451–469.

Cao, Z., Yang, Y., & Zhao, H. (2024). Head-wise shareable attention for large language models. *arXiv preprint arXiv:2402.11819*.

Chang, S.-J., Wang, J.-N., Wu, H., Qin, L., & Wang, N. (2024). Investigation into the crystal morphology and acidity of LLM-105 in recrystallization. *Journal of Energetic Materials*, *42*(2), 249–272.

Chatterjee, S., Bhattacharya, M., Lee, S.-S., & Chakraborty, C. (2023). Can artificial intelligence-strengthened ChatGPT or other large language models transform nucleic acid research? *Molecular Therapy–Nucleic Acids*, *33*, 205–207.

Chaudhury, S., Rakhra, M., Memon, N., Sau, K., & Ayana, M. T. (2021). Breast cancer calcifications: Identification using a novel segmentation approach. *Computational and Mathematical Methods in Medicine*, *2021*, 1–13.

Chen, C., Feng, X., Zhou, J., Yin, J., & Zheng, X. (2023). Federated large language model: A position paper. *arXiv preprint arXiv:2307.08925*.

Chen, Y., Qian, X., & Zeng, T. A preliminary study on the capability boundary of LLM and a new path towards AGI.

Collins, L. G., Haines, C., Perkel, R., & Enck, R. E. (2007). Lung cancer: Diagnosis and management. *American Family Physician*, *75*(1), 56–63.

De Bondt, R., Nelemans, P., Hofman, P., Casselman, J., Kremer, B., Engelshoven, J., & Beets-Tan, R. (2007). Detection of lymph node metastases in head and neck cancer: A meta-analysis comparing US, USgFNAC, CT and MR imaging. *European Journal of Radiology*, *64*(2), 266–272.

Dery, L., Kolawole, S., Kagey, J.-F., Smith, V., Neubig, G., & Talwalkar, A. (2024). Everybody prune now: Structured pruning of LLMS with only forward passes. *arXiv preprint arXiv:2402.05406*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Doll, R., & Hill, A. B. (1954). The mortality of doctors in relation to their smoking habits. *British Medical Journal*, *1*(4877), 1451.

Dou, Q., Chen, H., Jin, Y., Lin, H., Qin, J., & Heng, P.-A. (2017). Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning. In *Medical image computing and computer assisted intervention- MICCAI 2017: 20th international conference, Quebec City, QC, Canada, September 11–13, 2017, proceedings, part III 20* (pp. 630–638). Springer.

El-Baz, A., Beache, G. M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., Soliman, A., & Abdollahi, B. (2013). Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. *International Journal of Biomedical Imaging*, *2013*, 1–46.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are Gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

Fink, M. A., Bischoff, A., Fink, C. A., Moll, M., Kroschke, J., Dulz, L., Heußel, C. P., Kauczor, H.-U., & Weber, T. F. (2023). Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. *Radiology*, *308*(3), 231362.

Folio, L. R. (2012). *Chest imaging: An algorithmic approach to learning* (Vol. 1, pp. 1–147). Springer Science & Business Media.

Fukunaga, K. (1993). Statistical pattern recognition. In *Handbook of pattern recognition and computer vision* (pp. 33–60). World Scientific.

Gao, D., Wang, H., Li, Y., Sun, X., Qian, Y., Ding, B., & Zhou, J. (2023). Text-to-sql empowered by large language models: A benchmark evaluation. *arXiv preprint arXiv:2308.15363*.

Gao, P., Han, J., Zhang, R., Lin, Z., Geng, S., Zhou, A., Zhang, W., Lu, P., He, C., Yue, X., & Li, H. (2023). Llama-adapter v2: Parameter-efficient visual instruction model. *arXiv preprint arXiv:2304.15010*.

Gayap, H. T., & Akhloufi, M. A. (2024). Deep machine learning for medical diagnosis, application to lung cancer detection: A review. *BioMedInformatics*, *4*(1), 236–284.

Ge, Y., Hua, W., Ji, J., Tan, J., Xu, S., & Zhang, Y. (2023). Openagi: When llm meets domain experts. *arXiv preprint arXiv:2304.04370*.

Gillioz, A., Casas, J., Mugellini, E., & Abou Khaled, O. (2020). Overview of the transformer-based models for nlp tasks. In *2020 15th conference on computer science and information systems (FedCSIS)* (pp. 179–183). IEEE.

Gindi, A., Attiatalla, T. A., & Sami, M. M. (2014). A comparative study for comparing two feature extraction methods and two classifiers in classification of earlystage lung cancer diagnosis of chest x-ray images. *Journal of American Science*, *10*(6), 13–22.

Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., Petrella, F., Spaggiari, L., & Rosell, R. (2015). Non-small-cell lung cancer. *Nature Reviews Disease Primers*, *1*(1), 1–16.

Gruetzemacher, R., & Gupta, A. (2016). Using deep learning for pulmonary nodule detection & diagnosis.

Guldbrandt, L. M., Møller, H., Jakobsen, E., & Vedsted, P. (2017). General practice consultations, diagnostic investigations, and prescriptions in the year preceding a lung cancer diagnosis. *Cancer Medicine*, *6*(1), 79–88.

Gupta, S., Garg, A., Bishnoi, V., & Goel, N. (2023). Pulmonary nodules binary classification using cnn and lstm. In *2023 10th international conference on signal processing and integrated networks (SPIN)* (pp. 174–179). IEEE.

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *2*(4), 100089.

Hamilton, Z., Naffakh, N., Reizine, N. M., Weinberg, F., Jain, S., Gadi, V. K., Bun, C., & Nguyen, R. H.-T. (2023). *Relevance and accuracy of ChatGPT-generated NGS reports with treatment recommendations for oncogene-driven NSCLC*. American Society of Clinical Oncology.

Han, J., Zhang, R., Shao, W., Gao, P., Xu, P., Xiao, H., Zhang, K., Liu, C., Wen, S., Guo, Z., & Lu, X. (2023). Imagebind-LLM: Multi-modality instruction tuning. *arXiv preprint arXiv:2309.03905*.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., & Wang, Y. (2021). Transformer in transformer. *Advances in Neural Information Processing Systems*, *34*, 15908–15919.

Handa, P., Chhabra, D., Goel, N., & Krishnan, S. (2023). Exploring the role of ChatGPT in medical image analysis. *Biomedical Signal Processing and Control*, *86*, 105292.

Handa, P., Gupta, E., Muskan, S., & Goel, N. (2023). A review on software and hardware developments in automatic epilepsy diagnosis using eeg datasets. *Expert Systems*, *40*(9), 13374.

Haver, H. L., Lin, C. T., Sirajuddin, A., Yi, P. H., & Jeudy, J. (2023). Evaluating ChatGPT's accuracy in lung cancer prevention and screening recommendations. *Radiology: Cardiothoracic Imaging*, *5*(4), 230115.

He, K., Mao, R., Lin, Q., Ruan, Y., Lan, X., Feng, M., & Cambria, E. (2023). A survey of large language models for healthcare: From data, technology, and applications to accountability and ethics. *arXiv preprint arXiv:2310.05694*.

Henschke, C. I., McCauley, D. I., Yankelevitz, D. F., Naidich, D. P., McGuinness, G., Miettinen, O. S., Libby, D. M., Pasmantier, M. W., Koizumi, J., Altorki, N. K., & Smith, J. P. (1999). Early lung cancer action project: Overall design and findings from baseline screening. *The Lancet*, *354*(9173), 99–105.

Henzler, T., Konstandin, S., Schmid-Bindert, G., Apfaltrer, P., Haneder, S., Wenz, F., Schad, L., Manegold, C., Schoenberg, S., & Fink, C. (2012). Imaging of tumor viability in lung cancer: Initial results using 23na-mri. In *RöFo-Fortschritte Auf dem Gebiet der Röntgenstrahlen und der Bildgebenden Verfahren* (pp. 340–344). ©Georg Thieme Verlag KG.

Hong, K., Dai, G., Xu, J., Mao, Q., Li, X., Liu, J., Chen, K., Dong, H., & Wang, Y. (2023). Flashdecoding++: Faster large language model inference on gpus. *arXiv preprint arXiv:2311.01282*.

Hosseini, M., & Horbach, S. P. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review*, *8*(1), 4.

Howlader, N., Forjaz, G., Mooradian, M. J., Meza, R., Kong, C. Y., Cronin, K. A., Mariotto, A. B., Lowy, D. R., & Feuer, E. J. (2020). The effect of advances in lung-cancer treatment on population mortality. *New England Journal of Medicine*, *383*(7), 640–649.

Hu, Q., Souza, L. F.d. F., Holanda, G. B., Alves, S. S., Silva, F. H.d. S., Han, T., & Reboucas Filho, P. P. (2020). An effective approach for ct lung segmentation using mask region-based convolutional neural networks. *Artificial Intelligence in Medicine*, *103*, 101792.

Hu, Z., Lan, Y., Wang, L., Xu, W., Lim, E.-P., Lee, R. K.-W., Bing, L., & Poria, S. (2023). LLM-adapters: An adapter family for parameter-efficient fine-tuning of large language models. *arXiv preprint arXiv:2304.01933*.

Hub, M., Handa, P., Nautiyal, D., Chhabra, D., Dhir, M., Saini, A., Jha, S., Mangotra, H., Pandey, N., Thakur, A., & Singh, T. (2024). *Auto-WCEBleedGen Version V1 and V2: Challenge, datasets and evaluation*. Authorea Preprints.

Hub, M., Handa, P., Saini, A., Dutta, S., Pathak, H., Choudhary, N., Goel, N., & Dhanao, J. K. (2024). Auto-pcos classification challenge.

Husain, S. A., & Shigeru, E. (2000). Use of neural networks for feature based recognition of liver region on ct images. In *Neural networks for signal processing X. Proceedings of the 2000 IEEE signal processing society workshop (cat. No. 00TH8501)* (Vol. 2, pp. 831–840). IEEE.

Iannantuono, G. M., Bracken-Clarke, D., Floudas, C. S., Roselli, M., Gulley, J. L., & Karzai, F. (2023). Applications of large language models in cancer care: Current evidence and future perspectives. *Frontiers in Oncology*, *13*.

Inan, H., Upasani, K., Chi, J., Rungta, R., Iyer, K., Mao, Y., Tontchev, M., Hu, Q., Fuller, B., Testuggine, D., & Khabsa, M. (2023). Llama guard: LLM-based input-output safeguard for human-AI conversations. *arXiv preprint arXiv:2312.06674*.

Islam, S., & Walker, R. C. (2013). Advanced imaging (positron emission tomography and magnetic resonance imaging) and image-guided biopsy in initial staging and monitoring of therapy of lung cancer. *Cancer Journal (Sudbury, Mass.)*, *19*(3), 208.

Islami, F., Goding Sauer, A., Miller, K. D., Siegel, R. L., Fedewa, S. A., Jacobs, E. J., McCullough, M. L., Patel, A. V., Ma, J., Soerjomataram, I., Flanders, W. D., Brawley, O. W., Gapstur, S. M., & Jemal, A. (2018). Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA: A Cancer Journal for Clinicians*, *68*(1), 31–54.

Islami, F., Torre, L. A., & Jemal, A. (2015). Global trends of lung cancer mortality and smoking prevalence. *Translational Lung Cancer Research*, *4*(4), 327.

Jassim, M. M., & Jaber, M. M. (2022). Systematic review for lung cancer detection and lung nodule classification: Taxonomy, challenges, and recommendation future works. *Journal of Intelligent Systems*, *31*(1), 944–964.

Jayakumar, T., Farooqui, F., & Farooqui, L. (2023). Large language models are legal but they are not: Making the case for a powerful legalLLM. *arXiv preprint arXiv:2311.08890*.

Ji, J., Liu, M., Dai, J., Pan, X., Zhang, C., Bian, C., Sun, R., Wang, Y., & Yang, Y. (2023). Beavertails: Towards improved safety alignment of LLM via a human-preference dataset. *arXiv preprint arXiv:2307.04657*.

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., & Wen, Q. (2023). Time-LLM: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*.

Joublin, F., Ceravola, A., Deigmoeller, J., Gienger, M., Franzius, M., & Eggert, J. (2023). A glimpse in ChatGPT capabilities and its impact for AI research. *arXiv preprint arXiv:2305.06087*.

Kalla, D., & Smith, N. (2023). Study and analysis of ChatGPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology*, *8*(3).

Kareem, H. F., AL-Husieny, M. S., Mohsen, F. Y., Khalil, E. A., & Hassan, Z. S. (2021). Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset. *Indonesian Journal of Electrical Engineering and Computer Science*, *21*(3), 1731.

Ke, P., Wen, B., Feng, Z., Liu, X., Lei, X., Cheng, J., Wang, S., Zeng, A., Dong, Y., Wang, H., & Tang, J. (2023). Critiquellm: Scaling LLM-as-critic for effective and explainable evaluation of large language model generation. *arXiv preprint arXiv:2311.18702*.

Khawaja, M. A., Aziz, M. Z., & Iqbal, N. (2004). Effectual lung segmentation for cad systems using ct scan images. In *8th international multitopic conference, 2004. Proceedings of INMIC 2004* (pp. 49–54). IEEE.

Koyama, H., Ohno, Y., Seki, S., Nishio, M., Yoshikawa, T., Matsumoto, S., & Sugimura, K. (2013). Magnetic resonance imaging for lung cancer. *Journal of Thoracic Imaging*, *28*(3), 138–150.

Kumari, A., Kumari, A., Singh, A., Singh, S. K., Juhi, A., Dhanvijay, A. K. D., Pinjar, M. J., Mondal, H., & Dhanvijay, A. K. (2023). Large language models in hematology case solving: A comparative study of ChatGPT-3.5, google bard, and microsoft bing. *Cureus*, *15*(8), e43861.

Lai, Z., Zhang, H., Wu, W., Bai, H., Timofeev, A., Du, X., Gan, Z., Shan, J., Chuah, C.-N., Yang, Y., & Cao, M. (2023). From scarcity to efficiency: Improving clip training via visual-enriched captions. *arXiv preprint arXiv:2310.07699*.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT–does digital watermarking offer a solution? *International Journal for Educational Integrity*, *19*(1), 10.
Lanyi, M. (2000). Differential diagnosis of microcalcifications. In *Radiological diagnosis of breast diseases* (pp. 89–136). Springer.
Laskaridis, S., Kateveas, K., Minto, L., & Haddadi, H. (2024). Melting point: Mobile evaluation of language transformers. *arXiv preprint arXiv:2403.12844*.
Lee, N., Laine, A. F., Mrquez, G., Levsky, J. M., & Gohagan, J. K. (2009). Potential of computer-aided diagnosis to improve CT lung cancer screening. *IEEE Reviews in Biomedical Engineering*, *2*, 136–146.
Lemjabbar-Alaoui, H., Hassan, O. U., Yang, Y.-W., & Buchanan, P. (2015). Lung cancer: Biology and treatment options. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1856*(2), 189–210.
Li, C., Su, X., Fan, C., Han, H., Xue, C., & Zheng, C. (2023). Quantifying the impact of large language models on collective opinion dynamics. *arXiv preprint arXiv:2308.03313*.
Li, J., Tang, Z., Ding, Y., Wang, P., Guo, P., You, W., Qiao, D., Chen, W., Fu, G., Zhu, Q., & Zhou, G. (2023). Openba: An open-sourced 15b bilingual asymmetric seq2seq model pre-trained from scratch. *arXiv preprint arXiv:2309.10706*.
Li, L., Zhang, Y., & Chen, L. (2023). Prompt distillation for efficient llm-based recommendation. In *Proceedings of the 32nd ACM international conference on information and knowledge management* (pp. 1348–1357).
Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, *13*(4), 410.
Luo, Y., Kong, Q., Xu, N., Cao, J., Hao, B., Qu, B., Chen, B., Zhu, C., Zhao, C., Zhang, D., & Feng, F. (2023). Yayi 2: Multilingual open-source large language models. *arXiv preprint arXiv:2312.14862*.
Ma, X., Fang, G., & Wang, X. (2023). LLM-pruner: On the structural pruning of large language models. *arXiv preprint arXiv:2305.11627*.
MacRedmond, R., Logan, P., Lee, M., Kenny, D., Foley, C., & Costello, R. (2004). Screening for lung cancer using low dose CT scanning. *Thorax*, *59*(3), 237.
Makridakis, S., Petropoulos, F., & Kang, Y. (2023). Large language models: Their success and impact. *Forecast*, *5*(3), 536–549.
Mao, R., Chen, G., Zhang, X., Guerin, F., & Cambria, E. (2023). GPTEval: A survey on assessments of ChatGPT and GPT-4. *arXiv preprint arXiv:2308.12488*.
Mascaux, C., Tsao, M.-S., & Hirsch, F. R. (2018). Genomic testing in lung cancer: Past, present, and future. *Journal of the National Comprehensive Cancer Network*, *16*(3), 323–334.
Maurer, A. (2021). An early prediction of lung cancer using CT scan images. *Journal of Computing and Natural Science*, *1*, 39–44.
Menghani, G. (2023). Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys*, *55*(12), 1–37.
Mhaske, D., Rajeswari, K., & Tekade, R. (2019). Deep learning algorithm for classification and prediction of lung cancer using ct scan images. In *2019 5th international conference on computing, communication, control and automation (ICCUBEA)* (pp. 1–5). IEEE.
Montazeri, M., Afraz, A., Farimani, R. M., & Ghasemian, F. (2021). Natural language processing systems for diagnosing and determining level of lung cancer: A systematic review. *Frontiers in Health Informatics*, *10*(1), 68.
Mridha, M., Prodeep, A. R., Hoque, A. M., Islam, M. R., Lima, A. A., Kabir, M. M., Hamid, M. A., & Watanobe, Y. (2022). A comprehensive survey on the progress, process, and challenges of lung cancer detection and classification. *Journal of Healthcare Engineering*, *2022*(1), 5905230.
Mumtaz, U., Ahmed, A., & Mumtaz, S. (2024). LLMS-healthcare: Current applications and challenges of large language models in various medical specialties. *Artificial Intelligence in Health*, *1*(2), 16–28.
Naik, N., Khandelwal, A., Joshi, M., Atre, M., Wright, H., Kannan, K., Hill, S., Mamidipudi, G., Srinivasa, G., Bifulco, C., & Piening, B. (2023). Applying large language models for causal structure learning in non small cell lung cancer. *arXiv preprint arXiv:2311.07191*.
Nam, J. G., Park, S., Hwang, E. J., Lee, J. H., Jin, K.-N., Lim, K. Y., Vu, T. H., Sohn, J. H., Hwang, S., Goo, J. M., & Park, C. M. (2019). Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology*, *290*(1), 218–228.
Niu, C., & Wang, G. (2023). CT multi-task learning with a large image-text (lit) model. *bioRxiv*, 2023–04.
Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*.
Norlund, T., Doostmohammadi, E., Johansson, R., & Kuhlmann, M. (2023). On the generalization ability of retrieval-enhanced transformers. *arXiv preprint arXiv:2302.12128*.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., & Chou, R. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906.
Panagoulias, D. P., Palamidas, F. A., Virvou, M., & Tsihrintzis, G. A. (2023). Evaluating the potential of llms and ChatGPT on medical diagnosis and treatment. In *2023 14th international conference on information, intelligence, systems & applications (IISA)* (pp. 1–9). IEEE.
Papaleo, E., Ranzani, V., Tripodi, F., Vitriolo, A., Cirulli, C., Fantucci, P., Alberghina, L., Vanoni, M., De Gioia, L., & Coccetti, P. (2011). An acidic loop and cognate phosphorylation sites define a molecular switch that modulates ubiquitin charging activity in Cdc34-like enzymes. *PLoS Computational Biology*, *7*(5), 1002056.
Parveen, S. S., & Kavitha, C. (2012). A review on computer aided detection and diagnosis of lung cancer nodules. *International Journal of Computers and Technology*, *3*(3), 393–400.
Penedo, G., Malartic, Q., Hesslow, D., Cojocaru, R., Cappelli, A., Alobeidli, H., Pannier, B., Almazrouei, E., & Launay, J. (2023). The refinedweb dataset for falcon LLM: Outperforming curated corpora with web data, and web data only. *arXiv preprint arXiv:2306.01116*.
Peng, B., Li, C., He, P., Galley, M., & Gao, J. (2023). Instruction tuning with GPT-4. *arXiv preprint arXiv:2304.03277*.
Phillips, G. (2023). Evaluating the carousel format of sunderland online's LLM programme. *The Law Teacher*, *57*(4), 560–564.
Poerner, N., Waltinger, U., & Schütze, H. (2019). E-BERT: Efficient-yet-effective entity embeddings for BERT. *arXiv preprint arXiv:1911.03681*.
Prazanowska, K. H., & Lim, S. B. (2023). An integrated single-cell transcriptomic dataset for non-small cell lung cancer. *Scientific Data*, *10*(1), 167.
Preethi, D., & Ganapathy, K. (2022). Novel lung cancer detection using ann classifier in comparison with decision tree to measure the accuracy, sensitivity, specificity and precision. In *2022 2nd international conference on innovative practices in technology and management (ICIPTM)* (Vol. 2, pp. 528–534). IEEE.
Rahsepar, A. A., Tavakoli, N., Kim, G. H. J., Hassani, C., Abtin, F., & Bedayat, A. (2023). How AI responds to common lung cancer questions: ChatGPT vs google bard. *Radiology*, *307*(5), 230922.
Rajbhandari, S., Rasley, J., Ruwase, O., & He, Y. (2020). Zero: Memory optimizations toward training trillion parameter models. In *SC20: International conference for high performance computing, networking, storage and analysis* (pp. 1–16). IEEE.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ravichandiran, S. (2021). *Getting started with Google BERT: Build and train state-of-the-art natural language processing models using BERT*. Packt Publishing Ltd.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. In *Internet of things and cyber-physical systems*.

Rebedea, T., Dinu, R., Sreedhar, M., Parisien, C., & Cohen, J. (2023). Nemo guardrails: A toolkit for controllable and safe LLM applications with programmable rails. *arXiv preprint arXiv:2310.10501*.

Ribar, L., Chelombiev, I., Hudlass-Galley, L., Blake, C., Luschi, C., & Orr, D. (2023). Sparq attention: Bandwidth-efficient LLM inference. *arXiv preprint arXiv:2312.04985*.

Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and open-AI models: A preliminary review. *Future Internet*, *15*(6), 192.

Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2024). Llms in e-commerce: A comparative analysis of GPT and llama models in product review evaluation. *Natural Language Processing Journal*, *6*, *100056*.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Rydzewski, N. R., Dinakaran, D., Zhao, S. G., Ruppin, E., Turkbey, B., Citrin, D. E., & Patel, K. R. (2024). Comparative evaluation of LLMS in clinical oncology. *NEJM AI*, *1*(5), 2300151.

Sainz, O., Campos, J. A., García-Ferrero, I., Etxaniz, J., Lacalle, O. L., & Agirre, E. (2023). NLP evaluation in trouble: On the need to measure LLM data contamination for each benchmark. *arXiv preprint arXiv:2310.18018*.

Setio, A. A. A., Traverso, A., de Bel, T., Berens, M. S., Van Den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M. E., Geurts, B., Gugten, R., Heng, P. A., Jansen, B., de Kaste, M. M. J., Kotov, V., Lin, J. Y. H., Manders, J. T. M. C., Sóñora-Mengana, A., García-Naranjo, J. C., … Jacobs, C. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis*, *42*, 1–13.

Shariaty, F., & Mousavi, M. (2019). Application of cad systems for the automatic detection of lung nodules. *Informatics in Medicine Unlocked*, *15*, 100173.

Sharma, R. (2022). Mapping of global, regional and national incidence, mortality and mortality-to-incidence ratio of lung cancer in 2020 and 2050. *International Journal of Clinical Oncology*, *27*(4), 665–675.

Sharma, R., Tripathi, M., D'Souza, M., Jaimini, A., Saw, S. K., Singh, D., Mishra, A., & Mondal, A. (2009). The importance of 18F-FDG PET/CT, CT and x-rays in detecting primary stage III A lung cancer and the incidence of extra thoracic metastases. *Hellenic Journal of Nuclear Medicine*, *12*(1), 22–25.

Shayegani, E., Mamun, M. A. A., Fu, Y., Zaree, P., Dong, Y., & Abu-Ghazaleh, N. (2023). Survey of vulnerabilities in large language models revealed by adversarial attacks. *arXiv preprint arXiv:2310.10844*.

Shen, Y.-C., Chen, C.-H., & Tu, C.-Y. (2021). Advances in diagnostic bronchoscopy. *Diagnostics*, *11*(11), 1984.

Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., & Catanzaro, B. (2019). Megatron-lm: Training multi-billion parameter language models using model parallelism. *arXiv preprint arXiv:1909.08053*.

Shum, H.-Y., He, X.-D., & Li, D. (2018). From Eliza to Xiaoice: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*, 10–26.

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, *73*(1), 17–48.

Silvestri, G. A., Gould, M. K., Margolis, M. L., Tanoue, L. T., McCrory, D., Toloza, E., & Detterbeck, F. (2007). Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines. *Chest*, *132*(3), 178–201.

Singha, A., Cambronero, J., Gulwani, S., Le, V., & Parnin, C. (2023). Tabular representation, noisy operators, and impacts on table structure understanding tasks in LLMS. *arXiv preprint arXiv:2310.10358*.

Sommer, G., Tremper, J., Koenigkam-Santos, M., Delorme, S., Becker, N., Biederer, J., Kauczor, H.-U., Heussel, C. P., Schlemmer, H.-P., & Puderbach, M. (2014). Lung nodule detection in a high-risk population: Comparison of magnetic resonance imaging and low-dose computed tomography. *European Journal of Radiology*, *83*(3), 600–605.

Song, C. H., Wu, J., Washington, C., Sadler, B. M., Chao, W.-L., & Su, Y. (2023). Llm-planner: Few-shot grounded planning for embodied agents with large language models. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2998–3009).

Spark, H., & Luckner, R. (2022). Simplified vortex methods to model wake vortex roll-up in real-time simulations for fuel-saving formation flight. *CEAS Aeronautical Journal*, *13*(3), 763–778.

Sreevaatsav, B., Srikanth, C., Vamshi, B. K., Deepesh, G. V., & Sai, K. S. (2023). Llm for nvidia sdks. In *2023 international conference on emerging techniques in computational intelligence (ICETCI)* (pp. 176–179). IEEE.

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM – A tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249.

Sur, C. (2020). RBN: Enhancement in language attribute prediction using global representation of natural language transfer learning technology like google BERT. *SN Applied Sciences*, *2*(1), 22.

Suzgun, M., Melas-Kyriazi, L., & Jurafsky, D. (2022). Prompt-and-rerank: A method for zero-shot and few-shot arbitrary textual style transfer with small language models. *arXiv preprint arXiv:2205.11503*.

Suzuki, K. (2012). Pixel-based machine learning in medical imaging. *Journal of Biomedical Imaging*, *2012*, 1.

Tam, A. L., Lim, H. J., Wistuba, I. I., Tamrazi, A., Kuo, M. D., Ziv, E., Wong, S., Shih, A. J., Webster, R. J., III, Fischer, G. S., Nagrath, S., Davis, S. E., White, S. B., & Ahrar, K. (2016). Image-guided biopsy in the era of personalized cancer care: Proceedings from the society of interventional radiology research consensus panel. *Journal of Vascular and Interventional Radiology: JVIR*, *27*(1), 8–19.

Tang, H., Zhang, C., & Xie, X. (2019). Nodulenet: Decoupled false positive reduction for pulmonary nodule detection and segmentation. In *Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part VI 22* (pp. 266–274). Springer.

Tang, R., Chuang, Y.-N., & Hu, X. (2023). The science of detecting llm-generated texts. *arXiv preprint arXiv:2303.07205*.

Teng, Z., Ning, R., Liu, J., Zhou, Q., & Zhang, Y. (2023). Glore: Evaluating logical reasoning of large language models. *arXiv preprint arXiv:2310.09107*.

Tetko, I. V., Karpov, P., Van Deursen, R., & Godin, G. (2020). State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nature Communications*, *11*(1), 5575.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., & Li, Y. (2022). Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239*.

Tod, A. M., Craven, J., & Allmark, P. (2008). Diagnostic delay in lung cancer: A qualitative study. *Journal of Advanced Nursing*, *61*(3), 336–343.

Travis, W. D., Dacic, S., Wistuba, I., Sholl, L., Adusumilli, P., Bubendorf, L., Bunn, P., Cascone, T., Chaft, J., Chen, G., & Chou, T. Y. (2020). IASLC multidisciplinary recommendations for pathologic assessment of lung cancer resection specimens after neoadjuvant therapy. *Journal of Thoracic Oncology*, *15*(5), 709–740.

Tsou, C.-H., Lor, K.-L., Chang, Y.-C., & Chen, C.-M. (2015). Anatomy packing with hierarchical segments: An algorithm for segmentation of pulmonary nodules in ct images. *Biomedical Engineering Online*, *14*, 1–17.

Tu, H., Zhao, B., Wei, C., & Xie, C. (2023). Sight beyond text: Multi-modal training enhances LLMS in truthfulness and ethics. *arXiv preprint arXiv:2309.07120*.

Umapathi, L. K., Pal, A., & Sankarasubbu, M. (2023). Med-halt: Medical domain hallucination test for large language models. *arXiv preprint arXiv:2307.15343*.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer.

Walton, L., McNeill, R., Stevens, W., Murray, M., Lewis, C., Aitken, D., & Garrett, J. (2013). Patient perceptions of barriers to the early diagnosis of lung cancer and advice for health service improvement. *Family Practice*, *30*(4), 436–444.

Wang, S., Chung, F.-L., Wu, J., & Wang, J. (2014). Least learning machine and its experimental studies on regression capability. *Applied Soft Computing*, *21*, 677–684.

Wei, A., Haghtalab, N., & Steinhardt, J. (2023). Jailbroken: How does LLM safety training fail? *arXiv preprint arXiv:2307.02483*.

Weizenbaum, J. (1966). Eliza—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Westreich, D. (1990). A practical method for computing the exponential of a matrix and its integral. *Communications in Applied Numerical Methods*, *6*(5), 375–380.

Winkels, M., & Cohen, T. S. (2018). 3D G-CNNS for pulmonary nodule detection. *arXiv preprint arXiv:1804.04656*.

Wu, C., Zhang, X., Zhang, Y., Wang, Y., & Xie, W. (2023). PMC-llama: Further finetuning llama on medical papers. *arXiv preprint arXiv:2304.14454*.

Wu, K., Wu, E., Cassasola, A., Zhang, A., Wei, K., Nguyen, T., Riantawan, S., Riantawan, P. S., Ho, D. E., & Zou, J. (2024). How well do LLMs cite relevant medical references? An evaluation framework and analyses. *arXiv preprint arXiv:2402.02008*.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Zhang, S., Zhu, E., Li, B., Jiang, L., Zhang, X., & Wang, C. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*.

Wu, S., Fei, H., Qu, L., Ji, W., & Chua, T.-S. (2023). Next-GPT: Any-to-any multimodal LLM. *arXiv preprint arXiv:2309.05519*.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122–1136.

Xia, X., & Zhang, R. (2023). A novel lung nodule accurate segmentation of PET-CT images based on convolutional neural network and graph model. *IEEE Access*, *11*, 34015–34031.

Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., Compas, C., Martin, C., Flores, M. G., Zhang, Y., & Magoc, T. (2022). Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records. *arXiv preprint arXiv:2203.03540*.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, *31*(7), 1235–1270.

Zhang, C., Bauer, S., Bennett, P., Gao, J., Gong, W., Hilmkil, A., Jennings, J., Ma, C., Minka, T., Pawlowski, N., & Vaughan, J. (2023). Understanding causality with large language models: Feasibility and opportunities. *arXiv preprint arXiv:2304.05524*.

Zhang, G., Luo, Y., Zhu, D., Xu, Y., Sun, Y., & Lu, J. (2018). Spatial pyramid dilated network for pulmonary nodule malignancy classification. In *2018 24th international conference on pattern recognition (ICPR)* (pp. 3911–3916). IEEE.

Zhang, H., Li, X., & Bing, L. (2023). Video-llama: An instruction-tuned audio-visual language model for video understanding. *arXiv preprint arXiv:2306.02858*.

Zhang, J., Chen, Y., Niu, N., & Liu, C. (2023). A preliminary evaluation of ChatGPT in requirements information retrieval. *arXiv preprint arXiv:2304.12562*.

Zhang, J., Krishna, R., Awadallah, A. H., & Wang, C. (2023). EcoAssistant: Using LLM assistant more affordably and accurately. *arXiv preprint arXiv:2310.03046*.

Zhang, X., Yu, B., Yu, H., Lv, Y., Liu, T., Huang, F., Xu, H., & Li, Y. (2023). Wider and deeper LLM networks are fairer LLM evaluators. *arXiv preprint arXiv:2308.01862*.

Zhang, Z., Zheng, C., Tang, D., Sun, K., Ma, Y., Bu, Y., Zhou, X., & Zhao, L. (2023). Balancing specialized and general skills in LLMS: The impact of modern tuning and data strategy. *arXiv preprint arXiv:2310.04945*.

Zhao, D., Liu, Y., Yin, H., & Wang, Z. (2023). An attentive and adaptive 3D CNN for automatic pulmonary nodule detection in CT image. *Expert Systems with Applications*, *211*, 118672.

Zheng, S., Shen, Z., Pei, C., Ding, W., Lin, H., Zheng, J., Pan, L., Zheng, B., & Huang, L. (2021). Interpretative computer-aided lung cancer diagnosis: From radiology analysis to malignancy evaluation. *Computer Methods and Programs in Biomedicine*, *210*, 106363.

Zou, H., Zhao, Q., Bariah, L., Bennis, M., & Debbah, M. (2023). Wireless multi-agent generative AI: From connected intelligence to collective intelligence. *arXiv preprint arXiv:2307.02757*.

AUTHOR BIOGRAPHIES

Arushi Garg is a recent graduate of B.Tech in Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India. She is currently associated with Adobe, Bangalore, India as a Member of Technical Staff-software engineering. Her research area of interest includes artificial intelligence, computer vision, medical image processing, deep learning, transformers and large

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

language models. She has authored 6 conference articles. She was the AWS Cloud Captain in the year 2023, and GGSoC program manager in the year 2024, and has actively participated in training and placement cell at her university.

Smridhi Gupta is a recent graduate of B.Tech in Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India. She is currently associated with Citi, Pune as a technology analyst. She has actively participated in the research internship from Artificial Intelligence Institute of South Carolina and was a GWC scholar, flipkart, and Millennium Fellow, United Nations Academic Impact. Her research area of interest includes artificial intelligence, computer vision, medical image processing, deep learning, transformers and large language models and she has authored 5 conference articles.

Soumya Vats is a recent graduate of B.Tech in Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India. She is currently associated with Deutsche bank as a graduate analyst. Her research area of interest includes artificial intelligence, computer vision, medical image processing, deep learning, transformers and large language models and she has authored 2 conference articles. She has actively participated in the training and placement cell and technical and non-technical clubs at her university.

Dr. Palak Handa, Ph.D. is a working as a Scientist at the Research Centre for Medical Image Analysis and Artificial Intelligence (MIAAI), Department of Medicine, Danube Private University, Austria since June 2024. She completed her M.Tech. in Very Large-Scale Integrated Circuits (VLSI) from Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, New Delhi, India and Ph.D. in Computer Vision and Medical Image Processing from Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi, India. Her research area of interest includes medical image analysis, computer vision, artificial intelligence, biomedical signal processing and CAD simulations. She has developed and open-sourced 10 biomedical datasets with different biomedical modalities including video capsule endoscopy, colonoscopy, electroencephalography, and ultrasound imaging. She has authored 12 journal articles, 20 conference articles, 8 book chapters, and 2 granted national patents of international repute. She has co-organized 3 biomedical challenges in collaboration with famous international conferences and association namely ICIP 2024, Abu Dhabi, CVIP 2023, IIT Jammu, and IEEE Delhi Section and EDS Society.

Dr. Nidhi Goel, Ph.D. is working as a full-Professor at the Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University, Delhi, India. She was the head of the ECE Department for two consecutive terms (2018-2023). Prior to joining IGDTUW, she was associated with Delhi Technological University (formerly, DCE), Delhi, India. She has a teaching and research experience of approximately 20 years. She pursued her research interest and completed her PhD from the Indian Institute of Technology Roorkee, India. Her area of interests includes medical image processing, deep learning, artificial intelligence applications, multimedia security, computer vision, soft computing techniques. She has organized several biomedical challenges, workshops, has granted patents, research project grants, and published more than 80 research articles, editorials, and chapters in journals and conferences of international repute.

How to cite this article: Garg, A., Gupta, S., Vats, S., Handa, P., & Goel, N. (2024). Prospect of large language models and natural language processing for lung cancer diagnosis: A systematic review. *Expert Systems*, e13697. https://doi.org/10.1111/exsy.13697

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13697 by Delhi Technological University, Wiley Online Library on [21/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

PUBLIC SECTOR BANKS IN INDIA: A REVIEW OF LITERATURE

K. V. Bhanu Murthy
Former Dean Faculty of Commerce and Business
Delhi School of Economics
University of Delhi
Currently Professor
Delhi Technological University
Email: bhanumurthykv@yahoo.com

Amit Kumar Singh
Department of Commerce
Delhi School of Economics
University of Delhi
Email: aksingh1@commerce.du.ac.in

Sonia Kaushik
Associate Professor
Department of Commerce
Bharati College
University of Delhi
Email: soniakaushik@bharati.du.ac.in

Abstract
This study aims to gain insight on various issues related to the public sector banking industry in India. For this study, forty research papers and studies related to market structure, conduct and performance of public sector banks in India have been examined. These studies made important contributions to the comparative analysis of profitability of public sector banking industry with respect to private and foreign banks, but none of them provided a rigorous analysis of the impact of reforms on the profitability of public sector banking industry over time. Issues explored in these studies are related to interest and non-interest income, bank size, technology issues, and formation of non-performing assets. In many studies, DEA, stochastic frontier analysis, ratio, regression, graphs and charts, production function have been used. In very few studies, the S-C-P framework has been used

in the banking industry in India. Many of the studies have compared public sector banks with old private, new private and foreign banks. None of these studies have examined public sector banks at bank level. But now questions arise whether with the adaptation of liberal economic policies and market model, other objectives than profit are irrelevant. Secondly, given their own frame of reference, can they be dubbed as good or bad?

Keywords: Banks, Investment decisions, Market structure, Advertising, Investment.

JEL Keywords: G21, G11, L10, M37, E22

1. Introduction

In the development of every field of a study, a stage is reached when it becomes indispensable to take stock of the whole work done in the past. A critical review of literature needs to precede the research work. This helps to focus on the areas which did not receive adequate attention. The literature review must be arranged in terms of a theoretical framework used in the study so that the link between the theoretical framework used in the study and the review of literature becomes clear enough. This leads us to classify the extant literature in the banking industry in three categories relating to structure, conduct and performance because in the present study the Structure-Conduct-Performance paradigm has been used as a theoretical framework. Moreover, there is a large and growing volume of literature of varying quality in the field of Indian banking industry. Therefore, a brief review of this literature and committee recommendations on the banking sector has been made.

The research questions for this article are as follows:

1. How has liberalization affected the public sector banking industry in India?
2. Are there strategic groups within public sector banks in India?
3. How has the entry of private sector banks led to changes in the public sector banking industry?
4. What have been the changes in the market structure of public sector banks in India?
5. Which conduct variables can better measure the behavior of public sector banks?
6. How have market structure and conduct affected performance of public sector banks in India?

The literature review helps to focus on the areas which are not explored enough. The literature review has been arranged in terms of a theoretical framework that will be used in the proposed study so that the link between the theoretical framework used in the study and the review of literature becomes clear enough. This leads us to classify the literature review in the banking industry, in categories relating to structure, conduct and performance. Keeping these research questions in view, the theme of this paper is to lay out the literature that explores the public sector banking industry in detail. For this paper, we have explored forty research papers.

A brief introduction and research questions of this chapter (Section 1.1) is followed by the theoretical framework in section 1.2. Section 1.3 provides the literature review on the structure, conduct and performance of PSBs. Section 1.4 summarizes the research gap and conclusion of this paper is highlighted in section 1.5.

2. Theoretical Framework

Within the broader industrial organisation approach, there are three strands that are used to establish characteristics of an industry. The S-C-P framework (figure 1.) which is the most appropriate to study the banking industry has been used to study the nature of the banking market in terms of market structure, conduct and performance.

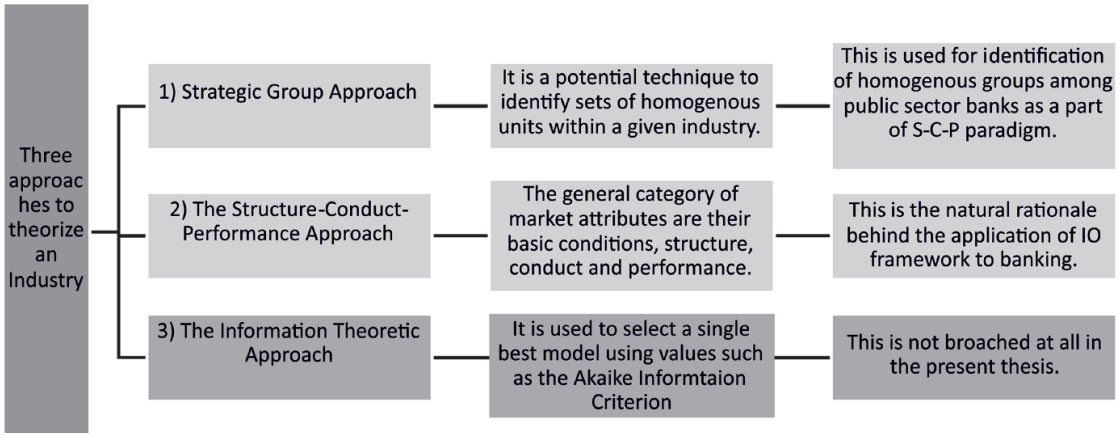


Figure 1: S-C-P Framework

Source: Authors' own compilation

The S-C-P paradigm assumes that there are certain attributes which are given and are referred to as basic conditions. Furthermore, there is a chain of causation that logically determines all the other attributes in a sequence as follows: In the chain, basic conditions are the primary determinants of the market or industry structure. In the next step, the structure of the industry influences the conduct or behavior of the participants in the industry. Conduct, in turn, influences performance finally. The schema of traditional S-C-P paradigm was like the flowchart below (figure 2):



Figure 2: Flowchart of S-C-P paradigm

Source: Authors' own compilation

3. Literature review

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
	A. Literature Review on Structure					
1	(Murthy & Deb, 2014) "Measurement and Determinants of Competition in Private Banking Industry in India during 1992-2002"	To analyse the limitations of concentration as a measure of competition.	1992-2002	Concentration and competition	Three-equation simultaneous equation model is used in the research through Tobit technique.	A fall in concentration ratio is equated with a rise in competition. An analysis of concentration ratio needs to be supplemented by an analysis of changes in the identities of leading firms.
2	(Bajaj, 2008) "Indian Public Sector Banking Industry: 1992-2006- An Impact of Reforms"	The focus was mainly on the impact of deregulation on public sector banks in India.	1992-2006	Assets, deposit, and advances	SCP framework	Share of public sector banks was slowly decreasing over the time because of the entry of new private banks and foreign banks.
3	(Gupta, 2013) "A comparative study of public and private banking industry in India"	It assesses the impact of banking reforms and competition in public and private sectors.	1996-2010	Net profit, operating efficiency, and absolute spread.	Use of Panel data research methodology involving 27 public sector banks and 19 private sector banks over fifteen years.	Competition has emerged in the form of new banks vs. incumbents rather than in the form of public vs. private banks. Operating efficiency and better conduct in terms of asset management, technology, advertisement, and absolute spread are the keys to performance.

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
4	(Amel & Rhoades, 1988) "Strategic groups in Banking"	This paper tested the strategic groups hypothesis using cluster analysis.	1978, 1981, and 1984	U.S. Treasury securities; state and local government securities; real estate loans; construction loans; farm loans; residential loans; multifamily nonfarm loans; commercial and industrial (C & I) loans; and loans to individuals	The paper uses Analytical approach and non-hierarchical clustering algorithm.	The results indicate that approximately six strategic groups exist in banking and are stable over time Implications of the results are (1) intra industry profit differences may be due to strategic groups rather than efficiency differences, (2) markets may generally be defined too broadly, (3) investigations for collusion need to focus on homogeneous groups in an industry rather than the whole industry, and (4) there is no simple strategy choice for banks between retail and wholesale banking.
5	(Hannan, 1991) "Bank commercial loan markets and the role of market structure: Evidence from surveys of commercial lending"	To estimate the relationship between market structure and various aspects of bank conduct and performance as implied by the structure-conduct-performance (S-C-P) paradigm and thereby assess the most commonly tested relationships in this large literature.	1984-1986	Market concentration, maturity of the loan, size of loan, total assets, market population, business failure rate, and average hourly earnings of nonsupervisory manufacturing employees in the market.	Secondary data has been used by the researcher. For evaluating data, coefficient and regression is used.	Greater price rigidity in concentrated markets is offered as a possible reason for the observed differences in the relationship between loan rates and market structure over time.
6	(Heggestad & Rhoades, 1976) "An Analysis of changes in Bank Market Structure"	It tests the market structure-stability relationship in commercial banking.	1966 and 1972	Controllable variables are bank mergers, bank holding company acquisitions, branching, and holding company laws while non-controllable variables are initial market structure, market growth, and size of banks.	Analysis based on 228 markets and over 2,000 firms was conducted to determine whether elements of market structure have a systematic influence on stability.	The greater the level of concentration in a market, the greater is the stability of firms. Thus, the results are generally consistent with expectations.

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
	B. Literature review on Conduct					
7	(Arasu, Sridevi, Nageswari, & Ramya, 2019) "A Study on Analysis of Non-Performing Assets and its Impact on Profitability	It investigates the impact of the level of Non-Performing Assets of private and public sector banks on profitability.	2014-2018	Profitability and Non-performing assets	The secondary data is collected for NPAs of 10 banks. The study has used mean, standard deviation, Pearson cross-correlation, and linear regression model.	The trend of Gross & Net NPA and profitability of both public and private sector banks are found significant.
8	(Numer & Devika, 2019) "A Study on comparative Analysis of Non-Performing Assets in Selected Private Sector Banks in India"	It tries to study the trend of Gross NPA and Net NPA of Axis Bank, Federal Bank, IndusInd Bank only.	2014-2018	GNPA (Gross Non-performing Assets), and NNPA (Net Non-performing Assets)	The paper has used ratio analysis, trend analysis analysing data and Pearson correlation coefficient tool.	On concluding three private sector banks for the comparative is based on the sales report and their market values in the banking sectors.
9	(Kambar, 2019) "A study on the consolidation and merger of public sector banks in India: Issues and challenges"	It attempts to understand mergers and consolidation.	2016-2017	Deposits, Gross Advances, Gross NPAs, and Net Profits/Loss	The study has used secondary data based on descriptive analysis methods.	A steady rise in the NPA's of banks affects not only the banking sector but the country's economy. creating a bad bank remains a politically volatile idea and is difficult to implement.
10	(Kapadia & Madhav, 2019) "NPAs in Indian Scheduled Commercial Banks: Origination and Impact on Economy"	To investigate the trend in NPAs.	2006-2017	GNPA (Gross Non-performing Assets), and NNPA (Net Non-performing Assets)	The study is descriptive in nature.	The prudential norms laid down by the RBI must equip the banking system to reduce if not completely to keep away from the problem of NPAs.
11	(Kumari, Singh, & Sharma, 2017) "Impact of Non-Performing Assets (NPAs) on Financial Performance of Indian Banking Sector"	It examines the relationship between NPAs and financial performance (ROA) of selected public and private sector banks.	2013-2017	Return on assets, Gross Non-performing Assets, Net Non-performing Assets, and Age	It is an empirical study regression model.	The study found a significant and positive impact of GNPA and NNPA on financial performance (ROA) of public sector banks.
12	(Raj, Jain, Bansal, & Verma, 2018) "Non-Performing Assets: A Comparative Study of SBI & ICICI Bank from 2014-2017"	To study whether there exists any linear relationship between Net profit and Net NPA in case of both the banks.	2014-2017	NPAs (Non-performing Assets), and Net Profit	The data is collected using secondary data analysis and the techniques used involve correlation analysis using SPSS.	On comparing the two banks based on the effect on its profitability, SBI has higher NPAs as compared to ICICI bank because of its public nature.

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
13	(Rayudu & Rajesh, 2018) "A Study on Non-Performing Assets in Selected Public and Private Sector Banks in India"	To study the status of Non-Performing Assets in selected public and Private sector banks.	2015-2017	NPAs (Non-performing Assets)	This study used techniques of mean, standard deviation, and two-way analysis of variance (ANOVA).	The government, Reserve Bank of India and management of banks should make more reforms in the financial sector especially in banking to control and avoid the bad loans and advances.
14	(Mishra & Pawaskar, 2017) "A Study of Non-Performing Assets and its Impact on Banking Sector"	The present study tries to understand the NPAs sector wise and the recovery through various channels.	2011-2016	NPAs (Non-performing Assets)	Financial Ratios	The conclusion of the study is that the Bank needs to have a better credit appraisal system so as to prevent NPAs from occurring.
15	(Alagarsamy, 2017) "Performance of non-performing assets (NPA) in state bank of India"	It analyses the performance of NPA in state banks of India.	2012-2017	GNPA (Gross Non-performing Assets)	Trend Analysis	The researcher concluded that the government and bankers have to make a serious effort to decrease the NPA otherwise it thoroughly damages the bank profit of the banker and also is not good for the developing countries.
16	(Mittal & Suneja, 2017) "The problem of Rising Non-performing Assets in Banking Sector in India: Comparative analysis of public and private sector banks"	It analyses the impact of NPAs on banks' performance and the reasons for mounting NPAs in banks in India.	2005-2016	GNPA (Gross Non-performing Assets), and NNPA (Net Non-performing Assets)	It is an analytical and descriptive study that analyses data using bar graphs and figures.	The government is taking many steps to reduce the problem of NPAs, but banks should also have to be more proactive to adopt a structured NPAs policy to prevent the non-performing assets and should follow stringent measures for its recovery.
17	(Bag & Islam, 2017) "Non-performing assets a biggest challenge in banking sector- a comparative study between India and Bangladesh banking sector"	To understand the concept of non performing assets of Indian Bank as well as Bangladesh Bank.	2011-2016	ROE, ROA, and NPA (Non-performing Asset)	The regression analysis has been performed to estimate the effect of NPAs on the profitability of banks in India.	The Conclusion of study is that the growth of NPAs in Indian public sector banks as well as Bangladesh State owned banks are dramatically upward rising while private sector banks in both the countries are more conscious about the NPAs as well as total advances.
18	(Laveena & Guleria, 2016) "A Study of Non-Performing Assets of Public Sector Banks in India"	This study tries to compare performance of public sector banks with private sector banks in India.	2011-2015	Net profit and ROA (Return on Asset)	This paper has computed NNPA to know the risk level of banks.	The NPA level of the banking system is still at a high level compared to the international standards.

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
19	(Singh, 2016) "A Study of Non-Performing Assets of Commercial Banks and it's recovery in India"	It studies the status of Non-Performing Assets of Indian Scheduled Commercial Banks in India.	2000-2014	GNPA (Gross Non-performing Assets) and NNPA (Net Non-performing Assets)	The paper has used charts and graphs to represent data.	The conclusion of the study shows that the extent of NPA is comparatively very high in public sector banks as compared to the foreign banks.
20	(Durafe & Singh, 2016) "Cyclical Behavior of Public and Private Sector Banks: A Comparative Study of Non-performing Assets"	To study the cyclical behavior of public sector bank groups and private sector bank groups in India with a focus on non-performing assets.	1999-2013	NPA (Non-performing Asset), credit inclination, Loan maturity, unsecured lending, cost condition, GDP, and Inflation adjusted GDP deflator.	The study has used secondary data and the methods of Pearson correlation coefficient and multiple regression analysis for the analysis of data.	The findings indicate that when GDP increases, NPA decreases and when GDP decreases NPAs increase.
21	(Garg, 2016) "A study on management of non-performing assets in Context of Indian banking system"	To study the preventive mechanism for NPA and Compromise settlement scheme.	2010-2015	NPA (Non-performing Asset), profitability, liquidity, and credit loss.	The research has used a combined approach that involves both the features of analytical and descriptive research designs.	The conclusion of the study based on the problem of NPA has arisen due to the basic objective behind lending of public sector banks that is social welfare.
22	(Shiralashetti & Poojari, 2016) "Non-Performing Assets: A Case Study of Syndicate Bank"	To analyse the gross, net NPA, and sector wise NPA.	2011-2015	Net profit and Gross NPA	The data has been analysed by using percentage, t-test, regression and one way ANOVA.	Banks must take more care in avoiding any account becoming NPA by taking proper preventive measures in an efficient manner.
23	(Bandyopadhyay, 2016) "An Analysis of the Non-Performing Assets of some selected public sector banks in India."	To examine the overall trends of NPAs and to explore the dynamicity of NPA as the variable under study over time in selected PSBs.	1996-2012	NPA (Non-performing Asset) and strategic banking variables.	The study has used an empirical approach for analysing the movement of NPA and its impact on PSBs based on Secondary data.	The study has clearly demonstrated that NPA in PSBs are on the rise for the last three years and is expected to rise still further in next three years.
24	(Rao & Patel, 2015) "A study on non-performing assets management with reference to public sector banks, private sector banks and foreign banks in India"	To analyse and interpret various NPA related ratios for aggregates of Public Sector, Private Sector and Foreign Banks in India.	2009-2013	NPA (Non-performing Asset) and GNPA (Gross Non-performing Asset),	The paper has used Descriptive research design that includes tools like ratio analysis, application of least square method, and use of SPSS. The paper has used secondary data for analysis of NPA.	The conclusion of the study in case of ANOVA test is found, Ratio of Gross NPA to Gross Advances for Public sector, Private Sector and Foreign Banks does not have significant difference from 2009 to 2013, and it does not reject our null hypothesis.

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
25	(Joseph & Prakash, 2014) "A Study on Analysing the Trend of NPA Level in Private Sector Banks and Public Sector Banks"	To identify the trend in NPA level.	2008-2013	NPA (Non-performing Asset), Standard advances, Sub-standard advances, and doubtful advances	Ratios	The conclusion of the study is that compared to private sector banks, public sector banks are more in the NPA level.
26	(Ibrahim & Thangavelu, 2014) "A Study on the Composition of Non-Performing Assets (NPAs) of Scheduled Commercial Banks in India"	To review the trend of Gross NPAs of public, private, and foreign banks of Scheduled Commercial Banks in India.	2007-2012	NPA (Non-performing Asset) and loss assets	The paper has collected secondary data for analysis and used regression analysis for finding conclusions.	The conclusion of the study is that the RBI and the Government of India have taken innumerable steps to reduce the volume of NPAs of the Scheduled Commercial Banks.
27	(Das & Dutta, 2014) "A Study on NPA of Public Sector Banks in India"	The objective of the study is to find out whether there is any difference in the NPA occurrence between the various banks during the period of the study.	2008-2013	Net non-performing assets	For the purpose of this study, secondary data has been used and the researcher has used analysis of variance (ANOVA) one way.	The study finds out that there is no significant difference between the means of NPA of the banks at five percent level of significance.
28	Siraj K.K(2014) "A Study on Non-Performing Assets Of Public Sector Banks In India With Special Reference To State Bank Of Travancore"	To analyse the trend of the NPA in Public Sector Banks in India vis-à-vis State Bank of Travancore.	2001-2012	Variables used are total advances, capital, reserve and surplus, deposits, advances-sector wise, investments, total assets, Interest expended, operating expenses, provision and contingencies, total expenses, and profit or loss.	The study is both a descriptive and analytical study. It has used both primary and secondary data. The statistical tools used include averages, ratio, Exponential Growth Rate, Correlation-Regression, ANOVA, F-test, t-test, Sobel test, Welch statistics.	Significant differences are observed between SBI and SB of Travancore.

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
	C. Literature review on Performance					
29	(Sarkar & Sarkar, 1998) "Does Ownership Always Matter? -Evidence from the Indian Banking Industry"	To investigate the effect of ownership by examining public and private sector banks.	1994-95	Explained variables includes ROA (Net profit/ assets), OPR (Operating profit/ assets), NIM (Net interest margin/assets), OCR (Operating cost/assets), and SER (Staff expense/ asset) while explanatory variables are assets, investment in Govt securities/ assets, Noninterest income/ assets, rural branches/urban branches, and priority sector advances/ assets.	Mean and standard Deviation	It is found privatization is likely to be an effective policy for improving the performance of ailing public-sector enterprises in emerging economies like India.
30	(Armijo & Jha, 1997) "Centre-State relations in India and Brazil: Privatisation of Electricity and Banking"	This paper examined the political factors involved in privatisation of the banking and electricity sectors of two nations: India and Brazil.	The period of study relates to the early 1990s	Center-state relations and the gradual process of national economic liberalization	It is a descriptive paper and uses secondary data.	They arrived at the conclusion that the leftist ideology and beliefs of particular politicians may be a factor in shaping politicians' attitudes toward privatization, but it is unlikely to be the most crucial one.
31	(Mohan, Privatisation: Theory and Evidence, 2001) "Privatisation: Theory and Evidence"	It attempts with the help of survey theoretical and empirical work on privatization to induce a measure of caution in statements about the likely efficacy of privatization.	1961-1990	Sales assets, and number of employees	The paper has used four performance measures that include return on assets, return on sales, net income, and return on equity.	The paper concludes that privatisation is seen to work best when allied with competition or deregulation.
32	(Subramanian, 2014) "Financial Performance Analysis – A Case Study"	It makes an attempt to evaluate the financial performance of private sector banks in India and rank them based on each variable	2003-2006	Business per employee, return on assets, profit per employee, capital adequacy, credit deposit ratio, operating profit and percentage of net non-performing asset to net advance	Regression	The paper concludes that the Indian banking system has undergone a drastic change since liberalization. The new generation private sector banks have best used the technology; utilize the manpower in an effective manner

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
33	(Prasad & Ghosh, 2005) "Competition in Indian Banking"	This paper evaluated the validity of the claim that competition in the Indian banking sector has increased since the inception of the financial sector reforms in 1992.	1996-2004	Number of banks, Total assets, total deposits, and total credits	The research paper has used the PANZAR-ROSSE test for assessing competition in banking.	The empirical evidence reveals that the Indian banking system operates under competitive conditions and earns revenues as if under monopolistic competition.
34	(Murthy & Deb, 2008) "Operationalizing and Measuring Competition: Determinants of Competition in Private Banking Industry in India"	This paper provides a theoretical framework of competition that can be applied to banking also. Discussing the evolution of competition, the study proves how ordinary Structure Conduct and Performance Approach (S-C-P paradigm) is unsuitable in judging market dynamics.	1995-2002	It provides a methodology to arrive at a market form in banking industry through an analysis of all the aspects of basic conditions, structure, conduct and performance	Correlation coefficient, chi square test, and simultaneous equation.	This paper argues that sustained growth and dynamics of industry is not price led. Growth arises out of changing basic conditions and dynamics arises out of sharing the new market created by basic conditions
35	(Tabi & Joseph, 2006) "Profitability Performance in the Banking Industry of CFA Countries: _ The Case of Commercial Banks in Cameroon"	This study tested empirically the Structure - Performance (S- P) hypothesis within the context of the Cameroonian Commercial banking system.	1987-1999	Independent variable includes Index of market concentration, Management of bank's capital, Bank's loan portfolio, Total advances to total deposit, Bank size, Expense control, and period of study while dependent variables include ROA, ROC, and ROE	Three accounting measures of bank's performance were utilised: return on capital (ROC), return on assets (ROA) and return on equity (ROE).	The results indicate that the market concentration power is of paramount importance in the determination of bank profitability the positive effects of bank's size, time and savings deposits to total deposit ratio and the 1994 CFA devaluation cannot be emphasized

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
36	(Neuberger, 1998) "Direct Banking - A Demand Pull and Technology Push Innovation"	His paper reviews the literature by focusing on the links between theoretical and empirical research.	1994-1995	Market share is the independent variable while price data as dependent variable	The paper has used SCP (Structure-conduct-performance) Paradigm for conducting theoretical and empirical research.	The results are to alternative assumptions about the type of competition or structural variables affected by deregulations and technological progress. He concludes, more research needs to be done which differentiates between different product and geographic markets and accounts for banking size, portfolio structure and regulatory settings.
37	(Srinivas & Kumar, 2010) "Is Structure-Conduct-Performance a case of structure–performance? Structural Equation Modelling of Indian industries"	This paper conceptualised a theoretical model Based on Bain's work for the Structure-Conduct-Performance (S-C-P) paradigm.	1991-2005	Structure variable is HHI and entry barrier; conduct variables are advertisement as proportion of sales, research and development Expenditure, debt–equity ratio, and distribution expenditure; and performance variable include profit after tax	The paper has collected data for 124 industries and the technique of Structural Equation Modelling (SEM).	The results indicate that direct and total effects of industry structure on firms' profitability are positive and significant while the indirect effects are insignificant. It is observed that in the presence of entry barriers, industry concentration has no significant effect on firms' profitability.
38	(Mohan, 2005) "Bank Consolidation Issues and Evidence"	It states that India's public sector banks lack a compelling rationale for consolidation.	1996-2004	Assets, deposits, profits, non-interest income, and operating costs	The study is a descriptive study and has used secondary data.	The performance of PSBs, measured by the appreciation in stock values, has also been very impressive. Also, greater size is the key to further performance improvement.

S. No.	Name of the Author & Title of the study	Objective of the study	Time Period of the study	Variables of the Study	Data & Methodology	Main Findings
39	(Kaur & Kaur, 2010) "Impact of Mergers on the Cost Efficiency of Indian Commercial Banks"	It examined the cost efficiency of Indian commercial banks	1990-2008	Input variables are labor, loanable funds, and physical capital while the output variables are non-interest income, net-interest income, and advances	The paper has analysed the cost efficiency of Indian commercial banks by using a non-parametric Data Envelopment Analysis Technique.	This study suggests that over the entire study period average cost efficiency of public sector banks was found to be 73.4 and for private sector banks is 76.3 percent. The findings of this paper suggest that to some extent the merger programme has been successful in the Indian banking sector.
40	(Gupta & Murthy, 2012) "Demise of Development Financial Institutions in India: A Critical Appraisal"	It evaluated the views in favor and in against demise of DFIs in India and concluded that commercial banks are not in a position to fill the gap of long-term financing being taken care of DFIs earlier.	1999-2005	NPA, Profit after tax, Margin, and Capital adequacy ratio	The research paper is a descriptive study and has used secondary data from the bank's annual reports, journals, RBI website.	It has been concluded that Development finance institutions were capable of bearing the risk of undertaking huge projects.

4. Research Gap

These studies made important contributions to the comparative analysis of profitability of public sector banking industry with respect to private and foreign banks, but none of them provided a rigorous analysis of the impact of reforms on the profitability of public sector banking industry over time. Issues explored in these studies are related to interest and non-interest income, bank size, technology issues, and formation of non-performing assets. At times, public sector banks have faced criticism that they are inefficient, uncompetitive, and a burden on exchequer. Many of the studies have compared public sector banks with old private, new private and foreign banks. But now questions arise whether with the adaptation of liberal economic policies and market model, other objectives than profit are irrelevant. Secondly, given their own frame of reference, can they be dubbed as good or bad? These issues are to be taken in the present study to have a deep insight into the functioning of public sector banks and to examine whether there are homogeneous groups among public sector banks.

5. Conclusion

From the discussion of literature on structure, conduct and performance, it broadly follows that the PSBs are facing stiff competition after the

entry of new private and foreign banks. Their share in the market is decreasing slowly and the share of private and foreign banks is increasing. The existing literature looks at concentration ratio in terms of asset, deposit and advances. Market structure needs to be analyzed by taking more variables like deposits plus advances to show barriers to entry, absolute advertisement expenditure as a proxy of selling cost, other income to show product differentiation, and concentration ratio which tells about the nature of the market. There is a strong need to check the presence of strategic groups amongst public sector banks in India because very often it is said that there are no strategic groups. Moreover, the State Bank of India (SBI) is the undisputed leader in the market, but we need to examine whether competitive rivalry has impacted SBI adversely. Also, we must investigate if there is presence of oligopolistic behavior in the market. In case of conduct it can be said that reforms have been successful in affecting change in the conduct of the banks. But banks still need to focus on various conduct variables to compete efficiently. The other conduct variables related to staff, operational expenditure, advertisement expenditure, other fixed assets, non-interest income, term loans to total advances, investment in G-sec etc. need to be examined in case of the public sector banking industry. There are some more general impressions about the public sector banking industry; they don't compete with new private banks which entered the Indian banking industry due to the historic accident which led to implementation of Narasimham Committee report I and II. Very often it is said that these banks are homogeneous in character and that is why they do not compete amongst each other. The extant literature shows that performance and productivity of the public sector banking industry has improved but banks level data needs to be analysed to have in-depth results to see how competitive forces are being unleashed in the industry. Moreover, one needs to check the impact of new policies and variables on profitability of the public sector banking industry because their profitability and efficiency is questioned at times.

Author Bionote: *Sonia Kaushik is Associate Professor, Department of Commerce, Bharati College, University of Delhi. She has teaching experience of more than 16 years in Income Tax Law and Practice, Investing in stock market, Banking and Insurance for the last many years. Her specialisation area is finance. She completed her Ph.D from Department of Commerce, Delhi School of Economics.*

References

Murthy, K. V., & Deb, A. T. (2014, July). Measurement and Determinants of Competition in Private Banking Industry in India during 1992-2002. *IUP Journal of Applied Finance, 20*(3), 72-93.

Bajaj, S. (2008). Indian Public Sector Banking Industry: 1992-2006- An Impact of Reforms. Unpublished M.Phil Dissertation.

Gupta, R. (2013). A comparative study of public and private banking industry in India. An unpublished Ph.D. thesis.

Amel, D. F., & Rhoades, S. A. (1988, November). STRATEGIC GROUPS IN BANKING. *The Review of Economics and Statistics, 70*(4), 685-689.

Hannan, T. H. (1991). Bank commercial loan markets and the role of market structure: Evidence from surveys of commercial lending. *Journal of Banking and Finance, 15*, 133-149.

Heggestad, A. A., & Rhoades, S. A. (1976, January). An analysis of changes in bank market structure. *Atlantic Economic Journal, 4*(3).

Arasu, B. S., Sridevi, P., Nageswari, P., & Ramya, R. (2019, June). A Study on Analysis of Non-Performing Assets and its Impact on Profitability. *International Journal of Scientific research in Multidisciplinary Studies, 5*(6), 1-10.

Numer, P., & Devika, P. (2019, April). A Study on comparative Analysis of Non-Performing Assets in Selected Private Sector Banks in India. *International Journal of Scientific and Research Publications, 9*(4), 303-320.

Kambar, P. (2019, June). A study on the consolidation and merger of public sector banks in India: Issues and challenges. *International Journal of Social Science and Economic Research, 4*(6), 4326-4329.

Kapadia, M., & Madhav, V. V. (2019, July). NPAs in Indian Scheduled Commercial Banks: Origination and Impact on Economy. *International Journal of Recent technology and Engineering, 8*(2S6), 538-541.

Kumari, R., Singh, P. K., & Sharma, V. C. (2017). Impact of Non-Performing Assets (NPAs) on Financial Performance of Indian Banking Sector. *6*(1), 122-133.

Raj, M., Jain, A., Bansal, S., & Verma, T. (2018, September). Non-Performing Assets: A Comparative Study of SBI&ICICI Bank from 2014-2017. *IOSR Journal of Business and Management, 20*(9), 78-84.

Rayudu, T. S., & Rajesh, P. (2018). A Study on Non-Performing Assets in Selected Public and Private Sector Banks in India. *International Journal of Scientific Research and Private Sector Banks in India, 7*(2), 360-368.

Mishra, M. M., & Pawaskar, J. R. (2017, March). A Study of Non-Performing Assets and its Impact on Banking Sector. *Journal for Research, 3*(1), 1-5.

Alagarsamy, T. (2017, August). Performance of non-performing assets (NPA) in state bank of India. *International Journal of Advanced research in Management and Social Sciences, 6*(8), 79-84.

Mittal, R. K., & Suneja, D. (2017, July). The problem of Rising Non-performing Assets in Banking Sector in India: Comparative analysis of public and private sector banks. *International Journal of Management, IT & Engineering, 7*(7), 384-398.

Bag, S., & Islam, S. (2017, November). Non-performing assets a biggest challenge in banking sector- a comparative study between India and Bangladesh banking sector. *ICTACT Journal of Management Studies, 3*(4), 620-624.

Laveena, & Guleria, K. S. (2016, April). A Study of Non-Performing Assets of Public Sector Banks in India. *International Journal of Research in Engineering, IT and Social Sciences, 6*(4), 26-34.

Singh, V. R. (2016, April). A Study of Non-Performing Assets of Commercial Banks and it's recovery in India. *Annual Research Journal of SCMS, 4*, 110-125.

Durafe, A., & Singh, M. (2016). Cyclical Behavior of Public and Private Sector Banks: A Comparative Study of Non-performing Assets. *Journal of Business and Management Studies, 1*(1), 14-25.

Garg, A. (2016, NOvember). A study on management of non-performing assets in Context of Indian banking system. *International Journal of Engineering technologies and Management Research, 3*(11), 15-25.

Shiralashetti, A., & Poojari, L. N. (2016, April). NON-PERFORMING ASSETS: A CASE STUDY OF SYNDICATE BANK. *Pacific Business Review International, 8*(10), 87-92.

Bandyopadhyay, G. (2016). An Analysis of the Non-Performing Assets of some selected public sector banks in India. In partial fulfillment of the degree of Ph.D. .

Rao, M., & Patel, A. (2015, March). A study on Non-Performing assets management with reference to public sector banks, private sector banks and foreign banks in India. *Journal of Management and Science, 5*(1).

Joseph , A. L., & Prakash, M. (2014, July). A Study on Analyzing the Trend of NPA Level in Private Sector Banks and Public Sector Banks. *International Journal of Scientific and Research Publications, 4*(7).

Ibrahim, M. S., & Thangavelu, R. (2014, March). A Study on the Composition of Non-Performing Assets (NPAs) of Scheduled. *Journal of Finance and Bank Management , 2*(1).

Das, S., & Dutta, A. (2014, Nov). A Study on NPA of Public Sector Banks in India. *Journal of Business and Management, 16*(11).

Sarkar, J., & Sarkar, S. (1998). Does Ownership Always Matter?—Evidence from the Indian Banking Industry. *JOURNAL OF COMPARATIVE ECONOMICS, 26*, 262-281.

Armijo, L. E., & Jha, P. S. (1997). Centre-State Relations in India and Brazil: Privatisation of Electricity and Banking. In S. Kahkonen, & A. Lanyi, *Institutions, Incentives and Economic Reforms in India* (Vol. 17, pp. 127-158). Sage Publications.

Mohan, T. R. (2001, December). Privatisation: Theory and Evidence. *Economic & Political Weekly, 36*(52), 4865-4871.

Subramanian, A. (2014). Financial Performance Analysis – A Case Study. *Star Commerce, 2*(7(2)), 10-18.

Prasad, A., & Ghosh, S. (2005, July). Competition in Indian Banking. *IMF Working Paper*.

Tabi, A., & Joseph, N. (2006). Market Structure and Profitability Performance in the Banking Industry of CFA Countries: The Case of Commercial Banks in Cameroon. *Journal of Sustainable Development in Africa, 8*(2), 1-14.

Neuberger, D. (1998). Industrial Organization of Banking: A Review. *International Journal of the Economics of Business, 5*(1), 97-118.

Srinivas, C., & Kumar, S. (2010). Is Structure-Conduct-Performance a case of structure–performance?: Structural Equation Modelling of Indian industries. *International Journal of Business and Emerging Markets, 2*(1), 3-22.

Mohan, T. R. (2005, March). Bank Consolidation Issues and Evidence. *Economic and Political Weekly, 40*(12), 1151-1152+1155-1161.

Kaur, P., & Kaur, G. (2010). Impact of Mergers on the Cost Efficiency of Indian Commercial Banks. *Eurasian Journal of Business and Economics, 3*(5), 27-50.

Gupta, R., & Murthy, K. B. (2012). Demise of Development Financial Institutions in India: A Critical Appraisal. *XIII Annual International Conference on Global Turbulence: Challenges & Opportunities* (pp. 1-37). Bangkok, Thailand: Delhi School of Professional studies and Research.

arXiv:2408.06722v1 [quant-ph] 13 Aug 2024

Quantum cloning transformation unlocks the potential of W class of states in a secret sharing protocol

Rashi Jain* and Satyabrata Adhikari†
Department of Applied Mathematics,
Delhi Technological University, Delhi-110042, India

One of the most challenging problems is to share a secret because the sender does not trust the receiver completely. Thus, the sender provides one part of the information to the receiver and shares the other part of the information to a third party on whom the sender can rely. The secret can be revealed when the receiver and the third party agree to cooperate. This is the essence of the secret-sharing protocol. A lot of studies have been done on it using the three-qubit GHZ state, and only a few works have involved the W state. In this work, we introduce a quantum secret sharing protocol exploiting a three-qubit W class of state shared between three parties, Alice (Sender), Bob (Mediator), and Charlie (Receiver). In the proposed protocol, the shared state parameters and the secret are linked in such a way that it is very difficult to factor them. We will show that these parameters can be factored out easily if the receiver uses a quantum cloning machine (QCM) and thus can retrieve the secret. We find that the protocol is probabilistic and have calculated the probability of success of the protocol. Further, we establish the relation between the success probability and the efficiency of the QCM. In general, we find that the efficiency of the constructed QCM is greater than or equal to $\frac{1}{3}$, but we have shown that its efficiency can be enhanced when the parameters of the shared state are used as the parameters of the QCM. Moreover, we derived the linkage between the probability of success and the amount of entanglement in the shared W class of state. We analyzed the obtained result and found that even a less entangled W class of state can also play a vital role in the proposed secret-sharing scheme.

I. INTRODUCTION

The transfer of confidential information plays a vital role in almost all industries today. It has now become a task of utmost importance to ensure that the information is sent to the desired receiver without any eavesdropping in between, hampering the system's security. Therefore, it is important first to construct a secret key to transmit the information later. Quantum cryptography [1] provides a basis for doing so. Quantum key distribution (QKD) [2, 3] and quantum secret sharing (QSS) [4–6] are the most efficient branches of quantum cryptography, allowing two or more entities to construct or distribute a secret. These uses the most powerful tools of quantum mechanics [7, 8], such as quantum entanglement [9, 10], quantum teleportation [11, 12] and quantum cloning [13–15].

Quantum cloning plays a vital role in quantum cryptography. In this regard, Wooters and Zurek [16], in 1982, stated that "an arbitrary quantum state cannot be cloned", famously known as the no-cloning theorem. This result makes one of the major differences between classical and quantum information processing. Since perfect cloning is not possible, an approximate QCM can be constructed [16–21]. The QCM proposed by Wooters and Zurek (W-Z copying machine) is given by [16]

$$\begin{aligned} |0\rangle_a|Q\rangle_x &\longrightarrow |0\rangle_a|0\rangle_b|Q_0\rangle_x \\ |1\rangle_a|Q\rangle_x &\longrightarrow |1\rangle_a|1\rangle_b|Q_1\rangle_x \end{aligned} \tag{1}$$

where $|Q\rangle_x$ denote the QCM state vector, subscript a denotes the original input state and subscript b denotes the copy mode of the input state. The copying machine state vector satisfy the following

$$\langle Q_0|Q_1\rangle = 0,\ \langle Q_i|Q_i\rangle = 1,\ i = 0, 1 \tag{2}$$

The W-Z copying machine is state-dependent [14, 22], i.e., the quality of the copies at the output of the quantum cloner depends on the input state. However, there are quantum cloners that are state-independent [23, 24]. Going towards in this direction, Buzek and Hillery [21] analyzed the possibility of approximately copying an arbitrary quantum state by using a universal quantum cloning machine (UQCM). UQCM is a type of quantum cloning machine in which the quality of the copy is independent of the input state. The following transformation describes the optimal UQCM

$$\begin{aligned} |0\rangle_a|0\rangle_b|Q\rangle_x &\longrightarrow \sqrt{\frac{2}{3}}|00\rangle_{ab}|0\rangle_x + \sqrt{\frac{1}{3}}|+\rangle_{ab}|1\rangle_x \\ |1\rangle_a|0\rangle_b|Q\rangle_x &\longrightarrow \sqrt{\frac{2}{3}}|11\rangle_{ab}|1\rangle_x + \sqrt{\frac{1}{3}}|+\rangle_{ab}|0\rangle_x \end{aligned} \tag{3}$$

where $|+\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$.

Later, other state-dependent and state-independent cloning transformations have been studied [25–27].

Due to the property that perfect copies of an arbitrary

* rashijain_23phdam07@dtu.ac.in
† satyabrata@dtu.ac.in

state cannot be made, therefore, one cannot distribute the shares of a quantum secret to two or more parties simultaneously. Thus, it makes the no-cloning theorem a fundamental constraint in quantum secret-sharing (QSS) schemes. In a secret sharing protocol, a sender distributes the secret message among all the participants, and each participant plays an equal and important role in successfully completing the protocol. This also ensures that if there is a dishonest member in the group, he may not be able to extract the secret without the cooperation of other participants.

The idea of classical secret sharing was first proposed by Shamir and Blakely [28, 29]. The quantum version of secret sharing has been studied in [4, 5]. Later, it has been shown that a secret is shared when a sufficient number of participants collaborate together [30–32]. Hillery et al. [4] used three-qubit maximally entangled GHZ states to split the information into two or more parts, and then only after collaboration from all the parties, the secret is shared. Thereafter, many multi-party secret-sharing protocols have been proposed [33–38]. Although most of the work on QSS schemes has been done using the GHZ states, but there have been only a few with W-states [39–42].

The motivation of this work is twofold: (i) to study the effect of using the W-class of states in QSS protocol and, (ii) to examine how the amount of entanglement contained in the W-class of states affects the probability of success of the QSS protocol.

The rest of the paper is organized as follows: In section II, we propose a symmetric QCM and study how better it can create a copy at the output of an arbitrary quantum state in the input. In section III, we propose a QSS protocol for a three-party system using a W-class of state. We, further, analyzed the proposed QSS scheme and calculated the probability of success of it. Moreover, we discussed in detail the relation between the probability of success of the QSS scheme and the efficiency of the proposed QCM. In section IV, the linkage between the amount of entanglement of the shared W class of state and the probability of success of the QSS scheme is studied. We finally conclude in section V.

II. SYMMETRIC QUANTUM CLONING TRANSFORMATION

In this section, we propose a symmetric QCM, which will be used in the later section, where the secret-sharing protocol will be discussed. Let us consider an arbitrary single-qubit state $|\chi\rangle^{in}$ that we would like to clone. The input state $|\chi\rangle^{in}$ can be expressed in the form

$$|\chi\rangle^{in} = x|0\rangle + y|1\rangle, \quad |x|^2 + |y|^2 = 1 \tag{4}$$

To make a copy of the input state, we propose the quantum cloning transformation, which is of the form

$$\begin{aligned} U|0\rangle_a|0\rangle_b|Q\rangle_c &= \big(|00\rangle_{ab} + p(|01\rangle_{ab} + |10\rangle_{ab})\big)|Q_0\rangle_c \\ U|1\rangle_a|0\rangle_b|Q\rangle_c &= \big(|11\rangle_{ab} + q(|01\rangle_{ab} + |10\rangle_{ab})\big)|Q_1\rangle_c \end{aligned} \tag{5}$$

where $|0\rangle_a$, $|0\rangle_b$ and $|Q\rangle_c$ denote the input qubit mode, blank qubit mode and cloning machine state vector respectively. The cloning machine state vector at the output is denoted by $|Q_0\rangle_c$ and $|Q_1\rangle_c$ respectively. Here we note that p and q may represent the complex parameters. Let us make the following assumption

$$\langle Q_0|Q_1\rangle = 0 \tag{6}$$

The unitarity of the transformation (5) gives

$$\begin{aligned} \langle Q_0|Q_0\rangle &= \frac{1}{1+2|p|^2} \\ \langle Q_1|Q_1\rangle &= \frac{1}{1+2|q|^2} \end{aligned} \tag{7}$$

The cloning transformation (5), when acts on an arbitrary single qubit quantum state, we obtain

$$\begin{aligned} &(x|0\rangle_a + y|1\rangle_a)|0\rangle_b|Q\rangle_c \\ &\quad = x|00\rangle_{ab}|Q_0\rangle_c + xp(|01\rangle_{ab} + |10\rangle_{ab})|Q_0\rangle_c + \\ &\quad\quad y|11\rangle_{ab}|Q_1\rangle_c + yq(|01\rangle_{ab} + |10\rangle_{ab})|Q_1\rangle_c \end{aligned} \tag{8}$$

For the reduced density operator ρ_{ab} after tracing out the machine state vectors, we have

$$\begin{aligned} \rho_{ab} =& |x|^2|00\rangle_{ab}\langle 00|\langle Q_0|Q_0\rangle + 2|x|^2|p|^2|+\rangle_{ab}\langle +|\langle Q_0|Q_0\rangle + \\ & |y|^2|11\rangle_{ab}\langle 11|\langle Q_1|Q_1\rangle + 2|y|^2|q|^2|+\rangle_{ab}\langle +|\langle Q_1|Q_1\rangle \end{aligned} \tag{9}$$

where $|+\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$.

The reduced density operator describing the original mode can be obtained by taking a partial trace over the copy mode, which gives

$$\begin{aligned} \rho_a =& |x|^2|0\rangle\langle 0|\langle Q_0|Q_0\rangle + 2|x|^2|p|^2(|0\rangle\langle 0| + |1\rangle\langle 1|)\langle Q_0|Q_0\rangle + \\ & |y|^2|1\rangle\langle 1|\langle Q_1|Q_1\rangle + 2|y|^2|q|^2(|0\rangle\langle 0| + |1\rangle\langle 1|)\langle Q_1|Q_1\rangle \\ =& \big[|x|^2(1+2|p|^2)\langle Q_0|Q_0\rangle + 2|y|^2|q|^2\langle Q_1|Q_1\rangle\big]|0\rangle\langle 0| + \\ & \big[2|x|^2|p|^2\langle Q_0|Q_0\rangle + |y|^2(1+2|q|^2)\langle Q_1|Q_1\rangle\big]|1\rangle\langle 1| \end{aligned} \tag{10}$$

Similarly, tracing out the mode a from (9), we get

$$\begin{aligned} \rho_b =& |x|^2|0\rangle\langle 0|\langle Q_0|Q_0\rangle + 2|x|^2|p|^2(|1\rangle\langle 1| + |0\rangle\langle 0|)\langle Q_0|Q_0\rangle + \\ & |y|^2|1\rangle\langle 1|\langle Q_1|Q_1\rangle + 2|y|^2|q|^2(|1\rangle\langle 1| + |0\rangle\langle 0|)\langle Q_1|Q_1\rangle \\ =& \big[|x|^2(1+2|p|^2)\langle Q_0|Q_0\rangle + 2|y|^2|q|^2\langle Q_1|Q_1\rangle\big]|0\rangle\langle 0| + \\ & \big[2|x|^2|p|^2\langle Q_0|Q_0\rangle + |y|^2(1+2|q|^2)\langle Q_1|Q_1\rangle\big]|1\rangle\langle 1| \end{aligned} \tag{11}$$

Since it is now clear that the two output states in the original mode and copy mode are described by the density operators ρ_a and ρ_b are identical, so we can say that

the prescribed transformation (5) describes a symmetric QCM.

To know how good the cloning machine is, i.e., how close is the output mode from the initial mode, we study the efficiency of the prescribed QCM. The efficiency of the cloning machine can be described by the Hilbert-Schmidt (HS) distance [43, 44], which can be defined as

$$D_i = \mathrm{Tr}\big[(\rho^{in} - \rho_i)(\rho^{in} - \rho_i)^\dagger\big], \quad i = a, b \tag{12}$$

where $\rho^{in} = |\chi\rangle^{in}\langle\chi|$ denotes the input density operator and can be expressed as

$$\rho^{in} = |x|^2|0\rangle\langle 0| + xy^*|0\rangle\langle 1| + yx^*|1\rangle\langle 0| + |y|^2|1\rangle\langle 1| \tag{13}$$

Now we are in a position to calculate D_a and D_b. Since the density operators ρ_a and ρ_b are equal so D_a and D_b will be equal. Therefore, the HS distance can be calculated as

$$\begin{aligned} D_a =& \mathrm{Tr}[(\rho^{in} - \rho_a)(\rho^{in} - \rho_a)^\dagger] \\ =& 2|x|^4\left(\frac{2|q|^4}{(1+|q|^2)^2} + \frac{2|p|^4}{(1+|p|^2)^2} - 1\right) + \\ & 2|x|^2\left(1 - \frac{4|q|^4}{(1+|q|^2)^2}\right) + \frac{4|q|^4}{(1+|q|^2)^2} \end{aligned} \tag{14}$$

It can be seen from the above expression (14) that we can make the proposed quantum cloning transformation state independent if we choose the parameters $|p|^2 = |q|^2 = 1$. However, in this scenario, the HS distance D_a will be equal to unity and may be considered very large, and thus, it will make the cloning transformation inefficient. For other values of $|p|^2$ and $|q|^2$, the HS distance D_a will be a function of the input state parameter $|x|^2$, and thus, the cloning transformation will be state-dependent. This indicates the fact that for some input states, the cloning transformation will be better, and for some, it will be worse. Thus, for state-dependent cloning transformation, it would be better to consider the average HS distance to determine the efficiency of the cloning transformation. The average HS distance can be defined as

$$\overline{D_a} = \int_0^1 D_a \mathrm{d}|x|^2 \tag{15}$$

To find the efficiency of the proposed cloning transformation (5), let us calculate the average HS distance. To do this, let us substitute (14) in (15), we get

$$\overline{D_a} = \frac{4}{3}\left[\frac{|q|^4}{(1+2|q|^2)^2} + \frac{|p|^4}{(1+2|p|^2)^2}\right] + \frac{1}{3} \tag{16}$$

We are now in a position to analyze the efficiency of the cloning transformation. Since the first term of (16) is always non-negative so $\overline{D_a} \geq \frac{1}{3}$. The equality is achieved only when $|p|^2 = |q|^2 = 0$ and in this case, the proposed cloning transformation reduces to the WZ quantum cloning transformation [16]. Thus, we may say that the prescribed quantum cloning transformation may not be an efficient quantum cloning transformation but we will show in the subsequent section that it will play a vital role in retrieving the secret in the QSS protocol.

III. SECRET SHARING PROTOCOL USING A CLASS OF W STATES AND QUANTUM CLONING TRANSFORMATION

A. Proposed QSS protocol

In this section, we introduce a QSS protocol using a W-class of state shared between three distant parties, namely Alice (A), Bob (B), and Charlie (C). We will show here that the W-class of state may be a good candidate for the secret sharing protocol even though it is not helpful in quantum teleportation protocol. Moreover, we will use a symmetric QCM to unmask the secret sent by the sender. The introduced protocol may be described in a few steps, which are given in the following subsections.

1. Construction of the composite system involving the single qubit in which the secret is encoded and the shared three-qubit state

Let us consider the secret sharing protocol between three parties: Alice (A), Bob (B), and Charlie (C). We assume that the sender, Alice, wishes to send the secret to the receiver, Charlie, via the mediator, Bob. To execute the protocol, Alice encodes the secret information in a single-qubit state of the form

$$|\phi\rangle_A = a|0\rangle_A + b|1\rangle_A, \quad |a|^2 + |b|^2 = 1 \tag{17}$$

Alice's task is to send the secretly prepared state to Charlie when Bob helps Charlie in this regard. We may note here a crucial point is that Bob can help Charlie only when he gets some partial information from Alice.

To start the protocol, Alice prepared the three-qubit W-class of states $|W\rangle_{ABC}$, which is of the following form

$$|W\rangle_{ABC} = \alpha|001\rangle_{ABC} + \beta|010\rangle_{ABC} + \gamma|100\rangle_{ABC} \tag{18}$$

where the normalization condition gives $|\alpha|^2 + |\beta|^2 + |\gamma|^2 = 1$. Alice keeps the first qubit to herself and distributes the second and third qubit to Bob and Charlie, respectively. We assume that the state parameters α, β and γ are known to Alice and Charlie only. The reason for this would be clear in the later section.

Since Alice wishes to communicate the secret $|\phi\rangle_A$ to Charlie via Bob using the shared W-class of state $|W\rangle_{ABC}$, so she constructs a composite system $|\psi\rangle_{AABC}$ by attaching the secret state $|\phi\rangle_A$ to the shared state $|W\rangle_{ABC}$. The four-qubit composite system is given by

$$|\psi\rangle_{AABC} = |\phi\rangle_A \otimes |W\rangle_{ABC} \tag{19}$$

2. Bell state measurement performed by Alice

In the composite system $|\psi\rangle_{AABC}$, the first two qubits are possessed by Alice, and the third and fourth qubits belong to Bob and Charlie, respectively.
Alice then performs a two-qubit Bell state measurement on her two qubits of the composite system $|\psi\rangle_{AABC}$ and as a consequence, Bob and Charlie share a two-qubit entangled state among themselves as per Alice's Bell state measurement result. It may be easy to find out that at this stage, it is not possible either for Bob or for Charlie to retrieve Alice's secret alone. Therefore, they need to cooperate with each other to get the secret. Once they agree to cooperate with each other, Alice chooses one of the two parties (either Bob or Charlie) to share her measurement result. Let us assume that Alice chooses Bob to send the Bell state measurement outcome. Alice then sends two classical bits only to Bob to inform him about her Bell state measurement outcome. According to the Bell state measurement outcome, the two-qubit entangled state shared between Bob and Charlie is given in the table below (Table 1).

3. Single qubit measurement performed by Bob

Depending upon Alice's two-qubit measurement result as $|\phi^+\rangle, |\phi^-\rangle, |\psi^+\rangle$ and $|\psi^-\rangle$, the shared state between Bob and Charlie can be re-expressed as

$$|\psi_1\rangle_{BC} = |0\rangle_B(a\alpha|1\rangle_C + b\gamma|0\rangle_C) + a\beta|1\rangle_B|0\rangle_C \quad (20)$$
$$|\psi_2\rangle_{BC} = |0\rangle_B(a\alpha|1\rangle_C - b\gamma|0\rangle_C) + a\beta|1\rangle_B|0\rangle_C \quad (21)$$
$$|\psi_3\rangle_{BC} = |0\rangle_B(a\gamma|0\rangle_C + b\alpha|1\rangle_C) + b\beta|1\rangle_B|0\rangle_C \quad (22)$$
$$|\psi_4\rangle_{BC} = |0\rangle_B(a\gamma|0\rangle_C - b\alpha|1\rangle_C) - b\beta|1\rangle_B|0\rangle_C \quad (23)$$

Since Alice chooses Bob to share her Bell state measurement outcome so Bob now performs a single-qubit measurement on any one of these states (20-23) in the computational basis $\{|0\rangle, |1\rangle\}$. Due to the single qubit measurement, the two cases might arise.
Case-I: If Bob obtains his measurement outcome as $|1\rangle$, then the protocol can be aborted, and it has to restart again.
Case-II: If Bob obtains the measurement outcome as $|0\rangle$, then the protocol may succeed and proceed in the following way. When Bob performs the single qubit measurement and if the measurement outcome is $|0\rangle$, then any of the four single qubit states, depending on the Bell state measurement outcome, collapses at Charlie's site. Bob then informs Charlie about the possible collapsed single qubit state by sending two classical bits. At this stage also, one may find that Charlie is unable to retrieve the secret sent by Alice unless $\alpha = \beta = \gamma = \frac{1}{\sqrt{3}}$. The single qubit measurement performed by Bob and the state collapsed at Charlie's end has been summarized in Table-II.

4. Charlie retrieves the secret using a QCM and performing Bell state measurement

In the previous step, if $\alpha \neq \gamma$, then we will show that Charlie can still retrieve the secret by using a QCM.
Thus, Charlie makes use of a QCM for secret extraction from the single-qubit state that he possesses. Since the shared state parameters α and γ are known to Charlie, so he uses them to design the QCM transformation (5) in the following way

$$\begin{aligned} |0\rangle|0\rangle|Q\rangle &\longrightarrow \big(|00\rangle + \alpha(|01\rangle + |10\rangle)\big)|Q_0\rangle \\ |1\rangle|0\rangle|Q\rangle &\longrightarrow \big(|11\rangle + \gamma(|01\rangle + |10\rangle)\big)|Q_1\rangle \end{aligned} \quad (24)$$

The unitarity of the transformation (24) gives

$$\begin{aligned} \langle Q_0|Q_0\rangle &= \frac{1}{1+2|\alpha|^2} \\ \langle Q_1|Q_1\rangle &= \frac{1}{1+2|\gamma|^2} \end{aligned} \quad (25)$$

To move ahead of the protocol, let us suppose that Bob gets the measurement outcome $|0\rangle$ when he performs a single qubit measurement on the state $|\psi_1\rangle$ given in (20). As a result, Charlie receives the state $a\alpha|1\rangle + b\gamma|0\rangle$ and then applies the QCM defined in (24) on the received state. He then obtains the following transformed state

$$(a\alpha|1\rangle + b\gamma|0\rangle)|0\rangle|Q\rangle \longrightarrow \frac{1}{\sqrt{2}}|\phi^+\rangle(a\alpha|Q_1\rangle + b\gamma|Q_0\rangle) + \frac{1}{\sqrt{2}}|\phi^-\rangle(b\gamma|Q_0\rangle - a\alpha|Q_1\rangle) + \sqrt{2}\alpha\gamma|\psi^+\rangle(a|Q_1\rangle + b|Q_0\rangle) \quad (26)$$

Charlie then perform measurement in the Bell basis and perform single qubit unitary transformation i.e. pauli matrices to retrieve the secret, which was sent by Alice. In a similar fashion, if Charlie received the state either in the form of $a\alpha|1\rangle_C - b\gamma|0\rangle_C$ or $a\gamma|0\rangle_C + b\alpha|1\rangle_C$ or $a\gamma|0\rangle_C - b\alpha|1\rangle_C$ after Bob's measurement then the unitary operation that has to be applied on the cloned state is given in the Table-III.

B. Analysis of the proposed QSS protocol

In the previous section, we find some cases where we have to abort the secret-sharing protocol in the midway. Thus, there is a possibility of failure of the proposed QSS protocol. Hence, we would like to analyze the probability of success of the proposed QSS protocol and also would

Alice's Bell state measurement outcome	Two-qubit entangled state between Bob and Charlie	Alice communicates with Bob by sending two C-bits
$\|\phi^+\rangle_{AA}$	$a\alpha\|01\rangle_{BC} + a\beta\|10\rangle_{BC} + b\gamma\|00\rangle_{BC}$	0_A0_A
$\|\phi^-\rangle_{AA}$	$a\alpha\|01\rangle_{BC} + a\beta\|10\rangle_{BC} - b\gamma\|00\rangle_{BC}$	1_A1_A
$\|\psi^+\rangle_{AA}$	$a\gamma\|00\rangle_{BC} + b\alpha\|01\rangle_{BC} + b\beta\|10\rangle_{BC}$	0_A1_A
$\|\psi^-\rangle_{AA}$	$a\gamma\|00\rangle_{BC} - b\alpha\|01\rangle_{BC} - b\beta\|10\rangle_{BC}$	1_A0_A

TABLE I. Alice performs a two-qubit Bell state measurement on the composite system (19). The measurement outcome is described in column 1. As a result of the measurement, Bob and Charlie share a two-qubit state given in column 2. Alice also sends two classical bits (C-bits) to Bob, describing her measurement outcome.

Shared state between Bob and Charlie	Bob's measurement outcome	State collapsed at Charlie's end
$\|0\rangle_B(a\alpha\|1\rangle_C + b\gamma\|0\rangle_C) + a\beta\|1\rangle_B\|0\rangle_C$	$\|0\rangle$	$a\alpha\|1\rangle_C + b\gamma\|0\rangle_C$
	$\|1\rangle$	$\|0\rangle_C$
$\|0\rangle_B(a\alpha\|1\rangle_C - b\gamma\|0\rangle_C) + a\beta\|1\rangle_B\|0\rangle_C$	$\|0\rangle$	$a\alpha\|1\rangle_C - b\gamma\|0\rangle_C$
	$\|1\rangle$	$\|0\rangle_C$
$\|0\rangle_B(a\gamma\|0\rangle_C + b\alpha\|1\rangle_C) + b\beta\|1\rangle_B\|0\rangle_C$	$\|0\rangle$	$a\gamma\|0\rangle_C + b\alpha\|1\rangle_C$
	$\|1\rangle$	$\|0\rangle_C$
$\|0\rangle_B(a\gamma\|0\rangle_C - b\alpha\|1\rangle_C) - b\beta\|1\rangle_B\|0\rangle_C$	$\|0\rangle$	$a\gamma\|0\rangle_C - b\alpha\|1\rangle_C$
	$\|1\rangle$	$\|0\rangle_C$

TABLE II. Bob performs single-qubit measurement on the shared two-qubit state obtained as a result of Alice's measurement in computational basis. The outcomes of Bob's measurement and the state collapsed at Charlie's end are given in columns 2 and 3, respectively.

like to investigate how the quality of the copying machine is related to the probability of success.

1. *Probability of success of the QSS scheme*

In the proposed QSS protocol, we have studied that Bob and Charlie after agreeing to co-operate with each other can decode Alice's secret using a QCM constructed by Charlie. However, there are cases when either Bob or Charlie have to abort the protocol. Thus, it is enough to consider only the cases when the extraction of the secret is successful at Charlie's end. By considering only those case where the QSS scheme is successful, we get the probability of success, which is given by

$$\begin{aligned} P_s =& \sum_{\alpha_A} \Big(P[\text{Two-qubit measurement outcome } \alpha_A] \times \\ & P[\text{single-qubit measurement outcome } |0\rangle_B] \times \\ & P[\text{two-qubit measurement outcome } |\psi^+\rangle_C] \Big) \\ =& 4|\alpha|^2|\gamma|^2 \end{aligned} \tag{27}$$

where $\alpha_A \in \{|\phi^+\rangle_A\langle\phi^+|, |\phi^-\rangle_A\langle\phi^-|, |\psi^+\rangle_A\langle\psi^+|, |\phi^-\rangle_A\langle\psi^-|\}$

2. *Linkage between the probability of success and the quality of the QCM*

In this section, we will investigate if anyway the quality of the copy emerging out of the quantum copying machine effects the probability of success of the proposed QSS scheme. In other words, we can ask the following question: Does there exist any relationship between the probability of success denoted by P_s and the average HS distance denoted by $\overline{D_a}$?
Recalling (16) with $|p|^2 = |\alpha|^2$ and $|q|^2 = |\gamma|^2$, $\overline{D_a}$ can be re-expressed as

$$\overline{D_a} = \frac{4}{3}\left[\frac{|\gamma|^4}{(1+2|\gamma|^2)^2} + \frac{|\alpha|^4}{(1+2|\alpha|^2)^2}\right] + \frac{1}{3} \tag{28}$$

Using (27) and (28), we can re-express the average HS distance $\overline{D_a}$ in terms of the probability of success P_s and we will denote it by $\overline{D_s}$ given as

State obtained by Charlie after Bob's single-qubit measurement	State obtained by Charlie after applying QCM	The state obtained if the Bell state measurement outcome is $\|\psi^+\rangle$	Unitary operation applied on the state obtained depending upon the two classical bits send by Bob
$a\alpha\|1\rangle_C + b\gamma\|0\rangle_C$	$\frac{1}{\sqrt{2}}\|\phi^+\rangle(a\alpha\|Q_1\rangle + b\gamma\|Q_0\rangle) + \frac{1}{\sqrt{2}}\|\phi^-\rangle(b\gamma\|Q_0\rangle - a\alpha\|Q_1\rangle)$ $+\sqrt{2}\|\psi^+\rangle\alpha\gamma(a\|Q_1\rangle + b\|Q_0\rangle)$	$a\|Q_1\rangle + b\|Q_0\rangle$	σ_z
$a\alpha\|1\rangle_C - b\gamma\|0\rangle_C$	$\frac{1}{\sqrt{2}}\|\phi^+\rangle(a\alpha\|Q_1\rangle - b\gamma\|Q_0\rangle) + \frac{1}{\sqrt{2}}\|\phi^-\rangle(-b\gamma\|Q_0\rangle - a\alpha\|Q_1\rangle)$ $+\sqrt{2}\|\psi^+\rangle\alpha\gamma(a\|Q_1\rangle - b\|Q_0\rangle)$	$a\|Q_1\rangle - b\|Q_0\rangle$	$\sigma_z\sigma_x$
$a\gamma\|0\rangle_C + b\alpha\|1\rangle_C$	$\frac{1}{\sqrt{2}}\|\phi^+\rangle(a\gamma\|Q_0\rangle + b\alpha\|Q_1\rangle) + \frac{1}{\sqrt{2}}\|\phi^-\rangle(a\gamma\|Q_0\rangle - b\alpha\|Q_1\rangle)$ $+\sqrt{2}\|\psi^+\rangle\alpha\gamma(a\|Q_0\rangle + b\|Q_1\rangle)$	$a\|Q_0\rangle + b\|Q_1\rangle$	I
$a\gamma\|0\rangle_C - b\alpha\|1\rangle_C$	$\frac{1}{\sqrt{2}}\|\phi^+\rangle(a\gamma\|Q_0\rangle - b\alpha\|Q_1\rangle) + \frac{1}{\sqrt{2}}\|\phi^-\rangle(a\gamma\|Q_0\rangle + b\alpha\|Q_1\rangle)$ $+\sqrt{2}\|\psi^+\rangle\alpha\gamma(a\|Q_0\rangle - b\|Q_1\rangle)$	$a\|Q_0\rangle - b\|Q_1\rangle$	σ_x

TABLE III. Charlie applies a QCM defined in (24) on the single-qubit state he has. He performs a two-qubit Bell state measurement on the state obtained in column 2, and if his measurement outcome is $|\psi^+\rangle$, he applies a unitary transformation according to the classical bits received from Bob.

$$\overline{D_s} = \frac{4}{3}\left[\frac{\left(1-|\beta|^2+\frac{P_s}{2}\right)^2+\frac{P_s^2}{4}-\frac{P_s}{2}}{(1+2|\gamma|^2)^2(1+2|\alpha|^2)^2}\right]+\frac{1}{3} \quad (29)$$

In the previous subsection, we have noted that $\overline{D_a} \geq \frac{1}{3}$ for any cloning machine parameter $|p|^2$ and $|q|^2$. But since in the proposed QSS protocol, $|p|^2$ and $|q|^2$ have been replaced by the three-qubit sharing state parameter $|\alpha|^2$ and $|\gamma|^2$ so it would be interesting to find out the possible range of $|\alpha|^2, |\beta|^2, |\gamma|^2$ and P_s for which we can make $\overline{D_s} \leq \frac{1}{3}$. To do this, we need to check whether there exists any range of $|\beta|^2$ and P_s for which the quantity $(1-|\beta|^2+\frac{P_s}{2})^2+\frac{P_s^2}{4}-\frac{P_s}{2}$ satisfies the inequality

$$\left(1-|\beta|^2+\frac{P_s}{2}\right)^2+\frac{P_s^2}{4}-\frac{P_s}{2} \leq 0 \quad (30)$$

Since the LHS of the inequality (30) is quadratic in P_s so by solving it, we get

$$-\left(\frac{1}{2}-|\beta|^2\right)-\frac{1}{2}\sqrt{12|\beta|^2-4|\beta|^4-7} \leq P_s \leq -\left(\frac{1}{2}-|\beta|^2\right)+\frac{1}{2}\sqrt{12|\beta|^2-4|\beta|^4-7} \quad (31)$$

The above inequality (31) is valid only when

$$\frac{3-\sqrt{2}}{2} \leq |\beta|^2 \leq 1 \quad (32)$$

When we vary $|\beta|^2$ in the interval $[\frac{3-\sqrt{2}}{2}, 1]$, we find from (31) that the probability of success P_s also varies within the interval $[0, 1]$, as it depends on the value of $|\beta|^2$.
We are now in a position to find the ranges for $|\alpha|^2$ and $|\gamma|^2$ with the help of normalization condition $|\alpha|^2+|\gamma|^2+|\beta|^2 = 1$ and using the lower bound of $|\beta|^2$. Therefore, we get

$$0 \leq |\alpha|^2+|\gamma|^2 \leq \frac{1}{\sqrt{2}}-\frac{1}{2} \quad (33)$$

Thus, if (32) and (33) are satisfied simultaneously, the average HS distance between the original and the copy would be less than $\frac{1}{3}$ and hence making it more efficient than Wootters-Zurek QCM. The relationship between $\overline{D_s}$ and P_s is shown in Fig.1.

IV. RELATIONSHIP BETWEEN CONCURRENCE FILL AND THE PROBABILITY OF SUCCESS OF THE SECRET SHARING SCHEME

In this section, we would like to establish the connection between the amount of entanglement in the shared three-qubit W class of state quantified by concurrence fill and the probability of success of the proposed secret sharing scheme. We investigate whether the amount of entanglement in the shared state impacts the probability of success of the secret-sharing scheme. In the proposed secret sharing scheme, we are using the W class of state to share the secret, and since tangle is unable to quantify the amount of entanglement in the W class of state, so we have to look for another measure that may quantify the amount of entanglement in a pure three-qubit W class of state.

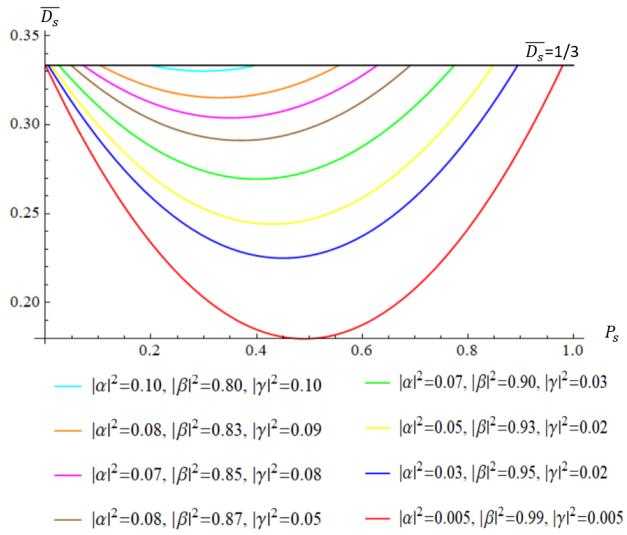


FIG. 1. The graph between $\overline{D_s}$ and P_s given by (29) is plotted for different values of $|\alpha|^2, |\beta|^2$ and $|\gamma|^2$ for the ranges given in (32) and (33). The straight line represents the value of $\overline{D_s}$ corresponding to the W-Z QCM.

A. Concurrence fill

Here, we will use an entanglement measure that may quantify the amount of entanglement in GHZ as well as the W class of state. To achieve this task, we use the *concurrence fill*, a measure of entanglement, defined as follows [45]

$$F_{ABC} = \left[\frac{16}{3}Q(Q - C^2_{A(BC)})(Q - C^2_{B(AC)})(Q - C^2_{C(AB)})\right]^{1/4} \tag{34}$$

where the quantities $Q, C_{A(BC)}, C_{B(AC)}$ and $C_{C(AB)}$ are given by

$$Q = \frac{1}{2}(C^2_{A(BC)} + C^2_{B(AC)} + C^2_{C(AB)}) \tag{35}$$

$$C_{i(jk)} = 4\ \det(\rho_i),\ \ i, j, k = A, B, C; i \neq j \neq k \tag{36}$$

The pre-factor 16/3 ensures the normalization of the concurrence fill, i.e., $F_{ABC} \in [0, 1]$. Now, we are in a position to quantify the amount of entanglement in a W class of state given in (18) using concurrence fill.
For a W class of state, we have

$$C^2_{A(BC)} = 4|\gamma|^2 - 4|\gamma|^4 \tag{37}$$
$$C^2_{B(AC)} = 4|\beta|^2 - 4|\beta|^4 \tag{38}$$
$$C^2_{C(AB)} = 4|\alpha|^2 - 4|\alpha|^4 \tag{39}$$

Using (37)-(39) and the normalization condition $|\alpha|^2 + |\beta|^2 + |\gamma|^2 = 1$, we have

$$Q = 2(1 - |\alpha|^4 - |\beta|^4 - |\gamma|^4) \tag{40}$$

The concurrence fill F_{ABC} can be re-expressed in terms of $|\alpha|^2$, $|\beta|^2$ and $|\gamma|^2$ as

$$F_{ABC} = \frac{4}{\sqrt[4]{3}}\Big((1 - |\alpha|^4 - |\beta|^4 - |\gamma|^4)(1 - |\beta|^4 - |\alpha|^4 + |\gamma|^4 - 2|\gamma|^2) (1 - |\alpha|^4 - |\gamma|^4 + |\beta|^4 - 2|\beta|^2)(1 - |\gamma|^4 - |\beta|^4 + |\alpha|^4 - 2|\alpha|^2)\Big)^{1/4} \tag{41}$$

Let us now calculate the value of F_{ABC} for some particular three-qubit state that belong to the W class of state. (i) When $|\alpha|^2 = |\beta|^2 = |\gamma|^2 = \frac{1}{3}$, therefore the value of F_{ABC} for W state is given as

$$F_{ABC} = 0.88889 \tag{42}$$

(ii) Let us consider another W class of state, which is defined as [46]

$$|W_n\rangle = \frac{1}{\sqrt{2+2n}}(|100\rangle + \sqrt{n}e^{i\gamma}|010\rangle + \sqrt{n+1}e^{i\delta}|001\rangle) \tag{43}$$

where $n = 1, 2, 3, \ldots$. γ and δ are denoting phases. Therefore, F_{ABC} for the above-defined W class of state is given as

$$F_{ABC} = 2\left(\frac{n^2(n^2+3n+1)}{3(1+n)^6}\right)^{1/4} \tag{44}$$

It can be observed from Fig. 2 that when the value of n increases, the amount of entanglement in the state $|W_n\rangle$ decreases.

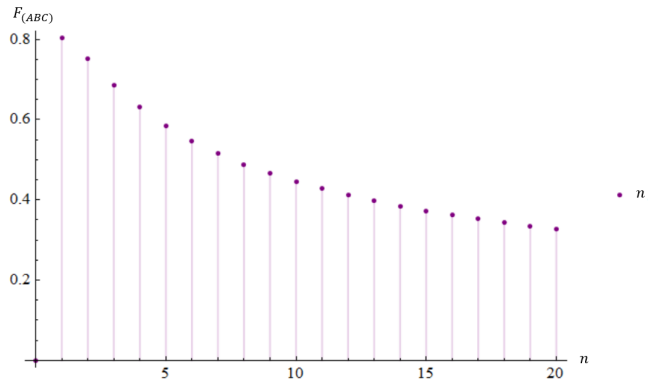


FIG. 2. Graph between F_{ABC} and n is drawn. As the value of n increases, the value of F_{ABC} decreases.

B. Relationship between the concurrence fill and the probability of success of the secret sharing scheme

In this section, we show how the probability of success of the proposed secret-sharing scheme varies with the amount of entanglement contained in the three-qubit W class of state shared between the three parties, Alice, Bob, and Charlie.

To start with, let us first recall P_s, the probability of success of the secret sharing scheme, which is given in (27). Now, we are in a position to express the values of $C^2_{A(BC)}$, $C^2_{B(AC)}$ and $C^2_{C(AB)}$ in terms of P_s and shared state parameters $|\alpha|^2, |\beta|^2$ and $|\gamma|^2$, which are given by

$$C^2_{A(BC)} = P_s\left(\frac{1-|\gamma|^2}{|\alpha|^2}\right) \tag{45}$$

$$C^2_{B(AC)} = 4|\beta|^2(1-|\beta|^2) \tag{46}$$

$$C^2_{C(AB)} = P_s\left(\frac{1-|\alpha|^2}{|\gamma|^2}\right) \tag{47}$$

Using (45)-(47), the expression for Q given in (35) reduces to

$$Q = P_s + 4|\beta|^2 - 4|\beta|^4 \tag{48}$$

Therefore, the concurrence fill F_{ABC} can then be re-expressed in terms of P_s and $|\beta|^2$ as

$$F_{ABC} = \left(\frac{64}{3}P_s^2|\beta|^4(P_s + 4|\beta|^2 - 4|\beta|^4)\right)^{1/4} \tag{49}$$

Since our interest lies in studying the effect of entanglement quantified by F_{ABC} on the success probability P_s, we can re-write (49) as a cubic equation in P_s, which is given by

$$P_s^3 + 4|\beta|^2(1-|\beta|^2)P_s^2 - \frac{3F^4_{ABC}}{64|\beta|^4} = 0 \tag{50}$$

Solving (50) using Cardon's method (given in Appendix-A), we get the solution as

$$P_s = \frac{3F^4_{ABC}}{64|\beta|^4} - \frac{128}{27}|\beta|^6(1-|\beta|^2)^3 \tag{51}$$

with the following restriction,

$$F_{ABC} \geq \left(\frac{16384}{81}|\beta|^{10}(1-|\beta|^2)^3\right)^{1/4} \tag{52}$$

Since we know that the success probability lies between [0,1] so using (51) and (52), we obtain the lower and upper bound of F_{ABC} as

$$\left(\frac{16384}{81}|\beta|^{10}(1-|\beta|^2)^3\right)^{\frac{1}{4}} \leq F_{ABC} \leq \left(\frac{8192}{81}|\beta|^{10}(1-|\beta|^2)^3 + \frac{64}{3}|\beta|^4\right)^{\frac{1}{4}} \tag{53}$$

Since the upper bounds of the concurrence fill F_{ABC} for the W class of state has been achieved for the state $|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$ which is given by 0.88889, so the upper bound of F_{ABC} given in (53) must be less than or equal to 0.88889. This gives the range of $|\beta|^2$ as $(0, 0.17)$. For the given value of $|\beta|^2$ lying in the interval $(0, 0.17)$, we have drawn $P_s - F_{ABC}$ graph, which is given in Fig.3. The following facts can be observed from Fig.3, which is given below.

Observation-1: For the value of $|\beta|^2 = 0.017$, we observe that P_s increases slowly but approaches to 1 for a very small range of F_{ABC}.

Observation-2: For $|\beta|^2 = 0.17$, we observe that P_s increases rapidly from 0 to 1 for a greater range of F_{ABC} as compared to the first case.

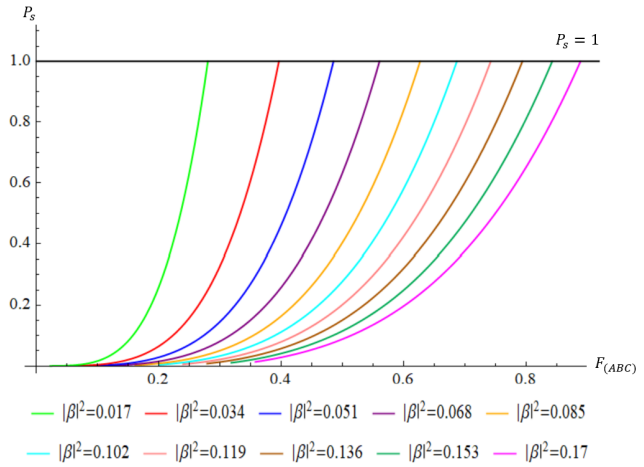


FIG. 3. The $P_s - F_{ABC}$ graph is given with respect to the value of $|\beta|^2$ lying in the interval $(0, 0.17)$ and F_{ABC} ranging according to the inequality (53) and less than equal to 0.88889.

V. CONCLUSION

To summarize, we have proposed a new QSS scheme between three parties, namely Alice, Bob, and Charlie, with the help of the W class of state and symmetric quantum cloning machine (QCM). In the proposed scheme, Alice wishes to send the secret information to Charlie via Bob, who has partial information of the secret, i.e., to retrieve Alice's secret, Bob has to cooperate with Charlie. We have assumed that the shared state parameters are known only to Alice and Charlie in the proposed QSS scheme. On the other hand, if the shared state parameters are known to all three parties and in case if Charlie's qubit goes to Bob's place, then in this situation, Bob (if he is dishonest) can take advantage of it. He (Bob) may be able to construct a QCM and then perform all the steps of the proposed secret-sharing scheme to find out Alice's secret without revealing the fact to Alice that Charlie is not present in the scheme. Therefore, it is necessary from the security point of view that only Alice and Charlie should know about the state parameters.

The quality of the QCM was analyzed by studying the HS distance D_a. When an unknown input state is fed into the introduced QCM, the average HS distance $\overline{D_a}$ is found to be greater than or equal to 1/3. Interestingly, if the parameters of the cloning machine tend towards 0, then the constructed QCM tends to the W-Z QCM. Further, we found that the proposed QCM may not be efficient for cloning in the sense that the average HS distance is not very small, but it plays a vital role in the introduced QSS scheme. Moreover, we observed that the probability of success of the proposed QSS scheme depends upon the parameters of the QCM. The ranges for success probability and state parameters were derived for which the QCM works better in the QSS scheme. In this regard, we have shown that there exist the values of the state parameters lying in the obtained ranges, for which $\overline{D_s} \leq 1/3$, making the constructed QCM a suitable choice for copying the input state for the proposed QSS protocol. Another interesting fact is that if we consider the shared state between the three parties of the form $|\psi\rangle = \alpha|001\rangle + \alpha|010\rangle + \beta|100\rangle$ with $2|\alpha|^2 + |\beta|^2 = 1$, then we find that the shared state $|\psi\rangle$ will be useful for secret sharing but the HS distance $\overline{D_s} \geq 1/3$. Thus, in this case, the constructed QCM is not useful in copying the input state, but it can be used in the proposed QSS scheme to retrieve the secret.

Moreover, we have studied the concurrence fill F_{ABC} that can be used as a measure of three-qubit entanglement for the shared W-class of state. We further derived the relationship between the probability of success (P_s) of the introduced QSS scheme and F_{ABC}. We found that as the value of $|\beta|^2$ increases, the value of F_{ABC} also increases from 0 to 0.88889. But in the perspective of P_s, we observed that as $|\beta|^2$ varies from 0.017 to 0.17, P_s approaches towards unity very slowly for a very small range of F_{ABC}, and then for a wider range of F_{ABC}, we found that P_s increases rapidly from 0 to 1.

[1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).
[2] A. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[3] C. H. Bennett and G. Brassard, in Proceedings of IEEE International Conference on Computers, Systems. and Signal Processing, Bangalore, India (IEEE, New York, 1984), p. 175
[4] M. Hillery, V. Buzěk, and A. Berthiaume, Phys. Rev. A **59**, 1829 (1999).
[5] R. Cleve, D. Gottesman, and H-K. Lo, Phys. Rev. Lett. **83**, 648 (1999).
[6] S. Bandyopadhyay, Phys. Rev. A **62**, 012308 (2000).
[7] L. E. Ballentine, Quantum mechanics: a modern development. World Scientific Publishing Company (2014).
[8] D. J. Griffiths, Introduction to quantum mechanics (1st ed.), New Jersey: Prentice Hall (1995).
[9] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[10] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, Adv. Opt. Photonics, **12**, 1012–1236 (2020).
[11] C.H. Bennett, G. Brassard, C. Cr´epeau, R. Jozsa, A. Peres and W.K. Wootters, Phys. Rev. Lett. **70**, (1895).
[12] F. Verstraete, and H. Verschelde, Phys. Rev. Lett. **90**, 097901 (2003).
[13] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, Rev. Mod. Phys. **77**, 1225 (2005).
[14] D. Bruß, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello and J. A. Smolin, Phys. Rev. A **57**, 2368 (1998).

[15] S. Adhikari, arXiv preprint arXiv:0902.1622 (2009).
[16] W. K. Wooters, and W. H. Zurek, Nature London, **299**, 802, (1982).
[17] R. Werner, Phys. Rev. A **5**, 1827, (1998).
[18] N. Gisin, and S. Massar, Phys. Rev. Lett. 79, 2153 (1997).
[19] N. Cerf, Jour. of mod. opt. **47**, 187–209, (2000).
[20] A. Kay, arXiv preprint arXiv:1407.4951, (2014).
[21] V. Bužek, and M. Hillery, Phys. Rev. A **54**, 1844 (1996).
[22] A. Chefles, and S.M.Barnett, Phys. Rev. A **60**, 136 (1999).
[23] V. Buzek, and M. Hillery, Phys. Rev. Lett. **81**, 5003 (1998).
[24] H. Fan, K. Matsumoto, and M.Wadati, Phys. Rev. A **64**, 064301 (2001).
[25] M. Ying, Phys. Lett. A **299**, 107 (2002).
[26] J. Fiurasek, S. Iblisdir, S. Massar, and N.J. Cerf, Phys. Rev. A **65**, 040302(R) (2002).
[27] D. Bruss, M. Cinchetti, G. M. D'Ariano, and C. Macchiavello, Phys. Rev. A **62**, 12302 (2000).
[28] A. Shamir, Commun. ACM **22**, 612 (1979).
[29] G. Blakely, Proc. AFIPS **48**, 313 (1979).
[30] D. Gottesman, Phys. Rev. A **61**, 042311 (2000).
[31] A. Karlsson, M. Koashi, and N. Imoto, Phys. Rev. A **59**, 162 (1999).
[32] S. Adhikari, A. K. Pati, I. Chakrabarty, and B. S. Choudhury, Quant. Inf. Proc. **6**, 197–219 (2007).
[33] S. Adhikari, I. Chakrabarty, and P. Agrawal, Quantum Inf. Comput. **12**, 253 (2012).
[34] N. Basak, N. Das, G. Paul, K. Nandi, and N. Patel, Quant. Inf. Proc. **11**, 393 (2023).
[35] F. Musanna, and S. Kumar, Quant. Inf. Proc. **19**, 348 (2020).
[36] L. Xiao, G. Lu Long, F.G. Deng, J.W. and Pan, Phys. Rev. A, **69**, 052307 (2004).
[37] Z.J. Zhang, Y. Li, and Z.X. Man, Phys. Rev. A, **71**, 044301 (2005).
[38] F.L. Yan, and T. Gao, Phys. Rev. A, **72**, 012304 (2005).
[39] J. Joo, Y. J. Park, J. Lee, J. Jang, and I. Kim, Jour. of the Kor. Phys. Soc., **46**, 763-768 (2005).
[40] L. L. Liu, C. W. Tsai, and T. Hwang, International Journal of Theoretical Physics, **51**, 2291-2306 (2012).
[41] C. W. Tsai, C. W. Yang, and N. Y. Lee, Mod. Phys. Lett. A, **34**, 1950213 (2019).
[42] G. D. Li, W. C. Cheng, Q. L. Wang, L. Cheng, Y. Mao, and H. Y. Jia, arXiv preprint arXiv:2402.02413, (2024).
[43] C. Witte, and M. Trucks, Phys. Lett. A **257**, 14 (1999).
[44] M. Ozawa, Phys. Lett. A **268**, 158 (2000).
[45] S. Xie, and J. H. Eberly, Phys. Rev. Lett. **127**, 040403 (2021).
[46] P. Agarwal, and A. K. Pati, Phys. Rev. A **74**, 062320 (2006).
[47] H. Krishan, Theory of Equations, Atlantic Publishers & Distributors (P) Ltd. (2022).

VI. APPENDIX

A. SOLUTION OF CUBIC EQUATION $P_s^3 + P_s^2 C_{B(AC)}^2 - \frac{3F_{ABC}^4}{64|\beta|^4} = 0$

In this section, we discuss the Cardon's method used for finding the roots of the cubic equation in P_s, which is given in (50). The roots so obtained can be used to determine the relationship between the probability of success (P_s) of the QSS scheme and the amount of entanglement quantified by F_{ABC} in the shared W class of state. Using (46), we can re-express (50) as

$$P_s^3 + P_s^2 C_{B(AC)}^2 - \frac{3F_{ABC}^4}{64|\beta|^4} = 0 \tag{54}$$

Cardon's method can be used to find the roots of the general cubic equation of the form

$$a_0 x^3 + 3a_1 x^2 + 3a_2 x + a_3 = 0 \tag{55}$$

where a_i $(i = 0, 1, 2, 3)$ denotes the real numbers. Comparing (55) with (54), we get

$$\begin{aligned} a_0 &= 1, \qquad a_1 = \frac{C_{B(AC)}^2}{3} \\ a_2 &= 0, \qquad a_3 = -\frac{3F_{ABC}^4}{64|\beta|^4} \end{aligned} \tag{56}$$

The cubic equation (55) can be re-expressed in the form

$$z^3 + 3Hz + G = 0 \tag{57}$$

where $H = a_0 a_2 - a_1^2$ and $G = a_0^2 a_3 - 3a_0 a_1 a_2 + 2a_1^3$. Substituting the values of a_0, a_1, a_2 and a_3 for H and G, we obtain,

$$\begin{aligned} H &= \frac{-C_{B(AC)}^4}{9} \\ G &= \frac{2}{27} C_{B(AC)}^6 - \frac{3F_{ABC}^4}{64|\beta|^4} \end{aligned} \tag{58}$$

Let us assume $P_s = u + v$. Cubing both sides, we get

$$P_s^3 - 3uvP_s - (u^3 + v^3) = 0 \tag{59}$$

Upon comparing (57) with (59), we have $u^3 v^3 = -H^3$ and $u^3 + v^3 = -G$. Considering u^3 and v^3 as the roots of the quadratic equation, which is given by

$$t^2 - (u^3 + v^3)t + u^3 v^3 = 0 \tag{60}$$

The discriminant D of the quadratic equation (60) is given by

$$D = \frac{9}{4096} \frac{F_{ABC}^8}{|\beta|^8} - \frac{12}{1728} \frac{F_{ABC}^4}{|\beta|^4} C_{B(AC)}^6 \tag{61}$$

Since the roots of the quadratic equation (60) are real numbers lying within the interval $[0, 1]$, so we must have $D \geq 0$. This gives

$$F_{ABC} \geq \frac{16384}{81} |\beta|^{10} (1 - |\beta|^2)^3 \tag{62}$$

Therefore, the roots of the quadratic equation (60) are given by

$$t = \frac{\frac{3F_{ABC}^4}{64|\beta|^4} - \frac{2}{27}C_{B(AC)}^6 \pm \sqrt{\frac{9}{4096}\frac{F_{ABC}^8}{|\beta|^8} - \frac{12}{1728}\frac{F_{ABC}^4}{|\beta|^4}C_{B(AC)}^6}}{2} \tag{63}$$

The two roots of (60) are u^3 and v^3, which can be given as

$$u^3 = \frac{\frac{3F_{ABC}^4}{64|\beta|^4} - \frac{2}{27}C_{B(AC)}^6 + \sqrt{\frac{9}{4096}\frac{F_{ABC}^8}{|\beta|^8} - \frac{12}{1728}\frac{F_{ABC}^4}{|\beta|^4}C_{B(AC)}^6}}{2}$$
$$v^3 = \frac{\frac{3F_{ABC}^4}{64|\beta|^4} - \frac{2}{27}C_{B(AC)}^6 - \sqrt{\frac{9}{4096}\frac{F_{ABC}^8}{|\beta|^8} - \frac{12}{1728}\frac{F_{ABC}^4}{|\beta|^4}C_{B(AC)}^6}}{2} \tag{64}$$

The three cube roots of u^3 are $u, u\omega, u\omega^2$ and those of v^3 are $v, v\omega, v\omega^2$, where $\omega = (-1 \pm i\sqrt{3})/2$ are the cube root of unity [47]. Therefore, we get

$$u = \frac{\frac{3F_{ABC}^4}{64|\beta|^4} - \frac{2}{27}C_{B(AC)}^6 + \sqrt{\frac{9}{4096}\frac{F_{ABC}^8}{|\beta|^8} - \frac{12}{1728}\frac{F_{ABC}^4}{|\beta|^4}C_{B(AC)}^6}}{2}$$
$$v = \frac{\frac{3F_{ABC}^4}{64|\beta|^4} - \frac{2}{27}C_{B(AC)}^6 - \sqrt{\frac{9}{4096}\frac{F_{ABC}^8}{|\beta|^8} - \frac{12}{1728}\frac{F_{ABC}^4}{|\beta|^4}C_{B(AC)}^6}}{2} \tag{65}$$

Finally, substituting the values of u, v from (65) and $C_{B(AC)}$ from (46) in $P_s = u + v$, we obtain

$$P_s = \frac{3}{64}\frac{F_{ABC}^4}{|\beta|^4} - \frac{128}{27}|\beta|^6(1-|\beta|^2)^3 \tag{66}$$

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Digital Object Identifier

RSHealth: A Ring Signature Scheme for Identity Anonymization and Transaction Privacy in Blockchain based E-Healthcare Systems

PUNAM PRABHA[1], KAKALI CHATTERJEE[2]
[1] Department of Computer Science & Engineering, National Institute of Technology, Patna-800005, Bihar, India (e-mail: punamp.phd18.cs@nitp.ac.in)
[2]Department of Computer Science & Engineering, National Institute of Technology, Patna-800005, Bihar, India (e-mail: kakali@nitp.ac.in)

Corresponding author: Punam Prabha (e-mail: punamp.phd18.cs@nitp.ac.in).

ABSTRACT

The current healthcare infrastructure faces significant challenges, including security vulnerabilities, privacy breaches, data inconsistencies, and overly accessible health records. These issues highlight the urgent need for comprehensive reforms in information management and patient data protection. To address the current issues, the present work proposes RSHealth framework, an e-healthcare ring signature scheme. It is a promising solution for ensuring anti-tampering of transaction fie and identity anonymization of sender healthcare stakeholder during transmission from main healthcare community to a branch healthcare community, maintaining compliance with data privacy regulation. The RSHealth scheme includes five functions: **TFSetup()**, **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()**. These functions respectively initialize system parameters, generate keys, create ring signatures, verify authenticity, and reveal signer identities by authorized receiver healthcare stakeholders. Extensive experimentation is performed for aforementioned framework to judge the latency and throughput; varying sender healthcare stakeholder and sizes of transaction. The performance analysis shows that increasing sender healthcare stakeholder and transaction sizes leads to increased throughput and decreased latency.

INDEX TERMS

Blockchain, E-healthcare system, Privacy, Ring Signature, Security

NOMENCLATURE

EHS	e-healthcare system
SL	system layer
IL	inter-network layer
BL	blockchain layer
CL	cloud layer
MHC	main healthcare community
BHC	branch healthcare community
SHS	sender healthcare stakeholder
RHS	receiver healthcare stakeholder
SSSL1	subsystem sublayer 1
SSSL2	subsystem sublayer 2
DPL	distributed patient ledger
m_f	medical facility
SHS_id	SHS identity
$xray_r$	x-ray report
$blood_r$	blood report
$urine_r$	urine report
sym	symptom
sym'	similar symptom
$antibody$	antibody test
$antigen$	antigen test
Tl_i	transaction list
p_trans_f	pseudo transaction file
$trans_file$	transaction file
V_{1i}	first verification parameter
V_{2i}	second verification parameter
$H_0(i)$	first hash function
$H_1(i)$	second hash function
$id*$	pseudo identity
$test_i$	list of test
$disease_i$	list of disease
nat	nucleic acid test
SHS_file	sender healthcare stakeholder file
Q	queue
$aids$	AIDS disease

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

sys_p	system parameter
sec_p	security parameter
ct_scan	CT scan
$S_{RSHealth,d}$	success for solving disease
id	original identity of SHS
$SHS(l)$	length of SHS's identity
IL_A	attacker on the inter-network layer
SL_C	challenger on the system layer

I. INTRODUCTION

All India Institute of Medical Sciences (AIIMS) stands as a cornerstone of India's healthcare system, storing the medical histories of many prominent individuals. The cyber-attack on November 23, 2022, highlighted the vulnerability of medical institutions to data breaches, emphasizing the need for robust security and privacy measures in remote patient treatment [1]. Key challenges include ensuring identity anonymity, anti-tampering, and authentication in accordance with organizational standards [2], [3], [4].

Due to inherent characteristics of decentralization, tamper resistance, anonymity, and public verifiability, blockchain has emerged as the foundational technology underpinning digital cryptocurrencies like Bitcoin and Ethereum [5], [6]. The exceptional technical attributes of blockchain extend its application beyond digital cryptocurrencies, finding widespread use in various other fields [7], [8]. In blockchain systems, the lack of centralized data processing and maintenance promotes transparency and quick consensus among nodes. Blockchain's capabilities extend to smart contracts, decentralized finance, supply chain transparency, secure identity verification, transparent voting systems, medical records management, intellectual property protection, peer-to-peer energy trading, virtual asset ownership, and platform interoperability [9].[10], [3], [11].

By providing data integrity, security, and transparency, blockchain's ledger management capabilities enhance data privacy frameworks [12]. Decentralized data storage reduces risks, while its immutable ledger and powerful cryptography secure sensitive information [13]. Blockchain helps organizations enhance stakeholder confidence and comply with regulations, improving their data privacy policies [14], [15]. However, this transparency can also pose challenges related to data privacy, potentially leading to information leakage [16]. Numerous blockchain privacy protection schemes have been proposed to enhance the anonymity of blockchain technology and safeguard the privacy of user identities and transaction data [8], [17].

Rivest et al [18] introduced the concept of signature with anonymity through the innovative ring signature-based e-healthcare system (EHS). This system allows a user to sign messages while remaining indistinguishable from other users within a specified group, referred as a ring [19]. The ring signature is a unique type of group signature [8].

An organization-level privacy protection mechanism for blockchain based transaction systems aims to facilitate the transformation of transactions between individual users to organizations. This transition not only enhances privacy but also establishes a balance between privacy protection and security supervision [15]. It urges blockchain based ring signature scheme which allow for anonymous yet accountable transactions within predefined members, thereby enhancing privacy while still allowing for oversight [8].

Implementing simplified cryptography and straightforward signature schemes in blockchain-based EHS can greatly enhance security and data privacy. Elgamal cryptography and its digital signature schemes ensure secure communication between parties [20], [21]. Efficiently addressing security concerns while preserving data privacy is crucial for managing medical data, especially in environments vulnerable to breaches or unauthorized access. Blockchain technology reduces network communication latency and provides rapid response times, making it effective for this purpose [8]. A blockchain-based EHS has been developed with four functional layers: the system layer (SL), inter-network layer (IL), blockchain layer (BL), and cloud layer (CL), arranged to facilitate medical treatment [22]. In medical emergencies or when facilities are lacking, the transaction file of sender healthcare stakeholder (SHS) can be transferred from the main healthcare community (MHC) to the branch healthcare community (BHC). Any SHS can join MHC after thorough verification by the BL. However, the current EHS lacks comprehensive facilities to address multiple health complexities simultaneously. Integrating Elgamal cryptography into a ring signature scheme within a blockchain-based EHS can enforce privacy measures while maintaining high security levels. This approach enables a more robust and flexible framework for secure, decentralized transactions.

A. OBJECTIVE

The primary objective of the proposed EHS is to securely facilitate the transfer of transaction files from SHS to a receiver healthcare stakeholder (RHS) through their registered organizations, while maintaining data privacy and security. For instance, if an SHS diagnosed with AIDS is registered with MHC and requires services beyond what MHC can offer, such as a CT scan, their transaction file must be transferred to BHC. This data transfer, particularly between SL and BL, risks disclosing the SHS's identity and transaction file privacy.

This work aims to achieve two key objectives within the EHS framework:

Implement Anti-Tampering: Ensure the integrity of medical data during transfer. Establish Identity Anonymity: Protect the SHS's identity from being disclosed during the data transfer process.

B. RESEARCH CONTRIBUTIONS

Ensuring the security of confidential personal information is a paramount concern during the transmission of transaction files from SHS to RHS. In this context, SHS aims to maintain identity anonymity during this transmission of transaction files. RSHealth is employed within the blockchain-based

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

EHS to oversee security and mitigate malicious activities at IL. The main research contributions of this paper are outlined as follows:

- Implementing a novel anti-tampering measure within blockchain based EHS, enhancing protection against malicious activities and unauthorized access.
- Implementing a novel data privacy protection framework, ensuring identity anonymization of SHS's sensitive information during file transmission.
- Analysing the performance of the implemented framework to demonstrate increased throughput and decreased latency as the number of SHS or transaction size increases.

C. ORGANIZATION OF THE PAPER

In Section II, an overview of background studies, encompassing layered architecture of EHS, and related works are discussed. Section III elaborates on the proposed RSHealth framework and its functions in blockchain based EHS for identity anonymization and anti-tampering of SHS's sensitive information. The security analysis, experimental analysis, and performance analysis of the proposed framework is evaluated in Section IV. Finally, Section V summarizes key findings and discusses implications for future research.

II. BACKGROUND

As technology continues to advance, the healthcare sector has evolved with the integration of cutting-edge technologies such as internet of thing (IoT), internet of medical thing (IoMT), blockchain technology. These systems enable simultaneous monitoring and control of various aspects of a patient's physiology, replacing standalone conventional devices. These modern medical device systems represent a distinctive category of EHS due to their embedded software, networking capabilities, and the complex physical dynamics.

A. E- HEALTHCARE SYSTEM

The proposed RSHealth scheme applied in EHS adopts a blockchain-empowered cloud architecture to maintain anonymization using ring signature throughout the life cycle of the transaction file. To achieve this objective, various members of the ring are defined in EHS, as depicted in the Figure 1.

- **System layer (SL):** In this layer, various interested SHS join a queue and register with the designated organization. This layer plays a crucial role from the initialization of transaction registration to the closure of transactions. SL is further divided into two sublayers:
  - **Subsystem sublayer 1 (SSSL1):** In this sublayer, SHS represents the patients, while RHS acts as the registered guardian of the patient. SHS is authorized with MHC, while the RHS is authorized with BHC. Transactions are sent by SHS via MHC, and received by RHS via BHC. RHS receives the file only upon successful completion of the life cycle of the transaction file movement process.
  - **Subsystem sublayer 2 (SSSL2):** MHC and BHC function as organizations where hospitals, nurses, lab technicians, and doctors generates an organization key for SHS and RHS respectively. MHC and BHC grants affiliation only upon receiving a request from SHS and RHS for the issuance of an organization certificate.
- **Inter-network layer (IL):** The transfer of transactions from SL to BL and vice versa is accomplished through IL. IL is responsible for creating duplicate files of SHS and forwarding them to BL for verification purposes.
- **Blockchain layer (BL):** In this layer, the blocks are interconnected and maintain a Distributed Patient Ledger (DPL) for the facilities available at BHC. BL verifies these facilities with the needed medical facilities of SHS, and send verified transactions relay signal to BHC. Additionally, BL makes use of smart contracts [which are specified in the appendix] to guarantee that the actions of SHS data processing are in accordance with the legal standards. BL also makes use of the cryptographic methods in order to guarantee the safety and integrity of data, which addresses important concerns in the healthcare industry. Through the use of BL, SHS is able to maintain ownership over their medical records and securely share them with healthcare professionals, thereby improving both their privacy and their data integrity.
- **Cloud layer (CL):** The verified transactions relay signal are stored in this layer. External transaction records, other than the original transaction file of SHS, are stored on a cloud storage server.

B. RELATED WORK

The domain of EHS emerges as a particularly anticipated area for advancement, especially in light of the COVID-19 pandemic and the subsequent post-pandemic landscape [3]. In this section, the overarching security and privacy of EHS, IoMT, and blockchain technology across diverse domains are focused in depth.

With the adoption of cloud-based solutions for the collection and storage of sensitive personal health information, there arises a necessity to share this data among various stakeholders within the network. Consequently, healthcare systems become susceptible to threats pertaining to confidentiality, integrity, and privacy. In this context, secure data access through such systems emerges as a significant challenge [23].

A blockchain-based privacy-preserving solution had been developed in order to solve the security issues present in cloud-assisted healthcare systems [12]. The work [10] is focused on deterministic smart contracts and their vulnerability to different attacks. Moreover, a trust-based consensus and optimization method is suggested in [9] to im-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

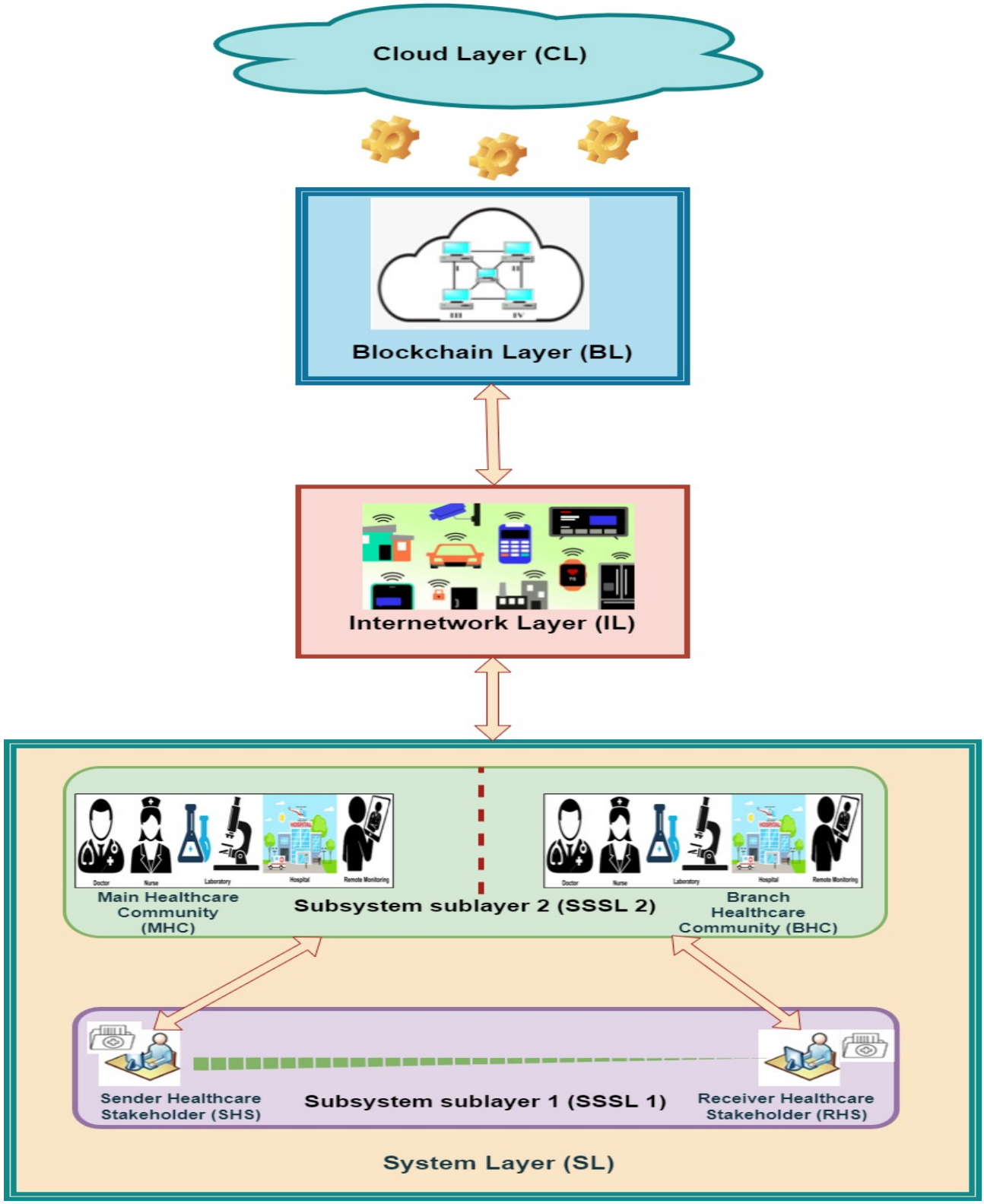


FIGURE 1: Layered architecture of e-healthcare system.

prove blockchain efficiency in IoT systems. Similarly, a blockchain-based framework is presented for protecting IoT data in smart city applications [24].

In [6], a blockchain-based system with improved encryption is described for safe electronic healthcare record storage and transfer in mobile cloud settings. A blockchain-enabled

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

access control and key management protocol is presented in [2] to protect IoMT-based EHS. In [25], improvement in the industrial IoT network security is suggested by combining blockchain technology with advanced data processing and encryption methods. In IoMT healthcare systems, Fortified-chain [13] gains efficiency, security, and privacy; but, it does not address blockchain energy consumption, system scalability for large-scale deployment, or interoperability problems with current healthcare infrastructure. Fortified-Chain 2.0 [14] addresses data privacy, security, and latency; However, blockchain energy consumption, and scalability issues still not been addressed.

The integration of artificial intelligence into the healthcare system aims to tackle ethical, legal, and technological challenges to enhance patient care [11]. The medical blockchain ensures patient privacy and data accuracy; however, it struggles with scalability, high computing overhead, and compatibility with current healthcare systems [16]. Solutions proposed in previous studies have not adequately addressed issues such as system compatibility, latency, and processing costs [5]. Additionally, privacy leaks and security vulnerabilities in blockchain technology pose significant barriers to its adoption in EHS [7]. Addressing these concerns is essential to improve the anonymity of blockchain transactions and protect sensitive identity and transaction data in EHS [26].

Numerous studies, as shown in Table 1, focus on improving security and privacy measures. In this context, the identity anonymity and anti-tampering features of ring signatures enhance privacy [17], [27], making them especially useful for electronic payments and auctions [28].

From the aforementioned study, the following research challenges have been identified:

- Despite the numerous advantages of blockchain-based EHS (such as: traceability, decentralization, security and privacy), a significant challenge remains in ensuring sufficient secrecy for SHS's data transmitted over inside and outside system layer.
- The inherent anonymity of blockchain provides a certain level of privacy protection, but it can also bypass the oversight of relevant organizations.
- A blockchain-based EHS necessitates the implementation of a RSHealth to establish a robust data privacy framework.
- In the context of blockchain-based ring signatures, the implementation of anti-tampering measures, and identity anonymization should be designed with consideration for an organization-friendly environment.

III. PROPOSED RSHealth: RING SIGNATURE SCHEME FOR BLOCKCHAIN BASED E-HEALTHCARE SYSTEM

RSHealth is incorporated into EHS for transaction privacy. At the initial stage, the RSHealth is utilized within SL to create blocks using transaction file of SHS, allows set of multiple parties namely, (SHS, MHC) and (BHC, RHS) to collectively sign a transaction without revealing the identity of SHS. Additionally, the ElGamal public key cryptography [20] and ElGamal digital signature scheme [21] are employed between SSSL1 and SSSL2 to generate organization keys, ensuring the security of the original credentials of SHS. After that, these facilitate validation in BL without disclosing which specific participant in SL produced the signature, thus preserving privacy. At the end, it enables authorized RHS to open the transaction file. These concepts are essential for formulating functions of RSHealth : **TFSetup()**, **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()**.

A. WORKING OF THE PROPOSED FRAMEWORK

In the proposed framework, RSHealth is designed with a focus on organization-friendly features to offer robust security supervision and privacy protection capabilities. When SHS needs to send a file via MHC, the broadcasted file only displays the record of SHS generated by MHC, rather than SHS's identity. Simultaneously, BL validates the record of SHS generated by MHC using a proposed RSHealth. After successful validation, a transaction relay signal is sent to SL via IL. In SL, BHC transfers the successful transaction relay information to the registered RHS. Meanwhile, additional information is stored on CL for potential future treatment purposes.

The operational functionalities of the proposed EHS as seen in Figure 2, are outlined as follows:

1) EHS is starting its activities from a point denoted as SSSL1.
2) One of the Interested SHS from a queue has sent a request to MHC for the issuance of an organization certificate using **TFSetup()**.
3) MHC applies **TFKeygen()** to issue a certificate of organization for SHS.
4) SHS anonymizes its identity using **TFRing()** and conducts a transaction with MHC.
5) MHC conceals the credentials of SHS and forwards them to BL via IL.
6) BL validates the transaction with assistance from DPL using **TFVerify()** and sends transaction relay information to BHC via IL.
7) BHC provides the required medical facilities and sends transaction relay information to RHS.
8) RHS obtains the transaction file from BHC using **TFOpen()**. The successful transaction relay information is stored on CL using Algorithm 1.

The proposed RSHealth is designed with certain assumptions addressed in the next subsection. Moreover, the security model of the proposed framework and RSHealth based transaction file privacy protection are described in-detail in the subsequent subsections.

B. ASSUMPTIONS

Definition 1 (If the length of SHS identity is large enough, then it is considered OK otherwise ignored.) Define a func-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

TABLE 1: Literature survey summary

Author's work	Year	Think about security	Think about privacy	Aim of the work	Limitation
Ma et al.[29]	[2017]	no	no	A detailed overview of the system architecture, data collection methods, analytics algorithms, and potential applications in EHS.	Lack of in-depth analysis.
Zhang et al. [30]	[2018]	no	no	The integration of various intelligent devices and systems within hospital settings to create interconnected networks that facilitate data exchange, analysis, and automation.	Risk of cyber security vulnerabilities.
Deepak et al. [31]	[2019]	yes	yes	Anonymity-based user authentication protocols are favored for addressing privacy preservation concerns within the IoMT ecosystem.	The solution involves a complex mechanism.
Alsubaei et al. [32]	[2019]	yes	no	A web-based IoMT security assessment framework is developed using ontological scenario-based approach, aiming to recommend security features tailored to IoMT environments.	Complexity involved in handling diverse stakeholders efficiently.
Zheng et al. [15]	[2020]	yes	yes	An organization-level privacy protection mechanism for blockchain-based transaction systems, facilitating the transition from individual user transactions to organizational transactions. This mechanism aims to strike a balance between privacy protection and security supervision	Not specifically designed for use by healthcare professionals.
Li et al. [8]	[2020]	yes	yes	The complete anonymity provided by ring signatures ensures both the security of data and user identity privacy in blockchain applications.	Not specifically tailored for use by healthcare professionals.
Hara et al. [19]	[2021]	yes	yes	A generic construction of ring signature with unconditional anonymity in the plain model.	Not specifically designed for use by healthcare professionals
Egala et al.[13]	[2021]	yes	yes	Develop a blockchain-based decentralized architecture for IoMT healthcare to enhance EHR security, privacy, and efficiency.	Fails to address security supervision facilities.
Wu et al. [33]	[2021]	yes	yes	An edge computing paradigm is employed to implement a policy and role-based access control scheme based on local differential privacy.	Lacks adequate security supervision measures.
Zheng et al. [5]	[2022]	yes	yes	A blockchain-based scheme for secure and privacy-preserving patient health information sharing to improve diagnosis accuracy in EHS.	Does not address interoperability, latency, or blockchain computational overhead issues.
Shreya et al. [23]	[2022]	yes	no	A lightweight authentication technique is used, incorporating a privacy-preserving schema through fully homomorphic encryption.	The anonymity of healthcare professionals' information is not adequately addressed.
zhang et al.[12]	[2022]	yes	yes	To tackle security issues in cloud-based EHS using blockchain-based privacy-preserving solution.	Vulnerable to breaches; doctors might collude to alter/sell patient data.
Padma et al.[10]	[2022]	yes	no	Emphasizes deterministic smart contracts and attacks.	Lack anonymization of patients' health details.
Lee et al.[16]	[2022]	yes	yes	Blockchain and smart contract-based electronic medical record sharing to secure data accuracy and patient privacy.	Lacks security oversight tools that are easy for organizations to use.
Kumar et al. [34]	[2023]	yes	yes	A reputation-based blockchain framework is suggested as a means to uphold patient data security while also safeguarding privacy.	Lacks organization-friendly security supervision mechanisms.
Wazid et al.[2]	[2023]	yes	no	Blockchain-based access control and key management for IoMT-based EHS	It ignores scalability, interoperability, and latency with healthcare infrastructure.
Egala et al.[14]	[2023]	yes	yes	Fortified-Chain 2.0, a decentralized blockchain-based EHS, improves data privacy, security, and latency.	Does not have any security supervision measures that are organization-friendly.
Mohamed et al.[35]	[2024]	yes	no	Developing a comprehensive fuzzy decision-making framework to determine the optimal blockchain provider within the IoMT domain.	Lack supervision mechanisms.
Padma et al.[24]	[2024]	yes	yes	Secure IoT data in smart city applications with blockchain.	Not relevant to the healthcare industry.
Verma et al.[6]	[2024]	yes	yes	Secure healthcare data storage and transfer in mobile cloud environments using blockchain and enhanced encryption.	Underestimation of scalability and interoperability under the supervision of an organization.
Padma et al.[9]	[2024]	yes	yes	Addresses network overhead and energy consumption in blockchain for IoT.	Not address latency and throughput challenges.
Bobde et al.[25]	[2024]	yes	no	Blockchain integration with strong encryption and data processing improves Industrial IoT network security.	Does not address encryption and blockchain computational overhead.
Williamson et al.[11]	[2024]	yes	no	Critically assess healthcare AI integration's ethical, legal, and technological challenges to improve patient care.	Does not address implementation difficulties like healthcare resource constraints.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

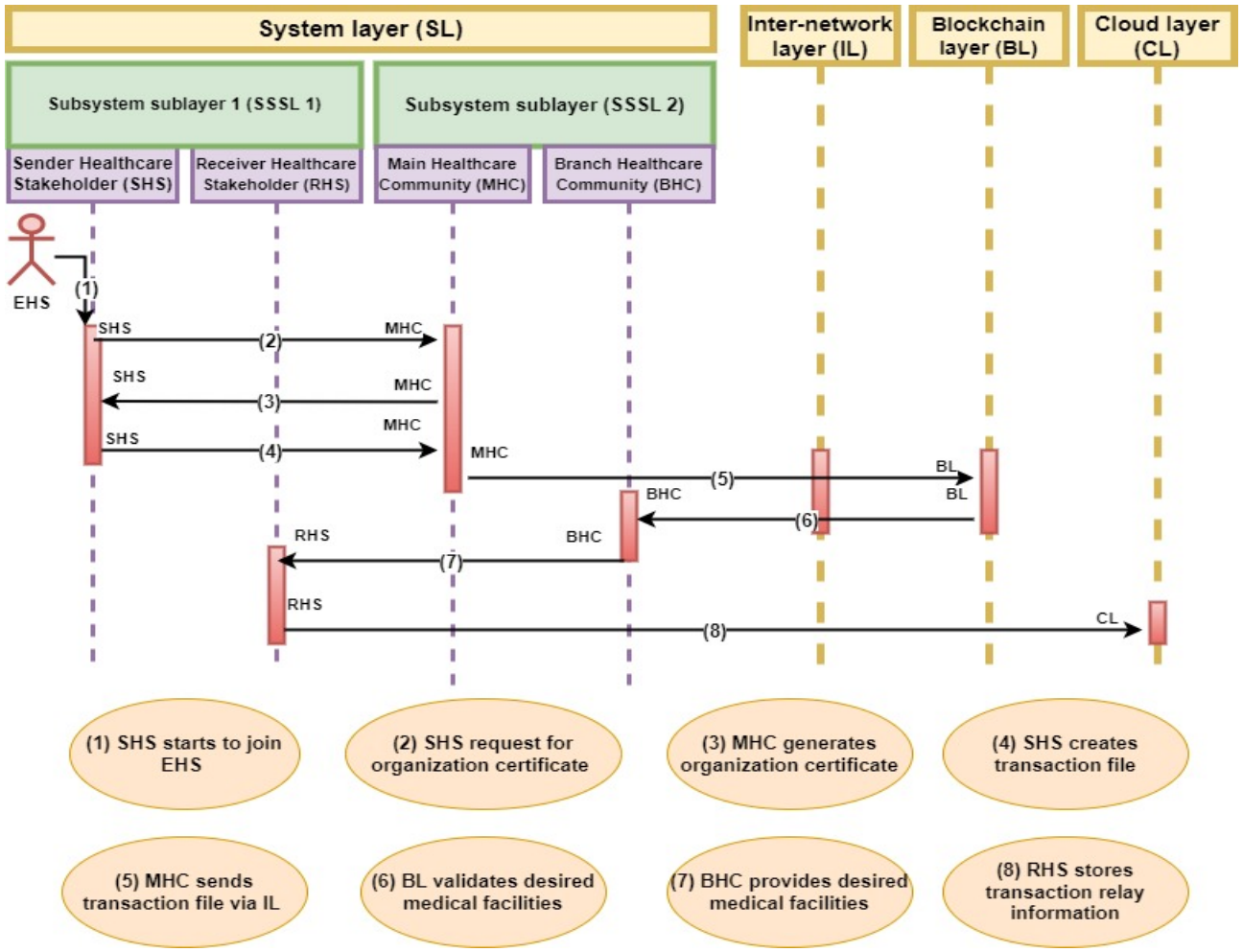


FIGURE 2: Sequence diagram of the proposed framework.

tion $shs(l)$. When l_0 exists for any $p > 0$, and $l \geq l_0$ exists, so that $shs(l) \leq 1/l^p$, then $shs(l)$ is ignored.

Definition 2 (Transaction file of SHS must be large enough and contains multiple transactions including tests and reports.) Transaction file $trans_file$ of SHS contains report $report$ including (xray report $xray_r$, blood report $blood_r$, urine report $urine_r$) and test $test$ including (nucleic acid test nat, Ct-scan ct_scan, antibody $antibody$, antigen $antigen$), where multiple reports are available. The probability of finding specific disease $disease$ in a limited time is $Adv_{disease,trans_file}(l) = P_r[disease(report, test) = m_f : m_f \in trans_file]$.

Definition 3 (If transaction file does not contain required medical facilities then ignored.) For all transaction file $trans_file$ within a limited time, there are some ignoring medical facilities. $Adv_{disease,p_trans_f}(l) \leq SHS_id$ exists, so the probability $report[disease(report, test) = m_f : m_f \in p_trans_f]$ is ignored.

C. SECURITY MODEL OF THE PROPOSED FRAMEWORK

In the proposed framework, the RSHealth considers attackers on the inter-network layer IL_A, which have access to the public keys of n SHSs and the private keys of some SHSs.

Gamer 1: the anti-tampering of RSHealth

For attacker on the inter-network layer IL_A and a challenger on the system layer SL_C, success in entering the blockchain-based EHS with RSHealth $S_{RSHealth,d}$ is determined by the probability as given in Game 1.

- **Phase 1:** Once the security parameter sec_p is set, the challenger on the system layer SL_C proceeds to complete Phase 1 to acquire the system parameters sys_p. Subsequently, SL_C forwards these system parameters sys_p to the attacker situated on the inter-network layer IL_A.
- **Phase 2:** Once the attacker on the inter-network layer IL_A selects any value, the challenger on the system layer SL_C supplies the corresponding hash value to IL_A.
- **Phase 3:** The attacker on the inter-network layer IL_A selects and requests SHS test $test_i$, prompting the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

challenger on the system layer SL_C to provide the corresponding test of SHS $test_i$.

- **Phase 4:** The attacker on the inter-network layer IL_A queries for SHS's disease $disease_i$, and the challenger on the system layer SL_C provides the corresponding disease of SHS $disease_i$.
- **Phase 5:** The attacker on the inter-network layer IL_A selects and submits SHS's identity $id*$, prompting the challenger on the system layer SL_C to provide the corresponding transaction file p_trans_f after applying the RSHealth to IL_A.
- **Phase 6:** At the end, the attacker on the inter-network layer IL_A generates the transaction file p_trans_f of pseudo identity $id*$ that satisfies the following two conditions:
  1) p_trans_f is a valid transaction file generated by the attacker on the inter-network layer IL_A.
  2) $id*$ does not available in the transaction file.

Definition 4 (Anti-tampering) If the attacker on the inter-network layer IL_A creates a maximum of q_T test queries, q_R hash queries for report, a maximum of q_D queries related to disease, and prepares a maximum of q_{TF} transaction files within the maximum time T, then the probability of success $S_{RSHealth,d}$ is at least single SHS, ensuring SHS file SHS_file = (T, q_T, q_R, q_D, q_{TF}, SHS) remain anti-tampered under adaptive selection message attack.

Gamer 2: the anonymity of RSHealth

Assuming $SHS = SHS_1, SHS_2, SHS_3, ..., SHS_n$ represents the n SHSs entities. Similar to Game 1, there exists an attacker on the inter-network layer IL_A and a challenger on the system layer SL_C.

- **Step 1:**The challenger on the system layer SL_C executes **Phase 1** of **Game 1** to compute system parameters sys_p and then transmits obtained result to the attacker on the inter-network layer IL_A.
- **Step 2:** The attacker on the inter-network layer IL_A adaptively makes polynomial limited-order queries q_T for transaction files $trans_file$.
- **Step 3:** The attacker on the inter-network layer IL_A outputs an identity $id*$, the test list T_l of n SHSs, two different tests antibody $antibody$, and antigen $antigen$ belonging to T_l, and sends all of them to the challenger on the system layer SL_C. The challenger on the system layer SL_C randomly selects a symptom sym from any test $test_i$ and returns SHS file SHS_file = **TFRing(**$id*$**,** Tl_i**,** $disease_i$**)** to the attacker on the inter-network layer IL_A.
- **Step 4:** Finally, the attacker on the inter-network layer IL_A outputs symptom $sym^{'}$ of any test $test_i$.
- **Step 5:** The attacker on the inter-network layer IL_A succeeds in this game if and only if sym = $sym^{'}$.

Definition 5 (Identity Anonymity) The probability of the attacker on the inter-network layer IL_A succeeding in Game 2 is given by $Success(IL_A) = P_r[sym = sym^{'}] = 1/2 + SHS$. The precondition for a SHS's transaction file to have full anonymity is that no attacker can win Game 2 with a non-negligible probability. This implies that for any polynomial-time attacker, the advantage of winning Game 2 to discover a SHS's identity is negligible.

D. TRANSACTION FILE PRIVACY PROTECTION BASED ON RING SIGNATURE

This section primarily discusses the utilization of RSHealth technology to devise a fully anonymous transaction file storage within the blockchain-based EHS, ensuring the confidentiality of SHS's identity during transmission from MHC to BHC. A smart contract is deployed in the BL to oversee the activities in the layer, and upon meeting predefined conditions, a preset instruction is activated to execute transaction file $trans_file$. The specific function, based on Algorithm 1, is constructed as follows:

1) **TFSetup(**$sec_p \rightarrow sys_p$**)**: This function is implemented at SSSL1. SHS takes a security parameter sec_p, which is a large prime number, as input. SHS then randomly checks the x-ray report $xray_r$ to ensure it is greater than least security parameter sec_p provided by EHS. Next, EHS identifies the disease $disease$ (AIDS $aids$) of every SHS in the queue $Q = \langle SHS_file^*, \times \rangle$ such that $1 \leq aids \leq xray_r$. Additionally, it collects the report $report$ of SHS that are primitive roots in the queue $Q = \langle SHS_file^*, \times \rangle$. Finally, it produces the system parameters $sys_p = (xray_r, aids, Q)$ as output.
2) **TFKeygen** $(sys_p \rightarrow [test_i, disease_i])$: Each SHS SHS_i $(1 \leq i \leq n)$ of MHC in the system layer applies the nucleic acid test nat based on the CT-scan ct_scan, disease $aids$, and x-ray report $xray_r$, following the (1):
$$nat = ct_scan^{aids} \% xray_r. \tag{1}$$
The public key of an SHS is defined as $(ct_scan, nat, xray_r) \in test_i$ and its private key is $aids \in disease_i$.
3) **TFRing([**id_i**,** $test_i$**,** $disease_i$**]** $\rightarrow$ $[SHS_id_i, Tl_i]$**)**: Here, if the transaction initiator SHS_k performs a random urine test $urine_t$ in the queue $Q = \langle SHS_file^*, \times \rangle$, then MHC calculates (2) and (3) as follows:
$$H_0(i) = H(SHS_id_i) \tag{2}$$
$$H_1(i) = [id_i - aids_i \times H_0(i)] \times urine_t_i^{-1} \% (xray_r_i - 1) \tag{3}$$
$$SHS_id_i = (ct_scan_i^{urine_t_i}) \% (xray_r_i) \tag{4}$$
$$Tl_i = id_i \times (nat_i^{urine_t_i}) \% (xray_r) \tag{5}$$
The original identity id of SHS_k is id_k. SHS's identity is anonymized by MHC, and the output contains transaction file $trans_file$ as:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

$(H_0(i), H_1(i), SHS_id_1, SHS_id_2, SHS_id_3, ...SHS_id_k, ...SHS_id_n, Tl_1, Tl_2, Tl_3, ...Tl_k...Tl_n)$.

4) **TFVerify(**$[V_1i, V_2i] \rightarrow [true, false]$**)**: Using(4) and (5), the transaction file is calculated by MHC. Then, MHC sent this transaction file $trans_file$ to BL via IL for verification, as mentioned in (6), and (7).

$$\left\{ \begin{array}{l} V_{1i} = nat_i^{H_0(i)} \times H_0(i)^{H_1(i)} \% (xray_r_i) \\ V_{2i} = (ct_scan_i^{id_i}) \% (xray_r_i) \end{array} \right\} \quad (6)$$

$$\sum_{i=1}^{n} V_{1i} = \sum_{i=1}^{n} V_{2i} \quad (7)$$

To calculate V_{1i} and V_{2i} using (6), BL first need to determine the values of V_{1i} and V_{2i} based on the required medical facilities i.e, ct-scan ct_scan and available medical facility in DPL i.e, nucleic acid test nat. Then, BL can verify whether (7) is true. If (7) is true, it implies that (4) and (5) have not been used by the attacker on the inter-network layer IL_A, indicating that the transaction file is valid. Conversely, if (7) is false, the transaction file is not valid.

5) **TFOpen(**$[SHS_id_i, Tl_i] \rightarrow id_i$**)**: After proper verification by BL, the valid transaction relay signal is transferred to BHC. BHC then provides the necessary medical facilities as per the requirement. Finally, valid transaction signal is sent to RHS. The original credentials of SHSs are again obtained by RHS, as given in (8).

$$id_i = Tl_i \times [(SHS_id_i)^{aids_i}]^{-1} \% (xray_r_i) \quad (8)$$

Algorithm 1: Proposed RSHealth

```
/* Intial setup using TFSetup()                    */
Input: sec_p
Output: sys_p
/* Key generation using TFKeygen()                 */
for i=1 to n do
    nat = ct_scan^aids % xray_r
    SHS_i = TFKeygen (sys_p → [test_i, disease_i]) ;
/* Transaction creation using TFRing()             */
for i=1 to n do
    Perform random urine test urine_t for transaction file T_file and
      then calculate:
    /* Hash calculation                            */
    H_0(i) = H(SHS_id_i)
    SHSID_i = (ct_scan_i^(urine_t_i)) % (xray_r_i)
    Tl_i = id_i × (nat_i^(urine_t_i)) % (xray_r)
/* Transaction verification using TFVerify()
   */
for i=1 to n do
    V_1i = nat_i^(H_0(i)) × H_0(i)^(H_1(i)) % (xray_r_i)
    V_2i = (ct_scan_i^(id_i)) % (xray_r_i)
    Σ_{i=1}^{n} V_1i = Σ_{i=1}^{h} V_2i
/* Transaction opened by RHSs using TFOpen()
   */
for i=1 to n do
    id_i = Tl_i × [(SHS_id_i)^(aids_i)]^(-1) % (xray_r_i)
```

IV. RESULTS ANALYSIS

In this section, the proposed RSHealth algorithm is evaluated in context of security analysis, execution analysis and performance analysis.

A. SECURITY ANALYSIS

A blockchain-based EHS with organization-level privacy protection relies on several security mechanisms, including blockchain technology, public key cryptosystem, and RSHealth. By integrating these security mechanisms into the design of a blockchain-based EHS establishes a robust framework for managing privacy of SHS's identity and security of transaction file. The proposed RSHealth guarantees accuracy, anonymity and anti-tampering features.

Lemma 4.1: **Assurance of accuracy.**

Proof: RHS receives V_1i, V_2i and uses identity id to verify the transaction list Tl_i according to (7), (9) , (10), (1), (2), (4), and (11). If it is true, then the RSHealth assures accuracy.

$\sum_{i=1}^{n} V_1i = \sum_{i=1}^{n} V_2i$

If $0 < V_1i < xray_r_i$ and $0 < V_2i < (xray_r_i - 1)$ exists then calculate:

$$V_1i = ct_scan_i^{id_i} \% (xray_r_i) \quad (9)$$

$$V_2i = (nat_i^{H_0(i)}) \times (H_0(i)^{H_1(i)}) \% (xray_r_i) \quad (10)$$

If V_1i is congruent to V_2i, the transaction file is accepted by RHS; otherwise, it is rejected.

$nat_i = ct_scan_i^{aids_i} \% (xray_r_i)$

$H_0(i) = H(SHS_id_i)$

$SHSID_i = (ct_scan_i^{urine_t_i}) \% (xray_r_i)$

$$\begin{aligned} &\sum_{i=1}^{n} V_2i \\ &= nat_i^{H_0(i)} \times H_0(i)^{H_1(i)} \% (xray_r_i) \\ &= ct_scan_i^{aids_i \times SHSID_i} \times SHS_id_i^{H_1(i)} \% (xray_r_i) \\ &= ct_scan_i^{aids_i \times SHSID_i} \times ct_scan_i^{urine_t_i \times H_1(i)} \% (xray_r_i) \\ &= ct_scan_i^{aids_i \times SHSID_i + urine_t_i \times H_1(i)} \% (xray_r_i) \\ &ct_scan_i^{id_i} \% (xray_r_i) \\ &\equiv \sum_{i=1}^{n} V_1i \end{aligned} \quad (11)$$

Here, ct_scan_i is available in DPL. Hence,

$$id_i = aids_i \times SHS_id_i + urine_t_i \times H_1(i)$$

or,

$$H_1(i) = [id_i - aids_i \times H_0(i)] \times urine_t_i^{-1} \% (xray_r_i - 1)$$

■

Lemma 4.2: **Assurance of anti-tampering.** In Game 1, the attacker situated on the inter-network layer IL_A has the capability to dynamically choose an identity for launching

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

an attack. Additionally, it is assumed that anti-tampering measures are successfully implemented within a reasonable time frame, with a probability of significant success.

Proof: If the challenger situated on the system layer SL_C receives any report, then the challenger's response to the attacker's query q_T is as follows:

- **Phase 1:** With a security parameter sec_p, the challenger operating on the system layer SL_C initiates Phase 1 to acquire the system parameters sys_p. Subsequently, the challenger on SL_C transmits these system parameters sys_p to the attacker situated on the inter-network layer IL_A.
- **Phase 2:**When the attacker positioned on the inter-network layer IL_A requests $H_0(i)$, the challenger situated on the system layer SL_C furnishes the corresponding hash value for $H_0(i)$.
- **Phase 3:** When the attacker operating on the inter-network layer IL_A requests SHS test $test_i$, the challenger located on the system layer SL_C supplies the corresponding SHS test $test_i$.
- **Phase 4:** If the attacker on the inter-network layer IL_A queries for the disease $disease_i$ of SHS, the challenger on the system layer SL_C ceases operation. Otherwise, the challenger furnishes the corresponding disease of SHS, denoted as $disease_i$.
- **Phase 5:** When the attacker situated on the inter-network layer IL_A selects and submits the original identity $id*$ of SHS, the challenger on the system layer SL_C supplies the corresponding transaction file p_trans_f after applying Algorithm 1 as specified in (1), (2), (3), (4) and (5).

  $$nat = ct_scan^{aids}\%xray_r.$$

  $$H_0(i) = H(SHS_id_i)$$

  $$H_1(i) = [id_i - aids_i \times H_0(i)] \times urine_t_i^{-1}\%(xray_r_i - 1)$$

  $$SHS_id_i = (ct_scan_i^{urine_t_i})\%(xray_r_i)$$

  $$Tl_i = id_i \times (nat_i^{urine_t_i})\%(xray_r)$$

  Ultimately, the original identity id^* is present in the transaction file $p_trans_f = (id_i*, SHS_id_1, SHS_id_2, ..., SHS_id_n, Tl_1, Tl_2, ..., Tl_n)$

- **Phase 6:** In conclusion, the attacker on the inter-network layer IL_A generates the transaction file p_trans_f for the pseudo identity $id*$. Thus, both the generated transaction files $trans_file$ and p_trans_f are considered valid.

  1) $trans_file = (id_i, SHS_id_1, SHS_id_2, ..., SHS_id_n, Tl_1, Tl_2, ..., Tl_n)$
  2) $p_trans_f = (id_i*, SHS_id_1, SHS_id_2, ..., SHS_id_n, Tl_1, Tl_2, ..., Tl_n)$

  Therefore, in Game 1, the attacker situated on the inter-network layer IL_A can indeed fabricate the p_trans_f. However, assuming that anti-tampering measures are effectively implemented within a reasonable time frame with a substantial probability of success, it becomes challenging for the attacker to tamper the $trans_file$. Even if the attacker randomly selects test such as urine test $urine_t$ to falsify nat, $H_0(i)$, $H_1(i)$, SHS_id_i, and Tl_i, the private key aids $aids_i$ of SHS is required for these calculations. Without knowing this private key, it is infeasible for the attacker to create id_i, rendering the forging of the $trans_file$ impossible. This indicates that the RSHealth effectively mitigates tampering attempts.

■

Lemma 4.3: **Assurance of SHS's identity anonymity.** The RSHealth ensures identity anonymity for all SHSs, guaranteeing that all reports have an equal probability of $P_r[sym = sym'] = 1/2$. This means that regardless of which SHS generates the transaction file, the likelihood of anonymity being preserved is consistent.

$$trans_file = (id_i, SHS_id_1, SHS_id_2, ..., SHS_id_n, Tl_1, Tl_2, ..., Tl_n).$$

Proof: In a manner similar to Game 1, both the attacker on the inter-network layer IL_A and the challenger on the system layer SL_C engage in Game 2.

- **Step 1:** The challenger on the system layer SL_C initiates Phase 1 of Game 1 to compute the system parameters sys_p and transmits it to the attacker situated on the inter-network layer IL_A.
- **Step 2:** The attacker positioned on the inter-network layer IL_A adaptively issues polynomial-limited-order queries q_T for transaction files.
- **Step 3:** The attacker on the inter-network layer IL_A provides an identity $id*$, a test list T_l comprising tests for n SHSs, and two distinct tests $antibody$ and $antigen$ both belonging to T_l. This information is then transmitted to the challenger on the system layer SL_C. The challenger randomly selects a symptom sym from any test $test_i$ and responds by sending the transaction file **TFRing**$(id, Tl_i, disease_i)$ to the attacker on the inter-network layer IL_A.
- **Step 4:** At the end, the attacker situated on the inter-network layer IL_A outputs the symptom sym' corresponding to any test $test_i$.
- **Step 5:** The attacker on the inter-network layer IL_A achieves success in this game if and only if $sym = sym'$.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

Prior to any SHS or RHS actively disclosing all the information contained within the transaction file, the original identity id of the created transaction file remains concealed from any third party. In the key generation **TFKeygen()**, and **TFOpen()** functions, knowledge of the secret disease $aids$ is imperative. In the signature **TFRing()**, and verification **TFVerify()** functions, the values $H_0(i)$ and $H_1(i)$ are based on SHS_id_i. Both SHS_id_i and Tl_i are derived by randomly selecting $urine_t_i$. The likelihood of an outsider of the blockchain-based EHS member correctly guessing the original identity of an SHS id suffering from AIDS or another disease does not exceed $1/(n+1)$, while the probability for an insider member of the blockchain-based EHS does not exceed $1/n$. Consequently, RSHealth ensures identity anonymity.

■

B. EXPERIMENTAL ANALYSIS

A system comprises an Intel(R) Core(TM) i5-8250u CPU @ 1.60 GHz with 8GB RAM and a 64-bit operating system, x-64 based processor is utilized for implementing the proposed RSHealth data privacy framework. The framework Algorithm 1 is executed step-by-step using the Remix IDE open source web and desktop application as demonstrated in Figure 3.

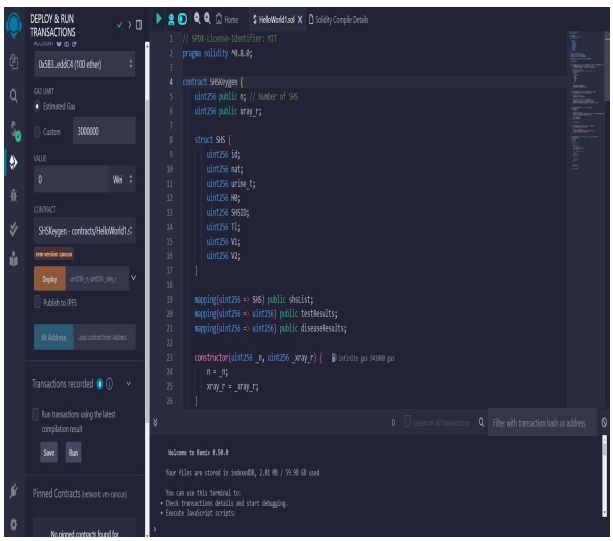


FIGURE 3: Execution of RSHealth.

TABLE 2: Computation time, transaction fee, and transaction file of SHS

Algorithms	Time(s)	Transaction_Fee (ETH)	Transaction file (KB)
TFSetup()	0.00503	0.00647853753848344	1808
TFKeygen()	0.00603	0.00647897154924584	1603
TFRing()	0.00945	0.00634729249825985	4061
TFVerify()	0.00753	0.00264783587424182	5427
TFOpen()	0.00802	0.00647753753848343	3205
RSHealth	0.014785	0.947859523641245625	6808

After deployment, computation time, transaction fee and size of transaction required for every functions of

TABLE 3: Normalization on computation time and transaction file

Algorithms	Time (msec)	Transaction_Fee (ETH)	Transaction file (MB)
TFSetup()	50	0.00647853753848344	2
TFKeygen()	60	0.00647897154924584	2
TFRing()	90	0.00634729249825985	4
TFVerify()	70	0.00264783587424182	5
TFOpen()	80	0.00647753753848343	3
RSHealth	100	0.947859523641245625	6

RSHealth **TFSetup()**, **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** along with RSHealth itself is listed in Table 2. Moreover, the normalized computation time, and transaction size are represented in Table 3. RSHealth is employed in a manner that facilitates its throughput or performance by estimating gas consumption during deployment of each function of RSHealth considering **TFSetup()**, **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** by varying number of SHSs SHS_i and sizes of transaction.

For instance, with 100 SHSs, the RSHealth requires less computation time for **TFSetup()** (50msec), **TFKeygen()** (60msec) , **TFRing()** (90 msec), **TFVerify()** (70msec), and **TFOpen()**(80m sec). Likewise with 200 SHSs, the RSHealth gains less computation time for **TFSetup()** (52msec), **TFKeygen()** (63msec) , **TFRing()** (91m sec), **TFVerify()** (74msec), and **TFOpen()**(85 msec). Simultaneously, with 300 SHSs, the RSHealth requires less computation time for **TFSetup()** (54msec), **TFKeygen()** (66msec) , **TFRing()** (92msec), **TFVerify()** (78msec), and **TFOpen()**(90 msec). Likewise, with 400 SHSs, the RSHealth requires less computation time for **TFSetup()** (56msec), **TFKeygen()** (69msec) , **TFRing()** (93msec), **TFVerify()** (82msec), and **TFOpen()**(95 msec). Finally, with 500 SHSs, the RSHealth reaches computation time for **TFSetup()** (58msec), **TFKeygen()** (72msec) , **TFRing()** (94msec), **TFVerify()** (86msec), and **TFOpen()**(100 msec).

In evaluating the RSHealth in terms of **TFSetup()**, **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()**, Figure 4 presents a layered graphical representation of the test results. This provides robust framework under the supervision of relevant organization, ensuring identity anonymity and anti- tampering measures are satisfied.

C. PERFORMANCE ANALYSIS

The performance of the proposed RSHealth is evaluated utilizing five functions: **TFSetup()**, **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** with varying sizes of transaction file $trans_file$ from 2MB to 6MB (2 MB, 3 MB, 4 MB, 5 MB, and 6 MB). The throughput and latency is calculated using (12), and (13) respectively.

$$Throughput = \sum_{i=1}^{n} (Gas \div Time) * trans_file_i \quad (12)$$

$$Latency = \sum_{i=1}^{n} (Gas * Time) \div SHS_i \quad (13)$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

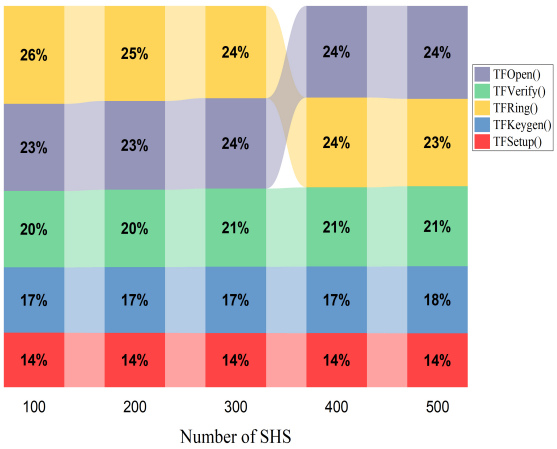


FIGURE 4: Gas consumption in each function of RSHealth.

As for example, if number of SHS is 100, size of transaction file is 2MB, gas consumption (as seen in Figure 4) and time spend by every functions of RSHealth **TFSetup()** , **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** are (14%, 50msec), (17%, 60msec), (26%, 90msec), (20%, 70msec), and (23%, 80msec) respectively then throughput is 2.85 MBps and latency is 73 msec using (12), and (13) respectively. Similarly, if number of SHS is 200, size of transaction file is 3MB, gas consumption (as seen in Figure 4) and time spend by every functions of RSHealth **TFSetup()** , **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** are (14%, 52msec), (17%, 63msec), (25%, 91msec), (20%, 74msec), and (25%, 85msec) respectively then throughput is 4.06MBps and latency is 38.39 msec using (12), and (13) respectively. Consequently, if number of SHS is 300, size of transaction file is 4MB, gas consumption (as seen in Figure 4) and time spend by every functions of RSHealth **TFSetup()** , **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** are (14%, 54msec), (17%, 66msec), (24%, 92msec), (21%, 78msec), and (24%, 90msec) respectively then throughput is 5.25 MBps and latency is 26.28 msec using (12), and (13) respectively. Similarly, if number of SHS is 400, size of transaction file is 5MB, gas consumption (as seen in Figure 4) and time spend by every functions of RSHealth **TFSetup()** , **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** are (14%, 56msec), (17%, 69msec), (24%, 93msec), (21%, 82msec), and (24%, 95msec) respectively then throughput is 6.32 MBps and latency is 20.47 msec using (12), and (13) respectively. Finally, if number of SHS is 500, size of transaction file is 6MB, gas consumption (as seen in Figure 4) and time spend by every functions of RSHealth **TFSetup()** , **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()** are (14%, 58msec), (18%, 72msec), (23%, 94msec), (21%, 86msec), and (24%, 100msec) respectively then throughput is 7.32 MBps and latency is 16.95 msec using (12), and (13) respectively. The tested results are shown in Table 4.

The proposed method RSHealth obtains high throughput as the number of SHS, and size of transaction file increases as shown in Figure 5. Also, it provides low latency when increasing the number of SHS, and transaction files as shown in Figure 6.

The comparative results provided in Table 5. The tested results as shown in Table 5 works as follows: while observing the overall framework performance in terms of throughput, RSHealth having superior outcomes resulting in Existing scheme 1 [15], and Existing scheme 2 [28] are shown in Figure 7.

Despite a substantial increase in the number of SHS from 100 to 500, computation time of RSHealth remain modest due to the employment of efficient cryptosystem. Additionally, Table 6 shows a uniqueness of the proposed framework in comparison with existing works.

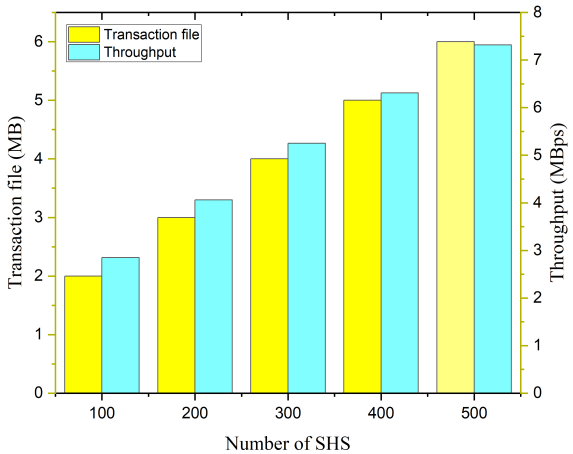


FIGURE 5: Throughput with increasing no. of SHS and transaction file.

TABLE 4: Throughput and latency of RSHealth.

S.N.	no. of SHS	Transaction file (MB)	Throughput (MBps)	latency (msec)
1.	100	2	2.85	73.00
2.	200	3	4.06	38.39
3.	300	4	5.25	26.28
4.	400	5	6.31	20.47
5.	500	6	7.32	16.95

TABLE 5: Comparison of RSHealth with Existing works [15] and [28]

No. of SHS	Transaction file (MB)	RSHealth (MBps)	Existing scheme1 [15]	Existing scheme2 [28]
100	2	2.85	1.23	1.10
200	3	4.06	2.10	2.20
300	4	5.25	2.40	2.30
400	5	6.31	3.21	3.05
500	6	7.32	3.35	3.48

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

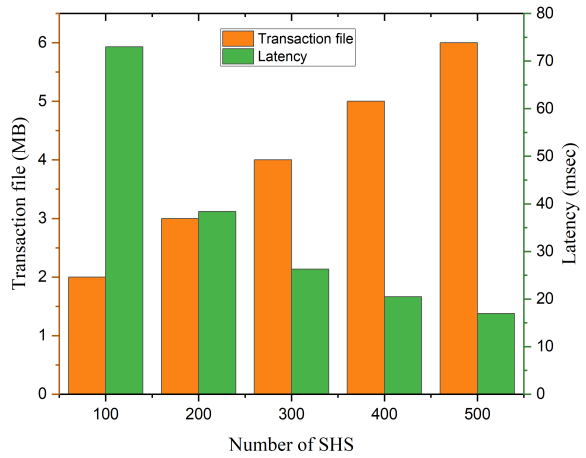


FIGURE 6: Latency with increasing no. of SHS and transaction file.

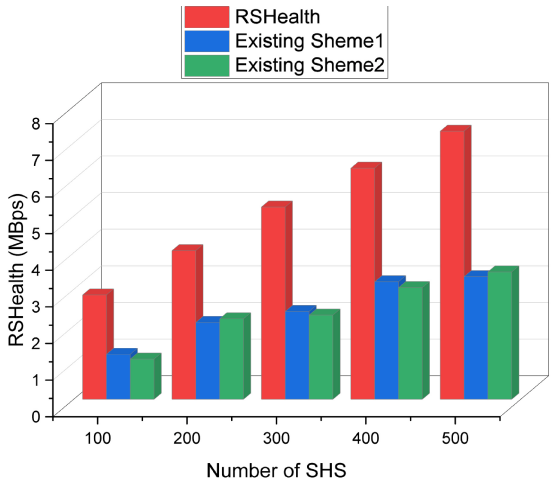


FIGURE 7: Performance of RSHealth with Existing schemes.

TABLE 6: Existing works with RSHealth

Paper	EHS	signature scheme	layer security	annonymity	anti-tampering
kumar et al. [34]	yes	no	yes	no	no
hara et al. [19]	no	ring	no	yes	no
li et al. [8]	no	ring	no	yes	yes
prabha et al. [22]	yes	no	yes	no	no
mamta et al. [4]	yes	no	yes	no	no
Existing scheme 1 [15]	no	group	no	yes	yes
Existing scheme 2 [28]	no	ring	no	yes	yes
RSHealth	yes	ring	yes	yes	yes

V. CONCLUSION AND FUTURE WORK

The present study focuses on the security and privacy of sensitive healthcare transaction files in a blockchain-based EHS. Specifically, it addresses the challenges of ensuring security and privacy when transmitting transaction files from MHC to BHC in the lack of necessary medical facilities at MHC. The RSHealth framework has five functions: **TFSetup()**, **TFKeygen()**, **TFRing()**, **TFVerify()**, and **TFOpen()**. A comprehensive security analysis confirms that RSHealth is highly effective in ensuring both anti-tampering and identity anonymity for SHS. The performance research shows that there is a higher rate of data transfer and a lower delay in processing as the number of SHS and the size of each transaction rise. One significant drawback of this study is the omission of IoT device authentication provided on IL. In order to successfully utilize the blockchain-based EHS, it is imperative to do additional research and apply the suggested framework to authenticate devices for the IL. These efforts are essential for improving and deploying the framework in blockchain-based EHS.

DECLARATION

- Ethical approval: Not applicable.
- Competing interests: The authors declare that they have no known competing financial interests or personal relationship that could have appeared to influence the work reported in this paper.
- Authors' contributions: The authors confirm contribution to the paper as follows: Punam Prabha: Conceptualization, Data curation, Formal analysis, Writing-original draft, Writing-review & editing, Validation. Kakali Chatterjee: Writing-review & editing, Validation.
- Funding: Not applicable.
- Availability of data and materials: Not applicable.

REFERENCES

[1] T. Navaneethakrishnan, R. Manohar, and V. Srinivasan, "Decoding ransomware attacks: Unmasking legal solutions in the IT ACT of 2000," Indian Journal of Law and Legal Research, vol. 1, 2023.
[2] M. Wazid and P. Gope, "BACKM-EHA: A novel blockchain-enabled security solution for IoMT-based e-healthcare applications," ACM Transactions on Internet Technology, vol. 23, no. 3, pp. 1–28, 2023.
[3] H. Habibzadeh, K. Dinesh, O. R. Shishvan, A. Boggio-Dandry, G. Sharma, and T. Soyata, "A survey of healthcare Internet of Things (HIOT): A clinical perspective," IEEE Internet of Things Journal, vol. 7, no. 1, pp. 53–71, 2019.
[4] B. Mamta, B. Gupta, K. Li, V. Leung, K. Psannis, and S. Yamaguchi, "Blockchain-assisted secure fine-grained searchable encryption for a cloud-based healthcare cyber-physical system. ieee/caa j autom sin," 2021.
[5] D. Zhang, S. Wang, Y. Zhang, Q. Zhang, and Y. Zhang, "A secure and privacy-preserving medical data sharing via consortium blockchain," Security and Communication Networks, vol. 2022, no. 1, p. 2759787, 2022.
[6] G. Verma, "Blockchain-based privacy preservation framework for healthcare data in cloud environment," Journal of Experimental & Theoretical Artificial Intelligence, vol. 36, no. 1, pp. 147–160, 2024.
[7] R. Gupta, P. Bhattacharya, S. Tanwar, N. Kumar, and S. Zeadally, "Garuda: A blockchain-based delivery scheme using drones for healthcare 5.0 applications," IEEE Internet of Things Magazine, vol. 4, no. 4, pp. 60–66, 2021.
[8] X. Li, Y. Mei, J. Gong, F. Xiang, and Z. Sun, "A blockchain privacy protection scheme based on ring signature," IEEE Access, vol. 8, pp. 76 765–76 772, 2020.
[9] A. Padma and M. Ramaiah, "GLSBIOT: GWO-based enhancement for lightweight scalable blockchain for IoT with trust based consensus," Future Generation Computer Systems, vol. 159, pp. 64–76, 2024.
[10] A. Padma and R. Mangayarkarasi, "Detecting security breaches on smart contracts through techniques and tools a brief review: Applications and challenges," in International Conference on Information and Management Engineering. Springer, 2022, pp. 361–369.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

[11] S. M. Williamson and V. Prybutok, "Balancing privacy and progress: A review of privacy challenges, systemic oversight, and patient perceptions in AI-Driven healthcare," Applied Sciences, vol. 14, no. 2, p. 675, 2024.
[12] G. Zhang, Z. Yang, and W. Liu, "Blockchain-based privacy preserving e-health system for healthcare data in cloud," Computer Networks, vol. 203, p. 108586, 2022.
[13] B. S. Egala, A. K. Pradhan, V. Badarla, and S. P. Mohanty, "Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control," IEEE Internet of Things Journal, vol. 8, no. 14, pp. 11 717–11 731, 2021.
[14] B. S. Egala, A. K. Pradhan, P. Dey, V. Badarla, and S. P. Mohanty, "Fortified-chain 2.0: Intelligent blockchain for decentralized smart healthcare system," IEEE Internet of Things Journal, 2023.
[15] H. Zheng, Q. Wu, J. Xie, Z. Guan, B. Qin, and Z. Gu, "An organization-friendly blockchain system," Computers & Security, vol. 88, p. 101598, 2020.
[16] J.-S. Lee, C.-J. Chew, J.-Y. Liu, Y.-C. Chen, and K.-Y. Tsai, "Medical blockchain: Data sharing and privacy preserving of EHR based on smart contract," Journal of Information Security and Applications, vol. 65, p. 103117, 2022.
[17] L. Wang, G. Zhang, and C. Ma, "A survey of ring signature," Frontiers of Electrical and Electronic Engineering in China, vol. 3, pp. 10–19, 2008.
[18] R. L. Rivest, A. Shamir, and Y. Tauman, "How to leak a secret," in Advances in Cryptology—ASIACRYPT 2001: 7th International Conference on the Theory and Application of Cryptology and Information Security Gold Coast, Australia, December 9–13, 2001 Proceedings 7. Springer, 2001, pp. 552–565.
[19] K. Hara and K. Tanaka, "Ring signature with unconditional anonymity in the plain model," IEEE Access, vol. 9, pp. 7762–7774, 2021.
[20] A. V. Meier, "The Elgamal cryptosystem," in Joint Advanced Students Seminar, 2005.
[21] R. A. Haraty, A. N. El-Kassar, and B. M. Shebaro, "A comparative study of Elgamal based digital signature algorithms," Journal of Computational Methods in Sciences and Engineering, vol. 6, no. s1, pp. S147–S156, 2006.
[22] P. Prabha and K. Chatterjee, "Design and implementation of hybrid consensus mechanism for IoT based healthcare system security," International Journal of Information Technology, vol. 14, no. 3, pp. 1381–1396, 2022.
[23] S. Shreya, K. Chatterjee, and A. Singh, "A smart secure healthcare monitoring system with internet of medical things," Computers and Electrical Engineering, vol. 101, p. 107969, 2022.
[24] A. Padma and R. Mangayarkarasi, "Blockchain based an efficient and secure privacy preserved framework for smart cities," IEEE Access, vol. 12, pp. 21 985–22 002, 2024.
[25] Y. Bobde, G. Narayanan, M. Jati, R. S. P. Raj, I. Cvitić, and D. Peraković, "Enhancing Industrial IoT network security through blockchain integration," Electronics, vol. 13, no. 4, p. 687, 2024.
[26] Y. Tang, F. Xia, Q. Ye, M. Wang, R. Mu, and X. Zhang, "Identity-based linkable ring signature on ntru lattice," Security and Communication Networks, vol. 2021, pp. 1–17, 2021.
[27] C. Gamage, B. Gras, B. Crispo, and A. S. Tanenbaum, "An identity-based ring signature scheme with enhanced privacy," in 2006 Securecomm and Workshops. IEEE, 2006, pp. 1–5.
[28] X. Zhang and C. Ye, "A novel privacy protection of permissioned blockchains with conditionally anonymous ring signature," Cluster Computing, vol. 25, no. 2, pp. 1221–1235, 2022.
[29] Y. Ma, Y. Wang, J. Yang, Y. Miao, and W. Li, "Big health application system based on health internet of things and big data," Ieee Access, vol. 5, pp. 7885–7897, 2016.
[30] H. Zhang, J. Li, B. Wen, Y. Xun, and J. Liu, "Connecting intelligent things in smart hospitals using NB-IOT," IEEE Internet of Things Journal, vol. 5, no. 3, pp. 1550–1560, 2018.
[31] B. D. Deebak, F. Al-Turjman, M. Aloqaily, and O. Alfandi, "An authentic-based privacy preservation protocol for smart e-healthcare systems in IoT," IEEE Access, vol. 7, pp. 135 632–135 649, 2019.
[32] F. Alsubaei, A. Abuhussein, V. Shandilya, and S. Shiva, "IoMT-SAF: Internet of medical things security assessment framework," Internet of Things, vol. 8, p. 100123, 2019.
[33] G. Wu, S. Wang, Z. Ning et al., "Blockchain-enabled privacy-preserving access control for data publishing and sharing in the internet of medical things," IEEE Internet of Things Journal, vol. 9, no. 11, pp. 8091–8104, 2021.
[34] A. Kumar and K. Chatterjee, "A lightweight blockchain-based framework for medical cyber-physical system," The Journal of Supercomputing, vol. 79, no. 11, pp. 12 013–12 041, 2023.
[35] M. Mohamed, A. Elsayed, and M. Voskoglou, "Collaboration of vague theory and mathematical techniques for optimizing healthcare by recommending optimal blockchain supplier," Sustainable Machine Intelligence Journal, vol. 6, pp. 6–1, 2024.

BIOGRAPHY



PUNAM PRABHA received her B.Tech degree in Computer Science & Engineering from Aryabhatta Knowledge University, Patna, Bihar, India, in 2013, and her M.Tech in Computer Science & Engineering with a specialization in Information Security from the Indian Institute of Technology (Indian School of Mines), Dhanbad, India, in 2016. Since 2018, she has been working as an Assistant Professor in the Computer Science & Engineering Department at the Department of Science, Technology, and Technical Education, Patna, Bihar, India. Additionally, she is pursuing a part-time Ph.D. in Computer Science Engineering from the National Institute of Technology, Patna, Bihar, India.

She has authored one Scopus-indexed journal paper and six international conference papers. She received best paper awards at "PEEIC-2019" and "FICTA-2022". Her research interests include information security, blockchain technology, and data privacy.



KAKALI CHATTERJEE is an Assistant Professor in Computer Science and Engineering Department of National Institute of Technology Patna, India. She has completed her Ph.D. from Delhi University (formerly Delhi College of Engineering) and M.Tech from Guru Gobind Singh Indraprastha University, Delhi. She has completed research project in Centre for Development of Advanced Computing, a R&D and Academic centre of Govt. of India. She has published many SCI/SCIE/SCOPUS indexed International Journals of Springer and Elsevier. She has attended & presented research papers in reputed conferences and also published some book chapters in Springer. She is also a joint principal investigator of ISEA Project (Phase-II), Govt of India. Her research interest is in the field of Information Security and Cryptography. Currently she is doing research in Healthcare security, Biometrics, Malware detection etc.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611

APPENDIX

A. SMART CONTRACT CODE IN SOLIDITY OF RSHEALTH

```
// SPDX-License-Identifier: MIT
pragma solidity ^0.8.0;

contract RSHealthComputation {
    uint256 public transactionFileSize = 2 * 1024 * 1024; // 2MB in bytes

    struct Computation {
        uint256 startTime;
        uint256 endTime;
        uint256 gasUsed;
    }

    mapping(string => Computation) public computations;

    event ComputationStarted(string functionName, uint256 startTime);
    event ComputationEnded(string functionName, uint256 endTime, uint256 gasUsed);

    modifier trackComputation(string memory functionName) {
        uint256 startGas = gasleft();
        uint256 startTime = block.timestamp;
        emit ComputationStarted(functionName, startTime);

        _;

        uint256 endTime = block.timestamp;
        uint256 gasUsed = startGas - gasleft();
        computations[functionName] = Computation(startTime, endTime, gasUsed);
        emit ComputationEnded(functionName, endTime, gasUsed);
    }

    function keygen(uint256 sys_p, uint256[] memory testResults, uint256[] memory diseaseResults)
    public trackComputation("Keygen") {
        uint256 n = testResults.length;
        uint256 xray_r = 123456; // Arbitrary value for example

        for (uint256 i = 0; i < n; i++) {
            uint256 nat = (sys_p * (i + 1)) % xray_r;
            // Perform Keygen computations...
        }
    }

    function createTransactionFile(uint256[] memory urineTests)
    public trackComputation("CreateTransactionFile") {
        uint256 n = urineTests.length;
        uint256 xray_r = 123456; // Arbitrary value for example

        for (uint256 i = 0; i < n; i++) {
            uint256 urine_t = urineTests[i];
            // Perform transaction file creation computations...
        }
    }

    function verifyTransactionFile(uint256[] memory testResults)
    public  trackComputation("VerifyTransactionFile") returns (bool) {
        uint256 n = testResults.length;
        uint256 xray_r = 123456; // Arbitrary value for example

        uint256 sumV1 = 0;
        uint256 sumV2 = 0;

        for (uint256 i = 0; i < n; i++) {
            uint256 V1 = (testResults[i] * (i + 1)) % xray_r;
            uint256 V2 = (testResults[i] * (i + 1)) % xray_r;

            sumV1 += V1;
            sumV2 += V2;
        }

        return sumV1 == sumV2;
    }
```


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611


```
    function openTransactionFile(uint256[] memory aids)
    public trackComputation("OpenTransactionFile") {
        uint256 n = aids.length;
        uint256 xray_r = 123456; // Arbitrary value for example

        for (uint256 i = 0; i < n; i++) {
            uint256 inv = modInverse((aids[i] * (i + 1)), xray_r);
            // Perform opening transaction file computations...
        }
    }

    function modInverse(uint256 a, uint256 m) internal pure returns (uint256) {
        int256 m0 = int256(m);
        int256 y = 0;
        int256 x = 1;

        if (m == 1) return 0;

        while (a > 1) {
            int256 q = int256(a) / int256(m);
            int256 t = int256(m);

            m = a % m;
            a = uint256(t);
            t = y;

            y = x - q * y;
            x = t;
        }

        if (x < 0) x += m0;

        return uint256(x);
    }
}
```

B. SMART CONTRACT CODE IN SOLIDITY OF EXISTING SCHEME 1

```
// SPDX-License-Identifier: MIT
pragma solidity ^0.8.0;

import "@openzeppelin/contracts/utils/math/SafeMath.sol";
import "@openzeppelin/contracts/access/Ownable.sol";

contract RSHealthSystem is Ownable {
    using SafeMath for uint256;

    // Define system parameters
    struct SystemParams {
        uint256 p;
        uint256 g1;
        uint256 g2;
    }

    SystemParams public params;

    constructor(uint256 _p, uint256 _g1, uint256 _g2) {
        params.p = _p;
        params.g1 = _g1;
        params.g2 = _g2;
    }

    // Define keys and certificates
    struct RegManagerKey {
        uint256 x;
        uint256 y;
        uint256 u;
        uint256 v;
    }

    struct OrgKey {
        uint256 h1;
        uint256 u1;
```


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611


```
        uint256 v1;
        uint256 w1;
        uint256 nu1;
        uint256 n1;
        uint256 g1;
    }

    struct UserKey {
        uint256 h;
        uint256 u;
    }

    struct OrgCertificate {
        uint256 sigma0;
        uint256 r;
    }

    struct UserCertificate {
        uint256 A;
        uint256 a;
    }

    mapping(address => RegManagerKey) public regManagerKeys;
    mapping(address => OrgKey) public orgKeys;
    mapping(address => UserKey) public userKeys;
    mapping(address => OrgCertificate) public orgCertificates;
    mapping(address => UserCertificate) public userCertificates;

    // Helper functions
    function modExp(uint256 base, uint256 exp, uint256 mod) internal pure returns (uint256) {
        uint256 result = 1;
        base = base % mod;
        while (exp > 0) {
            if (exp % 2 == 1) {
                result = (result * base) % mod;
            }
            exp = exp >> 1;
            base = (base * base) % mod;
        }
        return result;
    }

    // Key generation for RegManager
    function regManagerKeyGen(uint256 x, uint256 y) public onlyOwner {
        uint256 u = modExp(params.g2, x, params.p);
        uint256 v = modExp(params.g2, y, params.p);
        regManagerKeys[msg.sender] = RegManagerKey(x, y, u, v);
    }

    // Key generation for Org
    function orgKeyGen(
        uint256 h1,
        uint256 x1,
        uint256 y1,
        uint256 r1,
        uint256 nu1,
        uint256 p1,
        uint256 q1,
        uint256 g1
    ) public {
        require(h1 != 1, "h1 should not be 1");
        uint256 u1 = modExp(h1, x1, params.p);
        uint256 v1 = modExp(h1, y1, params.p);
        uint256 w1 = modExp(params.g2, r1, params.p);
        uint256 n1 = p1.mul(q1);
        uint256 lambda1 = lcm(p1.sub(1), q1.sub(1));
        require(gcd(L(modExp(g1, lambda1, n1.mul(n1)), n1), n1) == 1, "Invalid g1");

        orgKeys[msg.sender] = OrgKey(h1, u1, v1, w1, nu1, n1, g1);
    }

    // Key generation for User
    function userKeyGen(uint256 h, uint256 x) public {
        require(h != 1, "h should not be 1");
```


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611


```
        uint256 u = modExp(h, x, params.p);
        userKeys[msg.sender] = UserKey(h, u);
    }

    // Issue Org Certificate
    function issueOrgCert(address orgAddress, uint256 r) public onlyOwner {
        RegManagerKey storage regKey = regManagerKeys[msg.sender];
        OrgKey storage orgKey = orgKeys[orgAddress];
        uint256 sigma0=modExp(params.g1,regKey.x.add(orgKey.nu1).add(regKey.y.mul(orgKey.w1)),params.p);
        orgCertificates[orgAddress] = OrgCertificate(sigma0, r);
    }

    // Issue User Certificate
    function issueUserCert(address userAddress, uint256 a, uint256 r1) public {
        UserKey storage userKey = userKeys[userAddress];
        OrgKey storage orgKey = orgKeys[msg.sender];
        uint256 A = modExp(params.g1.div(userKey.u), r1.add(a), params.p);
        userCertificates[userAddress] = UserCertificate(A, a);
    }

    // Transaction Generation (simplified for example purposes)
    function generateTransaction(address receiver, uint256 amount) public {
        // Implement transaction generation logic
    }

    // Transaction Verification (simplified for example purposes)
    function verifyTransaction(bytes memory transaction) public view returns (bool) {
        // Implement transaction verification logic
        return true;
    }

    // Tracking (simplified for example purposes)
    function trackUser(address userAddress) public view returns (bool) {
        // Implement user tracking logic
        return true;
    }

    // Utility functions
    function lcm(uint256 a, uint256 b) internal pure returns (uint256) {
        return a.mul(b).div(gcd(a, b));
    }

    function gcd(uint256 a, uint256 b) internal pure returns (uint256) {
        while (b != 0) {
            uint256 temp = b;
            b = a % b;
            a = temp;
        }
        return a;
    }

    function L(uint256 x, uint256 n) internal pure returns (uint256) {
        return (x.sub(1)).div(n);
    }
}
```

C. SMART CONTRACT CODE IN SOLIDITY OF EXISTING SCHEME2

```
// SPDX-License-Identifier: MIT
pragma solidity ^0.8.0;

contract CARS {
    // Parameters
    uint256 public p;
    uint256 public q;
    uint256 public g;
    uint256 public h;
    address public CA;

    // Events
    event UserRegistered(address indexed user, uint256 publicKey);
    event TransactionSigned(address indexed user, bytes signature);
    event TransactionVerified(bool success);
```


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611


```
    event MisbehaviorDetected(address indexed user);

    // Structs
    struct User {
        uint256 publicKey;
        uint256 secretKey;
    }

    struct Signature {
        uint256 q;
        uint256[] rs;
        uint256 a;
        uint256 b;
        uint256 c;
        uint256 h;
    }

    // Mappings
    mapping(address => User) public users;
    mapping(bytes => bool) public transactions;

    // Constructor to initialize the system
    constructor(uint256 _p, uint256 _q, uint256 _g, uint256 _h) {
        p = _p;
        q = _q;
        g = _g;
        h = _h;
        CA = msg.sender;
    }

    // Modifier to restrict access to CA
    modifier onlyCA() {
        require(msg.sender == CA, "Only CA can call this function");
        _;
    }

    // Key Generation
    function registerUser(address user, uint256 secretKey) public onlyCA {
        uint256 publicKey = modExp(g, secretKey, p);
        users[user] = User(publicKey, secretKey);
        emit UserRegistered(user, publicKey);
    }

    // Signature generation function
    function signTransaction(bytes memory txData, uint256 u, uint256 v, uint256 w) public {
        require(users[msg.sender].publicKey != 0, "User not registered");

        uint256 a = modExp(h, u, p);
        uint256 b = modExp(a, v, p);
        uint256 c = modExp(a, w, p);
        uint256 hValue = (w + v * uint256(keccak256(abi.encodePacked(txData, a, c)))) % p;

        uint256 txHash = uint256(keccak256(abi.encodePacked(txData, a, b)));
        uint256 R = modExp(g, uint256(keccak256(abi.encodePacked(txData))), p);

        uint256[] memory rs = new uint256[](1);
        rs[0] = uint256(keccak256(abi.encodePacked(txData, txHash, R))) % q;

        uint256 rk=rs[0]-(uint256(keccak256(abi.encodePacked(txData)))*users[msg.sender].secretKey) % q;
        uint256 qValue=(uint256(keccak256(abi.encodePacked(txData))-rk*users[msg.sender].secretKey) % q;

        Signature memory sig = Signature(qValue, rs, a, b, c, hValue);
        transactions[txData] = true;

        emit TransactionSigned(msg.sender, abi.encode(sig));
    }

    // Verification function
    function verifyTransaction(bytes memory txData, bytes memory sigData) public returns (bool) {
        Signature memory sig = abi.decode(sigData, (Signature));

        uint256 txHash = uint256(keccak256(abi.encodePacked(txData, sig.a, sig.b)));
        uint256 R = modExp(g, sig.q, p);
        bool valid = (sig.h == (sig.c * uint256(keccak256(abi.encodePacked(txData, sig.a, sig.c))) % p)
```


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. 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/ACCESS.2024.3439611


```
        && (modExp(g, sig.q, p) * modExp(users[msg.sender].publicKey,
        uint256(keccak256(abi.encodePacked(txData))), p)) % p == R);

        emit TransactionVerified(valid);
        return valid;
    }

    // Misbehavior detection function
    function detectMisbehavior(address user, uint256 a, uint256 e, uint256 v) public onlyCA {
        uint256 b = modExp(a, e, p);
        bool misbehavior = (modExp(g, a + v * e, p) == (b * modExp(users[user].publicKey, e, p)) % p);

        if (misbehavior) {
            emit MisbehaviorDetected(user);
        }
    }

    // Modular exponentiation function
    function modExp(uint256 base, uint256 exp, uint256 mod) internal pure returns (uint256) {
        uint256 result = 1;
        while (exp > 0) {
            if (exp % 2 == 1) {
                result = (result * base) % mod;
            }
            base = (base * base) % mod;
            exp /= 2;
        }
        return result;
    }
}
```


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

Silicon
https://doi.org/10.1007/s12633-024-03125-2

RESEARCH

Sensitivity Investigation of Underlap Gate Cavity-Based Reconfigurable Silicon Nanowire Schottky Barrier Transistor for Biosensor Application

Anil Kumar[1] · Vijay Thakur[1] · Suraj Kumar[1] · Sumit Kale[1] · Kaustubh Ranjan Singh[1]

Received: 1 July 2024 / Accepted: 21 August 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract
This study investigates the sensitivity of Underlap Gate Cavity-based Reconfigurable Silicon Nanowire Schottky Barrier Transistor (UCG-RSiNW SBT) with an underlap gate-drain region for biosensing application. The featured unique reconfigurable capability enables the device to operate as either p-type or n-type, dependent on the applied bias polarity. The proposed biosensor incorporates a cavity beneath the control gate on the source side, facilitating the placement of both neutral and charged biomolecules with varying dielectric constant (K) values. Upon injection of biomolecules into the cavity, the device changes electrostatic characteristics, including modulation in threshold voltage, potential, electric field, and sub-threshold swing, I_{ON}, I_{ON}/I_{OFF} ratio. The threshold voltage (V_{TH}) Sensitivity of n-mode is enhanced by 97.91%, while that of p-mode is raised by 16% compared to conventional RFET biosensors. The drain current sensitivity and the linearity of proposed biosensor is enhanced upto the values of 2792 and 0.997 respectively in n-mode configuration whereas in p-mode configuration, the drain current sensitivity and the linearity comes out to be 968 and 0.995 respectively. These high sensitivity and linearity values make this biosensor superior to the existing state-of-the-art biosensors. The findings from this study provide valuable insights into the development of highly sensitive biosensors for applications in diverse fields, including healthcare and biotechnology.

Keywords Reconfigurable · Biosensors · Schottky barrier · Sensitivity

1 Introduction

The detection and quantification of biomolecules play a vital role in various fields, including medical diagnostics, environmental monitoring, and food safety. Conventional biosensors often rely on labeling techniques, attaching fluorescent or radioactive tags to biomolecules for detection. While these techniques have been widely used, they suffer from several limitations, including the requirement for labeling reagents, the potential for interference from labeling processes, and the limited shelf life of labeled biomolecules. Label-free biosensors offer an attractive alternative, providing real-time detection without the need for labeling. These biosensors utilize various physical or chemical principles to detect the presence or concentration of biomolecules directly, offering advantages in simplicity, speed, and cost-effectiveness [1–3].

The initial FET-based biosensor, known as the Ion-Sensitive FET (ISFET), demonstrated a high capability for sensing charged biomolecules. Nonetheless, ISFET faced a significant limitation as it was unable to detect neutral biomolecules effectively. To address this shortcoming, researchers extensively studied the dielectric modulation (DM) method to enhance FET-based biosensors for both charged and neutral biomolecule detection. Recently, innovative device technologies such as Tunnel FET (TFET), Junction-Less FET (JFET), and Impact Ionization FET (IMFET) have been explored for label-free biomolecule detection. However, TFET-based biosensors encounter challenges like low sensitivity, variability issues, and complex fabrication requirements due to their low on-state current, high thermal budget, and significant doping needs. Additionally, the benefits of TFET-based biosensors have only been demonstrated theoretically, with no practical devices reported so far. Similarly, biosensors based on JFET and

✉ Sumit Kale
sumitkale785@gmail.com

[1] Department of Electronics and Communication Engineering, Delhi Technological University, 110042 Delhi, New Delhi, India

Published online: 29 August 2024

IMFET also require high doping levels, leading to problems such as random dopant fluctuations, device variability, and fabrication complexity at the nanoscale, which ultimately diminishes their sensitivity for biosensor applications.

Silicon nanowires play an important role in developing highly sensitive biosensors due to their high surface-to-volume ratio, compatibility with existing technology, rapid response, versatility, and label-free detection. Silicon nanowire-based reconfigurable FET biosensors hold immense potential for detecting a wide range of biomolecules [4]. Several researchers, such as Vijay et al., Ajay et al., Sumeet et al., Priyanka et al., etc., have shown the RFET-based biosensor has a cavity in the middle of the gates. However, these researchers overlooked two crucial factors: the location of biomolecule immobilization and the ambipolar behaviour of the device. This seemingly inconspicuous detail can significantly impact the overall performance of the biosensor, particularly its sensitivity.

RFETs rely on Schottky junctions formed at the interface between the source/drain metal and the Silicon Channel. These junctions play a vital role in controlling the flow of electrons within the device. When biomolecules are immobilized close to these junctions, they can interact with the electric field, leading to changes in the Schottky barrier width and, consequently, the current flow. This interaction forms the basis of detection in RFET-based biosensors. However, if the biomolecules are immobilized too far from the source/drain interface, their interaction with the electric field becomes weaker. This results in reduced sensitivity, meaning the biosensor requires higher concentrations of biomolecules to produce a detectable response [5–11]. To overcome these challenges and achieve optimal performance, there is a need to focus on the strategic positioning of the biomolecule immobilization site. This typically involves creating a cavity region near the source interface, where the biomolecules are specifically bound. This approach offers several advantages, such as enhanced sensitivity, reduced Schottky barrier width, and an increased tunneling region.

In this paper, for the first time, we introduce the underlap gate cavity-based reconfigurable silicon nanowire Schottky barrier transistor (UGC-RSiNW SBT) as a labeled-free biosensor with the gate-drain underlap region. The purpose of the gate-drain underlap region is to reduce the ambipolarity of the proposed biosensor. The purpose of the underlap gate-drain region is to reduce the ambipolarity of the proposed biosensor. The presence of underlap gate-drain region modifies the energy band structure of the device, resulting in a wider tunneling region at the drain and channel interface, due to which the tunneling of the charge carriers from the channel to drain reduces and thus the ambipolar conduction reduces. The extension of the gate electrode over the drain electrode also allows for better electrostatic control of the channel by the gate maintaining the channel in a non-conductive state when the device is supposed to be "off," effectively reducing undesired charge carrier flow that contributes to ambipolar behavior. To investigate the performance of the proposed device, the energy band diagram, electric field, surface potential, V_{TH} sensitivity, SS sensitivity and the drain current sensitivity have been plotted. The linear performance of the proposed device is also evaluated as the V_{Th} Linearity. The device performance and biosensor sensitivity are compared with various other reported works. The analysis shows that the proposed device is a promising candidate for label-free biosensor applications.

2 Device Structure

The 2D configuration of the UCG-RSiNW SBT biosensor is illustrated in Fig. 1(a). The schematic showcases SiO_2 layers of thickness 10 nm, 20 nm, and 1 nm positioned at control gate dielectric, isolater between control gate and program gate, and the adhesive layer in the cavity, respectively. TiO_2, with a thickness of 10 nm is deposited as the program gate dielectric. Aluminum is the gate material with a work function of 4.2 eV, while $NiSi_2$ as the source/drain electrode, with a work function of 4.83 eV, are utilized. A cavity, with a thickness of 9 nm and length of 25 nm, is established beneath the control gate toward the source end, intended for

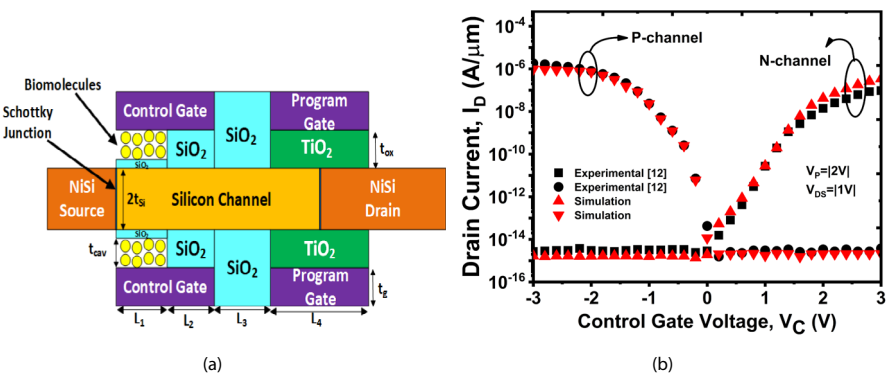


Fig. 1 (a) 2D cross-sectional view of the proposed device and (b) Calibration with the experimental results [12]

biomolecule placement. Other device dimension values are tabulated in Table 1.

The proposed device can work as an N-Channel as well as a P-Channel field effect transistor. In an N-type (P-type) configuration, the drain voltage (V_{DS}) is set at 0.8 (-0.8V) volts, while the program gate voltage (V_P) is maintained at 1.5 (-1.5V) volts. During experimentation, the control gate voltage (V_C) is varied within the range of -2 volts to 2 volts.

The simulated results are calibrated using experimental data for both N and P-mode setups [12]. By varying the V_C from -3V to +3V at $V_{DS} = \pm 1V$, the comparison in Fig. 1(b) shows a strong alignment between simulation and experimental outcomes. This close match between TCAD simulation results and actual data confirms the reliability of the ATLAS TCAD tool in assessing the biosensor device's performance. The biosensor's performance is assessed using the Silvaco ATLAS TCAD tool [13]. The model incorporates the Effective Bohm Quantum Potential (BQP) to address quantum effects, the Universal Schottky Tunneling model for metal-semiconductor junction tunneling [14, 15], and includes the Auger model, Fermi Dirac statistics, conmob model, consrh model, cvt model, fnord, Boltzmann, bqp.qdir=4, temperature=300, and non-local band-to-band (bbt.nonlocal) tunneling model [16, 17].

3 Results and Discussion

3.1 Energy Band Diagram

The existence of biomolecules in the cavity region significantly modulates the width of the Schottky barrier at the interface between the Source and Channel. In the n-type mode, the program gate voltage (V_P) is 1.5V, while the drain voltage(V_D) is 0.8 V. The control gate voltage is varied from -2 to 2 V. In this configuration, the dielectric constant of neutral biomolecules K=1, 2.63, 5, and 12. Our observations show that as the dielectric constant increased, the conduction band shifted downward, leading to a reduction in the tunneling width and an increase in the tunneling region (as depicted in Fig. 2(a)). This shift resulted in an increase in the tunneling rate of electrons from the source to the channel, ultimately reaching the drain. As a result, the on-current increased with increasing K, while the threshold voltage decreased. Simultaneously, the Schottky barrier width for holes at the drain side increased, which prevented hole injection into the Channel.

In Fig. 2(c), we depict the impact of charged biomolecules on the energy band diagram, considering Q_F values of 1E12 cm^{-2} and -1E12 cm^{-2} at K=12 in n-type mode. When dealing with positively charged biomolecules, we observed a reduction in tunneling width, indicating a downward shift in the conductance band. This shift increased the ON current. Conversely, for negatively charged biomolecules in n-type mode, the conduction band shifted upward, leading to an increase in tunneling width and a subsequent decrease in the ON current.

Similarly, in the p-type mode, where V_p=-1.5V and V_D=-0.8V are applied, we varied the control gate voltage from -2V to 2V while increasing the dielectric constant (K) values to K=1, 2.63, 5, and 12. Our observations showed that as the dielectric constant increased, the valence band shifted upward, as illustrated in Fig. 2(b). This upward shift in the valence band resulted in a reduction of the tunneling width for holes. Consequently, as k increased, a greater number of holes entered the Channel from the source and reached the drain, leading to an increase in the on-current. Simultaneously, the Schottky barrier width for electrons at the drain side increased. This increase in barrier width prevents electron injection into the channel. In Fig. 2(d), we depict the impact of charged biomolecules on the energy band diagram, considering Q_F (biomolecules charge density) values of 1E12 cm^{-2} and -1E12 cm^{-2} at a constant K=12 in n-type mode. The valence band shifts upward for negatively charged biomolecules in n-type mode, leading to a decrease in tunneling width and a subsequent increase in the ON current. For positive-charged biomolecules, we observe an increase in tunneling width, indicating a downward shift in the valance band. This shift results in a decrease in the ON current.

Table 1 Device Parameters of UGC-RSiNW SBT Biosensor

Parameters	Value
Source/Drain Metal work-function, $\phi_{s,d}$	4.83 eV
Gate Metal work-function, ϕ_g	4.2 eV
Channel Doping Concentration	10^{15} cm^{-3}
Si NW Channel Length, L	70 nm
Control Gate Length, L_{cg}=$L_1 + L_2$	40 nm
Program Gate Length over channel, L_4	20 nm
Isolator ($Si O_2$) Length, L_{Iso}=L_3	10 nm
Si NW Thickness, $2t_{Si}$	10 nm
$Si O_2$ Thickness under the cavity, t_{ud}	1 nm
$Si O_2/Ti O_2$ Thickness, t_{ox}	10 nm
Cavity Length, L_{cav}=L_1	25 nm
Cavity Thickness, t_{cav}	9 nm
Thickness of Program/Control Gate, t_g	10 nm
Source Length, L_S	10 nm
Drain Length, L_D	30 nm

3.2 Surface Potential

The dielectric constant of the biomolecules influences the gate capacitance of the proposed device. A higher dielectric constant increases the gate capacitance, which in turn affects the electric field distribution within the device. This change in

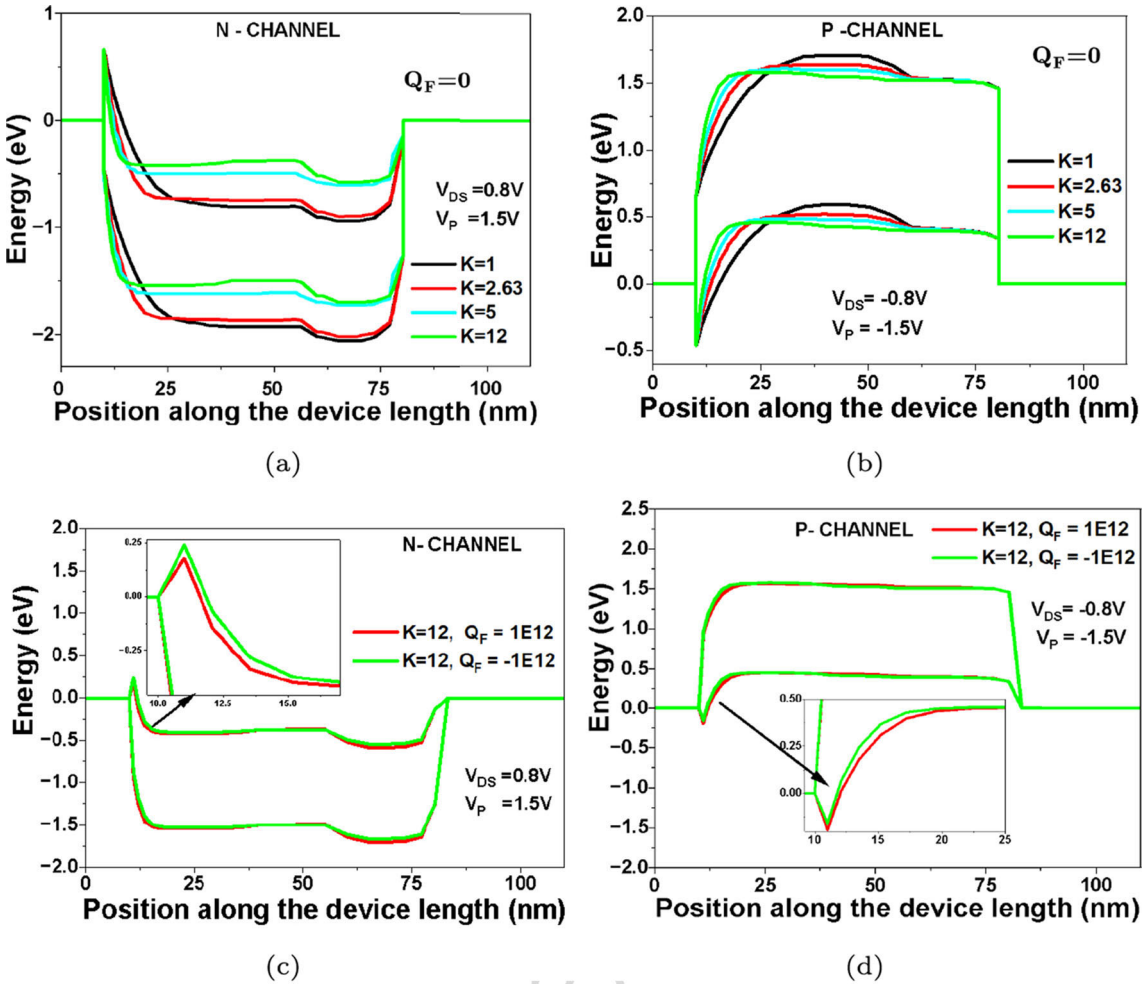


Fig. 2 Variation in Energy band diagram of proposed biosensor along the Channel length with a varying dielectric constant of neutral biomolecules when operating in (a) n-mode and (b) p-mode, and with a varying charge density of charged biomolecules for (c) n-mode and (d) p-mode

the electric field modifies the surface potential. Biomolecules with higher dielectric constants can also induce more polarization at the interface, leading to a shift in the surface potential. This happens because the increased polarization alters the charge distribution in the channel. In Fig. 3a, the potential curve rises in the cavity region with increasing dielectric constant values of neutral biomolecules in the n-type mode, resulting in a higher on-current due to the increased flow of electrons to the drain. Conversely, in the p-type mode, the potential decreases as the dielectric constant increases, leading to a greater number of holes reaching the drain and increasing the on-current as represented in Fig. 3b.

Upon further investigation, the surface potential variation of charged biomolecules with a charge density Q_F = (-1E12,1E12) cm^{-2} was examined when placed in the cavity at the fixed dielectric constant of biomolecules (K = 12). In the n-type mode, depicted in Fig. 3c, it was observed that the surface potential increases for positive charge density biomolecules compared to negative charge density biomolecules. Similarly, in the p-type mode, illustrated in Fig. 3d, a similar trend was observed. This is because of the interaction of the charged biomolecules with the charges in the semiconductor channel. Positively charged biomolecules will attract electrons (or repel holes), while negatively charged biomolecules will repel electrons (or attract holes). This interaction modifies the charge density in the channel, leading to a shift in the surface potential. The presence of charged biomolecules changes the local electric field in the vicinity of the gate. Positive charges will enhance the electric field, while negative charges will reduce it. This variation in the electric field directly affects the surface potential.

3.3 Electric Field

In n-type mode, as illustrated in Fig. 4a, when the dielectric constant of neutral biomolecules within the cavity increases (with values of K=1, 2.63, 5, and 12), there is a decrease in the tunneling width for electrons. This reduction prompts more electrons to enter the Channel, increasing the electric field

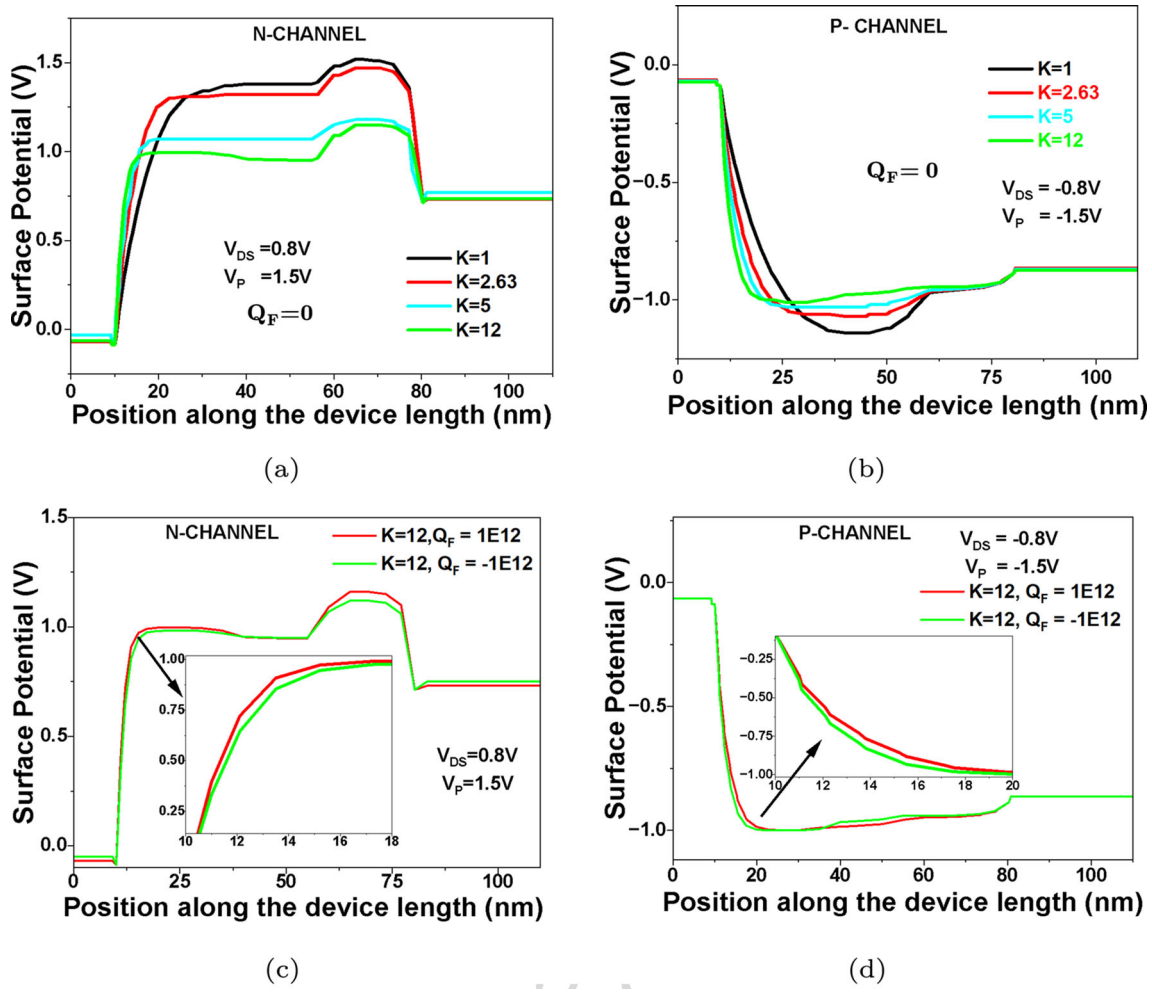


Fig. 3 Variation in Potential of proposed biosensor along the Channel length with varying dielectric constant of neutral biomolecules when operating in (a) n-mode and (b) p-mode, and with varying charge density of charged biomolecules for (c) n-mode and (d) p-mode

value. In p-type mode, as shown in Fig. 4b, when the dielectric constant (K) of biomolecules in the cavity increases, the tunneling width for holes decreases. This means more holes enter the Channel, leading to an increase in the electric field value.

The electric field variation for charged biomolecules (QF = 1E12, -1E12) cm^{-2} at a constant K value of 12 is illustrated in Fig. 4c and d. In N-type mode, as depicted in Fig. 4c, the electric field value is higher for positively charged biomolecules in comparison to negatively charged ones. Conversely, in P-type mode, the electric field value for negatively charged biomolecules surpasses that for positively charged biomolecules, as shown in Fig. 4d.

3.4 Drain Current

In Fig. 5a, in N-type mode, as the value of the dielectric constant (K=1, 2.63, 5, and 12) of neutral biomolecules increases, the tunneling width for electrons decreases. Consequently, more electrons enter the Channel, leading to an increase in drain current. Similarly, in Fig. 5b, for P-type mode, as the value of K of neutral biomolecules increases, the tunneling width for holes decreases. This results in more holes entering the Channel, causing an increase in drain current.

Figure 5c and d represent the variation of drain current with charged biomolecules (Qf = -1E12,1E12) cm^{-2} at a constant K = 12. In N-type mode, as shown in Fig. 5c, the drain current is higher for positive charge biomolecules, and lower for negative charge biomolecules, when compared to neutral biomolecules (Q_F = 0 cm^{-2}). Similarly, in P-type mode, as illustrated in Fig. 5d, the drain current for negatively charged biomolecules is higher. In contrast, for positively charged biomolecules, it is lower in comparison to neutral charged biomolecules, at K = 12. Table 2 illustrates the DC output parameters of the proposed biosensor.

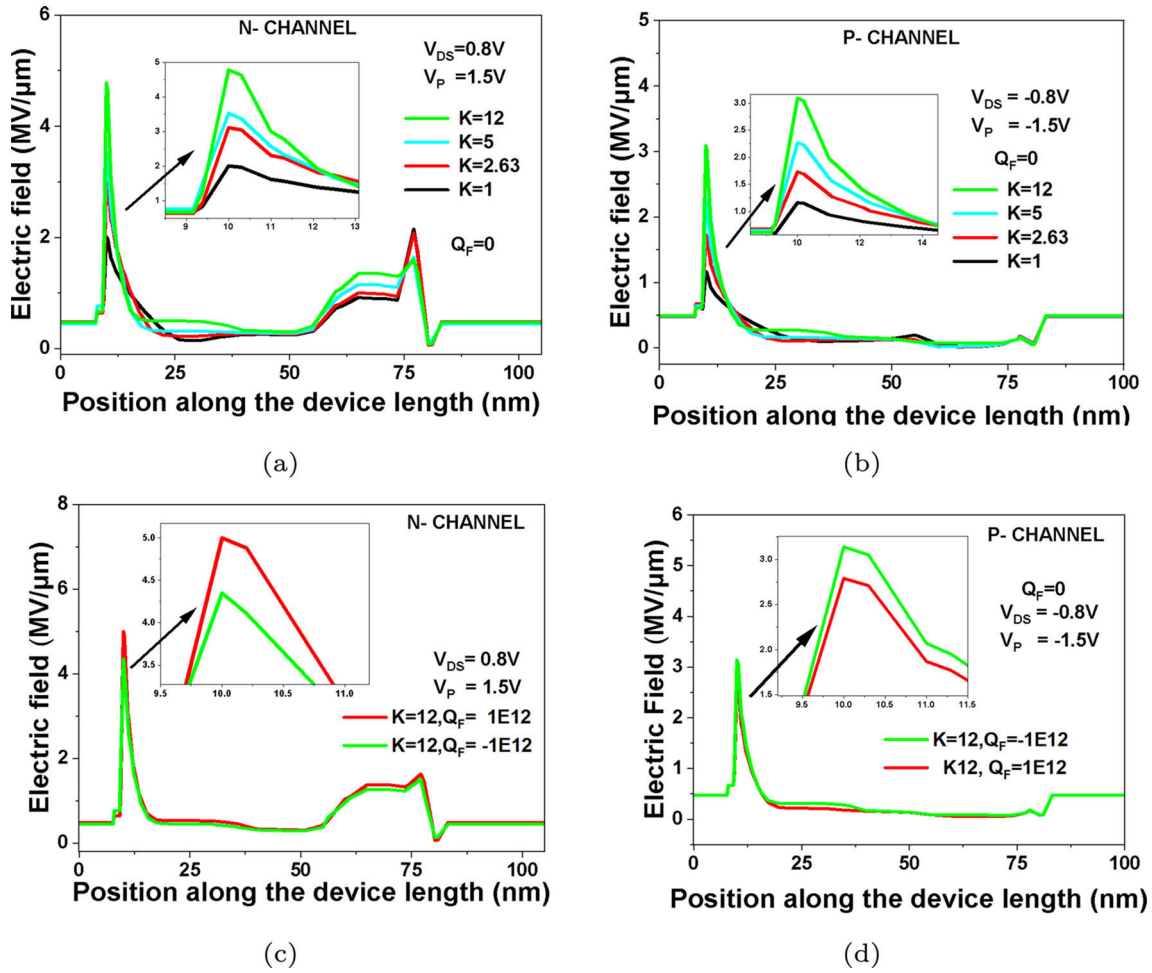


Fig. 4 Variation in Electric Field of proposed biosensor along the Channel length with varying dielectric constant of neutral biomolecules when operating in (a) n-mode and (b) p-mode, and with varying charge density of charged biomolecules for (c) n-mode and (d) p-mode

3.5 Sensitivity Analysis

Sensitivity stands as a pivotal parameter in biosensor evaluation. Its computation involves a comparison between the device characteristics when the cavities are occupied by biomolecules and when they are filled with air.

3.5.1 V_{TH} Sensitivity

To calculate the V_{TH} sensitivity, Eq. 1 is utilized [20, 21]

$$S_N = \left|\frac{V_{TH(K=1)} - V_{TH(K>1)}}{V_{TH(K=1)}}\right| \quad (1)$$

$$S_C = \left|\frac{V_{TH(neutral)} - V_{TH(charged)}}{V_{TH(neutral)}}\right| \quad (2)$$

Here, S_N is the V_{TH} sensitivity for neutral biomolecules and S_C is the V_{TH} sensitivity for charged biomolecules. $V_{TH(K=1)}$ is the threshold voltage when the cavity is filled with air. $V_{TH(K>1)}$ is the threshold voltage when the cavity is filled with neutral biomolecules. $V_{TH(neutral)}$ is the threshold voltage of neutral biomolecules at K=12. $V_{TH(charged)}$ is the threshold voltage of charged biomolecules at K=12.

Placing neutral biomolecules with varying dielectric constants (K values of 1, 2.63, 5, and 12) into the UCG-RSiNW SBT biosensor cavity, we calculated V_{TH} sensitivity using Eq. 1. The results showed that as the dielectric constant of biomolecules increased, the V_{TH} sensitivity also increased in both N-type and P-type modes, as illustrated in Fig. 6a and b. It was noted that in the N-type mode, the V_{TH} sensitivity reached its maximum at K=12, measuring 0.48, while in the P-type mode, it was recorded at 0.058 for the K=12.

We further investigated V_{TH} sensitivity using Eq. 2 by varying the charge density of biomolecules (QF=-1E13,-1E12,1E12,1E13) cm^{-2} at k=12 for both N-type and P-type modes. The results are illustrated in Fig. 6c and d. As the magnitude of the charge density increased, the V_{TH} sensitivity also exhibited an increase in both N-type and P-type modes. Specifically, for N-type mode, the maximum sensitivity was observed at Q_F= -1E13 cm^{-2}, reaching 1.2. In

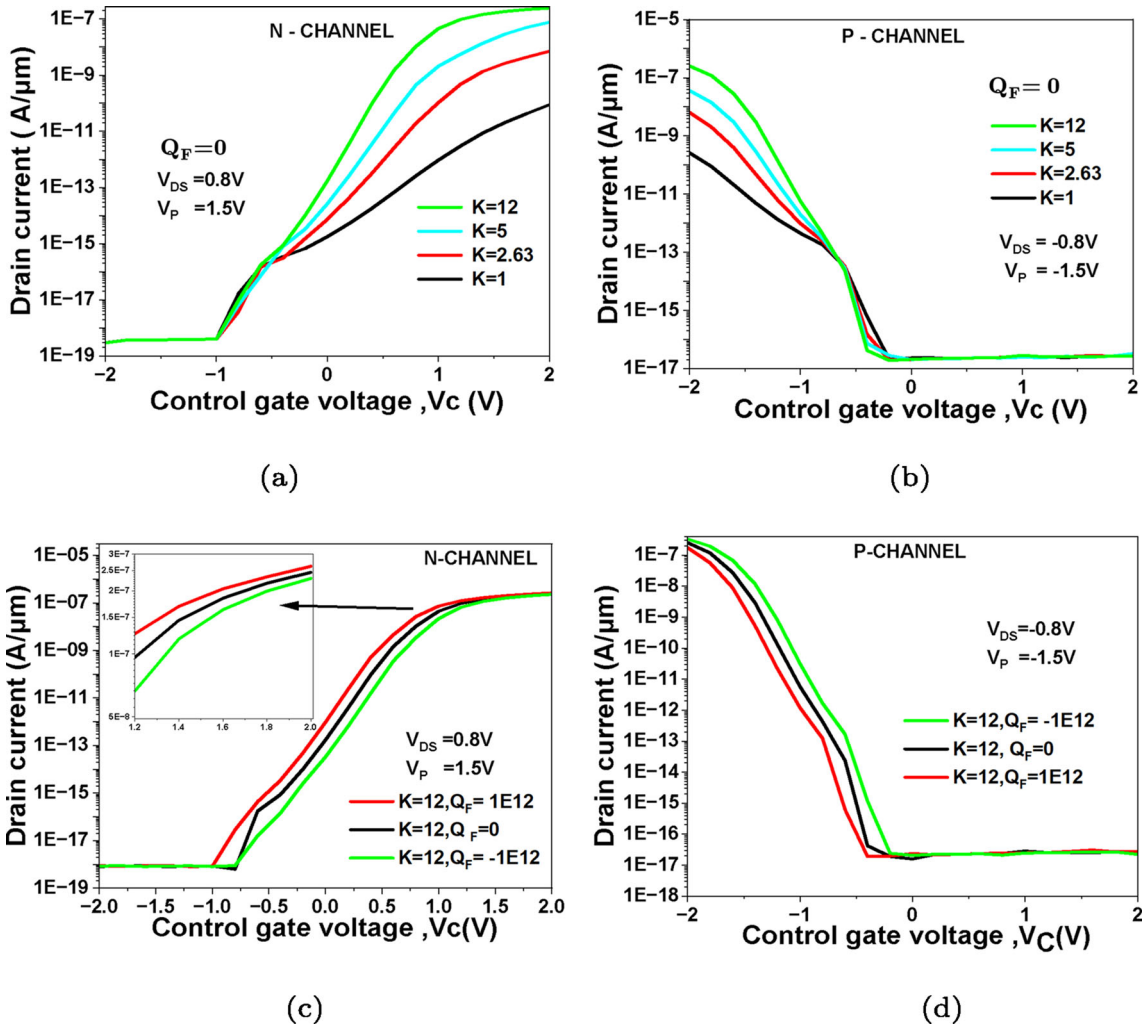


Fig. 5 Variation in Drain Current of the proposed biosensor with respect to control gate voltage Vc(V) with varying dielectric constant of neutral biomolecules when operating in (a) n-mode and (b) p-mode, and with varying charge density of charged biomolecules for (c) n-mode and (d) p-mode

contrast, for P-type mode, the maximum sensitivity occurred at Q_F=1E13 cm^{-2}, measuring 0.09.

3.5.2 Subthreshold Swing (SS) Sensitivity

To compute the subthreshold swing (SS) sensitivity of neutral biomolecules, we applied Eq. 3.

$$SS_N = |\frac{SS(k=1) - SS(k>1)}{SS(k=1)}| \tag{3}$$

$$SS_C = |\frac{SS_{neutral} - SS_{charged})}{SS_{neutral}}| \tag{4}$$

Here, SS_N is the Subthreshold Slope sensitivity for neutral biomolecules and SS_C is the Subthreshold Slope sensitivity for Charged biomolecules. $SS(K = 1)$ is the subthreshold slope(mv/dec) of air, $SS(K > 1)$ is the subthreshold slope(mv/dec) of neutral biomolecules. $SS(k = 12)_{neutral}$ is the Subthreshold Slope (mv/dec) for neutral biomolecules having dielectric constant, K=12. $SS(k = 12)_{charged}$ is the Subthreshold Slope (mv/dec) for charged biomolecules having dielectric constant, K= 12.

Placing neutral biomolecules with varying dielectric constants (K= 1, 2.63, 5, and 12) in the UCG-RSiNW SBT biosensor cavity, we used Eq. 3 to calculate SS sensitivity. In both N-type and P-type modes, we observed that as the dielectric constant increased, the SS sensitivity also increased, as depicted in Fig. 7a and b. It was noted that in the N-type mode, the maximum SS sensitivity was observed at k=12, reaching 0.75, while in the P-type mode, the maximum SS sensitivity at k=12 was recorded as 0.35.

We extended our investigation of V_{TH} sensitivity using Eq. 4, varying the charge density of biomolecules (Q_F=-1E13, -1E12, 1E12, 1E13) cm^{-2} at k=12 for both N-type and

Table 2 DC Output Parameters of UCG-RSiNW SBT Biosensor

Dielectric Constant	Q_F (cm^{-2})	N-Channel				P-Channel			
		V_{TH} (V)	I_{ON} (A)	I_{ON}/I_{OFF}	SS (mV/dec)	V_{TH} (V)	I_{ON} (A)	I_{ON}/I_{OFF}	SS (mV/dec)
1	0	1.60	8.8E-11	1.493E9	152	-1.72	2.64E-10	1.28E7	116
2.63	0	1.48	7.14E-9	2.06E13	91	-1.71	6.42E-9	3.25E8	85
5	0	1.43	7.68E-8	2.10E13	79	-1.67	3.63E-8	2.78E9	77
12	0	0.83	2.46E-7	1.26E14	36	-1.62	2.55E-7	1.34E10	72
12	1E12	0.71	2.62E-7	7.08E13	31	-1.70	1.76E7	7.61E11	86
12	1E13	1.03	6.60E-7	2.39E14	142	-1.77	1.02E-12	73420.2	93
12	-1E12	0.94	2.30E-7	1.65E15	30	-1.51	3.34E-7	5.15E12	92
12	-1E13	1.83	1.87E-8	5.05E12	26	-1.22	1.48E-6	1.36E11	98

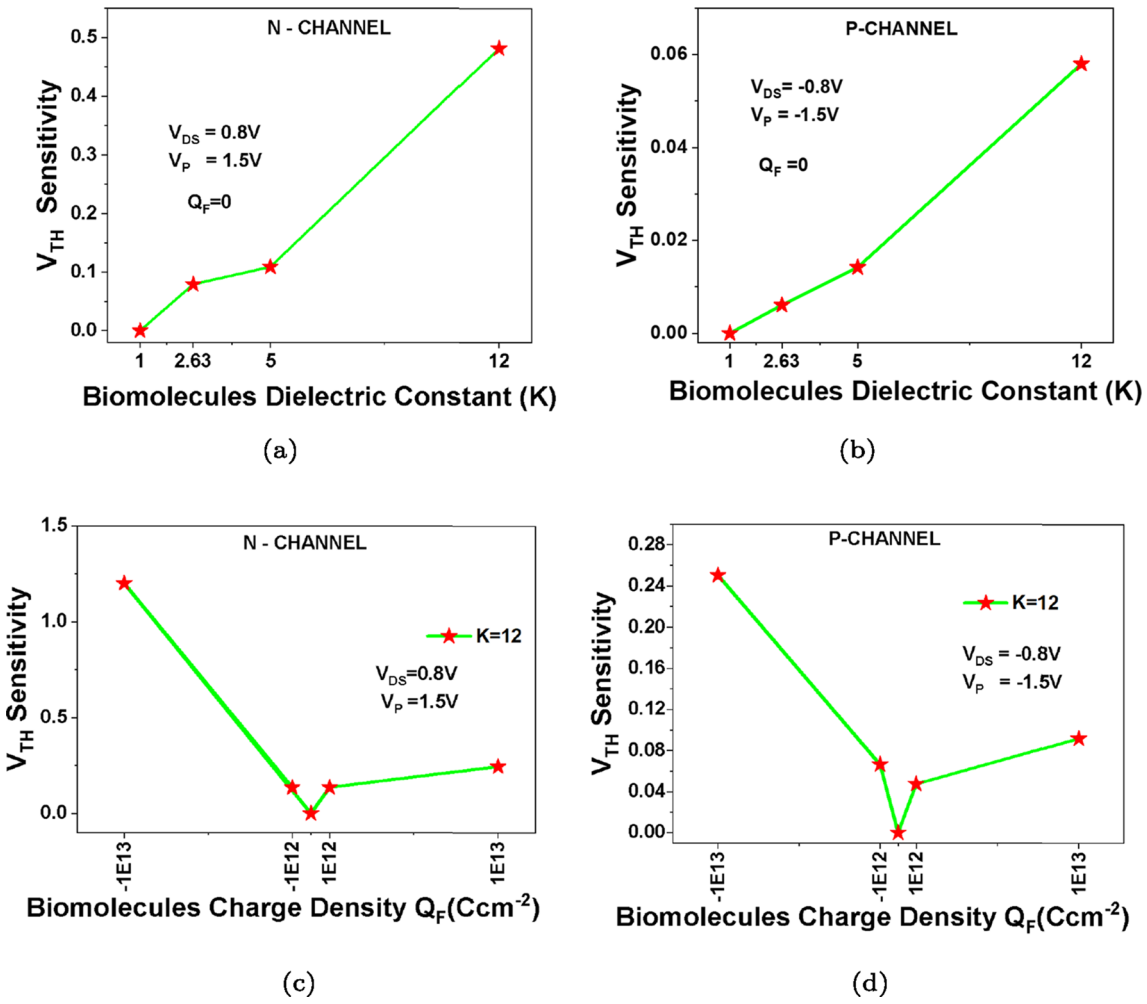


Fig. 6 V_{TH} sensitivity analysis of proposed biosensor for neutral biomolecules with varying dielectric constant (k) when operating in (a) n-mode and (b) p-mode, and with varying charge density of charged biomolecules for (c) n-mode and (d) p-mode

P-type modes. The results are depicted in Fig. 7c and d. As the magnitude of the charge density increased, the SS sensitivity also increased in both N-type and P-type modes. Specifically, in N-type mode, the highest SS sensitivity, reaching 2.87, was observed at Q_F=1E13 cm^{-2}. Conversely, in P-type mode, the maximum SS sensitivity occurred at Q_F= -1E13 cm^{-2}, measuring 0.35. Table 3 illustrates the sensitivity of the proposed biosensor with the reported biosensor.

3.5.3 Drain Current Sensitivity

The drain current sensitivity of the device can be calculated as [23]:

$$I_{ON}\ Sensitivity = |\frac{I_{ON(k=1)} - I_{ON(k>1)}}{I_{ON(k=1)}}| \quad (5)$$

Where $I_{ON(k=1)}$ and $I_{ON(k>1)}$ are the drain current of the device for air and some other biomolecules immobilized in the cavity respectively.

Figure 8 provides a comprehensive analysis of the sensitivity of the proposed device focusing on the impact of variation in the dielectric constant of the biomolecules on the drain current sensitivity. Figure 8(a) and (b) demonstrate that the Ids sensitivity increases as the dielectric constant rises, with the N-Channel configuration showing a higher sensitivity of approximately 2792 compared to the P-Channel configuration, which has a sensitivity of about 968.

3.6 Linearity Performance Analysis

Linearity is a property that indicates the precision of the measured response to a straight line, mathematically represented

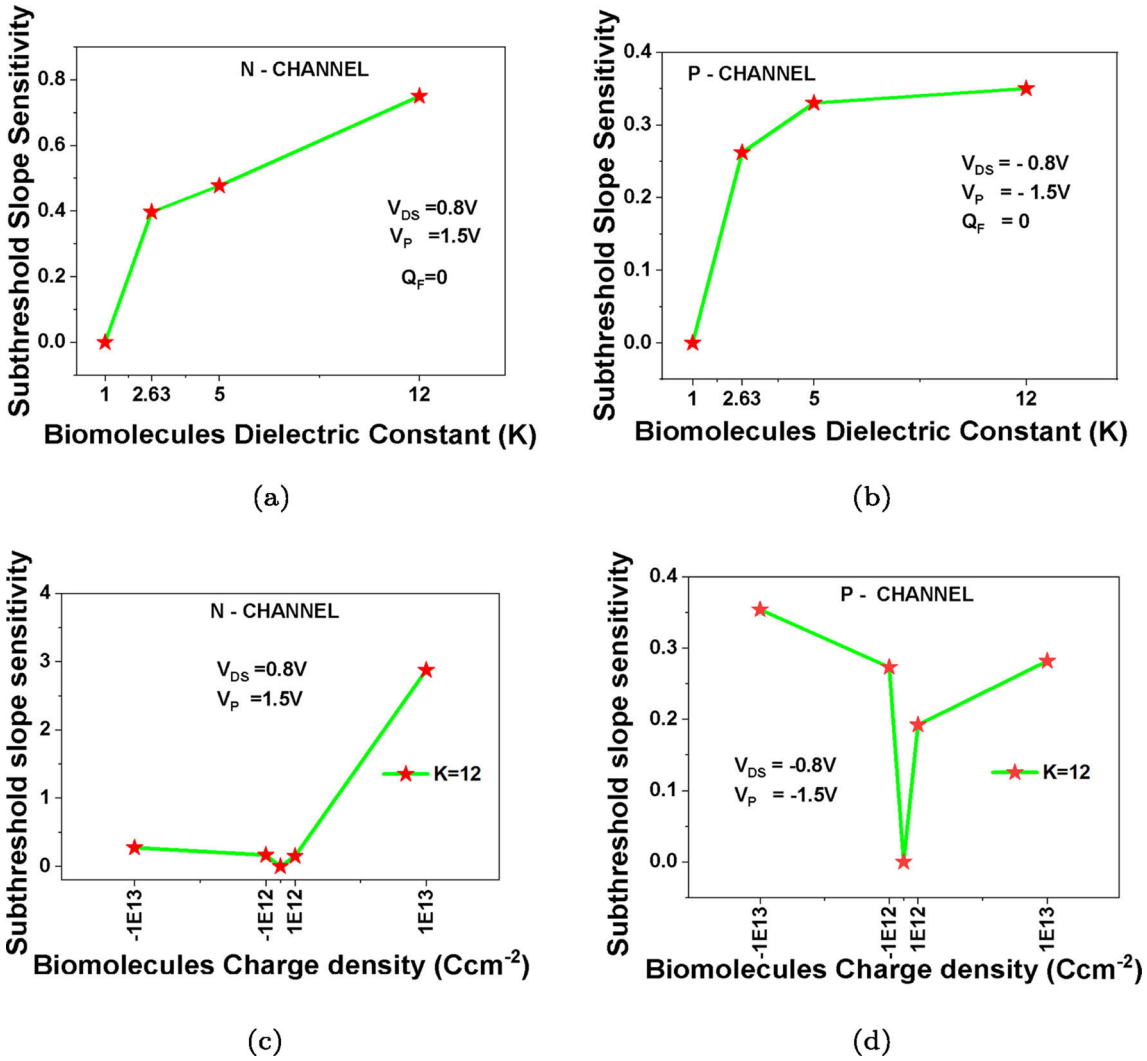


Fig. 7 SS sensitivity analysis of proposed biosensor for neutral biomolecules with varying dielectric constant (k) when operating in (a) n-mode and (b) p-mode, and with varying charge density of charged biomolecules for (c) n-mode and (d) p-mode

as

$$V_{Th} = slope(K > 1) + V_{Th}(K = 1) \tag{6}$$

Table 3 Sensitivity Comparison with the reported Si NW FET-based biosensors

SiNW FET Biosensors	V_{TH} Sensitivity
GAA Si Nanowire FET (K=12) [18]	0.157
NW-FET Biosensor (K=10) [19]	0.070
C-RFET (N-Channel, K=12) [10]	0.010
C-RFET (P-Channel, K=12) [10]	0.050
VS-NW-FET Biosensor (K=10) [19]	0.140
This Work (N-type K=12)	0.480
This Work (P-type K=12)	0.058

Slope(K>1)=slope at the different dielectric constants of the biomolecules. $V_{Th}(K = 1)$=threshold voltage at K=1,i.e., for air.

To investigate the linearity of the proposed device, we have calculated the Pearson coefficient correlation (r) which measures the degree of the linear relationship between two variables (K and VTh). Figure 9(a) and (b) represents the linear relationship between the threshold voltage and the dielectric constant of the biomolecules immobilized in the cavity for the device operating in N-channel and P-channel configurations respectively.

In N-channel configuration, the proposed device exhibits the strong negative correlation between the threshold voltage and the dielectric constant of the biomolecules with a Pearson coefficient of r=0.997 as the threshold voltage decreases with the increase in the value of the dielectric constant. Whereas,

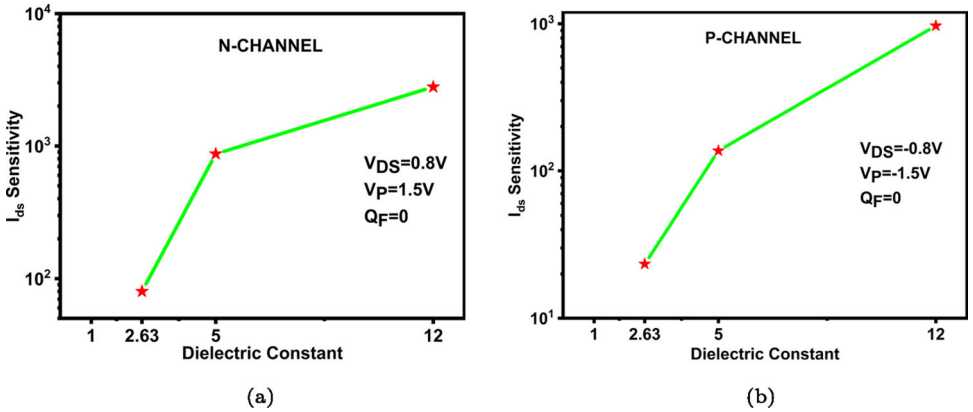


Fig. 8 Drain current sensitivity analysis of proposed biosensor for neutral biomolecules with varying dielectric constant (k) when operating in (a) n-mode and (b) p-mode

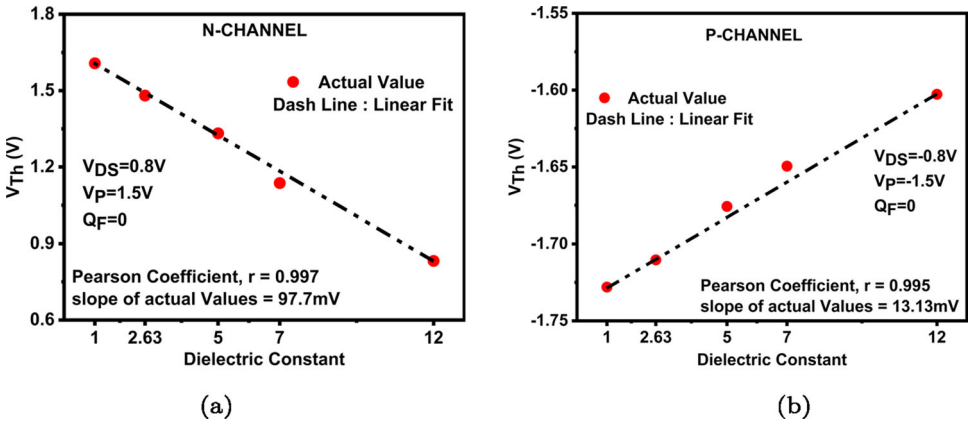


Fig. 9 Linearity analysis of proposed biosensor for neutral biomolecules with varying dielectric constant (k) when operating in (a) n-mode and (b) p-mode

Table 4 Linearity parameter Comparison of the SE R-Si NW SBT biosensor to state-of-art FET-based biosensors

S. No.	FET Based Biosensors	Linearity (Pearson Coefficient)
1.	Reconfigurable FET Biosensor for a Wide Detection Range and Electrostatically Tunable Sensing Response (nCFET) [22]	$r \geq 0.976$
2.	Dual Cavity Dielectric Modulated Ferroelectric Charge Plasma TFET as Biosensor [24]	$r \geq 0.8$
3.	Double Gate 6H Silicon Carbide Schottky Barrier FET [25]	$r \geq 0.98$
4.	Si NW Bio FET [26]	$r \geq 0.957$
5.	This work (N-Channel)	$r \geq 0.997$
6.	This work (P-Channel)	$r \geq 0.995$

in P-channel configuration, a strong positive correlation with the Pearson coefficient of r=0.995 is shown between the threshold voltage of the device and the dielectric constant of the biomolecules. The linearity of the proposed device is compared with the existing state-of-the-art biosensors as shown in the Table 4.

4 Conclusion

This work investigates the performance of UCG-RSiNW SBT as a label-free biosensor for detecting neutral and charged biomolecules. The proposed device outperformed the FET-based biosensors. For biomolecules possessing a neutral charge, the sensitivity of n-mode for V_{TH} is enhanced by 97.91%, while that of p-mode is raised by 16% compared to conventional RFET biosensors. The proposed biosensor offers high sensitivity for charged biomolecules too. For positively charged biomolecules, the biosensor exhibited V_{TH} sensitivity of 0.24 (0.09) and 2.87 (0.28), respectively, for n-mode and p-mode. On the other hand, negatively charged biomolecules showed V_{TH} and SS sensitivity of 1.2 (0.25) and 0.27 (0.35) respectively for n-mode and p-mode. The proposed biosensor demonstrated a high drain current sensitivity of approximately 2792 and 968 in n-mode and P-mode respectively. Similarly, a high degree of linearity, with a Pearson coefficient of -0.997 and 0.995 in N and P-channel configurations, respectively. is exhibited by the proposed biosensor. These observed sensitivities and linearity offer valuable insights into the performance characteristics of the biosensor, which are essential for its potential applications across various biosensing scenarios.

Author Contributions A.K. was responsible for the device's design and conducting simulations. S.K. focused on interpreting the obtained results and preparing figures. V.T. calibrated the results against experimental data and took the lead in crafting the initial version of the manuscript. S.K. and K.R.S. played a pivotal role in strategizing the simulation approach, providing guidance throughout the project, engaging in technical discussions, and refining the manuscript. All authors collaboratively reviewed and gave their approval for the final manuscript.

Funding This work is supported by Science and Engineering Research Board, Department of Science and Technology, Government of India under grant no. EEQ/2021/000831.

Data Availability No datasets were generated or analysed during the current study.

Code Availability Not applicable

Declarations

Competing Interests The authors declare no competing interests.

Ethics Approval The authors declare that they have no known competing interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to participate Not applicable

Consent for publication Not applicable

References

1. Poojari NN, Singh RR, Ahmed N, Kale S (2022) A Comparative Study of Short Gate and Full Gate Dielectric Modulated Schottky Barrier TFET for Biosensing Application. Silicon 14(18):12223–12233
2. Kumar A, Kale S (2023) A Comparative Analysis of Cavity Positions in Charge Plasma based Tunnel FET for Biosensor Application. IETE J Res 1–14
3. Kale S, Kondekar PN (2017) Design and investigation of dielectric engineered dopant segregated Schottky barrier MOSFET with NiSi source/drain. IEEE Trans Electron Devices 64(11):4400–4407
4. Kalra S, Kumar MJ, Dhawan A (2019) Reconfigurable FET biosensor for a wide detection range and electrostatically tunable sensing response. IEEE Sens J 20(5):2261–2269
5. Weber WM, Trommer J, Grube M, Heinzig A, König M, Mikolajick T (2014) Reconfigurable silicon nanowire devices and circuits: Opportunities and challenges. In: 2014 Design, Automation & Test in Europe Conference & Exhibition (DATE). IEEE, pp 1-6
6. Wu P, Reis D, Hu XS, Appenzeller J (2021) Two-dimensional transistors with reconfigurable polarities for secure circuits. Nature Electron 4(1):45–53
7. Bhattacharjee A, Dasgupta S (2018) A compact physics-based surface potential and drain current model for an S/D spacer-based DG-RFET. IEEE Trans Electron Devices 65(2):448–455
8. Biswas A, Rajan C, Samajdar DP (2022) A Novel RFET Sensor for Label-Free Biomolecule Detection. Silicon 14(15):9533–9541
9. Biswas A, Rajan C, Samajdar DP (2022) A Novel HM-HD-RFET Biosensor for Label-Free Biomolecule Detection. J Electron Mater 51(11):6388–6396
10. Saha P, Dash DK, Sarkar SK (2019) Nanowire reconfigurable FET as biosensor: Based on dielectric modulation approach. Solid-State Electron 161:107637
11. Kumar A, Kale S (2023) Spacer-Engineered Reconfigurable Silicon Nanowire Schottky Barrier Transistor as a Label-Free Biosensor. Silicon 1–14
12. Heinzig A, Slesazeck S, Kreupl F, Mikolajick T, Weber WM (2012) Reconfigurable silicon nanowire transistors. Nano Lett 12(1):119–124
13. SILVACO: Tcad device simulation software (2016) Atlas. Santa Clara, CA, USA
14. Kale S, Latha NH, Bramhane LK (2022) Design and Proposal of Double Pocket Schottky Barrier TFET with Dielectric Modulation for Biosensors Applications. Silicon 14(16):10957–10966
15. Latha NH, Kale S (2020) Dielectric modulated Schottky barrier TFET for the application as label-free biosensor. Silicon 12(11):2673–2679
16. Singh R, Kaim S, MedhaShree R, Kumar A, Kale S (2022) Dielectric engineered Schottky barrier MOSFET for biosensor applications: proposal and investigation. Silicon 14(8):4053–4062
17. Kale S (2023) Investigation of dual metal gate Schottky barrier MOSFET for suppression of ambipolar current. IETE J Res 69(1):404–409
18. Ashima, Dhandapani V, Raj B (2023) Design and Performance Assessment of Graded Channel Gate-All-Around Silicon

Nanowire FET for Biosensing Applications. Silicon 15(8):3535–3542
19. Liu FC, Li C, Guo JM, Jiang HF, You HL, Zhuang YQ (2020) A High Sensitivity Biosensor Based On Vertically Stacked Silicon Nanosheet-FET. In 2020 IEEE 15th international conference on solid-state and integrated circuit technology (ICSICT). IEEE, pp 1–3
20. Thakur V, Kumar A, Kale S (2024) Analytical modeling of spacer-engineered reconfigurable silicon nanowire Schottky barrier transistor for biosensing applications. Micro Nanostruc 207799
21. Kumar A, Kale S (2024) Dual-k Reconfigurable Silicon Nanowire Schottky Barrier Transistor for Biosensing Application. Presented at the 8^{th} IEEE Electron Devices Technology and Manufacturing (EDTM) Conference, Bangalore, India
22. Kalra S, Kumar MJ, Dhawan A (2019) Reconfigurable fet biosensor for a wide detection range and electrostatically tunable sensing response. IEEE Sens J 20(5):2261–2269
23. Kumar A, Kale S (2024) Noise and Sensitivity Analysis of the Dielectric Modulated Reconfigurable SiNW-SBT for Biosensor Applications. Micro Nanostruc 207923
24. Singh S, Singh S, Mohammed MK, Wadhwa G (2022) Dual cavity dielectric modulated ferroelectric charge plasma tunnel fet as biosensor: For enhanced sensitivity. IEEE Trans NanoBiosci
25. Rashid S, Bashir F, Khanday FA, Beigh MR (2022) Double gate 6H silicon carbide schottky barrier fet as dielectrically modulated label free biosensor. Silicon 1–12
26. Zhao W, Hu J, Liu J, Li X, Sun S, Luan X, Zhao Y, Wei S, Li M, Zhang Q et al (2022) Si nanowire bio-fet for electrical and label-free detection of cancer cell-derived exosomes. Microsys Nanoeng 8(1):57

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.

2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

Simulation-Based Evaluation of IDMT Characteristics for Overcurrent Relay Protection

Utkarsh Rishu
Electrical Engineering Department
Delhi Technological University
utkarshrishu.24@gmail.com

Anup Kumar Mandpura
Electrical Engineering Department
Delhi Technological University
kanup@dtu.ac.in

Rohan Pillai
Electrical Engineering Department
Delhi Technological University
Rohan92@dtu.ac.in

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10625963

***Abstract*—- In the realm of power system protection, the identified problem of reliable protection scheme underscores a crucial void in the design and execution of Inverse Definite Minimum Time (IDMT) overcurrent protection mechanisms. This paper highlights a critical gap in the design and implementation of IDMT overcurrent protection within power systems. Despite advancements in power system protection research and education, there is a lack of a systematic approach to designing and implementing numerical IDMT overcurrent relay characteristics. The absence of standardized methodologies and tools poses challenges to the reliability and effectiveness of protection schemes. The research aims to address this gap by developing and validating IDMT overcurrent relay characteristics following IEC 60255 standards. The objectives include creating a systematic methodology, simulating the designed characteristics using MATLAB/Simulink, and evaluating the performance in terms of speed, selectivity, and reliability. The software implementation involves a detailed Simulink model of an electrical system with an IDMT protection mechanism. The research emphasizes the need for a robust methodology to enhance the understanding and optimization of IDMT overcurrent protection in electrical power systems.**

***Index Terms*—IDMT Relay, MATLAB/Simulink, Inverse Time Characteristics, IEC Standard, Power System Protection**

I. INTRODUCTION

The Inverse Definite Minimum Time (IDMT) Protection Scheme is a crucial component in safeguarding power systems against faults and abnormalities. Operating on an inverse relationship between tripping time and fault current magnitude, IDMT protection relies on advanced numerical relays with programmable capabilities for precise configuration. This protection scheme effectively responds to diverse fault scenarios, enhancing overall power system reliability. The fundamentals of IDMT protection include current-dependent operation, characteristic curves, Plug Setting Multiplier (PSM), and Time Multiplier Setting (TMS). Paired with numerical relays, IDMT protection offers cost-effective and reliable solutions for over currents and short circuits in various power system components. The evolution of relay technology, from mechanical relays to smart relays and Intelligent Electronic Devices (IEDs), reflects the industry's need for faster, more dependable, and versatile switching solutions. However, deploying IDMT protection requires careful consideration of challenges, including coordination with other devices, sensitivity-stability balance, fault type discrimination, integration into communication networks, reliability and maintenance, cybersecurity, standard compliance, personnel training, and adaptability to system changes and upgrades. Successful deployment of the IDMT Protection Scheme contributes significantly to grid stability and the uninterrupted supply of electricity.

Power systems are vulnerable to various faults and abnormalities that can jeopardize the stability and reliability of the electrical grid. Safeguarding the equipment and ensuring the uninterrupted supply of electricity, protective relays play a crucial role in detecting and isolating faults. One such advanced protection scheme is the IDMT Protection Scheme. The primary objective of IDMT protection is to discriminate between normal operating conditions and abnormal situations, such as short circuits or overloads, and to initiate a quick and appropriate response to mitigate potential damage. In contrast to conventional protection techniques, IDMT protection is based on a characteristic curve that shows a negative correlation between operational time and fault current amplitude. When it comes to managing diverse types of faults, the IDMT protection method shines [1-3]. The protective relay's running duration reduces inversely with increasing fault current, enabling rapid reaction to serious faults while tolerating less severe irregularities to some extent.

Key components of the IDMT protection scheme include numerical relays, which utilize advanced algorithms and digital signal processing to measure and analyze electrical quantities accurately [4-5]. These relays are programmable, allowing for flexible and precise configuration based on specific system requirements. This protection scheme is widely employed in power distribution and transmission systems, as it balances sensitivity and selectivity. By employing IDMT protection with numerical relays, utilities can enhance the overall reliability of their power systems and reduce downtime associated with faults, contributing to the stability of the electrical grid [6-7]. Relay technology has evolved significantly since its inception in the early 19th century. Originally developed for telegraphy, relays have found applications in various fields including telecommunications, power systems, automation, and computing. IDMT relays are commonly used to protect against overcurrents and short circuits in power systems, such as feeders, transformers, motors, and distribution networks. IDMT protection is cost-effective and widely used for its simplicity and reliability. It provides coordination and selectivity in the protection scheme, enabling proper fault isolation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:57 UTC from IEEE Xplore. Restrictions apply.

II. FUNDAMENTALS OF INVERSE DEFINITE MINIMUM TIME (IDMT) RELAY

A protective relay known as IDMT protection, is used in power systems to identify and address unusual operating circumstances, such as over currents. Tripping time is inversely related to the size of the fault current; this is the concept upon which the IDMT protection relay functions. Higher fault currents result in faster tripping times, providing faster and more selective protection. The power system or protected equipment's current magnitude determines how IDMT protection relays work. The relay measures the current and calculates the operating time based on a specific characteristic curve [8-9].

For an IDMT relay, the characteristic curve shows how the amplitude of the fault current relates to the tripping time. The typical form of this curve is an inverse-time curve. The curve is obtained by parameters such as PSM and TMS [10-11] whose equation is given below:

$$t = \frac{K}{(\frac{I_f}{I_p})^a - 1} * TMS \tag{1}$$

where

- t = relay operation time in seconds
- k = constant for relay characteristics
- I_p = relay current setting value
- a = constant for representation of inverse time type
- I_f = actual secondary current value

PSM is a ratio that defines the multiple of the relay's nominal current setting at which it would operate. PSM is used to adjust the sensitivity of the relay. A lower PSM makes the relay more sensitive to lower fault currents. The characteristic curve is typically inverse, meaning that as the fault current increases, the tripping time decreases. The inverse relationship ensures faster tripping for higher fault currents, improving system selectivity and coordination [12-15]. IDMT relays are commonly used to protect against over currents and short circuits in power systems, such as feeders, transformers, motors, and distribution networks.

III. MODELLING OF IDMT RELAY

The MATLAB Sim Power System suite of tools is used for modelling of the relay. An overcurrent relay using IDMT capabilities is included in Simulink's extensive model. A three-phase electrical system delivering electricity to a load is shown in Fig.1. of the Simulink data. Table I shows the system parameters.

The three-phase IDMT over current protection tripping mechanism is shown in Fig.2. of the MATLAB/simulink model. The control mechanism has an output that is a logic 0 trip signal in the non-fault state and a logic 1 signal in the fault condition. After processing all the phase-specific trip signals, the OR block sends a Trip/no trip signal to the circuit breaker. Fig.2. shows that one of the phases is controlled by the input current during the relevant time. The current is converted to an RMS value over a 20-millisecond period by using the root-mean-square (RMS) block. Fixed threshold/pickup value is one of the two inputs of first relational operator. Ip is 2.6 A under these conditions.

TABLE I. SYSTEM PARAMETERS.

Sr.No.	Description	Values
1.	Source Voltage	415 V
2.	Source Resistance	0.00085485 Ohm
3.	Source Inductance	15.88 µH
4.	X/R ratio	5.833

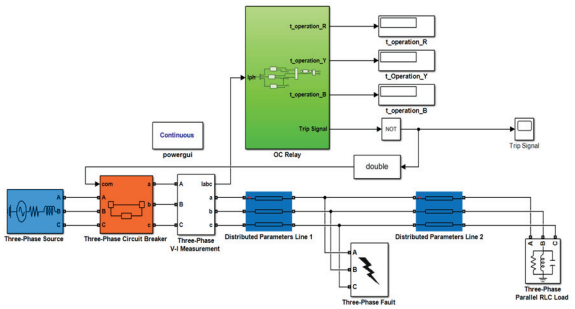


Fig. 1. Simulink model for IDMT protection

IV. SOFTWARE IMPLEMENTATION

The IDMT Overcurrent relay is guaranteed to operate after operating time 't' according to the IDMT characteristics that have been specified by second relational operator. For other lines, it uses a tripping mechanism that is similar. An OR block receives the output from all three control mechanisms and, in the event of a malfunction on any of the three lines, sends out a trip signal. An error situation occurs when the line current exceeds Ip, and the relational operator's output is 1.

- The IDMT characteristics' typical operating time might be determined via equation (1). Since the system is not interconnected hence Time Multiplier Setting (TMS) is equal to 1.
- The function

$$(\frac{I_f}{I_p})^a - 1 \tag{2}$$

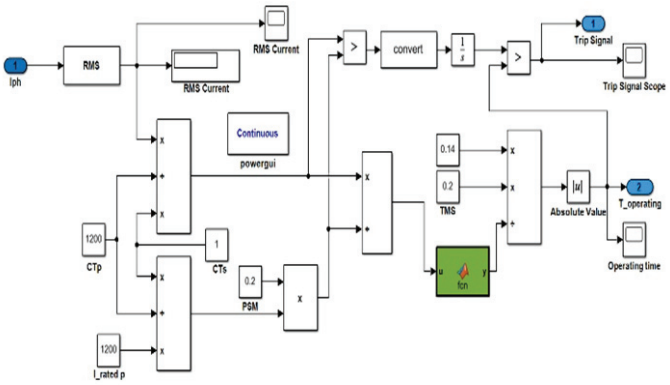


Fig. 2. Subsystem of Tripping Mechanism block (IDMT relay)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:57 UTC from IEEE Xplore. Restrictions apply.

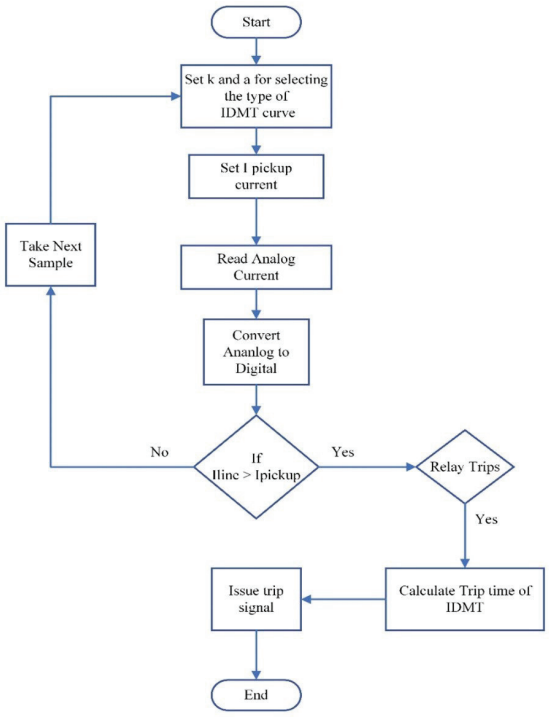


Fig. 3. Schematic of IDMT relay

is executed by a function block that is provided by the ratio

$$\frac{I_f}{I_p} \tag{3}$$

. The IDMT features are implemented via the dividing block.

- The IDMT overcurrent relay is implemented, as shown in Fig.3., which is a flow diagram.

V. SIMULATION RESULTS

The output waveforms for different relays and the results for the many IDMT attributes in the Simulink model are shown in Fig.4. to Fig.11. The signal for the circuit breaker to trigger is logic 1. Different locations and resistances of faults were used to alter the fault current value. The findings centered on the relay's operational duration which is being shown in Table II.

A. Normal Inverse Time Characteristics

Case-1: The result for Normal Inverse Time Characteristics; fault current I_f is obtained. Furthermore, 3.54 A and 11.05 s were determined to be the fault current and operational times, successively.

Case-2: The results for this Time Characteristics in second case, fault current I_f and operating time T_{op} are shown in Fig.5. The relevant values are 8.14 A and 5.94 s.

Relay trips to 1.0 form 0.0 as shown in Fig.4. and Fig.5. in the time frame of 11.05s and 5.94s respectively.

B. Very Inverse Time Characteristics

Case-1: Fault current I_f was determined to be 5.01 A and operating time T_{op} to be 16.25 s, as shown in Fig.6.

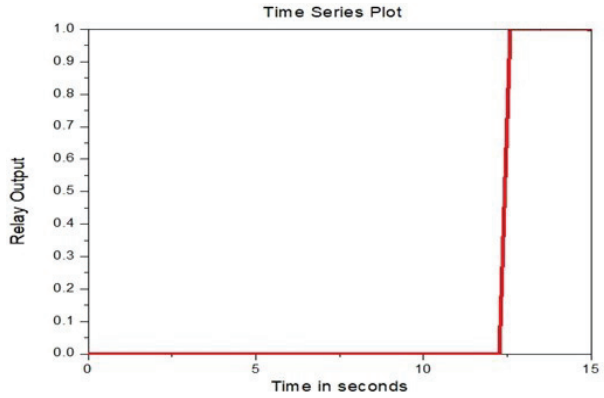


Fig. 4. Output for normal time inverse characteristics (Case 1)

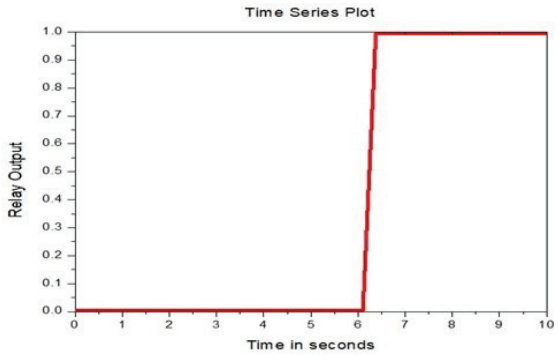


Fig. 5. Output for normal time inverse characteristics (Case 2)

Case-2: For Very Inverse Time Characteristics, the Operating time T_{op} is shown in the Figure 7 which is 5.68s and the fault current I_f is 9.01 A for the second case.

Relay trips to 1.0 form 0.0 as shown in Fig.6. and Fig.7. with operating time being 16.25s and 5.68s respectively.

VI. C. EXTREME INVERSE TIME CHARACTERISTICS

Case-1: Fig.8 displays the results for Extreme Inverse Time Characteristics, which were 5.45 A for fault current I_f and 40.67 s for operating time T_{op}.

Case-2: The operating T_{op} is 7.98 s as shown in the Fig.9. in second case of Extreme Inverse Time Characteristics, whereas fault current I_f for this operating time is 9.21 A.

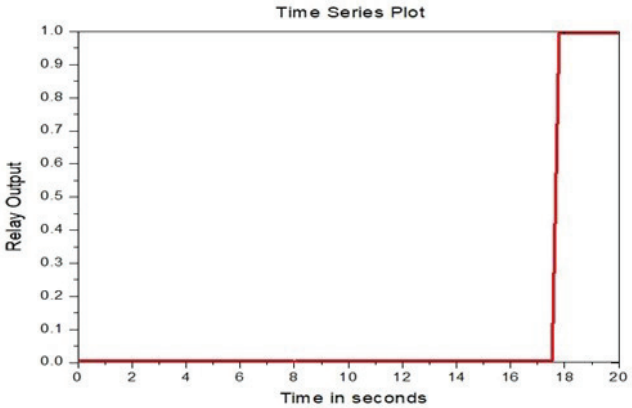


Fig. 6. Output for very inverse time characteristics (Case 1)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:57 UTC from IEEE Xplore. Restrictions apply.

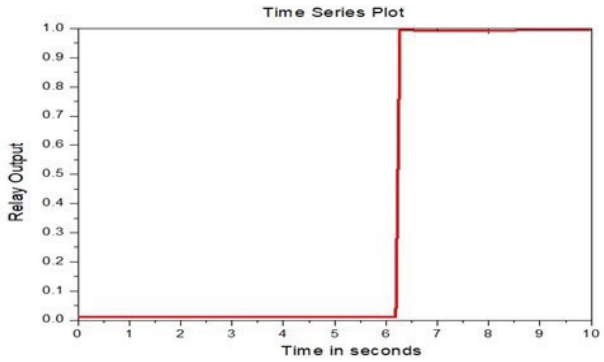


Fig. 7. Output for very inverse time characteristics (Case 2)

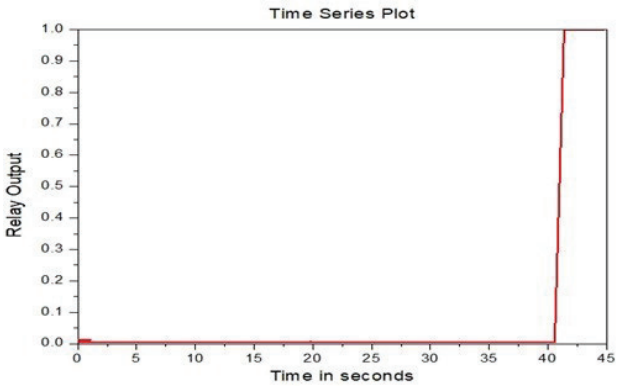


Fig. 8. Extreme inverse time characteristics: operating time (Case 1)

Here also, Relay trips to 1.0 form 0.0 with operating time being 40.67s and 7.98s as shown in Fig.8. and Fig.9. respectively.

VII. D. LONG INVERSE TIME CHARACTERISTICS

Case-1: In case of Long Inverse Time Characteristics, T_{op} or operating time obtained was 140 s, for which the fault current, comes out to be 5.98 A. The same is shown in Fig.10.

Case-2: A fault current of 8.56 A and an operating duration of 56.16 s were determined, as shown in Fig.11., for the Long Inverse duration Characteristics.

Tripping of the relay is obtained as shown in Fig.10. and Fig.11. for Long Inverse Time Characteristics where plot goes from 0.0 to 1.0 in the operating time of 140s and 56.16s respectively.

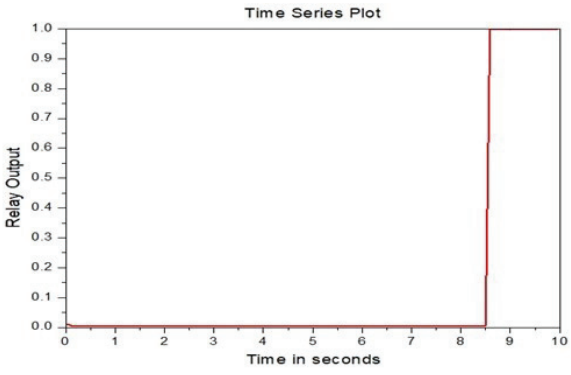


Fig. 9. Output for extreme inverse time characteristics (Case 2)

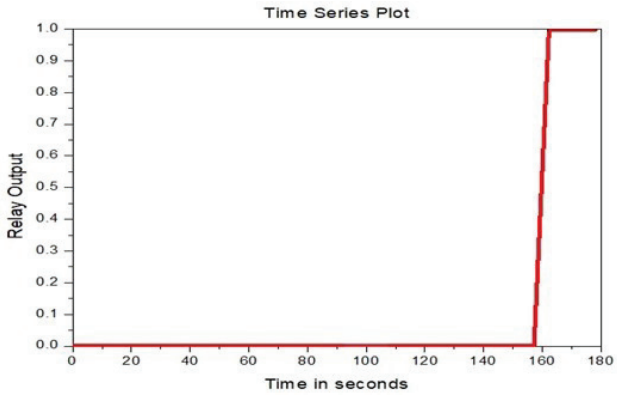


Fig. 10. Output for long inverse time characteristics (Case 1)

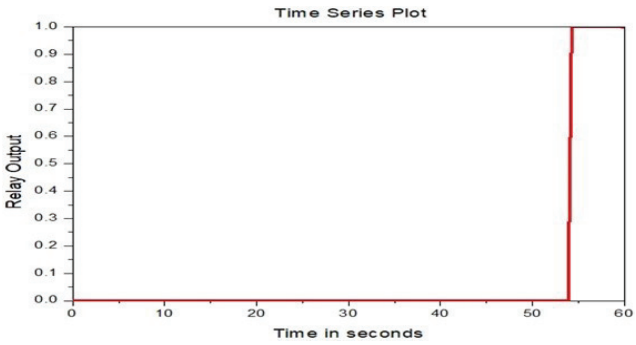


Fig. 11. Operating time for long inverse time characteristics (Case 2)

The graphical representation of the simulation data obtained is shown below in the Fig.12. which depicts the behaviour of the relay model as per the IEC standard of the protection scheme, that is trip time will be shorter for higher value of current multiplier. This proves out to be very helpful while designing the protection equipment for sever overloading conditions.

VIII. CONCLUSION AND FUTURE WORK

The simulation results presented in the content validate the effectiveness of the IDMT Protection Scheme in various fault scenarios. The results demonstrate the scheme's ability to respond to different fault currents with appropriate operating times, showcasing its versatility and reliability in realworld applications. Overall, the IDMT Protection Scheme, when carefully deployed and maintained, plays a vital role in enhancing the resilience and performance of power systems, ultimately contributing to a more robust and stable electrical grid. However, the successful deployment of the IDMT Protection Scheme comes with its set of challenges and considerations. These include coordination with other devices, sensitivity-stability balance, fault type discrimination, integration into communication networks, reliability and maintenance, cybersecurity, standard compliance, personnel training, and adaptability to system changes and upgrades. The future scope for the IDMT Protection Scheme envisions several promising avenues for advancement in power system reliability. Integration with AI holds the potential for enhancing decision-making, making the protection system.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:57 UTC from IEEE Xplore. Restrictions apply.

TABLE II. OBTAINED SIMULATION RESULT FOR DIFFERENT CHARACTERISTICS.

Sr. No	Name of the characteristics	Fault Current (I_f) in ampere	Operating Time (T_{op}) in seconds
1	Normal Inverse Characteristics	3.54 A	11.05 s
		8.14 A	5.94 s
2	Very Inverse Characteristics	5.01 A	16.25 s
		9.01 A	5.68 s
3	Extreme Inverse Characteristics	5.45 A	40.67 s
		9.21 A	7.98 s
4	Long Inverse Characteristics	5.98 A	140 s
		8.56 A	56.16 s

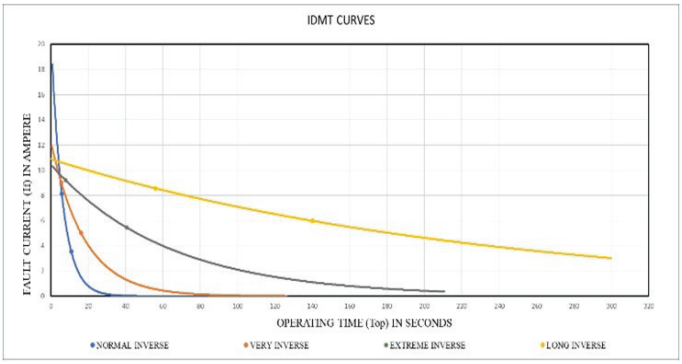


Fig. 12. Graphical Representation of IDMT Curves

more adaptive and self-learning. Cybersecurity measures need to evolve to safeguard IDMT relays from emerging digital threats. Advances in sensor technologies could improve fault detection and isolation, contributing to the scheme's sensitivity and responsiveness. Wide-area protection and communication, coupled with machine learning for fault pattern recognition, can foster faster and more coordinated responses across geographically distributed power systems.

REFERENCES

[1] Balasubramanian, Sujatha, J. Preetha Roselyn, and K. S. Srividya. "Adaptive Central Control Scheme and Investigation of Fault Current Contribution for Protection of Microgrid." In Journal of Physics: Conference Series, vol. 2335, no. 1, p. 012048. IOP Publishing, 2022.

[2] Tiwari, Ravi Shankar, and Om Hari Gupta. "Study of Combined Time & Current Grading Protection Scheme for Distribution System." In 2020 International Conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC), pp. 443-448. IEEE, 2020.

[3] Moshoeshoe, M., and S. Chowdhury. "Adaptive Overcurrent Protection Scheme with Zonal Approach for a Stand-Alone Renewable-based Microgrid." In 2021 IEEE PES/IAS PowerAfrica, pp. 1-5. IEEE, 2021.

[4] Noman, M., I. Ullah, M. A. Khan, A. Qazi, W. Farooq, A. Saqr, and A. Elsheikh. "Analysis of overcurrent protective relaying as minimum adopted fault protection for small-scale hydropower plants." International Journal of Environmental Science and Technology 21, no. 4 (2024): 4457-4470.

[5] Balasubramanian, Sujatha, J. Preetha Roselyn, and K. S. Srividya. "Adaptive Central Control Scheme and Investigation of Fault Current Contribution for Protection of Microgrid." In Journal of Physics: Conference Series, vol. 2335, no. 1, p. 012048. IOP Publishing, 2022.

[6] Tiwari, Ravi Shankar, and Om Hari Gupta. "Study of Combined Time & Current Grading Protection Scheme for Distribution System." In 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), pp. 443-448. IEEE, 2020.

[7] Noman, M., I. Ullah, M. A. Khan, A. Qazi, W. Farooq, A. Saqr, and A. Elsheikh. "Analysis of overcurrent protective relaying as minimum adopted fault protection for small-scale hydropower plants." International Journal of Environmental Science and Technology 21, no. 4 (2024): 4457-4470.

[8] Babu, G. Suresh, Sivasankaran Venugopal, and T. Murali Krishna. "Inverse Definite Minimum Time Relay Coordination in Radial Traction System." International Journal of Electrical Engineering and Technology 11, no. 4 (2020).

[9] Schreiber, Patrick, Tobias Boltz, Miriam Brosi, Bastian Haerer, Akira Mochihashi, A. I. Papash, Robert Ruprecht, Marcel Schuh, and AnkeSusanne Muller. "Effect of Negative Momentum Compaction Operation¨ on the Current-Dependent Bunch Length." In 12th International Particle Accelerator Conference: virtual edition, May 24th-28th, 2021, Brazil: proceedings volume/IPAC2021. Ed.: R. Picoreti, p. 2786. 2021.

[10] Mahmood, Jawad, and Ramzy Ali. "Pc/plc based simultaneous running of more than one laboratory experiments using wxpython and pymodbus." In Proceedings of 2nd International Multi-Disciplinary Conference Theme: Integrated Sciences and Technologies, IMDC-IST 2021, 7-9 September 2021, Sakarya, Turkey. 2022.

[11] Alaee, Pegah, and Turaj Amraee. "Optimal coordination of directional overcurrent relays in meshed active distribution network using imperialistic competition algorithm." Journal of modern power systems and clean energy 9, no. 2 (2020): 416-422.

[12] Saldarriaga-Zuluaga, S. D., J. M. Lopez-Lezama, and N. Mu´ noz-˜ Galeano. "Protection coordination considering different characteristic curves and a variable limit of the plug setting multiplier." Scientia et Technica 27, no. 3 (2022).

[13] Saldarriaga-Zuluaga, Sergio D., Jesus M. L´ opez-Lezama, and Nicol´ as´ Munoz-Galeano. "Adaptive protection coordination scheme in micro-˜ grids using directional over-current relays with non-standard characteristics." Heliyon 7, no. 4 (2021).

[14] Horsfall, D. J., D. C. Idoniboyeobu, C. O. Ahiakwo, and S. L. Braide. "Relay Coordination Analysis in SPDC Forcados Distribution Network for Operation, Planning and Future Expansion." Journal of Newviews in Engineering and Technology (JNET) 3, no. 4 (2021).

[15] Chang, Nana, Guobing Song, Junjie Hou, and Zhongxue Chang. "Fault identification method based on unified inverse-time characteristic equation for distribution network." International Journal of Electrical Power & Energy Systems 146 (2023): 108734.

[16] Selimaj, Antigona, Immanuel Hacker, Ronja Steinfurth, and Andreas Ulbig. "Optimal Time Coordination of Definite-Time Overcurrent Relays for Adaptive Protection Systems." In 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), pp. 1-5. IEEE, 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:55:57 UTC from IEEE Xplore. Restrictions apply.

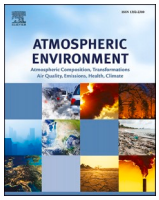
Atmospheric Environment 337 (2024) 120773



Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv



Size resolved particle contribution to vehicle induced ultrafine particle number concentration in a metropolitan curbside region

Kanagaraj Rajagopal [a], S. Ramachandran [b], Rajeev Kumar Mishra [a,*]

[a] Department of Environmental Engineering, Delhi Technological University, Delhi, India
[b] Space and Atmospheric Sciences Division, Physical Research Laboratory, Ahmedabad, India

HIGHLIGHTS

- Seasonal nanoparticle number concentration (PNC) in Delhi varies from 10^4 to 10^6 cm^{-3}.
- Particles in nucleation mode contribute >30% to PNC in summer, spring and monsoon.
- Accumulation mode particles contribute >35% to PNC in winter and autumn.
- Higher relative humidity favors coagulation resulting in higher concentration of N_{acc}.
- Correlation between N_{acc} and Carbon Monoxide is higher due to transport emissions.

ARTICLE INFO

Keywords:
Particle number concentration
Seasonal variation
Transport emissions
Urban pollution and meteorology
Implications

ABSTRACT

The concentrations and behavior of nano particles (10–1000 nm) in Delhi, a densely populated megacity, in different seasons (winter, spring, summer, monsoon, and autumn) are examined, for the first time. The concentration of particles is classified into four different sizes as N_{nuc} (10–30 nm, nucleation), N_{satk} (30–50 nm, small Aitken), N_{latk} (50–100 nm, large Aitken), and N_{acc} (100–1000 nm, accumulation mode), and the total (10–1000 nm) particle number concentration (PNC) as N_{total}. PNC ranges between 10^4 cm^{-3} and 10^6 cm^{-3} over Delhi during the year, and the highest concentration occurs in winter. Winter concentration is 2, 1.6 and 1.3 times higher than monsoon, summer, autumn and spring concentrations, respectively. N_{nuc}, N_{satk}, N_{latk} and N_{acc} and their respective contributions to total PNC exhibit significant seasonal variations. During winter N_{latk} and N_{acc} contribute more to total due to coagulation, with N_{acc} alone contributing >40% to total PNC. N_{nuc}, N_{satk}, and N_{latk} are higher in spring and summer during mid-day due to nucleation and/or ultrafine particle burst events. The direct primary emissions from engine exhaust produce a prominent double hump structure during morning and evening peak hours in winter and autumn. PNC and their contributions exhibit day-night variations as they are influenced by variations in emission sources and meteorological parameters (wind speed, relative humidity, temperature, solar radiation and boundary layer height) between day and night. Carbon monoxide correlates positively with N_{acc} in all seasons ($R^2 \geq 0.5$) as fossil fuel emission is a predominant source for gases and particles in the study environment. These quantitative results on seasonal variations of nano particles and air pollutants together with the knowledge on seasonal variations in meteorological parameters and atmospheric dynamics provide a foundation which can positively contribute better to the urban planning and devising mitigation measures aimed at improving air quality and public health.

1. Introduction

The air quality of any region depends on the quantity and quality of different particles present in the atmosphere (Yadav et al., 2021; Zhao et al., 2020). These particles in the atmosphere can cause several health impacts, including cardiovascular and pulmonary diseases (Kanawade et al., 2020; Kaur et al., 2005; Sharma et al., 2024; Schraufnagel, 2020). The particles present in the air vary based on size, shape, and composition, and their potential to create health impact also depends on these characteristics (Gao et al., 2015; Gurjar et al., 2010; Kumar et al., 2011;

* Corresponding author.
E-mail address: rajeevkumarmishra@dtu.ac.in (R.K. Mishra).

https://doi.org/10.1016/j.atmosenv.2024.120773
Received 7 May 2024; Received in revised form 15 August 2024; Accepted 21 August 2024
Available online 23 August 2024
1352-2310/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

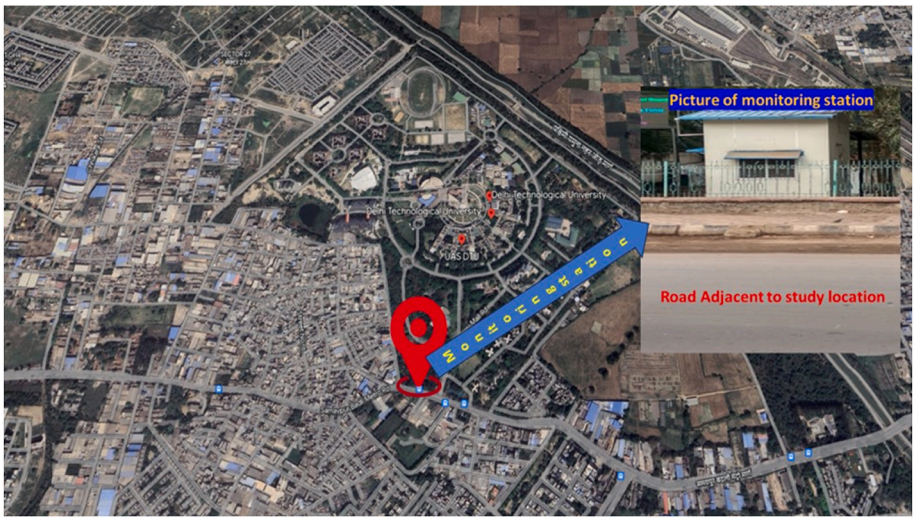


Fig. 1. Geographical location of the monitoring station with the picture of the monitoring station.

Ohlwein et al., 2019) Apart from particulate matter (PM), such as $PM_{2.5}$ and PM_{10}, particles of smaller size also are found/emitted in the atmosphere (Agarwal et al., 2020; Manojkumar and Srimuruganandam, 2021a,b). The role of smaller size particles in health and climate impacts is more crucial and higher than that of particulate matter (Liu et al., 2022; Mei et al., 2022; Wiebert et al., 2006). The understanding of these small size particles, especially the ultrafine particles, is less than that of traditional pollutants (Mullen et al., 2011; Quiros et al., 2013; Riesenfeld et al., 2000). This emphasizes the need to study these particles to understand their nature and characteristics, which will help initiate policy formulation and mitigation measures to reduce these harmful particle emissions(Khan et al., 2022).

The smaller size particles are measured in number concentration (Bhardawaj et al., 2017), whereas the larger particles are characterized by their mass concentration (Hussein et al., 2004; Wu et al., 2008). These small size particles are often categorized into three major size ranges: nucleation mode (N_{nuc} - 10 to 30 nm in diameter), Aitken mode (30–100 nm), and accumulation mode (100–1000 nm). The ultrafine particles (UFP) include particles in the nucleation and Aitken modes (10–100 nm) (Rajagopal et al., 2023). Since the study region is a curbside microenvironment where traffic sources play a major role, for a better interpretation, understanding and representation, the Aitken mode is further divided into two as small Aitken mode (N_{satk} - 30 to 50 nm) and large Aitken mode (N_{latk} - 50 to 100 nm). The Aitken mode particles in the atmosphere may condense and grow rapidly due to condensation of vapors present in the atmosphere (Griffin, 2013). This above classification of Aitken mode can help in determining the type of emissions as either freshly emitted traffic emissions or aged traffic emissions (Charron and Harrison, 2003). The growth of these particles can lead to the formation of accumulation mode particles with the existing particles in the atmosphere. The Aitken mode particles plays a crucial role in the formation of secondary particles such as cloud condensation nuclei (Pohlker et al., 2021). The atmospheric pollution can either inhibit or enhance the role of Aitken mode particles (Noble and Hudson, 2019), therefore, a better understanding of the Aitken mode particles is necessary not only for scientific purpose but also for mitigation. With these objectives, the Aitken mode particles are classified into two different modes. In recent years, global studies showed that ultrafine particles dominate the total number concentration (N_{total}), and their contribution can be up to 90 % in the total particle number concentration (PNC) (Kumar et al., 2008). The modes that contribute to mass and number concentrations in the nanoparticle size distribution vary. Further, their number concentrations need not necessarily follow the spatial and temporal dynamics as their mass concentration (Reche et al., 2011). The accumulation mode particles in an urban environment arise mostly from emissions from various anthropogenic sources such as traffic, industrial processes, biomass burning, and gas to aerosol processing (Gani et al., 2020; Ramachandran and Rupakheti, 2021). However, the UFP is generated mostly from traffic (Zhu et al., 2002), cooking, and from new particle formation events (Brines et al., 2015; Hofman et al., 2016). The dynamics of ultrafine particles in the atmosphere is not uniform and they exhibit variations depending on emission sources which vary as a function of season (Yadav et al., 2021). In an urban curbside environment, particle number concentration and dispersion depend on wind speed, direction, relative humidity (RH), and temperature (Kumar et al., 2008). Meteorology and seasonal factors, along with the local emission sources and other pollutants, influence the concentration of these particles in a particular region (Gómez-Moreno et al., 2011; Kozawa et al., 2012).

The study location Delhi, is one of the superiorly large highly populated and urbanized cities with a dense vehicle population in the roadside environment (Mohan et al., 2024a)The annual average pollutant concentration exceeds the country's recommended levels, and further, Delhi experiences nearly twice the particle number concentration than other cities in high-income countries (Gani et al., 2019; Mohan et al., 2024b). The region experiences varying climatic conditions, from hot summers (33.4 ± 5.2 °C) to cold winters (19 ± 4.5 °C), rainfall/precipitation during the monsoon season, and frequent pollution events. The study is based on the analysis of particles measured in the size range of 10–1000 nm over a year comprising different seasons. The objectives of the study include (i) investigating the seasonal profile of urban ultrafine particle emissions from the transportation sector in curbside environment, (ii) estimating the role of regional meteorology (wind, relative humidity, and boundary layer height) and trace gases (SOx, NOx, CO, O_3) in determining the concentration of these particles, and (iii) analyze the seasonal variation in size distribution and number concentration during the day and night when the magnitude of emissions are expected to vary depending on their sources. The results from the present study can help obtaining better insights into the diurnal and seasonal behavior of nanoparticles in urban curbside environments and can help understand the particle dynamics and behavior in similar other polluted urban cities. The results on particle concentration during different emission scenarios and seasons can further help in developing better mitigation strategies aimed at reducing the emission sources.

2. Monitoring site, period, instrumentation, and data analysis

The monitoring site is located northwest of Delhi, near a major

Table 1
Details of the instruments used to measure nanoparticle concentrations, gaseous pollutants, and meteorological parameters along with the country of production, detection limits, resolution, sensitivity, and accuracy of each instrument.

Sr. no	Parameters/ Instrument	Specifications
1.	Nanoparticles 10–1090 nm GRIMM (Germany)	Slip factor - 2.83 DMA length – 0.35 m, diameter – 0.040 m (inner), 0.026 m (outer). Optimum DMA temp – 20 °C Optimum DMA pressure - 101300 Pa Sampling port effective length – 0.38 m
2.	CO Monitor Horiba Ltd (Japan)	Limit of detection 0.05 ppm, zero drift ±0.1 ppm/ 24hr, linearity±1.0%, minimum interval-3 minutes
3.	SO_2 monitor Horiba Ltd (Japan)	Limit of detection 0.05 ppb, zero drift ±0.1 ppb/ 24hr, linearity±1.0%, minimum interval-5 minutes.
4.	$NO/NO_2/NO_X$ 2B technologies (USA)	Limit of detection <1 ppb, Nominal Flowrate 1.5 L/min, Resolution 0.1 ppb
5.	Ozone 2B technologies (USA)	Limit of detection <1.5 ppb, Nominal Flowrate 1.5 L/min, Resolution 0.1 ppb
6.	Meteorological parameters Aimil (India)	Accuracy better than + 5%, Sensitivity- 0.5%

Table 2
Seasonal classification (name and time frame), total number of days in each season, number of days for which data on pollutant characteristics are available and used in the present study.

Seasons	Time period	Total number of days in each season	Number of days for which data are available and used
Winter	December to mid-February	76	69
Spring	Mid-February to March	45	44
Summer	April to June	91	84
Monsoon	July to mid-September	77	41
Autumn	Mid-September to November	76	60

roadway curbside region. The road is identified as a major state highway with a high-density vehicle population of around 1300 vehicles/hour and is adjacent to the boundary of the Delhi Technological University campus. The measurements are done through the monitoring station established by Advance Air and Acoustics Research Laboratory (Fig. 1). The study region represents an urban setting dotted with a mixture of commercial, industrial, and educational establishments. The measurements are done in the curbside region, where the majority of emissions in highly polluted urban regoions are expected to be from the transportation sector, and residential areas located adjacent to the curbside environment.

GRIMM Aerosol Technik SMPS + C model 5403 (Sr. no. 54031801) is used to measure the atmospheric pollutants in the 10 nm–1090 nm size range at a scan interval of about 7 min. The measured particles are distributed in 45 different size segregated channels (Table 1). The SMPS consists of a Vienna-type LDMA (Long Differential Mobility Analyzer) and a butanol-based condenser for imaging. In addition, concentrations of gaseous pollutants and meterological parameters were measured, details of the instruments for measuring the above are given in Table 1. The sampling of air was done through a probe placed horizontally with a straight sampling tube that is < 1m in length. High humidity >60 % in the air sample is removed using a diffusion dryer. The calculation of size distribution for nanoparticles from the number concentrations is based on the estimates adopted from equation (9d) (Winklmayr et al., 1991). The measured data are classified on a seasonal basis into winter, spring, summer, monsoon, and autumn (Table 2), daytime (8 a.m.–8 p.m.) and nighttime (8 p.m.–8 a.m.). Out of 365 days in a year, 298 days of data were taken into account for the analysis, including all seasons (Fig. 2). The data measured by each instrument are cleaned, checked for quality control, and analyzed using validated software provided along with the instrument (81% of data is used for analysis after removing outliers). The study used Matlab2016a, R version (5.12.8), and IBM-SPSS for data analysis and statistical tests such as Levane and F-Test.

3. Results and discussion

3.1. Seasonal and temporal variation of particle number concentration (PNC)

The monitoring location in Delhi experiences widely varying

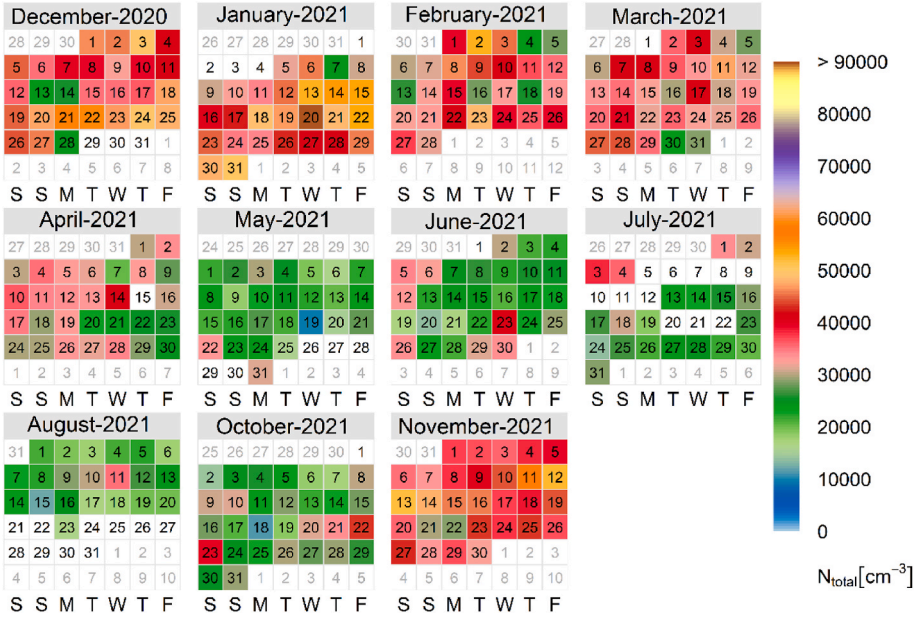


Fig. 2. Calendar plot showing the values of N_{total} concentration (particles in the 10–1000 nm size range) on the monitoring days from December 2020 to November 2021.

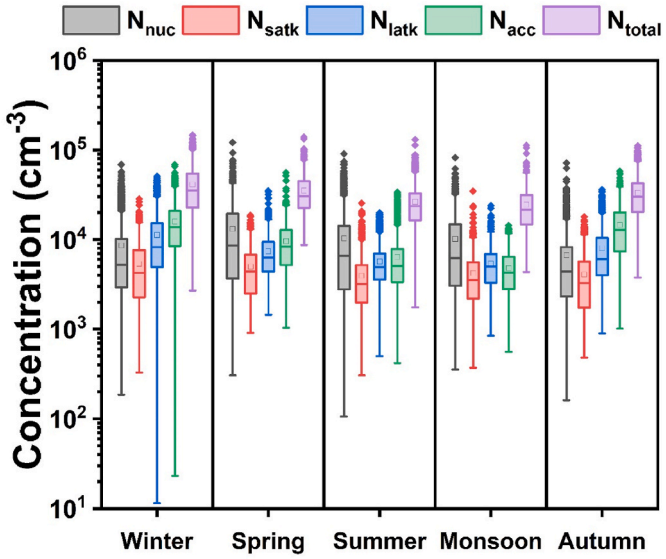


Fig. 3. Seasonal mean concentration (cm^{-3}) of particles in the nucleation (N_{nuc}, 10–30 nm), small Aitken (N_{satk}, 30–50 nm), large Aitken (N_{latk}, 50–100 nm), accumulation (N_{acc}, 100–1000 nm) size range and their total (N_{total}, 10–1000 nm) in different seasons in Delhi. The rectangular box represents data range of 25–75% in the box plot. The upper line represents the upper quartile value. The lower line represents a lower quartile value. The centerline of each box corresponds to the median value, and the unfilled square in each bar represents the mean value. The symbols (–) at the top and bottom of each box represent the maximum and minimum values.

concentration levels of particles in different seasons (Fig. 3). The total PNC or N_{total} corresponds to the sum of Nucleation - N_{nuc}, small Aitken - N_{satk}, large Aitken - N_{latk}, and accumulation - N_{acc} modes (total PNC/ $N_{total} = N_{nuc} + N_{satk} + N_{latk} + N_{acc}$). The concentration of total PNC is higher during winter, followed by spring, autumn, summer, and monsoon (Fig. 2). The average total PNC is $> 10^4$ cm^{-3} during the year (Fig. 3). The number concentration is higher during colder seasons (higher relative humidity and calm wind conditions) due to less dispersion when compared to hot seasons when wind speeds and ventilation coefficients are higher, which promote higher dispersion resulting in comparatively lesser concentrations (Fig. 3, Tables 3 and 4; Gani et al., 2021). The minimum and maximum PNC in different sizes varies in the range of 10^3 particles cm^{-3} (lowest in monsoon) to $>10^5$ particles $cm^{-3,}$ the highest in winter (Table 3). Results on a global study (Kumar et al., 2014) showed that particle concentrations can vary between 10^2 and 10^7 cm^{-3} in highly polluted megacities such as Delhi and Beijing (), consistent with the present study.

The seasonal mean PNC in different size bins of N_{satk} (5.3×10^3 cm^{-3}), N_{latk} (11.2×10^3 cm^{-3}), and N_{acc} (15.9×10 cm^{-3}) are the highest during the winter season giving rise to the highest N_{total} in winter (exceeding 40,000 particles cm^{-3}); interestingly, the seasonal behavior of N_{nuc} is different than the other sizes and N_{nuc} exhibits a peak during the spring season with a concentration of about 13.2×10^3 cm^{-3}. The higher N_{nuc} in spring occurs due to prevalent atmospheric conditions of moderate relative humidity and solar intensity, which facilitate the new particle formation process. The new particle formation process increases the concentration of N_{nuc} particles significantly as only very few anthropogenic sources directly emit particles in the nucleation size range. The N_{total} concentration in winter is ~2 times higher than monsoon and 1.6 times higher than summer (Table 3). A comparison of day and night time concentrations in different size bins reveal that the lower size range particle, N_{nuc}, is higher during the daytime (Fig. 4). Higher N_{nuc} concentration in daytime over urban regions represents fresh emissions and particle formation process , whereas higher N_{acc} mode particles in the nighttime occurs due to coagulation and emission from the traffic sources, especially heavy vehicles. The pattern and magnitude of N_{satk} and N_{latk} do not vary significantly between day and night (Fig. 4). The N_{total} is ~1.3 times lower in autumn and spring than winter. The lowest and highest concentrations in different sizes on seasonal scales are not uniform (Table 3). The PNC in bigger size range, N_{acc} and N_{latk}, is lower during the monsoon season due to wet removal leading to lower N_{total}, whereas PNC in smaller (N_{satk} and N_{nuc}) is relatively lower during the summer, monsoon and autumn seasons compared to spring due to new particle formation process and UFP bursts that occur during spring giving rise to higher N_{nuc} (Table 3).

Table 3
Statistical summary of particle number concentrations (PNC)[a] classified as a function of size into nucleation (N_{nuc}), small Aitken (N_{satk}), large Aitken (N_{latk}), accumulation (N_{acc}) and total (N_{total}) in different seasons over Delhi from the measurements made in 2020–2021.

Season	Parameter[a]	Minimum (cm^{-3})	Maximum (cm^{-3})	Mean ± 1σ (standard deviation) (cm^{-3})	Median (cm^{-3})
Winter	N_{nuc}	185	68626	8594 ± 9079	5201
	N_{satk}	327	28395	5335 ± 3842	4239
	N_{latk}	11	50832	11236 ± 8719	8248
	N_{acc}	23	68458	15905 ± 9645	13803
	N_{total}	2691	147182	41089 ± 23669	35134
Spring	N_{nuc}	306	121136	13234 ± 13287	8584
	N_{satk}	910	18674	4981 ± 3025	4390
	N_{latk}	1446	35027	7411 ± 4385	6326
	N_{acc}	1039	56033	9599 ± 5832	8337
	N_{total}	8694	139019	35226 ± 17909	30413
Summer	N_{nuc}	106	90646	10284 ± 10803	6556
	N_{satk}	305	25546	3928 ± 2734	3179
	N_{latk}	498	19861	5681 ± 3101	4916
	N_{acc}	418	33838	6411 ± 4808	5076
	N_{total}	1752	130433	26307 ± 13719	23702
Monsoon	N_{nuc}	353	82547	10161 ± 10078	6205
	N_{satk}	370	34664	4225 ± 2932	3546
	N_{latk}	843	23892	5397 ± 2811	5010
	N_{acc}	558	14345	4805 ± 2540	4257
	N_{total}	4337	112587	24590 ± 13256	21459
Autumn	N_{nuc}	161	71873	6667 ± 6955	4398
	N_{satk}	479	17876	4060 ± 2917	3263
	N_{latk}	897	36062	7990 ± 5789	6038
	N_{acc}	1012	58056	14610 ± 8940	12849
	N_{total}	3774	111316	33329 ± 17883	29893

[a] (N_{nuc}, nucleation (10–30 nm), N_{satk}, small Aitken (30–50 nm), N_{latk}, large Aitken (50–100 nm), N_{acc}, accumulation (100–1000 nm), and N_{total}, total PNC (10–1000 nm)).

Table 4
Statistical summary of gaseous pollutants in different seasons over Delhi during 2020–2021.

Season	Parameter[a]	Minimum	Maximum	Mean ± 1σ (standard deviation)	Median
Winter	NO_2 (μg/m^3)	5.0	488.0	124 ± 60.3	113.4
	NO (μg/m^3)	0.3	489.0	44.2 ± 55.3	23.5
	NO_x (ppb)	0.3	456.5	97.9 ± 60.4	81.1
	SO_2 (μg/m^3)	1.8	48.0	6.2 ± 3.6	4.8
	CO (mg/m^3)	0.2	10.2	1.8 ± 1.2	1.4
	Ozone (μg/m^3)	1.1	167.8	37.8 ± 20.6	34.7
	T (°C)	10.7	33.7	19.0 ± 4.5	18.4
	RH (%)	25.0	88.0	72.8 ± 17.5	80.6
	WS (ms^{-1})	0.1	9.0	1.0 ± 0.5	1.0
	SR (Wm^{-2})	4.3	278.4	40.5 ± 55.3	8.0
Spring	NO_2 (μg/m^3)	0.5	297.0	63.3 ± 36.4	52.2
	NO (μg/m^3)	5.1	291.5	23.6 ± 29.2	12.2
	NO_x (ppb)	11.4	220.0	46.6 ± 30.8	35.9
	SO_2 (μg/m^3)	2.6	41.3	5.6 ± 3.5	4.6
	CO (mg/m^3)	0.05	8.5	1.4 ± 1.1	1.2
	Ozone (μg/m^3)	0.4	172.1	31.6 ± 25.9	20.0
	T (°C)	16.9	42.7	26.6 ± 5.5	25.8
	RH (%)	13.8	87.9	54.5 ± 20	54.1
	WS (ms^{-1})	0.1	10.0	1.1 ± 0.9	1.0
	SR (Wm^{-2})	4.3	383.6	62.8 ± 83.6	7.8
Summer	NO_2 (μg/m^3)	15.5	197.6	51 ± 26	44.6
	NO (μg/m^3)	3.9	175.4	8.1 ± 6.2	6.9
	NO_x (ppb)	6.0	245.5	23.1 ± 10.8	20.6
	SO_2 (μg/m^3)	1.1	127.0	10.6 ± 11.3	7.1
	CO (mg/m^3)	0.1	4.8	1.1 ± 0.8	0.9
	Ozone (μg/m^3)	1.9	196.7	44.3 ± 33.6	34.4
	T (°C)	18.5	44.3	32.5 ± 5.2	32.0
	RH (%)	11.2	87.8	44.1 ± 18.1	42.1
	WS (ms^{-1})	0.2	11.7	1.4 ± 1.3	1.0
	SR (Wm^{-2})	3.4	468.1	91.4 ± 114.2	19.7
Monsoon	NO_2 (μg/m^3)	10.9	211.7	29.4 ± 16.4	24.8
	NO (μg/m^3)	2.6	36.7	8.7 ± 4.0	7.3
	NO_x (ppb)	4.2	71.9	17.1 ± 5.9	15.5
	SO_2 (μg/m^3)	4.6	49.0	8.1 ± 2.6	7.6
	CO (mg/m^3)	0.16	2.7	0.7 ± 0.4	0.6
	Ozone (μg/m^3)	0.5	184.7	37.4 ± 30.4	24.8
	T (°C)	27.4	44.8	33.0 ± 3.4	31.9
	RH (%)	24.9	87.8	71 ± 16.4	74.4
	WS (ms^{-1})	0.2	11.0	1 ± 0.8	0.9
	SR (Wm^{-2})	5.5	393.2	73.1 ± 96.9	18.0
Autumn	NO_2 (μg/m^3)	7.1	319.1	44.1 ± 39.2	32.5
	NO (μg/m^3)	4.8	187.3	13.9 ± 18.2	8.5
	NO_x (ppb)	3.7	318.5	31.7 ± 32.5	19.6
	SO_2 (μg/m^3)	4.3	17.5	7.9 ± 1.7	8.1
	CO (mg/m^3)	0.03	11.5	2.0 ± 1.5	1.6
	Ozone (μg/m^3)	0.8	163.2	35.1 ± 22.6	27.6
	T (°C)	15.6	40.3	25.9 ± 5.7	25.5
	RH (%)	23.1	87.9	63.5 ± 17.9	67.9
	WS (ms^{-1})	0.2	6.6	0.9 ± 0.4	0.9
	SR (Wm^{-2})	4.4	311.5	49.8 ± 68.1	7.4

[a] NO_2 - nitrogen dioxide, NO - nitric oxide or nitrogen monoxide, NO_x - nitrogen oxides, SO_2.sulfur dioxide, CO - carbon monoxide, T - Temperature, RH - relative humidity, WS - wind speed and SR - solar radiation.

3.2. Diurnal variation of particle number concentration

The diurnal variation of PNC in different size ranges clearly brings out the influence of traffic emissions in the urban curbside environment (Fig. 5). The N_{latk} and N_{satk} concentrations (Fig. 5b and c) exhibit a more prominently clear double hump structure during the morning (0700–1100 h) and evening peak (1600–2000 h) in all the seasons confirming the dominance of traffic emissions to PNC in these size ranges in all the seasons. The N_{latk} represents the particles in the size range of 50–100 nm, and most of the particles emitted from vehicular exhaust lie in this size range. In winter and autumn, N_{latk} is higher than the PNC in the other size ranges throughout the day. Further, N_{latk} is ~2 times higher in winter and autumn than in the other seasons. The N_{nuc} mode particles show a similar emission pattern in all seasons with peaks (Fig. 5a) during the active sunshine hours compared to the other hours of the day. This diurnal emission pattern of pollutants is induced by atmospheric and meteorological variations in boundary layer height, temperature, wind speed, relative humidity, solar radiation and ventilation coefficient (Fig. 6). The peak boundary layer height (BLH) (between 1200 and 1600 h) during winter is lowest (~500 m), whereas in summer the peak BLH increases by a factor of 4 going up to more than 2 km in Delhi (Fig. 6). In summer and spring N_{nuc} is higher during the active sunshine hours (Figs. 4 and 6) due to nucleation burst of particles, and new particle formation events, the details of the formation of particles due to the above and the processes involved are beyond the scope of the present study. The new particle formation events start as a result of clustering particles in the sub-3 nm size range and oxidative precursor vapors. As the measurements of these events are unavailable, these events could not be probed further in this study. The SO_2 concentration is low (Fig. 7) during spring as it gets utilized in the formation of H_2SO_4 vapors, which acts as a precursor for the formation of new particle in nucleation mode (Fig. 5a). The concentrations of other pollutants from vehicular emissions, such as nitrogen oxides (NO, NO_2 and NO_x) and carbon monoxide (CO) are also higher in winter (Fig. 7). During

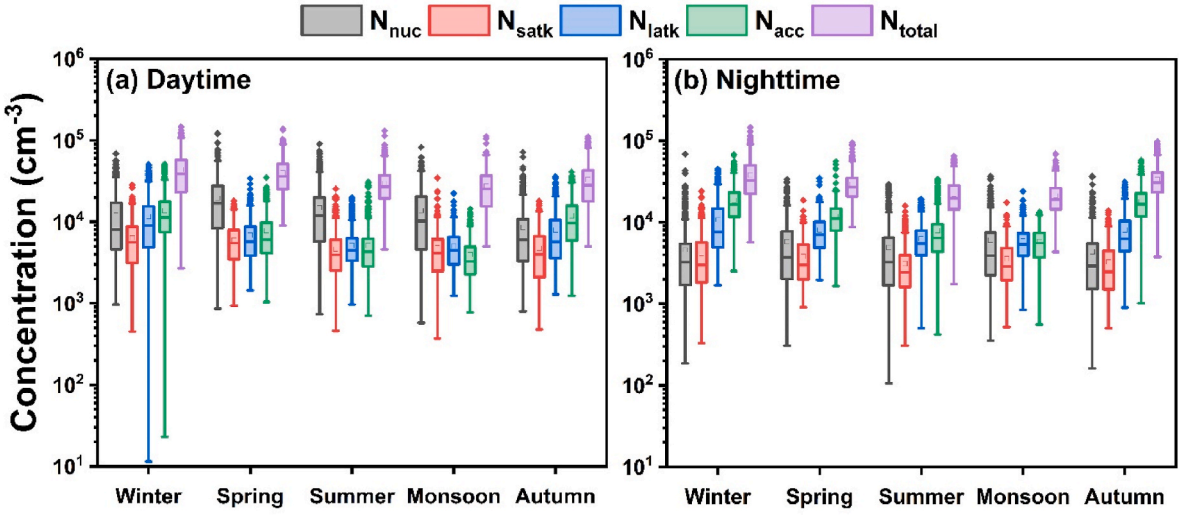


Fig. 4. (a) Daytime and (b) nighttime variation in the number concentration of particles (cm^{-3}) in the nucleation ($N_{nuc,}$ 10–30 nm), small Aitken (N_{satk}, 30–50 nm), large Aitken (N_{latk}, 50–100 nm), accumulation (N_{acc}, 100–1000 nm) and total (N_{total}, 10–1000 nm) during different seasons in Delhi. The rectangular box represents data range of 25–75% in the box plot. The upper line represents the upper quartile value. The lower line represents a lower quartile value. The centerline of each box corresponds to the median value, and the unfilled square in each bar represents the mean value. The symbols (−) at the top and bottom of each box denote the maximum and minimum values.

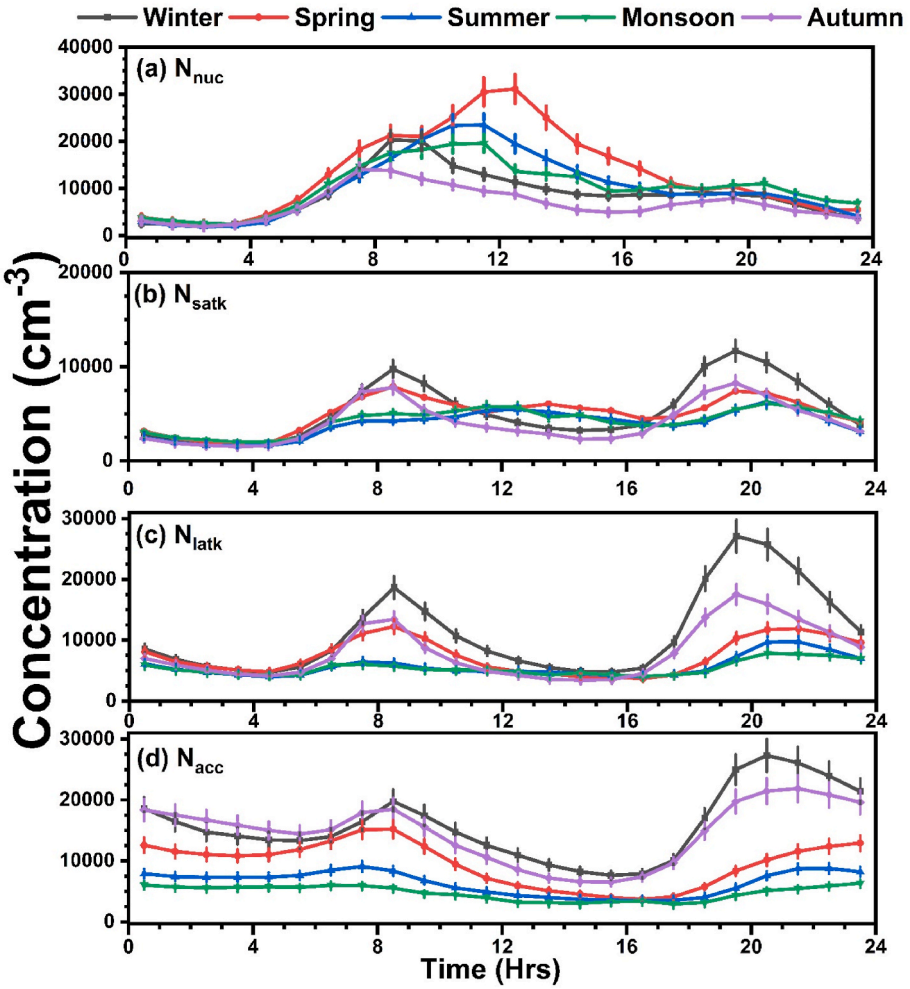


Fig. 5. Diurnal variation (hourly average) of (a) N_{nuc}, (b)N_{satk}, (c)N_{latk} and (d) N_{acc} (cm^{-3}) during winter, spring, summer, monsoon and autumn seasons in Delhi. Vertical bars correspond to ±1σ (standard deviation) from the mean.

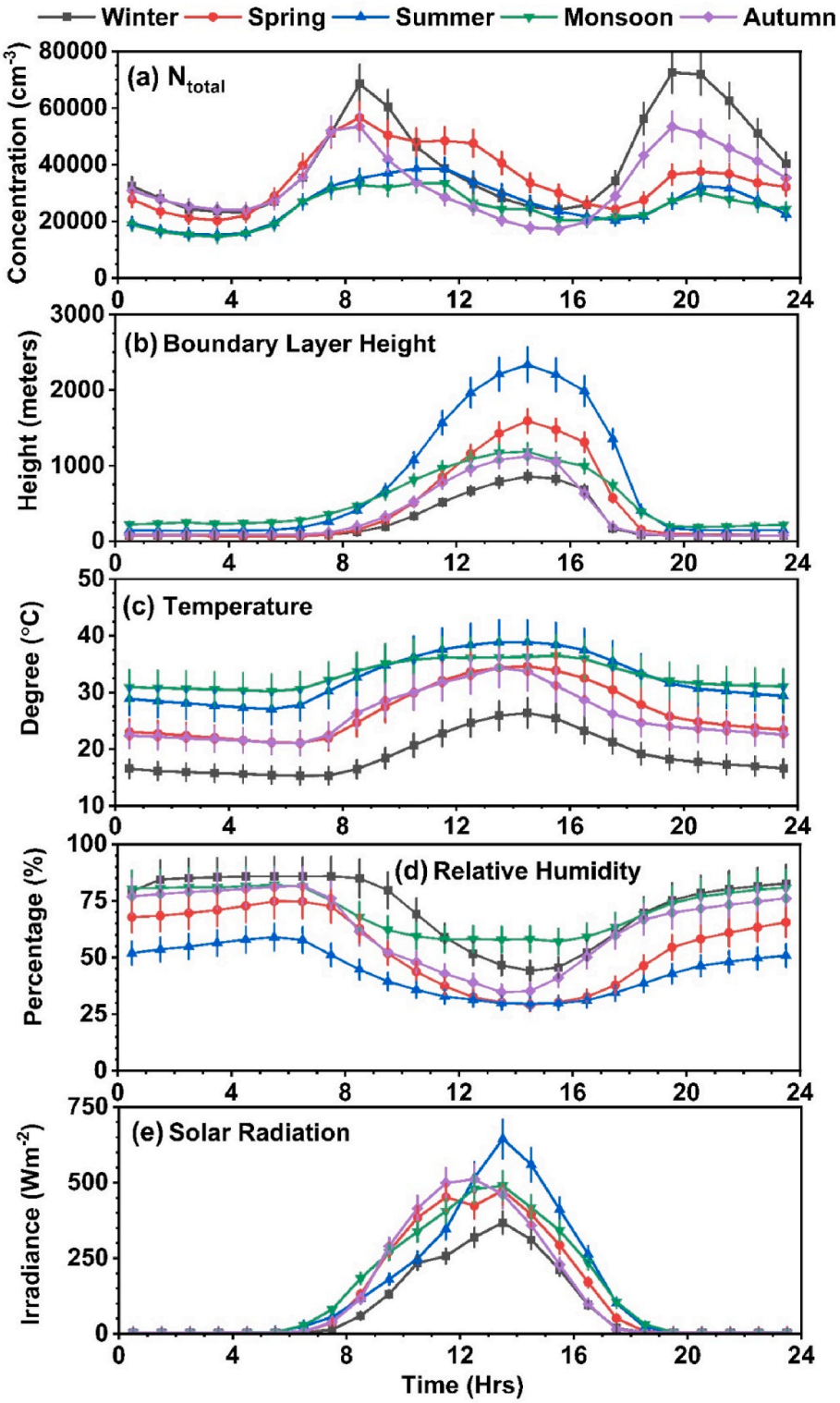


Fig. 6. Diurnal variation of (a) N_{total} concentration (cm^{-3}), (b) boundary layer height (m), (c) temperature (°C), (d) relative humidity (%) and (e) solar radiation (Wm^{-2}) during different seasons in Delhi. Vertical bars represent ±1σ (standard deviation) from the mean.

summer, spring, and monsoon seasons, N_{nuc} increases whereas N_{acc} decreases due to reduced condensation of particles. The microphysical process of aerosol condensation is strongly influenced by relative humidity. Over the study region, RH is lower during summer and spring than winter (Fig. 6) which leads to lower coagulation of particles. The warmer summer and monsoon seasons are usually more ventilated than the colder winters (lower BLH). The solar radiation (SR) is higher in summer whereas RH is higher in winter (Fig. 6). The average diurnal concentration increases during morning peak hours irrespective of the season; however, in the evening hours, the peak concentration is different in different seasons because of the differences in BLH (Fig. 6b). During the warmer periods (summer and monsoon seasons), a higher BLH results in a better ventilation, enabling dilution of particles resulting in lower N_{total}. A lower BLH, higher RH, lower temperature and

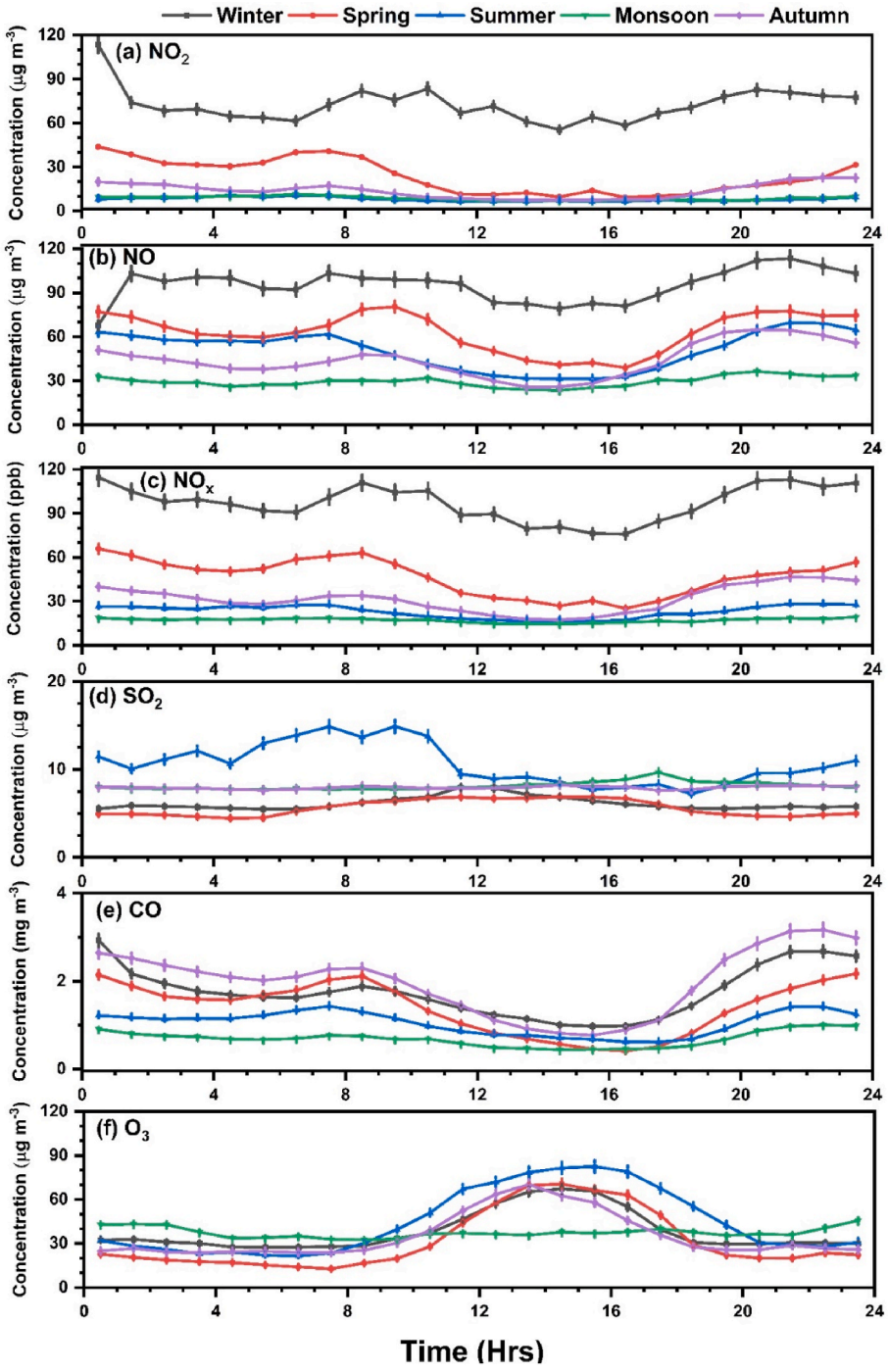


Fig. 7. Diurnal variation (based on hourly averages) of gaseous pollutants (a) nitrogen dioxide (NO_2 μg/m^3), nitric oxide or (b) nitrogen monoxide (NO μg/m^3), (c) nitrogen oxides (NO_x ppb), (d) sulfur dioxide (SO_2 μg/m^3), (e) carbon monoxide (CO mg/m^3), (f) ozone (O_3 μg/m^3) during winter, spring, summer, monsoon and autumn seasons over Delhi. Vertical bars denote ±1σ (standard deviation) from the mean.

lower SR helps trapping the pollutants near surface and result in significantly higher concentrations (>50,000 particles per cm^3) during winter whereas a higher BLH, lower RH, higher temperature and higher SR during summer (and in monsoon) results in at least a factor of 2-lower concentrations than winter (Fig. 6). Thus, this synergistic analysis of PNC in different size ranges along with local emissions of gaseous precursors, formation processes, and meteorological parameters reveals that the PNC is governed by local emissions of gaseous precursors and meteorological parameters, and the diurnal variations in PNC exhibit seasonal changes due to variations in these parameters.

3.3. Diurnal and seasonal variations of local emissions in the roadside environment: Correlation between gaseous pollutants and nano particles

The major source of nanoparticle emissions over the monitoring location is from the roadside emissions, especially vehicular exhaust. Apart from nanoparticles, various gaseous pollutants are also emitted due to different processes in the urban environment. The gaseous pollutants, such as CO and nitric oxide (NO), are released in the urban environment from combustion processes such as burning fossil fuels and biomass/biofuels. The diurnal variation of PNC and gases such as CO (and NO) shows a similar pattern, suggesting that the sources of these

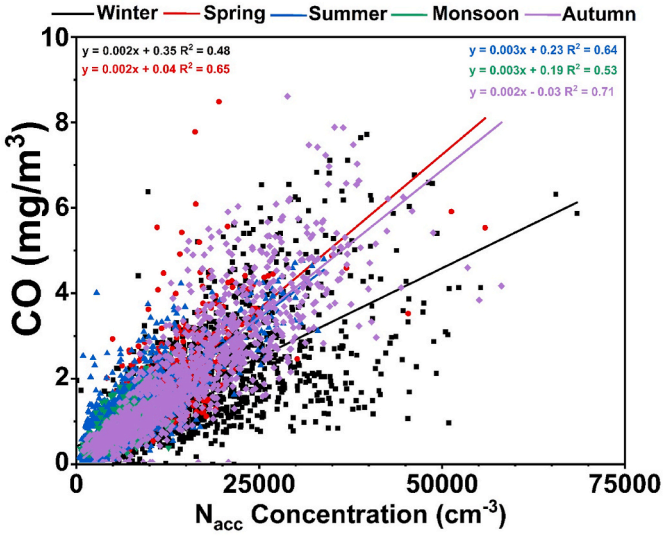


Fig. 8. Correlation between carbon monoxide (CO, mg/m^3) and particle number concentration in accumulation mode (N_{acc}, cm^{-3}) corresponding to the five seasons over Delhi. The equations of correlation obtained for different seasons are given in respective colors in the figure.

gaseous and nanoparticle emissions are the same, i.e., transportation sources in the monitoring location. The increase in ozone (O_3) concentration during the sunshine hours in mid-day and afternoon and reduced concentration of NOx in the presence of nitrogen dioxide (NO_2) as a precursor compound due to the photochemical reaction is evident during spring, summer, and autumn when ozone formation by volatile organic compounds (VOCs) and NO_x is a common phenomenon in urban regions (Fig. 7) (e.g., Nelson et al., 2021). In winter, the concentration of SO_2 is relatively higher in the afternoon, suggesting that formation of H_2SO_4 aiding new particle formation process is inhibited during this period due to unfavorable relative humidity and solar intensity (Fig. 7). In contrast, during summer and spring, SO_2 concentration in mid day is lower due to the use of sulfur compounds as a precursor material for the new particle formation events. During summer and spring, N_{nuc} is also higher (Fig. 5), due to the emissions and the new particle formation events that contribute to the increase in particle number concentrations in the roadside environment. The particles in the accumulation mode size range occur due to condensational growth and coagulation processes. The particles in accumulation mode relatively have longer residence times compared to nucleation mode particles. As a result, the contribution of N_{acc} in particulate matter formation, such as $PM_{2.5}$ and $PM_{10,}$ is proportionately higher (Apte et al., 2011). To explore the relation between the gaseous pollutant emitted due to combustion process over an urban environment and accumulation mode particles which have higher residence time, carbon monoxide (CO) is correlated with N_{acc} in all the seasons. CO is found to correlate positively well with N_{acc} in all the seasons in Delhi, with the coefficient of determination (R^2) between CO and N_{acc} being $\geq$0.48 in all the seasons (Fig. 8), whereas the R^2 between CO and N_{latk} (50–100 nm) is $\leq$ 0.40, confirming that vehicle-based emissions are the predominant source for PNC and other gaseous pollutants in the study region.

3.4. Influence of meteorology on particle size distribution: relative humidity and temperature

The concentration of particles in urban regions exhibit seasonal variability during the day due to the differences in meteorological parameters and emission sources (Sengupta et al., 2022). The particle concentration becomes higher during winter when the temperature is minimum and relative humidity is maximum (Fig. 6). The seasonal mean temperature during winter is 19 °C during which the total particle number concentration is the highest (Fig. 6, Tables 3 and 4). The lowest concentration of particles occurs in the monsoon season due to wet removal of particles by precipitation. Solar radiation and relative humidity are other important factors that aid particle and secondary pollutant formation. During the afternoon hours in almost all the seasons, SR increases and RH decreases (Fig. 6). During spring, summer and autumn in the afternoon hours, the increase in solar radiation increases the secondary ozone formation and decreases the particle number concentration due to dispersion. The average solar radiation during the peak sunshine hours is about 870 Wm^{-2} during summer (Fig. 6) and SR is lowest in winter (535 Wm^{-2}) (Fig. 6). Relative humidity is <50% during summer and is the highest in winter at 95%. During spring, monsoon and autumn, the temperature and relative humidity are higher, and are in the 25^o C to 32^o C, and 54–71% range, respectively. In general, the particle concentration peaks when the daily mean temperature is minimum and slumps when the daily mean temperature is high. This is due to the influence of the atmospheric boundary layer that varies with surface temperature; hence, the dispersion also varies accordingly (Gani et al., 2021). In winter, the particle concentration usually peaks due to temperature inversion; in summer, it is lower due to the strong air exchange. During nighttime, the surface temperature is lower, and the atmospheric BLH is also low, leading to a higher accumulation of particles.

The analysis reveals that temperature and relative humidity influence the size distribution of particles largely through the processes of condensation and coagulation, and new particle formation events (Figs. 9 and 10). The concentration of particles in the lower size range, especially the nucleation mode, is significantly higher during low RH (<30%) (Fig. 9a) and high temperature (T > 30 °C) (Fig. 10c) conditions that prevail during spring in the study area which favors the particles to undergo the natural gas-to-particle conversion mechanism. The mid-temperature range (20^{o}C- 30 °C) and the mid-RH (30%–60%) act as a transition period during which the lower size particles undergo physical transformation in the atmosphere and their size distribution gradually changes, increasing from smaller to the higher size. In the above RH and temperature regimes, the particle concentration varies between 1.5 and 3.0 $\times$ 10^4 cm^{-3} in all the three modes (N_{nuc}, N_{satk}, and N_{latk}) irrespective of the seasons. In the lower temperature (<20 °C) and higher RH (>60%) scenario, the concentrations of particles are higher in accumulation mode size range (Figs. 9c and 10a). Under higher RH condensation process begins which increases the number concentrations in the accumulation mode, and this phenomenon is evident in the winter and autumn seasons, followed by the spring season. The analysis clearly indicates that apart from the intensity of emission sources over the study region, the particle number concentration and size distribution exhibit significant variations based on the prevalent meteorology, in particular relative humidity and temperature. Further, it is clear that the meteorological parameters not only influence the concentration of particles but also modulate the proportion of particles in different size ranges depending on meteorological conditions, thereby influencing the N_{total}.

3.5. Size resolved particle contribution

The N_{total} concentration or total PNC is made up of contributions from all the size fractions (Total PNC/N_{total} = N_{nuc} + N_{satk} + N_{latk} + N_{acc}). Apart from changes in concentration, the contribution of different size fractions to N_{total} also varies from season to season due to changes in nanoparticle emission sources, patterns, and regional meteorology (Fig. 11). During the winter season, when N_{total} is the highest, N_{acc} contributes the maximum at 41% followed by N_{latk} 26% to N_{total}. N_{acc} contribution is higher in this season as a result of higher rate of coagulation in the atmosphere due to high relative humidity and lower temperatures. Similarly, in autumn, N_{acc} accounts for 45% of N_{total}. N_{acc} contributes $\geq$30% during winter and autumn to the total. During spring, summer, and monsoon, the contribution of N_{nuc} is 33%, 34%, and 36%,

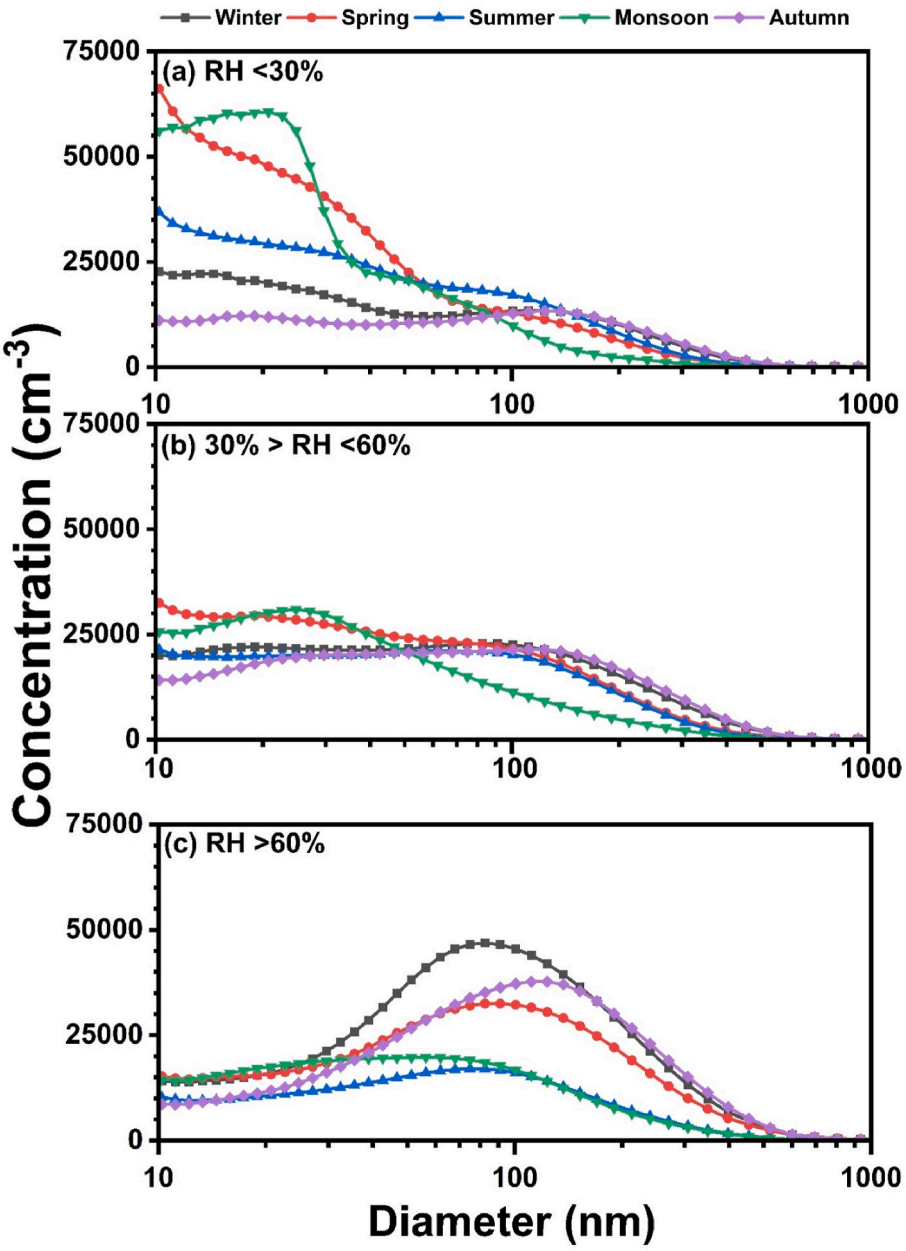


Fig. 9. Particle number size distribution characterized for different conditions of relative humidity (RH) as (a) RH<30%, (b) 30%>RH<60%, and (c) RH>60% over Delhi in different seasons of the study period.

respectively to N_{total}. The percentage contribution of N_{satk} and N_{latk} does not vary significantly during the year in Delhi, and their contributions lie in the 12–17% range, and 22–26% range, respectively (Fig. 11). This indicates that the contribution from the emission sources of particles in the size range of 30–100 nm is uniform throughout the year in Delhi. The smaller N_{nuc} and the larger N_{acc} exhibit a higher variation in their respective contributions to the total, varying from 20% to 36% and from 23% to 45%, respectively. Overall, the contribution of N_{satk} to total PNC remains more or less the same during all seasons. N_{latk} particles contribute about 25% to the total PNC in all seasons (Gani et al., 2019; Patel et al., 2021).

The percentage contributions by different size particles to N_{total} during daytime and nighttime (Fig. 11) vary and the results clearly indicate the distinct role of diurnal variation in emission sources. The % contributions of nucleation mode followed by accumulation contribute >60% to N_{total} during daytime with their sum contribution being the highest in spring season (67%) (Fig. 11). During the warm periods of spring, summer and monsoon, N_{nuc} contributes ~45% to N_{total} during daytime due to the prevalent meteorology which enables freshly emitted exhaust emissions to remain in the atmosphere longer. Whereas in autumn and winter N_{nuc} contributes <30% to N_{total} in daytime. In stark contrast, N_{acc} dominates the contribution (50% or more) to N_{total} during nighttime in winter and autumn as a result of coagulation of emitted particles as the prevalent atmospheric conditions (higher RH, lower temperature and BLH) favor coagulation (Fig. 11b). The N_{satk} contributes <20% during both day and night time in all the seasons. The contribution from N_{satk} and N_{latk} during day and night do not exhibit significant seasonal variabilities – N_{satk} contributes 14–17% (daytime) and 10–16% (nighttime), whereas N_{latk} contributes 18–25% (daytime) and 23–29% (nighttime), respectively. The variations in the contribution of different size particle to N_{total} suggests that the contribution from different sources either enables or inhibits due to the participation of precursor gases and prevailing meteorology, which influence the formation mechanism of nanoparticles in urban regions. Thus, it is clear

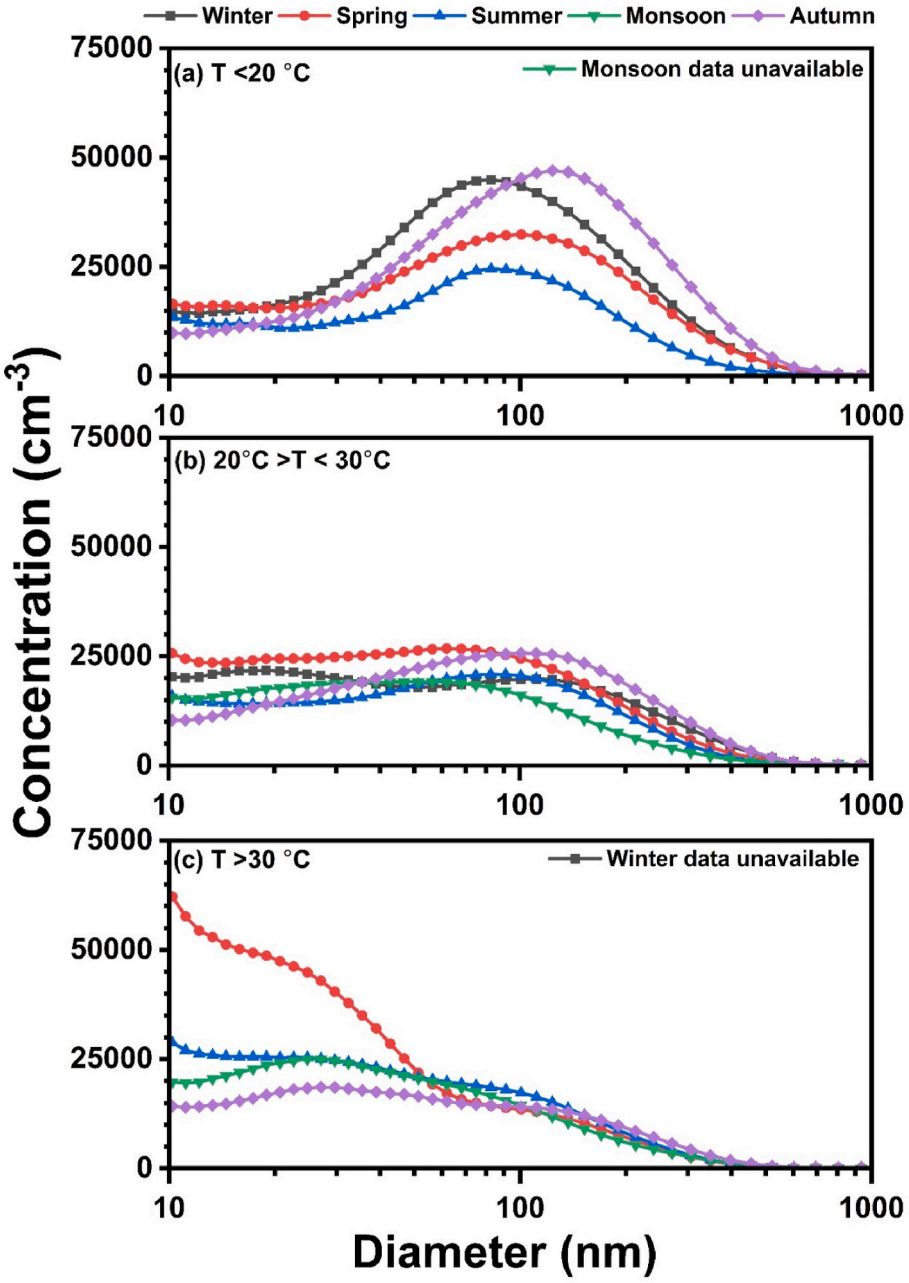


Fig. 10. Particle number size distribution classified as a function of temperature (T, °C) over Delhi as (a) T < 20 °C, (b) 20°C > T < 30 °C, and (c) T > 30 °C in different seasons of the study period.

that the variation in N_{total} concentration on seasonal scales occur as a result of differences in contribution from varied size ranges, sources, and meteorological conditions which vary diurnally and seasonally. Further, such complex mixture of the formation processes in different size ranges makes mitigating poor air quality conditions in these regions difficult.

3.6. Diurnal variation of particle size distribution: Heat map

A heat map represents the evolution of particle number size distribution with the time in different seasons (Fig. 12). In winter, the intensity of accumulation mode particles is high during evening hours, especially from 1800 to 2200 h due to coagulation of particles and lower BLH (Fig. 12a). In spring, the concentration is high between 0800 and 1600 h due to gas-to-particle formation mechanism wherein the concentration of particles of size <100 nm attain a peak between 1100 and 1300 h due to new particle formation events (Sebastian et al., 2021) (Fig. 12). The time periods, and favorable meteorological conditions for the new particle formation events reported in earlier studies (e.g., Gani et al., 2021) are similar as in the present study. The scenario is similar in summer (Fig. 12c), however, a complete new particle formation does not occur owing to the emissions from other sources such as vehicular exhaust which suppress the new particle formation process, which is termed a UFP burst. A UFP burst refers to a disturbed or incomplete new particle formation process. During the nucleation process, accumulation mode particle concentration starts increasing gradually from the beginning of the new particle formation event until the end of the process due to the growth of particles from smaller to larger sizes after the nucleation events. This is defined as a complete new particle formation process. The new particle formation process can last for a few hours in a day. In an urban region like Delhi, emissions from different complex

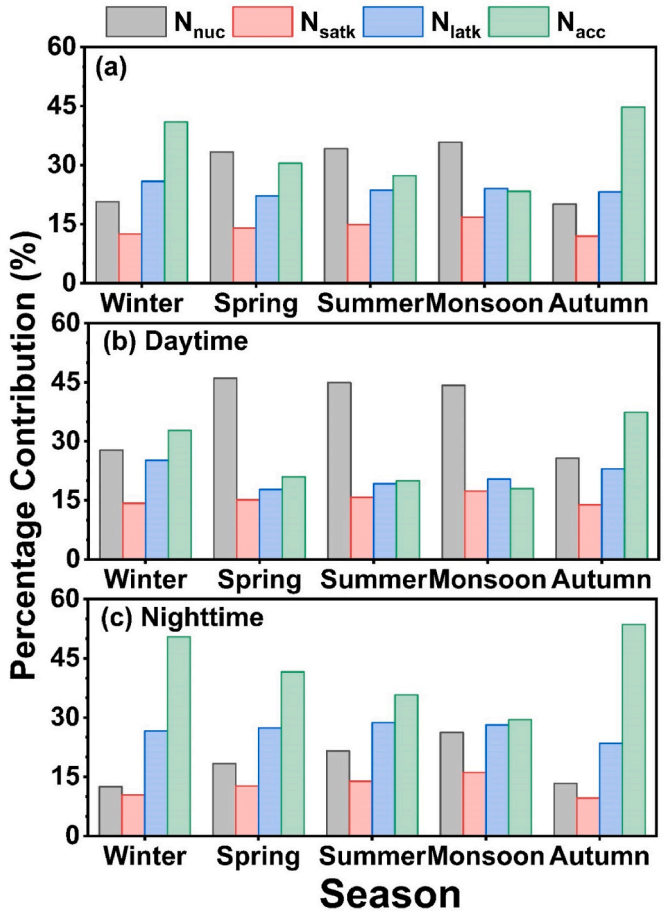


Fig. 11. Percentage contribution (%) of N_{nuc}, N_{satk}, N_{latk}, N_{acc} to total concentration (N_{total}) from winter to autumn in Delhi (a) throughout the day (24h), (b) in daytime (0800–2000 h) and (c) in nighttime (2000–0800 h).

sources often disturb/inhibit the above formation process. During the monsoon season, the burst in smaller size particles is interrupted by precipitation events. The particle concentration changes between 0800 and 1200 h based on sources and prevailing meteorological conditions (Fig. 12). The sources, and the formation of particles by nucleation and other secondary processes govern the magnitude of geometric mean diameter of particles, which is derived by analyzing the particle size distribution and the concentrations. In winter and autumn, the concentration of particles with a diameter >700 nm (that falls in accumulation mode) is higher, whereas the concentration of particles that are ≤100 nm covering the nucleation, small Aitken, and large Aitken sizes are significantly lower (Fig. 12). During spring and summer the concentration of particles <700 nm are significantly higher than the smaller size particles, both of which are consistent with the formation mechanism, sources, and meteorological conditions during these seasons. In summary, the heat map which provides a complete picture of formation process and concentrations of particles in different seasons on a diurnal scale, is consistent with the sources of particles and formation, and the prevailing meteorological conditions over the urban region Delhi.

4. Conclusions

The concentration and behavior of nano-sized pollutants/particles (10–1000 nm) in different seasons (winter, spring, summer, monsoon, and autumn) in mega city Delhi, are analyzed using the data collected from December 2020 to November 2021, for the first time. The concentration of particles is classified into four different sizes as N_{nuc} (10–30 nm, nucleation particles), N_{satk} (30–50 nm, small Aitken particles), N_{latk} (50–100 nm, large Aitken particles), and N_{acc} (100–1000 nm, accumulation mode particles), and the total particle number concentration (PNC) as N_{total} (or total PNC). The nano-sized pollutants exhibit significant diurnal and seasonal variations due to differences in environmental conditions, emission sources, regional meteorology, and atmospheric dynamics. PNC ranges between 10^4 cm^{-3} and 10^6 cm^{-3} over Delhi during the year, and the highest concentration is found during winter. On a seasonal scale, the concentration in winter is ~2 times higher than in monsoon, 1.6 times higher than in summer, and about 1.3 times higher than in autumn and spring. The diurnal variation analysis

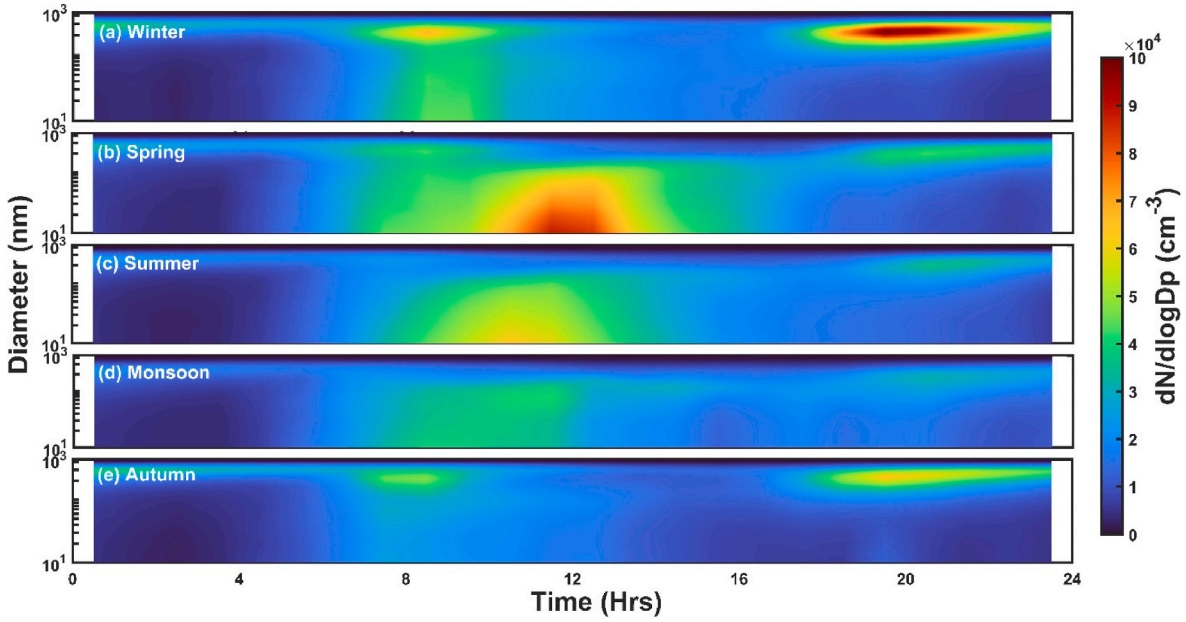


Fig. 12. Heat map of particles in the 10–1000 nm size range (dN/dlogDp, cm^{-3}) as a function of time of the day in different seasons – winter, spring, summer, monsoon, and autumn, respectively over Delhi.

shows distinct characteristics of seasonal, primary, and secondary emissions throughout the day. The direct primary emissions from engine exhaust exhibit a prominent double hump structure during morning and evening peak hours in winter and autumn. The double hump structure is the same for PNC and other gaseous pollutants related to vehicular emissions during these seasons. Carbon monoxide (CO) concentration correlates positively with N_{acc} (100–1000 nm) in all seasons ($R^2 \geq 0.5$), suggesting that emissions from fossil fuel combustion is the predominant source for gases and particles in the near-road curbside environment. Ozone concentration increases multi-fold during the peak sunshine hours due to photochemical reactions with nitrogen oxides (NO_x) and volatile organic compounds (VOCs) under the influence of sunlight and heat in all the seasons except monsoon. The study also reveals that apart from differences in primary direct emissions, the particles undergo secondary transformation (through coagulation and condensation) due to the prevalent meteorological conditions that differ in different seasons.

The concentrations of particles in different size ranges (N_{nuc}, N_{satk}, N_{latk} and N_{acc}) and their respective contributions to total PNC exhibit significant seasonal variations. During winter, the larger size particles (large Aitken and accumulation) contribute more to the total, with , accumulation mode alone contributing >40% to the total PNC. In winter, the concentration of accumulation mode particles (>500 nm) is higher in the evening hours due to the coagulation of smaller particles. N_{nuc}, $N_{satk,}$ and N_{latk} (<100 nm) are higher in spring and summer during mid-day due to nucleation and/or ultrafine particle (UFP) burst events. The concentrations in the nucleation mode (smaller size) and accumulation (larger size) exhibit higher deviations in their respective contributions to the total, with N_{nuc} contributing 20%–36% and N_{acc} accounting for 23%–45% in all the seasons. In contrast, the concentrations in the medium size exhibit lesser variability as the contribution of N_{satk} to N_{total} varies from 12% to 17%, and in case of N_{latk} its contribution to N_{total} varies from 22% to 26%, respectively during the year. . The particle number concentrations and their contributions are influenced by differences in different emission sources and due to variations in other meteorological parameters such as wind speed, relative humidity, temperature, solar radiation and boundary layer height during day and night. The percentage contributions by different size particles to N_{total} during daytime and nighttime vary and the results clearly indicate the distinct role of diurnal variation in emission sources. The % contributions of nucleation mode followed by accumulation contribute >60% to N_{total} during daytime with their sum contribution being the highest in spring season (67%). During the warm periods of spring, summer and monsoon, N_{nuc} contributes ~45% to N_{total} during daytime due to the prevalent meteorology which enables freshly emitted exhaust emissions to remain longer in the atmosphere. Whereas in autumn and winter N_{nuc} contributes <30% to N_{total} in daytime. In stark contrast, the concentration of large particles, N_{acc} and $N_{latk,}$ dominate N_{total} during nighttime, and in particular in winter and autumn N_{acc} and N_{latk} contribute 77% to N_{total} due to coagulation of emitted particles as the prevalent atmospheric conditions (higher RH, lower temperature and BLH) favor coagulation, with N_{acc} alone contributing ≥50% in both the seasons. The contribution of N_{satk} and N_{latk} is less to N_{total} during day and night, and they do not exhibit significant seasonal variabilities – N_{satk} contributes 14–17% (daytime) and 10–16% (nighttime), whereas N_{latk} contributes 18–25% (daytime) and 23–29% (nighttime), respectively.

This study provides insights into the behavior, dynamics, contribution, and characteristics of air pollutants in the nano-size range in the atmosphere. The study concludes that the dynamics and profiling of nanoparticles on the environment and human health are not only related to the concentration of direct primary emissions but also depend on the gaseous precursors, meteorological conditions, and atmospheric dynamics of/and over a particular region. The results provide quantitative insights into the number concentrations of particles (air pollutants) in different size ranges from nucleation to accumulation and their behavior in a roadside urban environment of a highly polluted urban city on seasonal scales, which will be useful in devising control strategies aimed to improve air quality, public health, environment, and climate. Concentrations of particles from nucleation to accumulation modes, and their total are significantly higher in Delhi in winter than other seasons. Air quality across north India is found to worsen during winter due to increases in aerosols, and rising relative humidity (Paulot et al., 2022). Further, due to rapid industrialization and growth in population improving air quality is seen to be a big challenge for India (Xie et al., 2024). The winter $PM_{2.5}$ concentrations decreased in 2022 by 20%, out of which one-half of the reduction was attributed to favorable meteorological conditions in the last 5 years, which are unlikely to persist in a warming world (Xie et al., 2024). An estimated 1.7 million premature mortalities in 2019 in India were attributed to severe surface-level air pollution according to a Global Burden of Disease study. Further, an estimated 15% of deaths in South Asia were linked to health effects of air pollution exposure (World Air Quality Report, 2020). These quantitative results on seasonal variations of air pollutants together with the knowledge on seasonal variations in meteorological parameters and atmospheric dynamics (e.g., lower BLH in winter favors trapping the pollutants near surface as opposed to summer) form a basis which can be effectively utilized while devising mitigation measures to enhance the positive effect of improving the air quality and public health.

Ethical approval & consent to participate

Ethical approval is not applicable, and the work has not been published anywhere previously in any form, and it is not in consideration for publication elsewhere.

Consent to participate

All authors approve the publication.

Funding

Not Applicable.

CRediT authorship contribution statement

Kanagaraj Rajagopal: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **S. Ramachandran:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Rajeev Kumar Mishra:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The measurements reported in the study were carried out in Advance Air and Acoustics Research Laboratory, Department of Environmental Engineering, Delhi Technological University.

References

Agarwal, A., Kaushik, A., Kumar, S., Mishra, R.K., 2020. Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period. Air Qulaity, Atmosphere & Health. https://doi.org/10.1007/s11869-020-00881-z.

Apte, J.S., Kirchstetter, T.W., Reich, A.H., Deshpande, S.J., Kaushik, G., Chel, A., Marshall, J.D., Nazaroff, W.W., 2011. Concentrations of fine, ultrafine, and black carbon particles in auto-rickshaws in New Delhi, India. Atmos. Environ. 45 (26), 4470–4480. https://doi.org/10.1016/j.atmosenv.2011.05.028.

Bhardawaj, A., Habib, G., Kumar, A., Singh, S., Nema, A.K., 2017. A review of ultrafine particle-related pollution during vehicular motion, health effects and control. Journal of Environmental Science and Public Health 1 (4), 268–288. https://doi.org/10.26502/jesph.96120024.

Brines, M., Dall'Osto, M., Beddows, D.C.S., Harrison, R.M., Gómez-Moreno, F., Núñez, L., Artíñano, B., Costabile, F., Gobbi, G.P., Salimi, F., Morawska, L., Sioutas, C., Querol, X., 2015. Traffic and nucleation events as main sources of ultrafine particles in high-insolation developed world cities. Atmos. Chem. Phys. 15 (10), 5929–5945. https://doi.org/10.5194/acp-15-5929-2015.

Charron, A., Harrison, R.M., 2003. Primary particle formation from vehicle emissions during exhaust dilution in the roadside atmosphere. Atmos. Environ. 37 (29), 4109–4119. https://doi.org/10.1016/S1352-2310(03)00510-7.

Gani, S., Bhandari, S., Patel, K., Seraj, S., Soni, P., Arub, Z., Habib, G., Hildebrandt Ruiz, L., Apte, J.S., 2020. Particle number concentrations and size distribution in a polluted megacity: the Delhi Aerosol Supersite study. Atmos. Chem. Phys. 20 (14), 8533–8549. https://doi.org/10.5194/acp-20-8533-2020.

Gani, S., Bhandari, S., Seraj, S., Wang, D.S., Patel, K., Soni, P., Arub, Z., Habib, G., Hildebrandt Ruiz, L., Apte, J.S., 2019. Submicron aerosol composition in the world's most polluted megacity: the Delhi Aerosol Supersite study. Atmos. Chem. Phys. 19 (10), 6843–6859. https://doi.org/10.5194/acp-19-6843-2019.

Gani, S., Chambliss, S.E., Messier, K.P., Lunden, M.M., Apte, J.S., 2021. Spatiotemporal profiles of ultrafine particles differ from other traffic-related air pollutants: lessons from long-term measurements at fixed sites and mobile monitoring. Environ. Sci. J. Integr. Environ. Res.: Atmosphere 1 (7), 558–568. https://doi.org/10.1039/d1ea00058f.

Gao, Y., Yang, T., Jin, J., 2015. Nanoparticle pollution and associated increasing potential risks on environment and human health: a case study of China. Environ. Sci. Pollut. Control Ser. 22 (23), 19297–19306. https://doi.org/10.1007/s11356-015-5497-0.

Griffin, R.J., 2013. The sources and impacts of tropospheric particulate matter. Nature Education Knowledge 4 (5), 1.

Gómez-Moreno, F.J., Pujadas, M., Plaza, J., Rodríguez-Maroto, J.J., Martínez-Lozano, P., Artíñano, B., 2011. Influence of seasonal factors on the atmospheric particle number concentration and size distribution in Madrid. Atmos. Environ. 45 (18), 3169–3180. https://doi.org/10.1016/j.atmosenv.2011.02.041.

Gurjar, B.R., Jain, A., Sharma, A., Agarwal, A., Gupta, P., Nagpure, A.S., Lelieveld, J., 2010. Human health risks in megacities due to air pollution. Atmos. Environ. 44 (36), 4606–4613. https://doi.org/10.1016/j.atmosenv.2010.08.011.

Hofman, J., Staelens, J., Cordell, R., Stroobants, C., Zikova, N., Hama, S.M.L., Wyche, K. P., Kos, G.P.A., van der Zee, S., Smallbone, K.L., Weijers, E.P., Monks, P.S., Roekens, E., 2016. Ultrafine particles in four European urban environments: results from a new continuous long-term monitoring network. Atmos. Environ. 136, 68–81. https://doi.org/10.1016/j.atmosenv.2016.04.010.

Hussein, T., Puustinen, A., Aalto, P.P., Mäkelä, J.M., Hämeri, K., Kulmala, M., 2004. Atmospheric Chemistry and Physics Urban aerosol number size distributions. Atmos. Chem. Phys. 4. www.atmos-chem-phys.org/acp/4/391/.

Kanawade, V.P., Tripathi, S.N., Chakraborty, A., Yu, H., 2020. Chemical characterization of sub-micron aerosols during new particle formation in an urban atmosphere. Aerosol Air Qual. Res. 20 (6), 1294–1305. https://doi.org/10.4209/aaqr.2019.04.0196.

Kaur, S., Nieuwenhuijsen, M., Colvile, R., 2005. Personal exposure of street canyon intersection users to PM2.5, ultrafine particle counts and carbon monoxide in Central London, UK. Atmos. Environ. 39 (20), 3629–3641. https://doi.org/10.1016/j.atmosenv.2005.02.046.

Khan, S., Gurjar, B.R., Sahu, V., 2022. Deposition modeling of ambient particulate matter in the human respiratory tract. Atmos. Pollut. Res. 13 (10), 101565 https://doi.org/10.1016/j.apr.2022.101565.

Kozawa, K.H., Winer, A.M., Fruin, S.A., 2012. Ultrafine particle size distributions near freeways: effects of differing wind directions on exposure. Atmos. Environ. 63, 250–260. https://doi.org/10.1016/j.atmosenv.2012.09.045.

Kumar, P., Fennell, P., Britter, R., 2008. Measurements of particles in the 5-1000 nm range close to road level in an urban street canyon. Sci. Total Environ. 402 (1), 437–447. https://doi.org/10.1016/j.scitotenv.2007.10.013.

Kumar, P., Gurjar, B.R., Nagpure, A.S., Harrison, R.M., 2011. Preliminary estimates of nanoparticle number emissions from road vehicles in megacity Delhi and associated health impacts. Environ. Sci. Technol. 45 (13), 5514–5521. https://doi.org/10.1021/es2003183.

Kumar, P., Morawska, L., Birmili, W., Paasonen, P., Hu, M., Kulmala, M., Harrison, R.M., Norford, L., Britter, R., 2014. Ultrafine particles in cities. Environ. Int. 66, 1–10. https://doi.org/10.1016/j.envint.2014.01.013.

Liu, X., Hadiatullah, H., Khedr, M., Zhang, X., Schnelle-Kreis, J., Zimmermann, R., Adam, T., 2022. Personal exposure to various size fractions of ambient particulate matter during the heating and non-heating periods using mobile monitoring approach: a case study in Augsburg, Germany. Atmos. Pollut. Res. 13 (7), 101483 https://doi.org/10.1016/j.apr.2022.101483.

Manojkumar, N., Srimuruganandam, B., 2021a. Health benefits of achieving fine particulate matter standards in India – a nationwide assessment. Sci. Total Environ. 763 https://doi.org/10.1016/j.scitotenv.2020.142999.

Manojkumar, N., Srimuruganandam, B., 2021b. Investigation of on-road fine particulate matter exposure concentration and its inhalation dosage levels in an urban area. Build. Environ. 198 (February), 107914 https://doi.org/10.1016/j.buildenv.2021.107914.

Mei, L., Yan, S., Li, Y., Jin, X., Sun, X., Wu, Y., Liang, Y., Wei, Q., Yi, W., Pan, R., He, Y., Tang, C., Liu, X., Cheng, J., Su, H., Xu, Q., 2022. Association between short-term PM1 exposure and cardiorespiratory diseases: evidence from a systematic review and meta-analysis. Atmos. Pollut. Res. 13 (1), 101254 https://doi.org/10.1016/j.apr.2021.101254.

Mohan, V., Kumar Soni, V., Mishra, R.K., 2024a. Analysing the impact of day-night road traffic variation on ultrafine particle number size distribution and concentration at an urban site in the megacity Delhi. Atmos. Pollut. Res. 15 (4) https://doi.org/10.1016/j.apr.2024.102065.

Mohan, V., Soni, V.K., Mishra, R.K., 2024b. Geographical variability of ultrafine particle concentrations in urban and background regions in India. Urban Clim. 56 https://doi.org/10.1016/j.uclim.2024.102066.

Mullen, N.A., Liu, C., Zhang, Y., Wang, S., Nazaroff, W.W., 2011. Ultrafine particle concentrations and exposures in four high-rise beijing apartments. Atmos. Environ. 45 (40), 7574–7582. https://doi.org/10.1016/j.atmosenv.2010.07.060.

Nelson, B.S., Stewart, G.J., Drysdale, W.S., Newland, M.J., Vaughan, A.R., Dunmore, R. E., Edwards, P.M., Lewis, A.C., Hamilton, J.F., Acton, W.J., Hewitt, C.N., Crilley, L. R., Alam, M.S., Åahin, Ü.A., Beddows, D.C.S., Bloss, W.J., Slater, E., Whalley, L.K., Heard, D.E., et al., 2021. In situ ozone production is highly sensitive to volatile organic compounds in Delhi, India. Atmos. Chem. Phys. 21 (17), 13609–13630. https://doi.org/10.5194/acp-21-13609-2021.

Noble, S.R., Hudson, J.G., 2019. Effects of continental clouds on surface Aitken and accumulation modes. J. Geophys. Res. Atmos. 124, 5479–5502. https://doi.org/10.1029/2019JD030297.

Ohlwein, S., Kappeler, R., Kutlar Joss, M., Künzli, N., Hoffmann, B., 2019. Health effects of ultrafine particles: a systematic literature review update of epidemiological evidence. Int. J. Publ. Health 64 (4), 547–559. https://doi.org/10.1007/s00038-019-01202-7. Springer International Publishing.

Patel, K., Bhandari, S., Gani, S., Campmier, M.J., Kumar, P., Habib, G., Apte, J., Hildebrandt Ruiz, L., 2021. Sources and dynamics of submicron aerosol during the autumn onset of the air pollution season in Delhi, India. ACS Earth Space Chem. 5 (1), 118–128. https://doi.org/10.1021/acsearthspacechem.0c00340.

Paulot, F., Naik, V., Horowitz, L.W., 2022. Reduction in near-surface wind speeds with increasing CO2 may worsen winter air quality in the Indo-Gangetic Plain. Geophys. Res. Lett. 49, e2022GL099039 https://doi.org/10.1029/2022GL099039.

Pohlker, M.L., Zhang, M., Campos Braga, R., Kruger, O.O., Poschl, U., Ervens, B., 2021. Aitken mode particles as CCN in aerosol- and updraft-sensitive regimes of cloud droplet formation. Atmos. Chem. Phys. 21 (15), 11723–11740. https://doi.org/10.5194/acp-21-11723-2021.

Quiros, D.C., Lee, E.S., Wang, R., Zhu, Y., 2013. Ultrafine particle exposures while walking, cycling, and driving along an urban residential roadway. Atmos. Environ. 73, 185–194. https://doi.org/10.1016/j.atmosenv.2013.03.027.

Rajagopal, K., Ramachandran, S., Mishra, R.K., 2023. Roadside measurements of nanoparticles and their dynamics in relation to traffic sources in Delhi: impact of restrictions and pollution events. Urban Clim. 51 (October 2022), 101625 https://doi.org/10.1016/j.uclim.2023.101625.

Ramachandran, S., Rupakheti, M., 2021. Trends in physical, optical, and chemical columnar aerosol characteristics and radiative effects over South and East Asia: satellite and ground-based observations. Gondwana Res. https://doi.org/10.1016/j.gr.2021.09.016.

Reche, C., Querol, X., Alastuey, A., Viana, M., Pey, J., Moreno, T., Rodríguez, S., González, Y., Fernández-Camacho, R., de La Campa, A.M.S., de La Rosa, J., Dall'Osto, M., Pré, t, A.S.H., Hueglin, C., Harrison, R.M., Quincey, P., 2011. New considerations for PM, Black carbon and particle number concentration for air quality monitoring across different European cities. Atmos. Chem. Phys. 11 (13), 6207–6227. https://doi.org/10.5194/acp-11-6207-2011.

Riesenfeld, E., Chalupa, D., Gibb, F.R., Oberdörster, G., Gelein, R., Morrow, P.E., Utell, M.J., Frampton, M.W., 2000. Ultrafine particle concentrations in a hospital. Inhal. Toxicol. 12 (Suppl. 2), 83–94. https://doi.org/10.1080/08958378.2000.11463201.

Schraufnagel, D.E., 2020. The health effects of ultrafine particles. Exp. Mol. Med. 52 (3), 311–317. https://doi.org/10.1038/s12276-020-0403-3. Springer Nature.

Sharma, M., Khare, M., Mishra, R.K., 2024. Air quality changes in Delhi due to open waste burning: an accidental fire in Bhalswa landfill. Int. J. Environ. Sci. Technol., 0123456789 https://doi.org/10.1007/s13762-023-04921-w.

Sebastian, M., Kanawade, V.P., Pierce, J.R., 2021. Observation of sub-3nm particles and new particle formation at an urban location in India. Atmos. Environ. 256 https://doi.org/10.1016/j.atmosenv.2021.118460.

Sengupta, A., Govardhan, G., Debnath, S., Yadav, P., Kulkarni, S.H., Parde, A.N., Lonkar, P., Dhangar, N., Gunwani, P., Wagh, S., Nivdange, S., Jena, C., Kumar, R., Ghude, S.D., 2022. Probing into the wintertime meteorology and particulate matter (PM2.5 and PM10) forecast over Delhi. Atmos. Pollut. Res. 13 (6), 101426 https://doi.org/10.1016/j.apr.2022.101426.

Wiebert, P., Sanchez-Crespo, A., Seitz, J., Falk, R., Phillipson, K., Kreyling, W.G., Möller, W., Sommerer, K., Larsson, S., Svartengren, M., 2006. Negligible clearance of ultrafine particles retained in healthy and affected human lungs. Eur. Respir. J. 28 (2), 286–290. https://doi.org/10.1183/09031936.06.00103805.

Winklmayr, W., Reischl, G.P., Lrndneat, A.0, Berner, A., 1991. A new electromobility spectrometer for the measurement of aerosol size distributions in the size range from 1 to 1000 nm. J. Aerosol.I 22 (3).

World Air Quality Report, 2020. https://www.iqair.com, 41.

Wu, Z., Hu, M., Lin, P., Liu, S., Wehner, B., Wiedensohler, A., 2008. Particle number size distribution in the urban atmosphere of Beijing, China. Atmos. Environ. 42 (34), 7967–7980. https://doi.org/10.1016/j.atmosenv.2008.06.022.

Xie, Y., Zhou, M., Hunt, K.M.R., Mauzerall, D.L., 2024. Recent PM2.5 air quality improvements in India benefited from meteorological variation. Nat. Sustain. https://doi.org/10.1038/s41893-024-01366-y.

Yadav, S.K., Kompalli, S.K., Gurjar, B.R., Mishra, R.K., 2021. Aerosol number concentrations and new particle formation events over a polluted megacity during the COVID-19 lockdown. Atmos. Environ. 259 (March), 118526 https://doi.org/10.1016/j.atmosenv.2021.118526.

Yadav, S., Sharma, R., Kumar, S., Agarwal, A., Mohan, V., Mishra, R.K., Shukla, A., 2021. Urban air pollution reduction: evidence from phase-wise analysis of COVID-19 pandemic lockdown. Arabian J. Geosci. https://doi.org/10.1007/s12517-021-07777-x.

Zhao, J., Birmili, W., Wehner, B., Daniels, A., Weinhold, K., Wang, L., Merkel, M., Kecorius, S., Tuch, T., Franck, U., Hussein, T., Wiedensohler, A., 2020. Particle mass concentrations and number size distributions in 40 homes in Germany: indoor-to-outdoor relationships, diurnal and seasonal variation. Aerosol Air Qual. Res. 20 (3), 576–589. https://doi.org/10.4209/aaqr.2019.09.0444.

Zhu, Y., Hinds, W.C., Kim, S., Shen, S., Sioutas, C., 2002. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. Atmos. Environ. 36.

Electrical Engineering
https://doi.org/10.1007/s00202-024-02654-8

ORIGINAL PAPER

Solar PV integrated simplified multilevel inverter configuration for power quality improvement using multilayer Gamma filter

Praveen Bansal[1] · **Alka Singh**[2]

Received: 23 September 2023 / Accepted: 29 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Power quality (PQ) issues have intensified due to the rapid integration of renewable sources into the utility grid. An effective control strategy is imperative to address these problems. This paper proposes a novel approach by replacing conventional 2-level inverters with a simplified 5-level multilevel inverter (SMLI) as a shunt active power filter (SAPF) unit. The SMLI exhibits superior performance, including low total harmonic distortion in voltage, reduced electromagnetic interference, and enhanced system flexibility. Further, a multilayer Gamma filter is utilized to control SAPF and extract fundamental component from nonlinear load. The developed control not only improves the power factor at the supply end but also resolves other PQ issues. The novelty of the paper is the innovative utilization of the Gamma filter and closed loop implementation of SMLI system which can operate effectively in two modes, viz. during day and night. Furthermore, the system is also tested under partial shading conditions. The solar PV array integrated at the DC link of the SMLI supplies the load during the daytime. However, at night, when the solar PV array's power is unavailable, the load's demand is met solely by the grid. The SAPF functioning ensures an improvement in power quality. The proposed and developed Gamma filter control action performs quick adaptive estimation under diverse load profiles. Furthermore, reduced memory requirement during operation guarantees optimal performance under dynamic circumstances involving fluctuating solar irradiation and load parameters. The OPAL-RT real-time simulator has been used to both simulate and experimentally validate the results. The application of this research is extensively useful in grid-connected PV system for enhancing PQ issues in distribution systems with integrated renewable energy sources.

Keywords Active filter · Adaptive control · Multilevel inverter · Nonlinear loads · Power quality

1 Introduction

Environmental issues and energy demand have been steadily rising along with improvement in living standards of people worldwide. The quest for environmentally friendly alternative energy sources has garnered global research interest. Due to their abundance, accessibility, and safety, distributed energy sources such as solar photovoltaic (PV), wind, hydro, and others are becoming increasingly popular. These renewable energy systems offer a sustainable solution to meet the growing energy needs while mitigating the adverse effects of conventional fossil fuels on the environment. [1–3].

Energy conversion and transmission devices that are connected to the grid serve as a bridge between the grid and renewable energy sources (RESs). Focus on maximum power point tracking (MPPT) algorithms, inverter topologies, grid synchronization, filter design, and islanding detection techniques for grid-tied PV inverters has been presented [4]. Several PQ issues in grid-connected solar photovoltaic systems have been presented under various conditions [5]. The PQ problems mainly result due to the nonlinear characteristics of the main loads connected to the RES-interfaced grid system. The system gets burdened with the harmonic current and reactive power demand of nonlinear loads, which reduces the system's efficiency. These nonlinear loads also cause voltage distortion and poor power factor. Several methods have

✉ Praveen Bansal
pbansal444@gmail.com

Alka Singh
alkasingh.eed@gmail.com

[1] Centre for Internet of Things, Madhav Institute of Technology and Science, Gwalior 474001, India

[2] Department of Electrical Engineering, Delhi Technological University, New Delhi, Delhi 110042, India

Published online: 25 August 2024

been suggested for improving PQ problems in distribution systems and development of improved grid synchronization techniques is discussed in [6].

In order to assure dependability, security, and efficiency of power supply [7], it is pertinent to maintain excellent power quality levels in modern electrical networks. A number of Flexible AC Transmission System (FACTS) devices [8, 9] are currently being developed and used for enhancing the quality of the electrical grid power. A large family of devices, including active and hybrid power filters [10, 11], Static compensators (STATCOM) [12, 13], Static VAR Compensators (SVC), Unified Power Flow/Quality Controllers (UPFC/UPQC), and Dynamic Voltage Restorers (DVR) [14], have been developed. These configurations have been implemented using a basic voltage source inverter (VSI) unit to inject the compensating current which eliminates harmonics and enhances power quality. In order to generate compensating currents that are 180^{o}out of phase with the harmonics, the SAPF is controlled to eliminate all harmonic components of load.

The invention of inverters initially commenced with conventional 2-level inverter topologies, which suffered acute voltage stress across circuit components in addition to high total harmonic distortion (THD) due to the low ratings of the individual components [15]. Compared to the multilevel inverters (MLI), conventional 2-level inverters have drawbacks. Due to the poor waveform quality of the square wave output voltage produced by 2-level inverters, the equipment is subjected to severe harmonic distortion and voltage stress. Additionally, conventional 2-level inverters encounter higher switching losses, constrained voltage ranges, increased switch stress, and restrictions on high-voltage applications. MLIs, on the other hand, offer better waveform quality, reduced switching losses, higher voltage levels, lower stress on switches, and are more suitable for high-voltage applications [16]. Due to these benefits, multilevel inverters are recommended for applications that require larger power outputs, better voltage quality, and increased system effectiveness [17].

MLI topologies are determined by the type, quality, and high/medium power requirement of the circuit and are categorized as Switched Capacitor (SC) [18], Cascade H-Bridge (CHB), Flying Capacitor (FC), Neutral Point Clamp (NPC) [19] and Flying Capacitor (FC) [20] multilevel inverters. For NPC and FC designs, a significant problem is the voltage in Direct Current (DC) link capacitors brought on by the limited switching states. While in CHBs, requiring several DC sources increases the topology's cost and size. But the main drawbacks of the multilevel configurations are their circuit complexity, which necessitates a large number of power switches that must be commutated in a precise order by a special (and complicated) modulator circuit. They also necessitate a large number of auxiliary dc levels, which are typically provided by a bulky array of capacitive voltage dividers or by independent supplies. Another element that raises the complexity of the modulator circuit is how to maintain a proper balance of the different dc link voltages. Thus, the paper targets a simplified inverter configuration [21] and this topology employed in the power stage offers a considerable advantage in terms of lower component count and reduced layout complexity. The topology utilizes no extra diodes or capacitors and reduces the number of main power switches needed by roughly 60%. This leads to reduced overall cost of the system, and detailed comparison of SMLI with other topologies is also discussed in Sect. 3.

In the proposed system, two PV arrays have been interconnected at the SMLI using single-stage single-phase grid-connected system. The single stage offers various advantages as compared to double stage conversion, viz. simple, cost-effective, high efficiency, scalability, seamless grid connections, space efficiency, decreased power losses, and operational flexibility. The proposed system can operate in two modes, viz. day and night. During daytime, the solar PV arrays are able to inject the active power to the loads/grid. In case, the load demand is insufficient, solar PV feeds extra power to the grid which results in grid current being out of phase with grid voltage. However, at night, the solar PV arrays are dysfunctional and SMLI acts as SAPF performing load compensation. It is controlled as a harmonic compensator unit and fulfills the reactive power requirement of the load.

Numerous studies on control methods are available in literature [22–30] for single phase and three-phase power systems which address harmonic detection and elimination techniques and ensure the grid current is sinusoidal in nature. The currently used grid-integrated photovoltaic (PV) inverter control methods need to be improved in order to guarantee safe and dependable operation. These fundamental extraction techniques can generally be divided into two categories from a theoretical standpoint: (a) frequency-domain techniques and (b) time-domain techniques. The satisfactory performance of SAPF depends on how accurately and effectively a control algorithm estimates the fundamental current component of load current. Hence, the selection of control algorithm/filter plays a crucial role in design. Literature review suggests various time domain algorithms, viz. Power Balance Theory (PBT) and Instantaneous Reactive Power Theory (IRPT) [22, 23]. These conventional algorithms are quite popular and have widely been used to estimate the fundamental load current component. Likewise, in the frequency domain, Kalman filter [24], discrete Fourier transform [25]-based algorithms have been applied. Table 1 shows the comparative analysis of various time domain, frequency domain, conventional and advanced filters proposed in the literature [26–34].

Table 1 Comparative analysis of various adaptive filter/algorithms for harmonic elimination

Name of algorithm/filter	Reference	Advantages	Disadvantages	Complexity	Convergence rate
Hilbert–Huang transform-based online differential relay algorithm	Biswal et al. [26]	Effective differential protection	Real-time processing required	High	High
Goertzel algorithm-based shunt active power filter with sliding mode controller	Karvekar et al. [27]	Robust control	Increased complexity	Medium	High
Back-propagation control algorithm for power quality improvement	Singh and Arya [28]	Effective PQ improvement	Requires training and iteration	High	High
Neural network-based conductance estimation control algorithm	Arya and Singh [29]	Effective estimation	Requires training and updates	Medium	Medium
Adaptive-RBFNN algorithm	Mittal et al. [30]	Adaptive learning capabilities	Higher computational requirements	High	Fast due to adaptive learning mechanisms
Gamma-filter: a new class of adaptive IIR filters	Principe et al. [31]	Adaptive filtering	Performance dependent on params	Medium	High
Projected kernel least mean p-power (PKLMP) Control technique	Manasi et al. [32]	Enhanced performance of grid integrated PV systems	Requires precise parameter tuning	Moderate	Moderate
Hardware-in-the-loop testing of seamless interactions of multi-purpose grid-tied PV inverter	Mansour et al. [33]	Facilitates testing/validation	Integration complexities	Medium	High
Delta-bar-delta neural-network-based control approach for power quality improvement	Shukl and Singh [34]	Effective for solar-PV systems	Requires parameter tuning	High	High

Several adaptive filtering algorithms have also been developed for the control of SAPF and are also tabulated. These adaptive algorithms are able to self-adjust their parameters in accordance with system dynamics and that their implementation is straightforward. However, there is a tradeoff to decide the correct value of step size in the algorithms for optimum performance.

Therefore, for accurate estimation of fundamental component of load current under solar irradiance changes and dynamic load conditions, the proposed system is controlled using Gamma filter [31]. The novelty of this paper is that it considers a reduced device count-based multilevel inverter for PQ improvement for a PV interfaced grid-connected system. The Gamma filter gives better dynamic performance and has combined characteristics of finite impulse response (FIR) and infinite impulse response (IIR).

The main novelties of the research are:

1. *Innovative Utilization of Gamma Digital Filter* The research introduces a novel application of the Gamma digital filter for power quality enhancement in grid-connected solar PV systems. This innovative approach has not been explored in previous literature and thus offers a unique solution to address PQ issues.
2. *Closed-Loop Simplified Multilevel Inverter (SMLI) as a SAPF Unit* This study fills a gap in existing literature by proposing the use of a simplified 5-level multilevel inverter (SMLI) as a shunt active power filter (SAPF) unit. Compared to conventional 2-level inverters, the

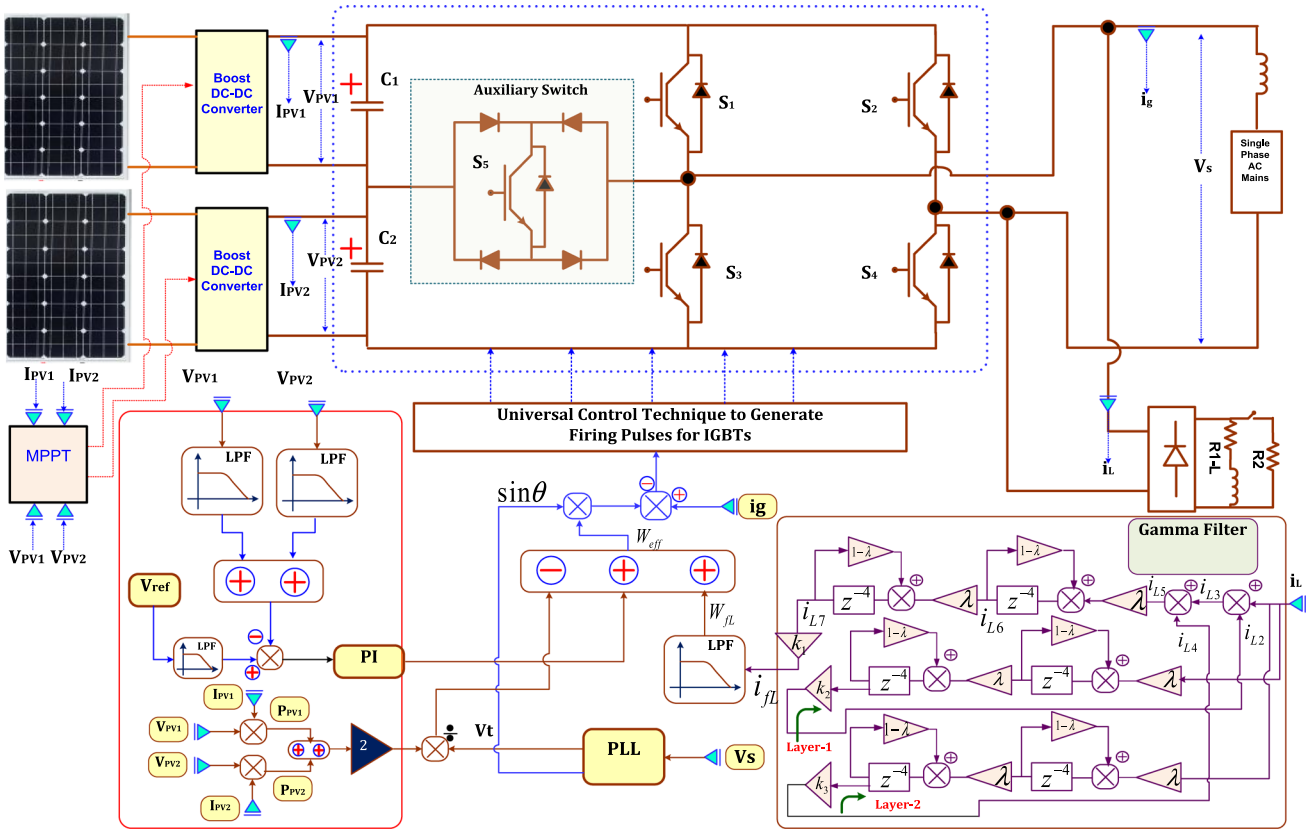


Fig. 1 Block diagram of Gamma filter-based SMLI

SMLI demonstrates superior performance, including low total harmonic distortion (THD) in voltage and reduced electromagnetic interference (EMI), enhancing system flexibility.

The major contributions of the paper are as follows:

1. The innovative utilization of the Gamma digital filter for PQ enhancement in grid-connected solar PV system is explored.
2. The closed-loop operation of SMLI multilevel inverter as a SAPF unit is discussed in detail.
3. Incorporation of a PV feed-forward term in the control approach to enhance dynamic response and improved operation of the Gamma filter control technique.
4. Presentation of simulation and experimental results for both daytime and night time scenarios, encompassing varying solar irradiance and dynamic load conditions as well as partial shading conditions.
5. Confirmation of system functionality in compliance with IEEE-519 and IEEE-1547 standards, with grid current harmonic distortion consistently below 5%.

The paper discusses the development of MATLAB/Simulink models controlled using Gamma digital filter [31], and validation of simulation results by experimental results obtained using OPAL-RT (model no. 4510) under various grid conditions.

The paper is organized into six separate sections. A thorough description of the system and layout of the system is given in Sect. 2. A comparison with traditional MLI and analysis of the Simplified Multilevel Inverter (SMLI) using the Universal Control Technique (UCT) is presented in Sect. 3. The paper explores the mathematical modeling of control algorithm in Sect. 4. The performance under dynamic solar irradiation is discussed in Sect. 5. In Sect. 6; the experimental results are summarized. A comparison with various filters is presented in Sect. 7. The results of the paper are finally concluded in Sect. 8.

2 System description

The generalized control block diagram of the proposed system is shown in Fig. 1. A single phase source of 110 V, 50 Hz grid feeds the nonlinear load. At the point of common coupling (PCC), a reduced device-based 5-level SMLI is connected with the grid using an interfacing inductor (L_f). Two DC link capacitors of voltage magnitude V_{DC1} and V_{DC2} are connected as DC link voltages in the SMLI configuration.

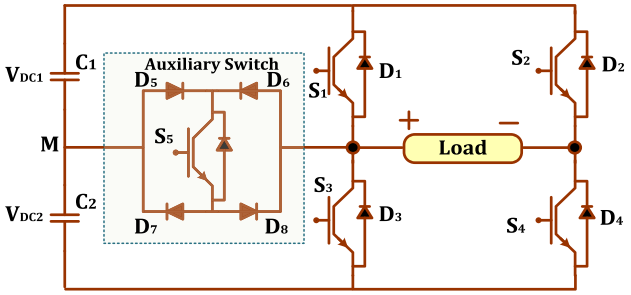


Fig. 2 Generalized block diagram of simplified multilevel inverter (SMLI)

These DC link capacitances have been connected to the PV array modules. The SMLI has been controlled using Universal Control Technique (UCT) discussed in the subsequent section. The SMLI is suitably controlled such that it is able inject a compensating current to eliminate the harmonics generated by the load current. The controller is realized using a Gamma filter which estimates the fundamental component of load current and further generates the reference current. This reference current is further compared with alternate phased opposition (APOD) pulse width modulation scheme for the generation of firing pulses for the SMLI power switches.

3 Analysis of simplified multilevel inverter (SMLI) with universal control technique (UCT)

The topology of simplified multilevel inverter is shown in Fig. 2. It consists of two symmetrical DC link voltages represented by DC link capacitors in proposed system, forming a voltage divider circuit. Power switch S_5 and diodes D_5–D_8 form an auxiliary circuit and S_1–S_4 represent insulated gate bipolar transistor (IGBT) with anti-parallel diodes D_1–D_4 as per the polarity shown in Fig. 2. The voltage magnitude of capacitors is equal, i.e., $V_{DC1} = V_{DC2}$; therefore, the topology is considered to be symmetrical configuration.

The most widely used conventional topology for 5-level output is cascaded H-Bridge MLI (CHB-MLI) because of its simple structure and quick assembly as compared to other conventional FCMLI and NPC-based converters. In case of higher level inverters, the dominant role is played by power switches, required gate drive circuit, volume occupied by the circuit and installation area. More complex circuits result in increased cost of the system and are also likely to suffer from reliability issues due to higher number of components. Hence, reduced device-based SAPF becomes an attractive alternative. In this paper, a SMLI has been used as SAPF unit and it is suitably controlled to inject compensating current at the PCC. A SMLI offers various advantages over conventional MLIs in terms of lower dv/dt stress across switches, lower THD in output voltage, reduced number of component requirement in terms of power switches, gate drive circuits and simplicity. The working principle of 5-level SMLI is depicted in Fig. 3, and the switching table is summarized in Table 2. The comparison table for different symmetrical configurations of 5-level MLIs is tabulated in Table 3.

The system cost can be calculated by using an index N_{cL}(count of components used in the system per voltage level), where N_{poles} = number of voltage levels; N_{aux} = number of Auxiliary switches; N_{IGBT} = number of power switches; N_D = number of diodes; N_{cap} = number of capacitors; N_{dc} = DC link sources; N_T = number of transformers. The detailed comparison table showing the total components needed in SMLI with conventional and other reduced device count MLI topologies is tabulated in Table 2

$$N_{cL} = \frac{N_{dc} + N_{cap} + N_T + N_{IGBT} + N_D + N_{aux}}{N_{Level}} \tag{1}$$

3.1 Comparison of 5-level SMLI with conventional and other RDC topologies

The parameter N_{cL} aids in obtaining the most cost-effective topology, which employs the ideal number of components and reduces the overall cost of the system. For the SMLI topology as compared to other MLI topologies, the N_{CL} index is found to be **3.4** as shown in Fig. 4. This value is the least as compared to other topologies indicating the proposed SMLI is cost effective and easy to implement.

3.2 Power loss analysis and capacitor boundary conditions

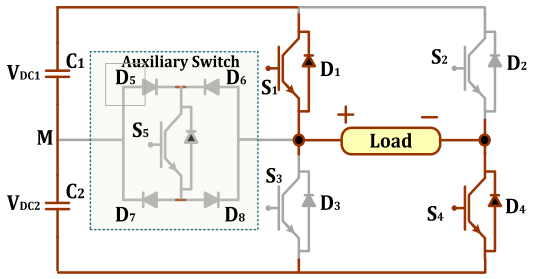
The maximum charge drawing capacity of capacitor can be calculated by [31]

$$\Delta Q_{Ck} = \int_{t_{1k}}^{t_{2k}} i_{out} \sin(2 f_o \pi t - \varphi)\, dt \tag{2}$$

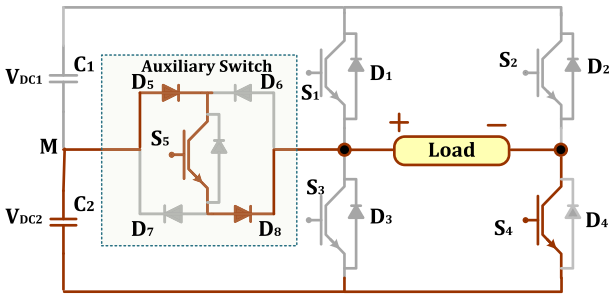
i_{out} represents the amplitude of load current and f_0 is the output frequency and the time period is considered to be t_{1k} to t_{2k}. The value of both the capacitors is same, due to the symmetrical configuration of inverter, and it is mathematically represented as

$$C_x = \frac{\Delta Q_{ck}}{\alpha V_{DC1}} \tag{3}$$

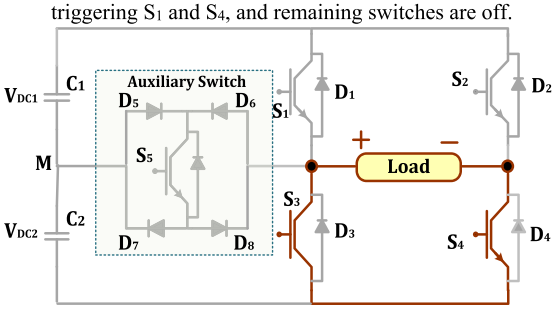
where α is considered to be acceptable voltage ripple and DC link voltage of capacitance is considered to be also equal, i.e., VDC1 = VDC2.



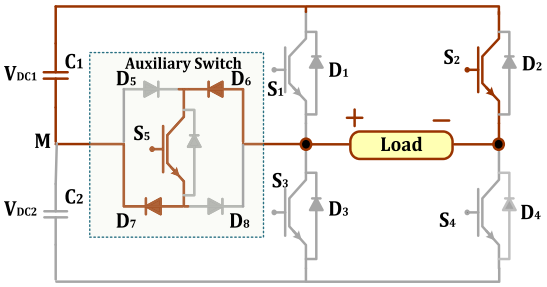
***Case A*: Maxima level output voltage ($2V_{DC}$):** It can achieved by triggering S_1 and S_4, and remaining switches are off.



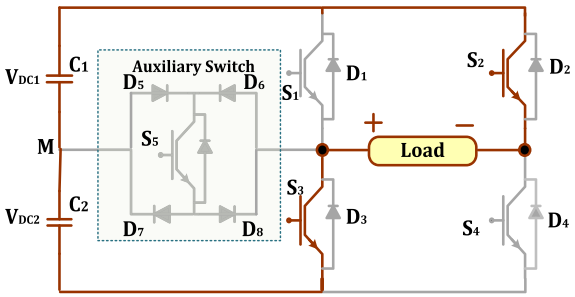
***Case B*: Positive Half level output voltage ($+V_{DC}/2$):** On triggering S_5 and S_4 ,diodes D_5 and D_8conduct while remaining switches are off.



Case c: Zero Level output voltage: By turning ON S_3 and S_4 and remaining switches are off then the output across load is found to be zero



Case d:Negative Half level output voltage ($-V_{DC}/2$): Turning ON switch S_5 connects load to the point M via diodes D_6 and D_7. Remaining switches are OFF



Case e: Minimal level output voltage ($-2V_{DC}$): Achieved by triggering S_2 and S_3, and remaining switches are off.

Fig. 3 Different switching combinations of the proposed SMLI

Table 2 Switching table of SMLI (5-Level)

Cases	S1	S2	S3	S4	S5	Case
a	√	x	x	√	x	2 Vdc
b	X	x	x	√	√	Vdc
c	X	x	√	√	x	0
d	X	√	x	x	√	− Vdc
e	√	x	x	√	x	− 2 Vdc
ON duration (times)	**2**	1	1	4	**2**	–

Table 3 Component count comparison of SMLI (5-Level)

Reference	N_{dc}	N_{cap}	N_{IGBT}	N_D	N_{poles}	N_{aux}	N_T	N_{cL}
CHB-MLI [18]	02	00	08	08	5	–	–	3.6
NPC [19]	01	04	08	20		–	–	6.6
FCMLI [20]	01	04	08	08		–	–	4.2
Hybrid MLI [35]	02	02	06	10		01	–	4.2
T-Type [36]	01	02	08	08		–	–	3.8
SMLI [21]	01	02	04	09		01	–	3.4

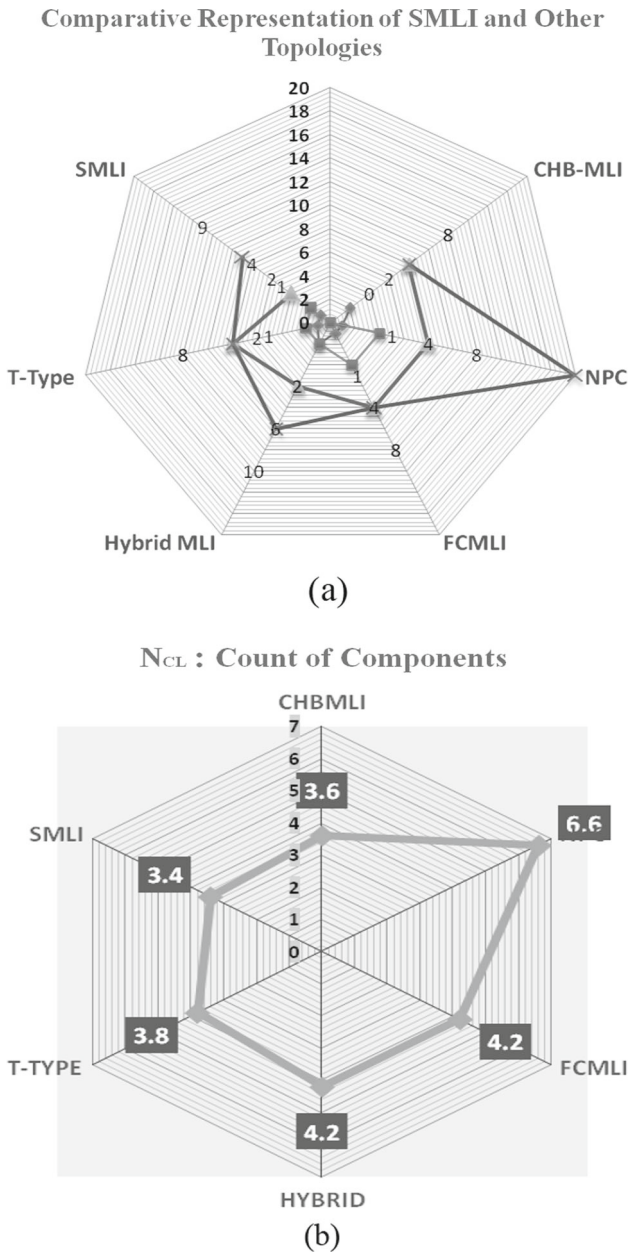


Fig. 4 N_{cL} analysis of SMLI and other topologies

The estimation of losses in the inverter can be done in three states: (i) ON State, (ii) OFF State and (iii) switching state of IGBTs. It is assumed to have a negative leakage current during OFF state; therefore, the switching losses are to be considered only for ON state. In SMLI, there are five IGBTs, four anti-parallel diodes and four diodes in the auxiliary circuit. As per the switching Table 2, the control of the SMLI developed using the Universal Control Technique (UCT) is discussed in the next section. The losses that occur in the system must be estimated. The generalized mathematical expression for the conduction of switches and diodes is represented as

$$P_{C,\mathrm{IGBT}} = \sum_{j=1}^{N_{\mathrm{IGBT}}} \frac{1}{2\pi} \int_0^{2\pi} \left(V_{\mathrm{ON,IGBT}}\, i(t) + R_{\mathrm{IGBT}}\, i^{\beta}(t)\right) \mathrm{d}t \tag{4}$$

$$P_{C,\mathrm{diode}} = \sum_{j=1}^{N_{\mathrm{Idiode}}} \frac{1}{2\pi} \int_0^{2\pi} \left(V_{\mathrm{ON,diode}}\, i(t) + R_{\mathrm{diode}}\, i^{2}(t)\right) \mathrm{d}t \tag{5}$$

Total conduction loss is defined as

$$P_{\mathrm{cond.}} = P_{c,\mathrm{IGBT}} + P_{c,\mathrm{Diode}} \tag{6}$$

Another loss taken into consideration is the switching loss, here individual switching loss across each power switch is added to get total switching losses for the SMLI configuration. Linear approximation of current and voltage is to be considered during switching period. The equation representing the switching losses is given as

$$P_{\mathrm{SW}} = \left[\sum_{x=1}^{N_{SW}} \left(T_{\mathrm{ON}} E_{\mathrm{ON}} + T_{\mathrm{off}}\, E_{\mathrm{off}}\right)\right] * f \tag{7}$$

T_{ON} and T_{OFF} are switching times; f is the switching frequency and energy losses during ON and OFF period are E_{ON} and E_{OFF}. The total losses in the system are estimated using the expression given below:

$$P_{\mathrm{cond.}} = P_{c,\mathrm{IGBT}} + P_{c,\mathrm{Diode}} + P_{\mathrm{SW}} \tag{8}$$

In this paper, SMLI has been controlled using UCT. This technique may be suitably applied to control different types of MLI. The number of voltage level is considered to be N_{level}, and in the proposed system, the minimum level is taken as $N_{\mathrm{level}} \geq 5$(including zero level). The generalized block diagram of the UCT is shown in Fig. 5.

3.3 Universal control technique (UCT

In UCT, a reference signal (generated from reference current) $T_{\mathrm{ref}}(t)$ is compared with triangular carrier modulating signal $T^{+}_{\mathrm{car},k}(t)$ for above zero reference and $T^{-}_{\mathrm{car},k}(t)$ for below zero reference when k is varied from 0 to N, depending upon the number of voltage levels. The output of the comparator is considered to be $T^{+}_{\mathrm{out},k}(t) = 1$ for all carriers' signals above zero reference ($T_{\mathrm{ref}}(t) \geq T^{+}_{\mathrm{carr},k}(t)$) otherwise output is (0). Similarly, for the carrier signals below zero reference ($T_{\mathrm{ref}}(t) \geq T^{-}_{\mathrm{carr},k}(t)$) the output is $T^{-}_{\mathrm{out},k}(t) = 0$, otherwise minus (−) one.

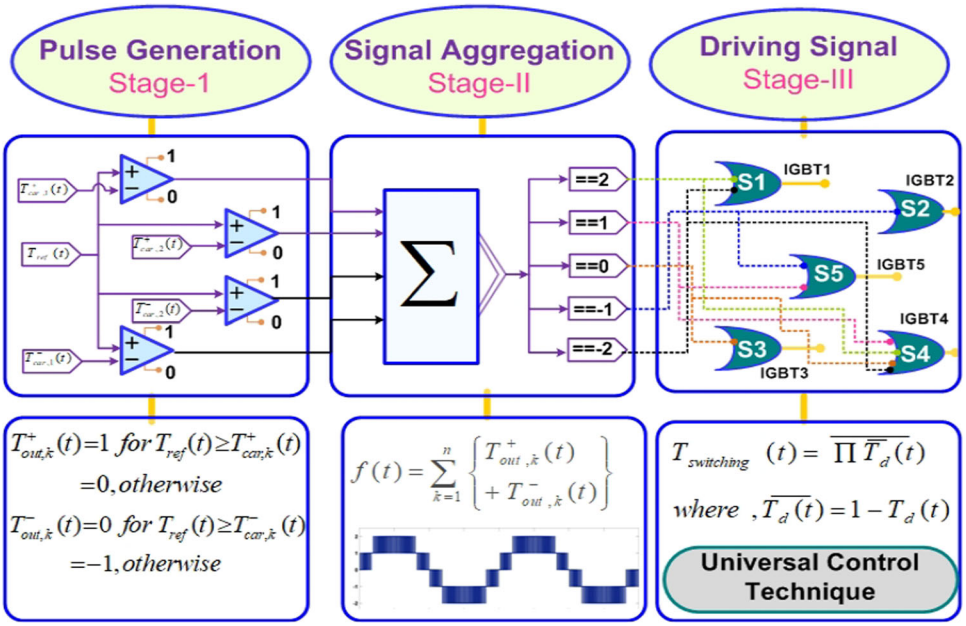


Fig. 5 Generalized block diagram of Universal control technique

The obtained signal is aggregated as $T_{\text{agg}}(t)$, giving summation of $\sum_{k=1}^{N}\left\{T^{+}_{\text{out},k}(t)+T^{-}_{\text{out},k}(t)\right\}$ The aggregated signals exhibit the expected output voltage level in the topology, which in this case is a 5-level output.

Now, to drive the switching signals of power switches the Boolean function $\overline{T}_d(t)$ is derived from the aggregated signal $T_{\text{agg}}(t)$, for a particular N-level, the switching function $T_{\text{swithcing}}(t)$ extracts the switching signal for particular switch whose stage is active (i.e. ON) and fed to OR gate and further sent to power switches for triggering.

4 Mathematical modeling of control algorithm and reference current generation

A multilayer adaptive Gamma Filter has been implemented in this paper to extract the fundamental component of load current. The Gamma filter combines the effect of infinite impulse response (IIR) and finite impulse response (FIR) characteristics. The filter involves a Gamma delay operator of $\frac{\lambda}{z-(1-\lambda)}$ with a gain feedback loop of $(1-\lambda)$. In time domain, the Gamma filter is represented as

$$T(n)=\sum_{k=0}^{n} w_k\, x_k(n) \tag{9}$$

where $x_k(n)=\{(1-\lambda)\,x_k(n-1)+\lambda\,x_{k-1}\,(n-1)\lambda=1, 2, 3.....k\}$ $x_0(n)\equiv x_k(n)$ is the input signal and $T(n)$ is the filter output signal, w_0 , w_1 , $w_3.....w_k$ and λ denote the adaptive filter parameters.

The Gamma filter transfer function for k^{th} layer is defined as

$$G_k(z)=\left\{\frac{\lambda}{z-(1-\gamma)}\right\}^{k} \tag{10}$$

In the proposed system, a fourth-order adaptive Gamma filter has been implemented to extract the fundamental component of load current. The filter is capable of eliminating inter and sub-harmonics of the input parameter. The expression for the fundamental load current (I_{fL}) with respect to the actual load current can be derived from the intermediate signals. In present case i_{L6} can be acquired from signal i_{L5}, represented as

$$i_{L6}=\frac{\lambda z^{-4}}{1+(1-\lambda)z^{-4}}*i_{L5} \tag{11}$$

Similarly, the layer 2 and layer 3 as shown in Fig. 6 will be utilized to acquire the intermediate signal i_{L5} as expressed as follows

$$i_{L5}=i_{L3}-i_{L4} \tag{12}$$

$$i_{L3}=i_{L}-i_{L2} \tag{13}$$

Likewise in the filter, the output current i_{L2} and i_{L4} can be evaluated from layer-2 and layer-4. Now the transfer function of fundamental load current (i_{Lf}) and actual load current (i_L) is represented as

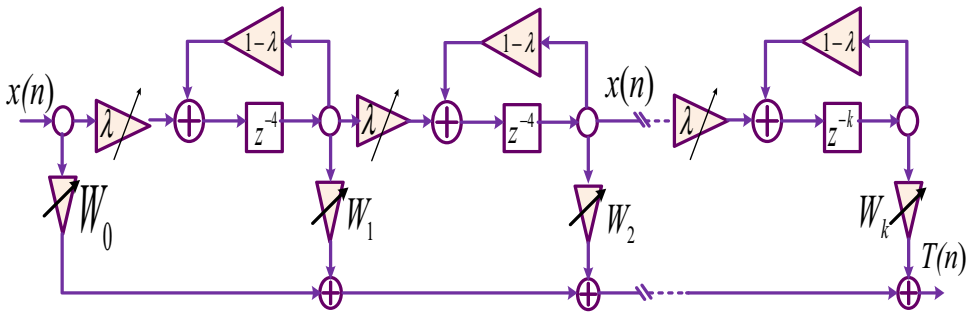


Fig. 6 Generalized representation of Gamma Filter

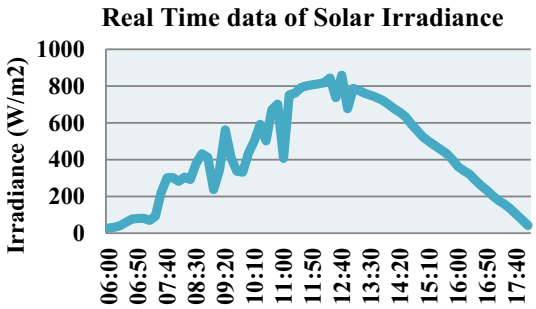


Fig. 7 Real-time solar radiation obtained in Delhi /NCR region

$$\frac{i_{Lf}}{i_L} = k_1 \left\{ \left(1 - \frac{k_2^2 \lambda^2 z^{-8}}{1 + (1-\lambda) z^{-4}}\right) \left(1 - \frac{k_3^2 \lambda^2 z^{-8}}{1 + (1-\lambda) z^{-4}}\right) \left(\frac{\lambda^2 z^{-8}}{1 + (1-\lambda) z^{-4}}\right) \right\} \tag{14}$$

For simplicity, it is assumed to have constant $k_1 = k_2 = k_3 = 1$, then the transfer function can be expressed as

$$\frac{i_{Lf}}{i_L} = \frac{\lambda^2 z^{-8} + 2\lambda^3 z^{-12} + (\lambda^4 - \lambda^3 - \lambda^2) z^{-16} + (2\lambda^5 - \lambda^4) z^{-20} + \lambda^6 z^{-24}}{(1 + (1-\lambda) z^{-4})^3} \tag{15}$$

The stability of Gamma filter depends upon the choice of parameter λ. When $\lambda = 1$, the Gamma filter gets converted into Widrow's adaptive transversal filter. Therefore, for a stable operation its range is considered to be $0 < \lambda < 2$. The obtained output current is considered to be active fundamental component of load current $w_{\mathrm{fL}}(\mathrm{A})$ and the unit vector template has been derived by sensing the grid voltage $V_g(V)$.The grid amplitude voltage $V_t(V)$ is calculated by computing the in-phase (v_{gp}) and quadrature (v_{gq}) voltage component given by

$$V_t = \sqrt{v_\alpha^2 + v_\beta^2} \tag{16}$$

The amplitude voltage $V_t(V)$ is further used to calculate the unit vector template as given by

$$u_p = \frac{v_{gp}}{v_t} = \sin \omega t \tag{17}$$

The obtained unit template is now used to generate reference current. Now, to control the DC link voltages of V_{DC1} and V_{DC2}, the sensed DC voltages of SMLI are added and compared with reference DC voltage V_{ref} of 200 V. The generated voltage DC error (V_{DE}) is fed to the proportional and integral controller to obtain output voltage W_{PDC} represented as

$$W_{PDC} = W_{DC}(n-1) + k_{pd} \{V_{DE}(n) - V_{DE}(n-1)\} + k_{id} V_{DE}(n) \tag{18}$$

where kpd and kpi represent the proportional and integral gains. For obtaining consistent and stable power from the PV arrays, feed forward term (FF), IPff(A) is used in calculations for compensating variations in solar irradiance. The FF term is derived by sensing the output power of both PV arrays and amplitude of sensed grid voltage, represented as

$$P_{PV} = P_{PV1} + P_{PV2} \tag{19}$$

$$W_{pff} = \frac{2 P_{PV}}{V_t} \tag{20}$$

Finally, the reference current component W_{ref} is computed by adding the estimated fundamental load current component W_{fL} with loss term W_{PDC} and subtracting the feed forward component of PV arrays W_{pff}, as represented by

Table 4 Simulation parameters

S. No	Parameters	Symbol	Value used in simulation
1	Single phase AC power supply	V_g (V)	110 V, 50 Hz
2	DC link capacitors	C_{DC1}, C_{DC2}	3000 μF each
3	Interfacing inductor	L_f	3 mH
4	Reference DC link voltage	E_{ref}	200 V
5	PI gains	k_p and k_i	1.22, 0.6
6	Open circuit voltage	V_{OC1} and V_{OC2}	70.6 V
7	Short circuit current	I_{SC1} and I_{SC2}	18.4 A

$$W_{\text{eff}} = W_{\text{PDC}} + W_{\text{fL}} - W_{\text{pff}} \tag{21}$$

Equation (21) is multiplied by the unit vector template to generate the reference current $i_g^*(A)$, given as

$$i_g^* = W_{\text{eff}} * u_p \tag{22}$$

The obtained reference current is subtracted from the actual source current which is further compared with phase shifted PWM scheme implemented with universal control scheme to generate firing pulses for IGBTs of SMLI.

5 Simulation outcome and discussion

The control method suggested in this study was developed and tested in the MATLAB/Simulink environment of version 2018b using built-in libraries. The specified model was simulated under typical utility/grid voltage settings (110 V-rms, 50 Hz). A single phase nonlinear load was taken into account in the simulation investigations. In addition, two solar photovoltaic (PV) arrays of rating 900 W each have been used. Through a SMLI, this PV configuration was directly connected to the DC-bus capacitors to provide active power to the source. Using coupling inductors, the SMLI is connected to the utility grid at the point of common coupling (PCC) and is suitably controlled to act as shunt compensator. The proposed control methodology has been developed to ensure power quality enhancement while seamlessly transferring the extracted solar power to the grid/load at unity power factor (UPF). The parameters used in the simulation are presented in Table 4.

The performance of the proposed controller was evaluated and assessed by testing the designed system under varying solar irradiance levels and dynamic load variations. The concurrent process of active power injection and power quality enhancement was monitored using diverse scope meters, and the results for the first few cycles were analyzed to identify significant findings.

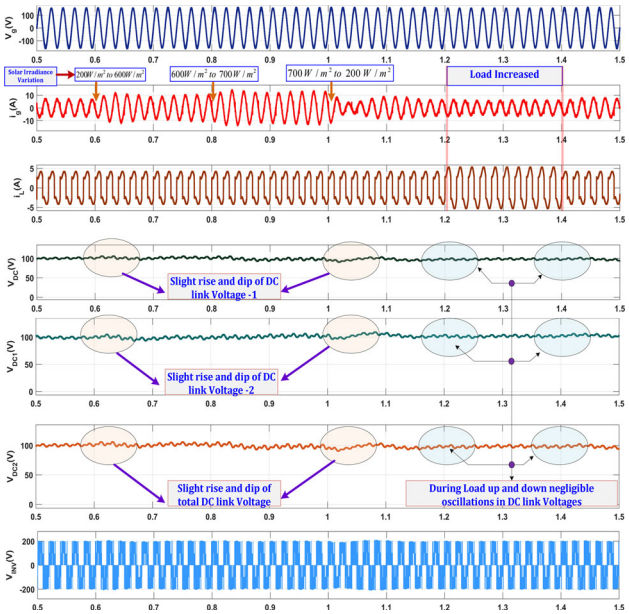


Fig. 8 Dynamic results of grid voltage V_g(V), grid current i_g(A), load current i_L(A), Dc link voltages V_{DC1}(V), V_{DC2}(V), total DC link voltage V_{DC}(V) and inverter output voltage V_{inv}(V) under solar irradiance and load variation during $t = 1.2$–1.4 s

5.1 Variable solar irradiance condition

In Mode I, i.e., during the daytime, real-time irradiation data have been taken into consideration for testing. The real-time data of solar irradiation, taken from 6:00 AM to 6:00 PM, are shown in Fig. 7 [37]. It is clearly observed that the obtained data have nonlinear characteristics. Therefore, the average reading of irradiation obtained from morning to evening is divided into four equal intervals of three hours each for the simulation. Table 5 shows the average irradiance readings for each of the four slots which are considered as 200 W/m^2, 600 W/m^2, 700 W/m^2 and 200 W/m^2, respectively.

The proposed controller's performance is assessed under varying solar irradiance conditions, as indicated in Fig. 8. The solar irradiance level experiences sudden changes during testing. It increases from 200 to 600 W/m^2 at $t = 0.3$ s, 600–700 W/m^2 at $t = 0.8$ s and then decreases back to 200 W/m^2 at $t = 1$ s. Figure 8 shows the waveforms of grid voltage V_g(V), grid current i_g(A), load current i_L(A) and total DC link voltage V_{DC}(V), individual DC link voltages V_{DC1} and V_{DC2}. Furthermore, the performance is also tested under dynamic load variation introduced from $t = 1.2$ to $t = 1.4$ s.

During grid integration, the source current traces become purely sinusoidal with a fast response. This faster response is also evident during variable solar irradiance conditions.

Table 5 Average value of real-time data of solar irradiation

S. No	Duration of interval	Average reading obtained	Used in simulation	Simulation time
1	06:00–09:00am	196.15 W/m^2	200 W/m^2	$t = 0.6$ s
2	09:00 am–12:00 noon	584.77 W/m^2	600 W/m^2	$t = 0.8$ s
3	12:00 noon–03:00 pm	711.55 W/m^2	700 W/m^2	t = 1
4	03:00–06:00 pm	274.22 W/m^2	200 W/m^2	–

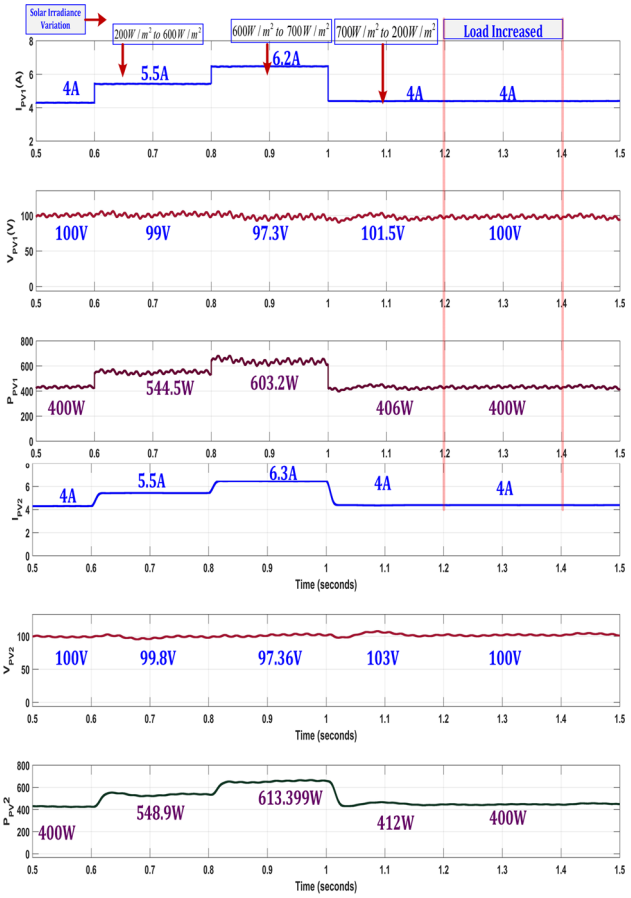


Fig. 9 Solar PV array parameters, viz. PV array currents I_{PV1}, I_{PV2}, P_{PV1} and I_{PV2}, V_{PV2} and P_{PV2} under solar irradiance variations and load changes

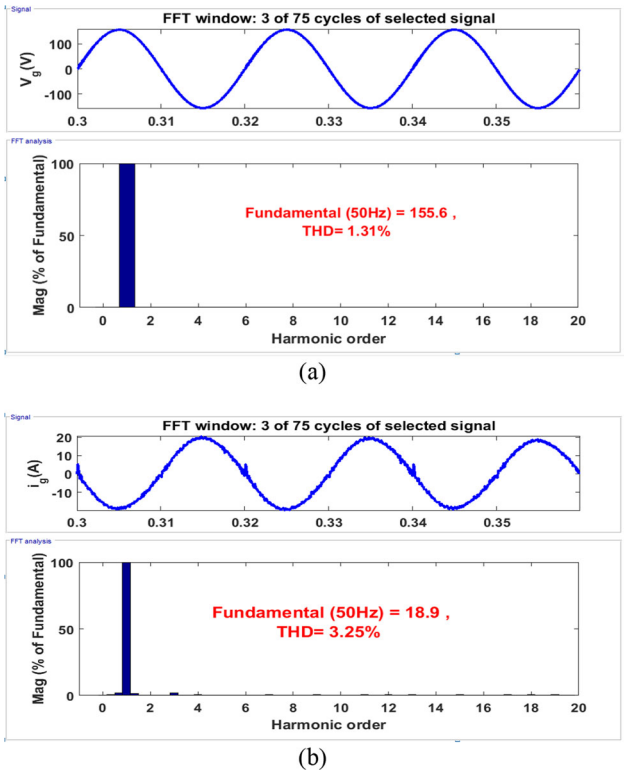


Fig. 10 THD (%) analysis. **a** Grid voltage V_g(V). **b** Grid current i_g(A)

The magnitude of DC-bus terminal voltage across both the capacitors remains stable across the operating region, exhibiting minimal overshoot/undershoot and quickly reaching the reference voltage level within two AC power cycles. The performance solar PV parameters (power, voltage and current) for both the arrays are presented in Fig. 9. Under varying solar PV irradiance and load, the performance of (I_{PV1}, V_{PV1}, P_{PV1}) and (I_{PV2}, V_{PV2}, P_{PV2}) is found to be satisfactory.

The THD analysis of various parameters such as grid voltage V_g(V) and grid current i_g(A) while considering fixed solar irradiance 1000 W/m^2 is shown in Fig. 10a,b. The THD of grid current is found to be 3.25% under normal and steady-state condition, and THD (%) in grid voltage is found to be 1.51%. The obtained THD(%) is within stipulated IEEE-1547 limits.

5.2 Variable solar irradiance under partial shading conditions

The system is now tested under the partial shading conditions by varying solar irradiance at different time intervals. In Pattern-1, as shown in Fig. 10, the solar irradiance varies in the range of [1000 W/m^2, 700 W/m^2, 500 W/m^2, 300 W/m^2]. In Pattern-2, the variation range is [450 W/m^2, 1000 W/m^2, 350 W/m^2, 300 W/m^2], as depicted in Fig. 11, along with the IV and PV characteristics.

Partial shading has a detrimental effect on the overall power output of the PV array. It can be observed that the maximum power generated (320 W) is likely lower than what the

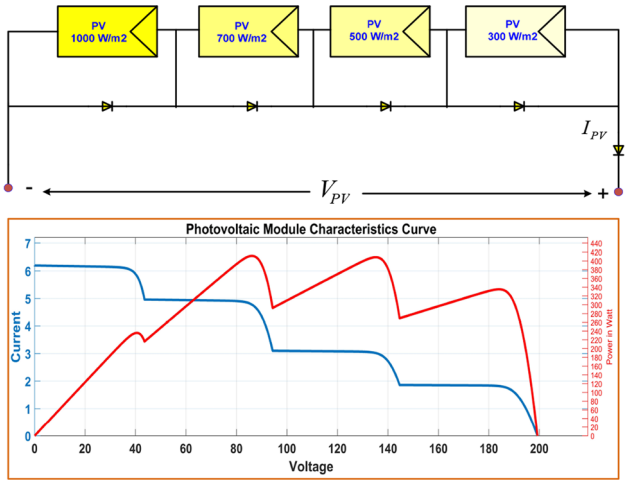


Fig. 11 Graph under partial condition of varying solar irradiance (Pattern-1)

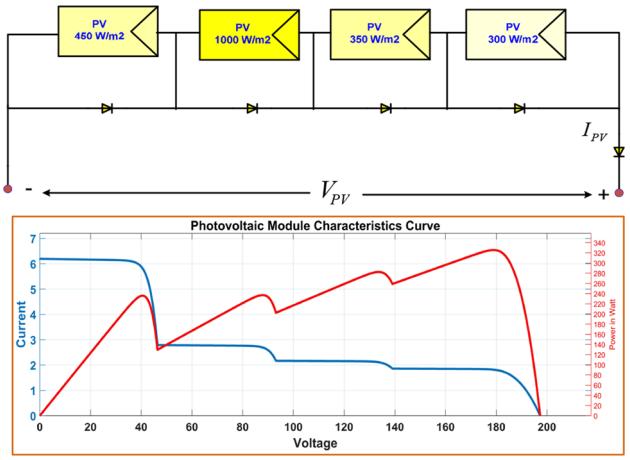


Fig. 12 Graph under partial condition of varying solar irradiance (Pattern-2)

array could produce under uniform conditions. To counteract power losses due to partial shading, several techniques are employed as mentioned in literature, such as bypass diodes that prevent overheating in shaded modules and maximum power point tracking (MPPT) systems that dynamically adjust the duty cycle to ensure the array operates at its MPP. The closed performance of the system under partial shading conditions is depicted in Figs. 13 and 14.

Under partial shading conditions, the solar irradiance varies in the range [1000 W/m^2 700 W/m^2 500 W/m^2 300 W/m^2] in pattern-1 and for pattern-2 the range is [450 W/m^2 1000 W/m^2 350 W/m^2 300 W/m^2] as shown in Figs. 13 and 14, respectively. The performance of parameters such as grid voltage vs(V), grid current i_g(A), load current current i_L(A), dc link voltage of capacitors V_{dc1}(V) & V_{dc2}(V) and total dc link voltages V_{dc} was analyzed. The

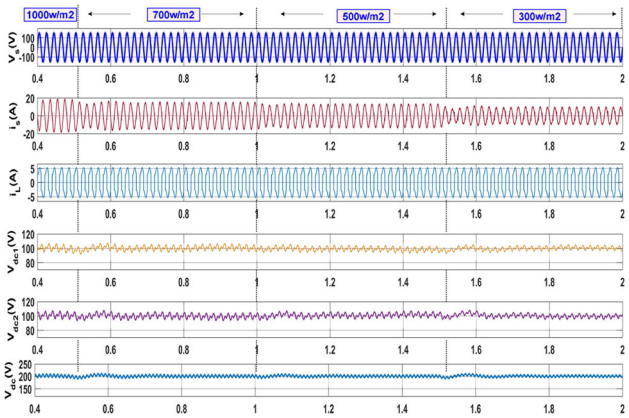


Fig. 13 Closed performance of the system under partial shading conditions with solar irradiance varying in range [1000 W/m^2 700 W/m^2 500 W/m^2 300 W/m^2], i.e., for Pattern-1

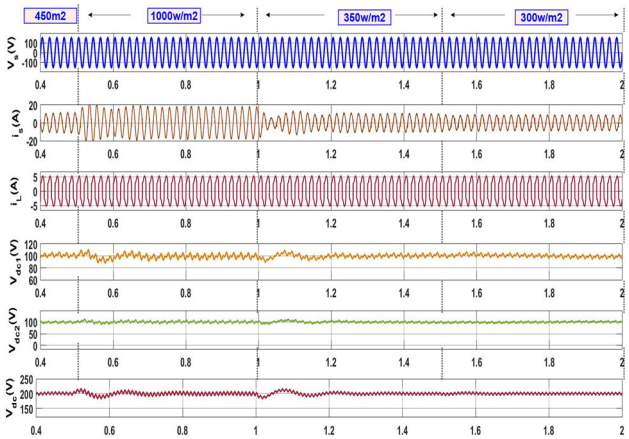


Fig. 14 Closed performance of the system under partial shading conditions with solar irradiance varying from range [450 W/m^2 1000 W/m^2 350 W/m^2 300 W/m^2], i.e., for Pattern-2

system works satisfactory under varying solar irradiance with some oscillations in the dc link voltages.

6 Experimental results and discussion

The proposed system was simulated in MATLAB-Simulink, and the findings were experimentally validated using OPAT-RT (OP-4510-model). Figure 15 depicts the photograph of the setup for experimental validation using OPAL-RT controller-based hardware in loop (HIL) system. OPAL-RT and MATLAB/Simulink perform extremely well together and can validate different control algorithms. It has two key components: the RT-Lab simulation environment and the HIL platform. The FPGA processor of the OPAL-RT engine allows for extremely fast signal processing.

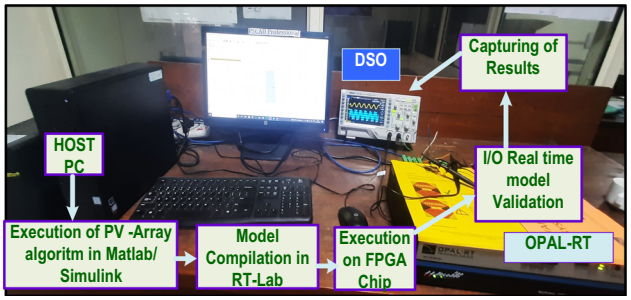


Fig. 15 Photograph of the experimental setup

6.1 Experimental results under dynamic load conditions

The performance of a single-stage PV-integrated system has been evaluated under dynamic conditions, including variations in load and solar irradiance, to capture dynamic results. The PV array supplies power to the grid when the load demand is low, observed by the grid current being 180° out of phase with grid voltage. Figure 16a displays the waveforms of grid voltage (V_g) and grid current (i_g), while Fig. 16 illustrates the waveforms of load current (i_L) and total DC link voltage (V_{DC}). It is noted that there are minimal variations (dips and rises) in DC link voltage during load fluctuations, indicating the effective performance of the Gamma filter and the prompt response of the PI controller in regulating DC link voltage. Similarly, Fig. 16c portrays the performance of the estimated fundamental load current component under dynamic load changes, with the proposed filter settling within 1–2 cycles to a steady-state value after dynamic load adjustments. Finally, the steady-state experimental results of grid voltage and inverter output voltage (V_{inv}) are depicted in Fig. 16d under a solar irradiance of 1000 W/m^2 and fixed load.

6.2 Experimental results under dynamic solar irradiance

Figure 17a, b shows the grid parameters, viz. grid active and reactive power, compensator active and reactive power under varying solar irradiance. The compensator not only supplies the active power to the grid, i.e., 1.489 kW but also meets

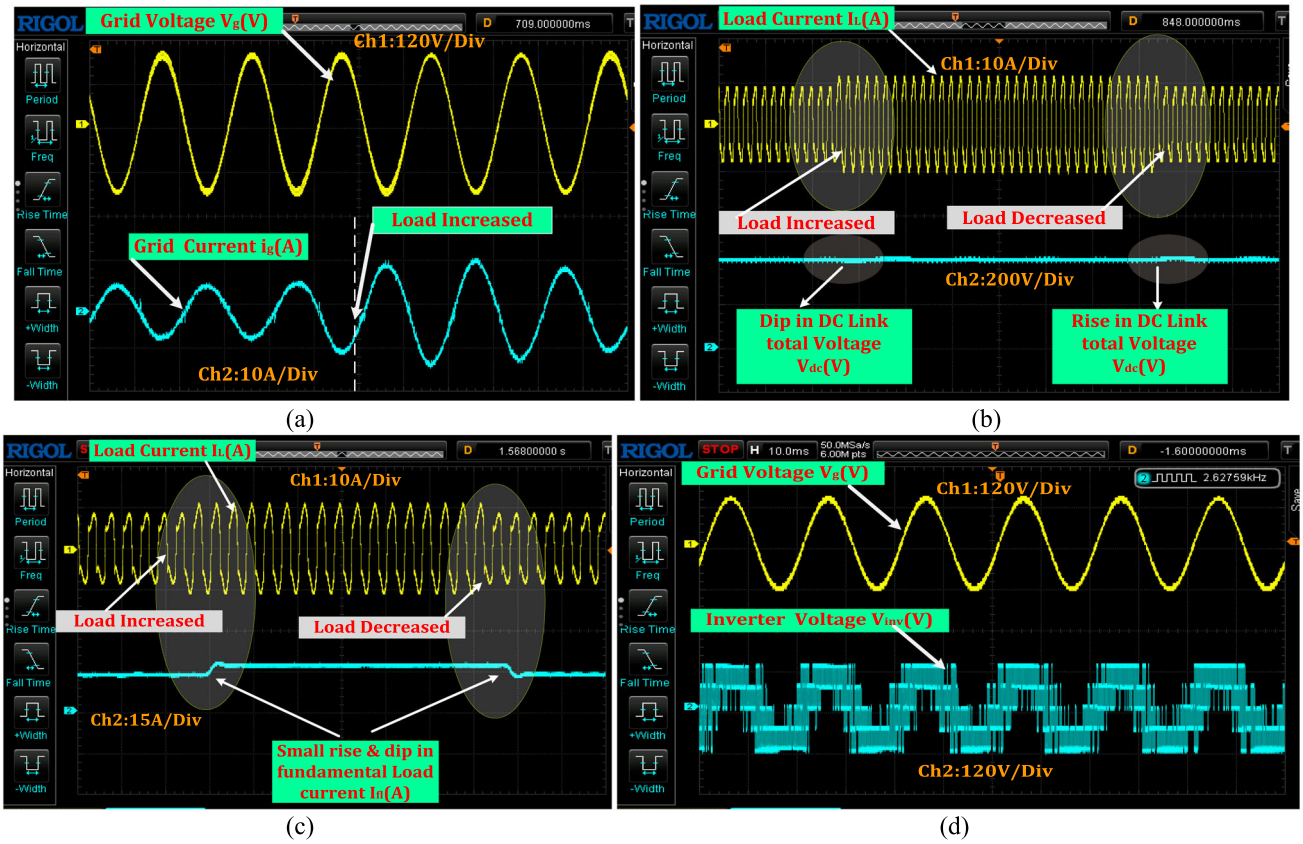


Fig. 16 Experimental results under dynamic load conditions. **a** Grid voltage V_g(V) w.r.t grid current i_g(A). **b** Load current i_L(A) w.r.t total DC link voltage V_{DC}(V). **c** Load current i_L(A) w.r.t fundamental estimated load current i_{fl}(A) and **d** Steady-state grid voltage V_g(V) w.r.t inverter voltage V_{inv}(V)

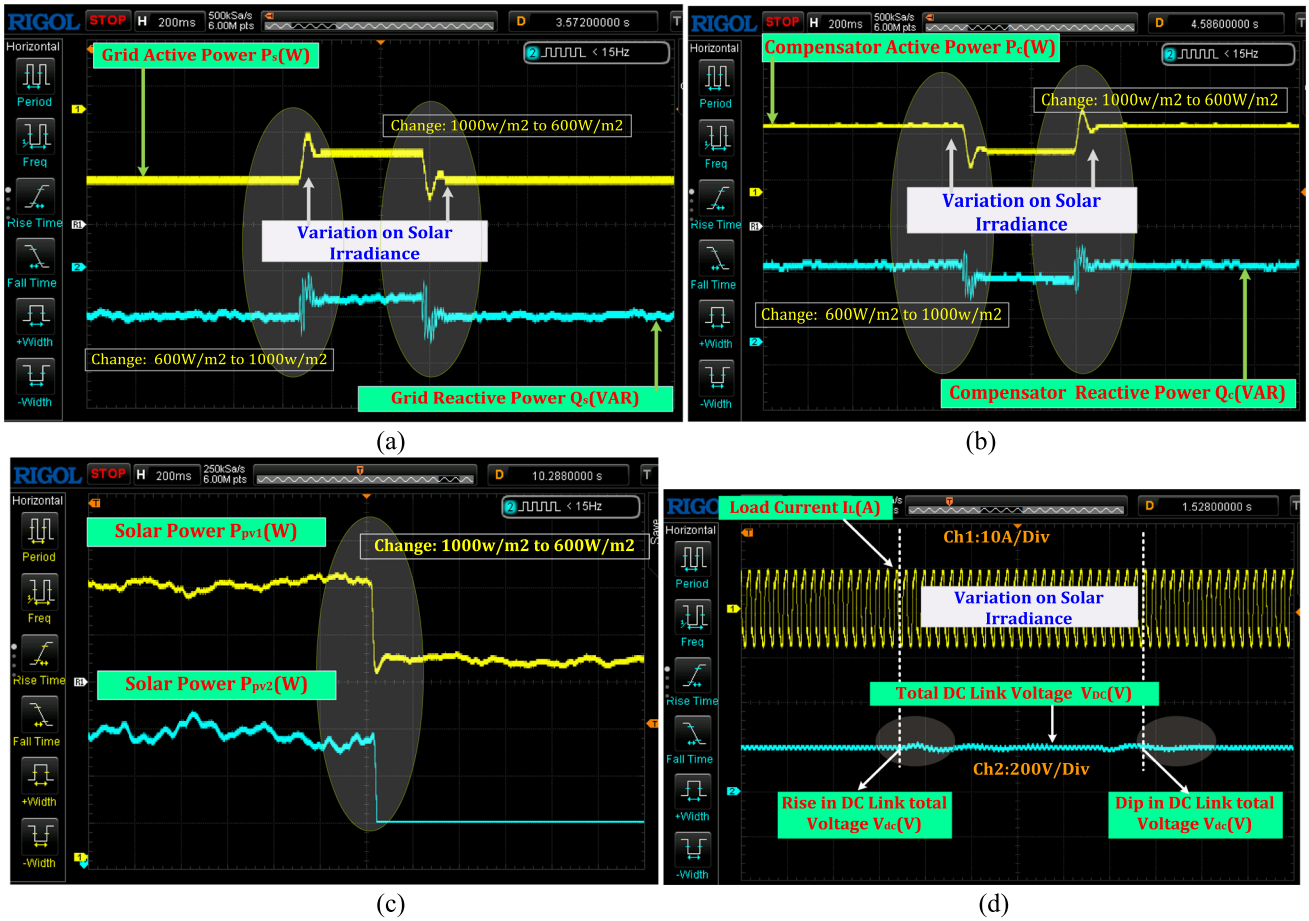


Fig. 17 Experimental results under solar irradiance. **a** Grid active power P_g(W) & grid reactive power Q_g(VAR). **b** Compensator active power P_c(W) & compensator reactive power Q_c(VAR). **c** Solar PV array output power P_{PV1} & P_{PV2} **d** Load current i_L(A) & DC link voltage V_{DC}(V)

Table 6 Power Flow under varying solar irradiance

Active and reactive power	$G = 1000$ W/m^2	Transition of $G =$ 1000–600 W/m^2	Transition of $G =$ 600–1000 W/m^2
Ps (W)	− 1489	− 643.8	− 1489
Qs (VAR)	3.183	6.353	3.183
Pc (W)	1818	1059	1831
Qc (VAR)	66.85	70.04	66.32
PL (W)	441.6	441.5	441.6
QL (VAR)	62.95	66.66	64.13

the 0.446 kW demand of load. In addition, the compensator also satisfies the reactive power demand of the load side. Table 6 shows the magnitude of grid, compensator and load power flow in the proposed system. During solar irradiance changing from 1000 to 600W/m^2, the active solar powers have decreased as shown in Fig. 17c. The change in the total DC link voltage of the compensator during dynamic variation of solar irradiance is depicted in Fig. 17d. This transient response of PV voltage and PV power indicates the proposed algorithm's effectiveness. The power flow analysis is tabulated in Table 7.

7 Comparative analysis of Gamma filter with other filters

The performance of the Gamma filter is now compared with Delta Bar Delta Neural Network filter and IIR filter for the estimation of fundamental active weight component of load

Table 7 Comparison of Gamma filter with other algorithms

S.No	Parameters	Gamma filter [31]	Delta-bar delta neural network filter[34]	IIR-filter [38]
1	Type of filter	Adaptive	Adaptive	Adaptive
2	Estimation of weights	Low overshoot of ± 0.2 A	Moderate overshoot of ± 0.5A	high overshoot of ± 2A
3	THD of supply current (A, rms) (under normal Grid)	3.52%	3.92%	4.2%
4	THD of load current (A, rms)	29.16%	29.16%	29.16%
5	Steady-state error	Less (± 0.3%)	Higher than gamma filter (± 1%)	Higher than both (± 2–3%)
6	Convergence	Faster (1 ~ 2 cycles) with negligible oscillations	Moderate (2 ~ 3cycles)	High (4 ~ 5cycles)
7	Sampling time	50 μs	60 μs	70 μs
8	Supply side power factor	~ 1(unity)	~ 0.99	~0.99
9	System complexity	Low	Moderate	High
	Multiplications	M + 2	2 M + 2	4 M + 1
	divisions	M + 4	2	1
	Additions	2 M	4 M + 1	3 M + 3
	operation required (Consider $M = 2$)	14	17	19

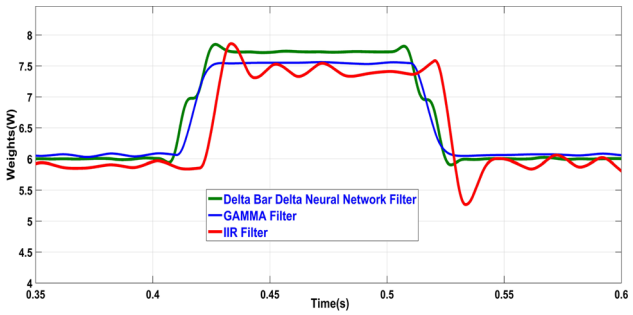


Fig. 18 Convergence analysis of various filter under dynamic load conditions

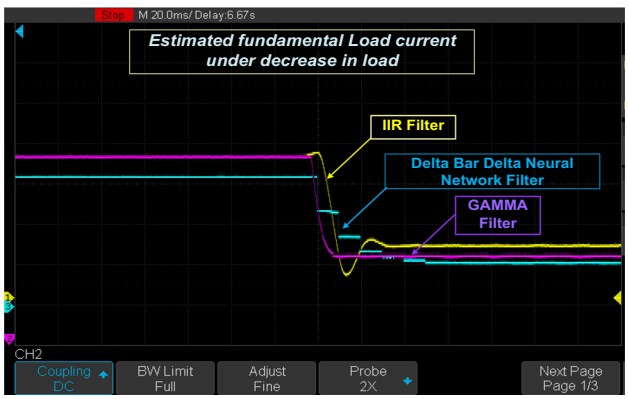


Fig. 19 Comparative analysis of experimental results for estimated fundamental weight with various filters under decrease in load

current under dynamic load variations. The detailed comparison is given in Table 7. The performance of Gamma filter shows a steady-state inaccuracy of less than 0.3% and a low overshoot of 0.2A and provides accurate weight estimate with minimum oscillations and rapid convergence in about 1–2 cycles as depicted in Fig. 18. Its processing efficiency is guaranteed by its modest complexity and 50 μs sample time. It also keeps the supply side power factor at or near unity.

The IIR Filter and the Delta-Bar Delta Neural Network Filter include adaptive features; however, they perform differently. The Delta-Bar Delta Neural Network Filter takes about two to three cycles to converge, but it does so with a substantial overshoot and a steady-state error that is larger than the Gamma filter. It retains a supply side power factor near 0.99 while operating with a slightly longer sampling time of 60 μs. However, out of the three filters, the IIR filter exhibits the largest steady-state inaccuracy and high overshoot. In comparison with the other filters, it operates with a higher complexity and exhibits slower convergence, needing 4∼5 cycles. This filter works at a sampling time of 70 μs, which is maximum. The experimental waveforms are depicted in Figs. 19 and 20 under dynamic load conditions.

8 Conclusion

Active power filters have become a popular choice for reducing the effects of undesirable harmonics caused by nonlinear

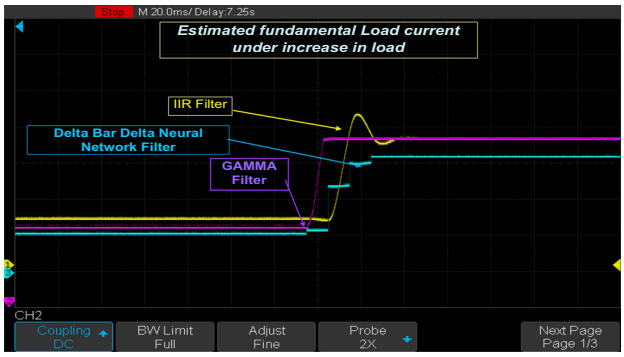


Fig. 20 Comparative analysis of experimental results for estimated fundamental weights with various filters under increase in load

loads and improving PQ issues. In this paper, a unique adaptive control topology based on Gamma filter is used for the closed loop control of SMLI. The SMLI configuration is based on reduced device count and shows NcL index of 3.4 as compared to available MLI topologies. The configuration is further successfully integrated with two PV arrays of 900 W rating each for active power injection also. A novel Gamma filter-based control technique has been formulated and tested on PV-interfaced SMLI. The Gamma filter control approach has shown the advantages of accurately extracting the fundamental component of load current as well as responding fast under dynamic changes. Rigorous testing is undertaken via MATLAB/SIMULINK simulations, and a wide spectrum of solar and load disturbance scenarios has been considered for performance testing. Furthermore, results with a real-time laboratory platform (OPAL-RT simulator) validate the simulation findings under dynamic variations in load and solar power irradiance and partial shading conditions. The outcomes of the performance point out the efficacy of the proposed Gamma adaptive control scheme as well as feasibility of using reduced device count-based new configuration for MLIs. The grid current THD standards laid out by IEEE-519 and IEEE-1547 without/with PV integration are followed strictly ensuring power quality improvement.

Author contributions The highlights of the paper are as follows: 1. The Gamma digital filter has been utilized for PQ improvement. This particular application controller has not been previously used to operate a grid-connected solar PV system effectively. 2. The implementation of closed loop simplified Multilevel Inverter acting as a SAPF unit is new and such work has not been presented to the best of the author's knowledge. 3. The use of PV feed-forward term in the control approach improves dynamic response while also speeding up the operation of the Gamma filter control technique. 4. Both daytime and night simulation and experimental results have been presented under varying solar irradiance and dynamic load conditions. 5. The system's functionality has also been confirmed in accordance with IEEE-519 & IEEE-1547 standards, and it has been found that the overall harmonic distortion of the grid currents is less than 5%.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Myneni H, Ganjikunta SK (2020) Energy management and control of single-stage grid-connected solar PV and BES system. IEEE Trans Sustain Energy 11(3):1739–1749
2. Babatunde OM, Munda JL, Hamam Y (2020) A comprehensive state-of-the-art survey on hybrid renewable energy system operations and planning. IEEE Access 8:75313–75346
3. Karthikeyan V, Rajasekar S, Das V, Karuppanan P, Singh AK (2017) Grid-connected and off-grid solar photovoltaic system. In: Islam F, Mamun K, Amanullah M (eds) Smart energy grid design for Island countries. Green energy and technology. Springer
4. Chatterjee S, Kumar P, Chatterjee S (2018) A techno-commercial review on grid connected photovoltaic system. Renew Sustain Energy Rev 81:2371–2397
5. Mahela OP et al (2021) Recognition of power quality issues associated with grid integrated solar photovoltaic plant in experimental framework. IEEE Syst J 15(3):3740–3748
6. Mohandes B, Moursi MSE, Hatziargyriou N, Khatib SE (2019) A review of power system flexibility with high penetration of renewables. IEEE Trans Power Syst 34:3140–3155
7. Salmeron P, Litran SP (2010) Improvement of the electric power quality using series active and shunt passive filters. IEEE Trans Power Del 25(2):1058–1067
8. Johal H, Divan D (2007) Design considerations for series-connected distributed FACTS converters. IEEE Trans Ind Appl 43(6):1609–1618
9. Divan D, Johal H (2007) Distributed facts—a new concept for realizing grid power flow control. IEEE Trans Power Electron 22(6):2253
10. Singh B, Al-Haddad K, Chandra A (1999) A review of active filters for power quality improvement. IEEE Trans Ind Electron 46(5):960–971
11. Heldwein ML, Ertl H, Biela J, Das D, Kandula RP, Munoz JA, Divan D, Harley RG, Schatz JE (2015) An Integrated controllable network transformer—hybrid active filter system. IEEE Trans Ind Appl 51(2):1692–1701
12. Mohod SW, Aware MV (2010) A STATCOM-control scheme for grid connected wind energy system for power quality improvement. IEEE Syst J 4(3):346–352
13. Wenchao S, Huang AQ (2010) Fault-tolerant design and control strategy for cascaded h-bridge multilevel converter-based STATCOM. IEEE Trans Ind Electron 57(8):2700–2708
14. Pradhan M, Mishra MK (2019) Dual P-Q theory based energy-optimized dynamic voltage restorer for power quality improvement in a distribution system. IEEE Trans Industr Electron 66(4):2946–2955
15. Gupta KK, Ranjan A, Bhatnagar P, Sahu LK, Jain S (2016) Multilevel inverter topologies with reduced device count: A review. IEEE Trans Power Electron 31(1):135–151. https://doi.org/10.1109/TPEL.2015.2405012
16. Bansal P, Singh A (2022) Nonlinear adaptive normalized Huber control algorithm for 5-level distribution static compensator. ElectrEng 104:1635–1648
17. Franquelo LG, Rodriguez J, Leon JI, Kouro S, Portillo R, Prats MA (2008) The age of multilevel converters arrives. IEEE Ind Electron Mag 2(2):28–39. https://doi.org/10.1109/MIE.2008.923519

18. Mak OC, Ioinovici A (1998) Switched-capacitor inverter with high power density and enhanced regulation capability. IEEE Trans Circuit Syst I Fundam Theory Appl 45(4):336–347. https://doi.org/10.1109/81.669056
19. Nabae A, Takahashi I, Akagi H (1981) A new neutral-point-clamped PWM inverter. IEEE Transactions on Industry Applications, IA 17(5):518–523. https://doi.org/10.1109/TIA.1981.4503992
20. Lai JS, Peng FZ (1995) "Multilevel converters—A new breed of power converters" Proc. Conf. Rec. IEEE Industry Appl. Conf Thirtieth IAS Annu Meeting 3:2348–2356
21. Ceglia G, Guzman V, Sanchez C, Ibanez F, Walter J, Gimenez MI (2006) A new simplified multilevel inverter topology for DC–AC conversion. IEEE Trans Power Electron 21(5):1311–1319. https://doi.org/10.1109/TPEL.2006.880303
22. Herrera RS, Salmeron P, Hyosung K (2008) Instantaneous reactive power theory applied to active power filter compensation: different approaches, assessment, and experimental results. IEEE Trans Ind Electron 55(1):184–196. https://doi.org/10.1109/TIE.2007.905959
23. Akagi H, Nabae A, Atoh S (1986) Control strategy of active power filters using multiple voltage-source PWM converters. IEEE Trans Ind Electron 22(3):460–465
24. Panigrahi R, Panda PC, Subudhi B (2014) A robust extended complex Kalman filter and sliding-mode control based shunt active power filter. Electric Power Comp Syst 5008:520–532. https://doi.org/10.1080/15325008.2013.871609
25. Platas-Garza MA, de la O Serna JA (2011) Dynamic harmonic analysis through Taylor-Fourier transform. IEEE Trans Instrum Meas 60(3):804–813. https://doi.org/10.1109/TIM.2010.2064690
26. Biswal S, Biswal M, Malik OP (2018) Hilbert Huang Transform based online differential relay algorithm for a shunt-compensated transmission line. IEEE Trans Power Deliv 33(6):2803–2811. https://doi.org/10.1109/TPWRD.2018.2827843
27. Karvekar S, Patil S, Mulla AM, Patil D (2014) Goertzel algorithm based shunt active power filter using sliding mode controller. Lect Note Eng Comput Sci 1:266–271
28. Singh B, Arya S (2014) Back-Propagation control algorithm for power quality improvement using DSTATCOM. IEEE Trans Ind Electron 61(3):1204–1212
29. Arya SR, Singh B (2014) Neural network based conductance estimation control algorithm for shunt compensation. IEEE Trans Ind Inf 10(1):569–577
30. Mittal S, Singh A, Chittora P (2024) Power quality enhancement in single phase two level/five level converters using adaptive-RBFNN algorithm. Electr Eng 21:1
31. Principe JC, de Vries B, de Oliveira PG (1993) The gamma-filter-a new class of adaptive IIR filters with restricted feedback. IEEE Trans Signal Process 41(2):649–656. https://doi.org/10.1109/78.193206
32. Pattnaik M, Badoni M, Tatte Y (2023) PKLMP based control technique for improving performance of multifunctional grid integrated PV system. Electric Power Syst Res 217:109096
33. Mansour AM, Arafa OM, Marei MI, Abdelsalam I, Aziz GAA, Sattar AA (2021) Hardware-in-the-loop testing of seamless interactions of multi-purpose grid-tied PV inverter based on SFT-PLL control strategy. IEEE Access 9:123465–123483. https://doi.org/10.1109/ACCESS.2021.3110013
34. Shukl P, Singh B (2020) Delta-bar-delta neural-network-based control approach for power quality improvement of solar-PV-interfaced distribution system. IEEE Trans Industr Inf 16(2):790–801. https://doi.org/10.1109/TII.2019.2923567
35. Abhilash T, Kirubakaran T, Somasekhar V. (2019) A seven-level hybrid inverter with DC-link and flying capacitor voltage balancing. Paper presented at: 2019 IEEE international conference on environment and electrical engineering and 2019 IEEE industrial and commercial power systems europe (EEEIC/I&CPS Europe). pp. 1–5
36. Mohammed MF, Qasim MA (2022) Single phase T-type multilevel inverters for renewable energy systems, topology, modulation, and control techniques: a review. Energies 15:8720
37. Rajput SK, Dheer DK (2021) Mathematical modelling and experimental validation for impact of high solar cell temperature on transformer loading and life. Renew Energy Focus 39:27–36
38. Chen S, Zhao Q, Ye Y, Qu B (2023) Using IIR filter in fractional order phase lead compensation PIMR-RC for grid-tied inverters. IEEE Trans Industr Electron 70(9):9399–9409. https://doi.org/10.1109/TIE.2022.3212433

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.

Speed Control of Three-Phase Induction Motor Using ANFIS and Flower Pollination Algorithm Optimized PI Controller

2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI) | 979-8-3503-6328-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/APCI61480.2024.10616561

Aayushi Aryam
Dept. of Electrical Engineering
Delhi Technological University
Delhi, India
aayushiaryam@gmail.com

Aayushi Singh
Dept. of Electrical Engineering
Delhi Technological University
Delhi, India
aayushisingh1808@gmail.com

Aditya Mohan
Dept. of Electrical Engineering
Delhi Technological University
Delhi, India
adityamohan03032001@gmail.com

Bhavnesh Jaint
Dept. of Electrical Engineering
Delhi Technological University
Delhi, India
bhavneshjaint@dtu.ac.in

***Abstract*— This paper presents a novel approach for the speed control of a three-phase induction motor by integrating an Adaptive Neuro-Fuzzy Inference System (ANFIS) with a Flower Pollination Algorithm (FPA) optimized Proportional-Integral (PI) controller. The primary objective is to improve the efficiency of induction motor speed control under various load conditions, ensuring robustness and adaptability. The ANFIS is employed to intelligently predict and adjust the motor speed in real-time, capitalizing on its ability to handle nonlinearities and uncertainties inherent in the motor dynamics. To further refine the control strategy, the PI controller's parameters are optimized using the Flower Pollination Algorithm, a nature-inspired optimization technique that mimics the pollination process of flowers. This optimization ensures that the PI controller works in perfect harmony with the FPA model, resulting in improved response times, minimized speed overshoot, and enhanced stability under dynamic load changes. Simulation results exhibit the superiority of the proposed method over traditional control strategies, offering significant improvements in speed control accuracy and system resilience. The FPA-PI outperforms the ANFIS based control and provides effective speed control of the motor.**

Keywords—: Adaptive Neuro-Fuzzy Inference System (ANFIS), Flower Pollination Algorithm (FPA), Proportional Integral (PI) Controller, Three-Phase Induction motor

I. INTRODUCTION

In the domain of industrial automation and control, three-phase induction motors are highly regarded for their durability, operational efficiency, and broad utilization. Yet, achieving precise control over their speed is a complex endeavor, exacerbated by the motors' nonlinear behavior and their vulnerability to fluctuating load demands.

A 3-phase induction motor's speed may be managed via indirect field-oriented controller, utilizing a real-time control system, was introduced by [1]. The primary achievement highlighted in this research is the effective deployment of the suggested control strategy, leading to precise control over the motor's speed. In another study, [2] developed a method for the real-time regulation of a 3-phase induction motor's speed, employing LabVIEW alongside fuzzy logic. The significant outcome of this research is the enhanced performance in speed control achieved by the implementation of the fuzzy logic control strategy, which showed superiority over conventional control methods. The design and implementation of PID controllers for enhancing the speed regulation of a 3-phase induction motor, utilizing the direct-axis current coordinate through IFOC, were the focus of [3]. The successful creation and application of these PID controllers, leading to superior speed control performance, stands out as the study's central achievement. In contrast, the implementation of a scalar control approach for the speed regulation of a three-phase induction motor was explored in [4]. The effective control of the motor's speed, achieved through the successful application of the scalar control method, was identified as the primary outcome of this research. Meanwhile, a method employing hysteresis space vector pulse width modulation for the speed regulation of a 3-phase induction motor was introduced by [5]. The main discovery of this investigation was the precise and efficient control over the motor's speed, attributed to the effective deployment of the proposed control strategy.

In the realm of driving systems and power electronics, the speed of 3-phase induction motors has been controlled utilizing various strategies. Among these, the technique of online tuning with genetic algorithms has been explored by [6]. This research centered on optimizing the speed control of three-phase induction motors through the application of genetic algorithms. The effectiveness of this proposed method was validated through experimental results presented in the study. It was concluded by the authors that the speed of three-phase induction motors can be effectively managed using genetic algorithms for online tuning, resulting in enhanced efficiency and performance.

Research by [7] explored the application of non-conventional power sources, such as PV system and wind, for controlling the speed of 3-phase induction motors, finding that renewable energy offers a viable, sustainable option for motor control. Additionally, [8][9] introduced a control system based on an ANFIS for speed regulation, demonstrating through simulations its accuracy and efficiency in managing motor speed. These studies contribute to developing eco-friendly and precise control strategies for induction motors.

979-8-3503-6328-9/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:12:14 UTC from IEEE Xplore. Restrictions apply.

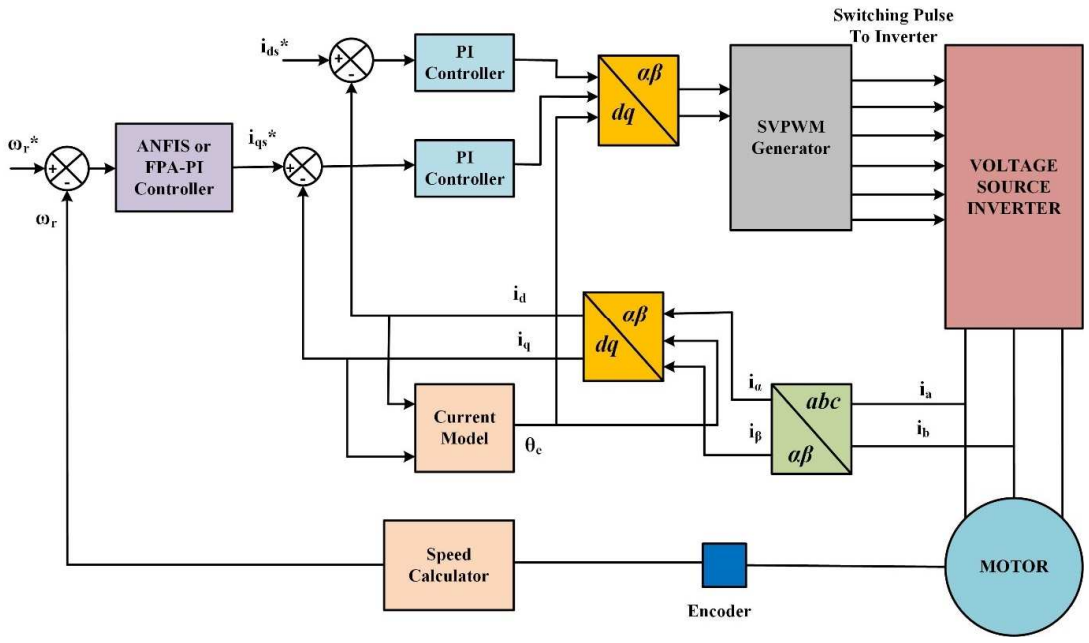


Fig. 1. Representation of Field control of Induction motor using ANFIS & FPA optimized PI Controller

Recent progress in controlling the speed of three-phase induction motors includes various methods like genetic algorithms, renewable energy, adaptive neuro-fuzzy systems, and advanced sliding mode controls. These approaches have significantly improved motor control systems' efficiency and reliability. Notably, an enhanced sliding mode control technique has shown superior robustness and performance through experiments in [10]. Additionally, integrating adaptive sliding-mode control with predictive torque control has proven to increase speed accuracy and responsiveness, confirmed by simulations and experiments in [11]. These advancements underscore the ongoing efforts to refine motor speed control, suggesting a promising direction for future innovative and efficient control solutions.

The Flower Pollination Algorithm (FPA) emerges as a promising solution for optimization in electrical power systems, inspired by natural pollination, illustrated by [12]. Highlighted in a review by [13], FPA excels in addressing complex issues like optimal power flow, economic dispatch, and voltage stability, showcasing its advantages and areas for improvement compared to other methods. The article serves as a key resource for understanding FPA's impact on power systems and suggests future research directions to enhance its application in this field. Fig.1 shows the Representation of Field control of Induction motor using ANFIS & FPA optimized PI Controller.

This study introduces a novel control strategy that combines the Adaptive Neuro-Fuzzy Inference System with a Flower Pollination Algorithm optimized Proportional-Integral (PI) controller. This method leverages ANFIS's adaptive learning to model and control motor speed accurately, while the FPA fine-tunes the PI controller for optimal performance. Aimed at improving precision, stability, and efficiency in controlling three-phase induction motors, this integrated approach addresses system nonlinearities and variable loads, promising a new standard in industrial motor speed control.

II. Speed Control of Three Phase Induction Motor

Equations (1) through (8) may be used to describe a 3-phase squirrel cage induction motor mathematically inside a synchronously rotating reference frame.

$$V_{qs} = R_s i_{qs} + \frac{d\varphi_{qs}}{dt} + \omega_e \varphi_{ds} \quad (1)$$

$$V_{ds} = R_s i_{ds} + \frac{d\varphi_{ds}}{dt} + \omega_e \varphi_{qs} \quad (2)$$

$$V_{qr} = R_r i_{qr} + \frac{d\varphi_{qr}}{dt} + (\omega_e - \omega_r)\varphi_{dr} \quad (3)$$

$$V_{dr} = R_r i_{dr} + \frac{d\varphi_{dr}}{dt} + (\omega_e - \omega_r)\varphi_{qr} \quad (4)$$

when both V_{qr} and V_{dr} are zero. The flux equation is provided by

$$\varphi_{qs} = L_{ls} i_{qs} + L_m (i_{qs} + i_{qr}) \quad (5)$$

$$\varphi_{qr} = L_{lr} i_{qs} + L_m (i_{qs} + i_{qr}) \quad (6)$$

$$\varphi_{ds} = L_{ls} i_{qs} + L_m (i_{ds} + i_{dr}) \quad (7)$$

$$\varphi_{dr} = L_{lr} i_{dr} + L_m (i_{ds} + i_{dr}) \quad (8)$$

Applied to the stator are the voltages V_{qs} and V_{ds}, with i_{ds}, i_{qs}, i_{dr}, i_{qr} representing the respective d and q axis currents for the stator and rotor. The components of stator and rotor flux are given by φ_{qs}, φ_{qr}, φ_{ds}, φ_{dr}, while R_s and R_r denote the resistances of the stationary and rotating coil, respectively. Inductances for the stationary and rotating coil are represented by L_{ls} and L_{lr}, with L_m indicating the mutual inductance. The equation for electromagnetic torque is subsequently presented.

$$T_e = \frac{3}{2}\frac{P}{2}\frac{L_m}{L_r}(\varphi_{dr} i_{qs} - \varphi_{qr} i_{ds}) \quad (9)$$

P denotes the total amount of pole in the motor. The q-axis element of the rotating field, φ_{qr}, would drop to zero once vector control was attained. So, the electromagnetic torque depends only on the stator current in the q-axis. By relying only on one parameter, this change streamlines the regulation of electromagnetic torque.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:12:14 UTC from IEEE Xplore. Restrictions apply.

$$T_e = \frac{3}{2}\frac{P}{2}\frac{L_m}{L_r}(\varphi_{dr} i_{qs}) \quad (10)$$

The inner current loops and the external speed control loop are the two control loops, divide the control mechanism. The estimation of rotor flux quantities is carried out using the rotor's angular velocity, stator current, and a computational rotor time constant, as described in equation (11).

$$\frac{d\theta_e}{dt} = \frac{1}{T_r}\frac{i_{qs}}{i_{ds}} + \omega_r \quad (11)$$

The rotor speed, ω_r , undergoes comparison with the commanded rotor speed, $\omega_r *$, and the discrepancy is handled by the PI speed controller. This controller is responsible for producing a reference for the stator q-axis current, i_{qs} . Following its processing via the Clark along with Park Transformation, the motor current's feedback is compared to the currents of reference for the d- and q-axes. After this comparison, the PI controller, based on the errors identified, formulates a voltage command signal. This signal is sent into the Space Vector Pulse Width Modulation (SVPWM) after undergoing Inverse Park Transformation (IPT), which produces the signal that switches for the Voltage Source Inverter (VSI). As a result, the speed of the induction motor may be controlled by adjusting the stator winding current. The equation [10] expresses the conversions between the synchronously spinning d-q frame, the fixed α-β frame, and the stationary a-b-c frame.

Clarke Transformation

$$\begin{bmatrix} I_\alpha \\ I_\beta \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ 0 & \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \end{bmatrix} \quad (12)$$

Inverse Clarke transformation

$$\begin{bmatrix} I_a \\ I_b \\ I_b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{-1}{2} & \frac{\sqrt{3}}{2} \\ \frac{-1}{2} & \frac{-\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} I_\alpha \\ I_\beta \end{bmatrix} \quad (13)$$

Park Transformation

$$\begin{bmatrix} I_d \\ I_q \end{bmatrix} = \begin{bmatrix} cos\theta_e & sin\theta_e \\ -sin\theta_e & cos\theta_e \end{bmatrix} \begin{bmatrix} I_\alpha \\ I_\beta \end{bmatrix} \quad (14)$$

Inverse Park transformation

$$\begin{bmatrix} I_\alpha \\ I_\beta \end{bmatrix} = \begin{bmatrix} cos\theta_e & sin\theta_e \\ -sin\theta_e & cos\theta_e \end{bmatrix} \begin{bmatrix} I_d \\ I_q \end{bmatrix} \quad (15)$$

A. Space Vector Pulse Width Modulation Technique

For the top three components in a three-phase voltage source inverter, SVPWM is a unique switching technique that generates pseudo-sinusoidal electrical currents across the stator phases. As shown in Fig. 2, when an upper transistor (S1, S3, S5, set to 1) is activated, the matching descending transistor (S2, S4, S6 set to 0) is immediately deactivated. As a result, the on and off modes of the higher transistors S1, S3, and S5 may be used to determine the output voltage. The relationship between the line-to-line voltage vector [Vab, Vbc, Vca]^T with the phase voltage vector [Va, Vb, Vc]^T and the switching variable vector [a, b, c]^T is established by equations (16)–(17).

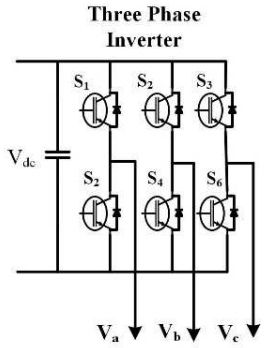


Fig. 2. Representation of Three phase Inverter

$$\begin{bmatrix} V_{ab} \\ V_{bc} \\ V_{ca} \end{bmatrix} = V_{dc} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (16)$$

$$\begin{bmatrix} V_{an} \\ V_{bn} \\ V_{cn} \end{bmatrix} = \frac{V_{dc}}{3} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (17)$$

In the configuration of a three-phase bridge inverter, 23 permissible switching states exist. As illustrated in Figure 3, six non-zero vectors (V1 to V6) form the axis of a hexagon and are responsible for delivering electric power to the motor, with each pair of adjacent non-zero vectors separated by an angle of 60 degrees. There are two null vectors (V0 and V7) at the origin, which do not apply any voltage to the load. The eight vectors with the labels V0 to V7 are known as the fundamental space vectors. Adjacent vectors V1 and V2 are carefully chosen to provide the phase-to-center voltage, or Vref.

III. ANFIS & FLOWER POLLINATION OPTIMIZED PI CONTROLLER

A. ANFIS

The Adaptive Neuro-Fuzzy Inference System (ANFIS) comprises five layers shown in Fig.3., where grid partitioning is utilized over methods such as scatter partitioning, tiered partitioning, and fuzzy c-means clustering to reduce the rule count to manageable numbers. This is achieved by dividing the entire input space into subsets specified by regions likely to contain the input vectors.

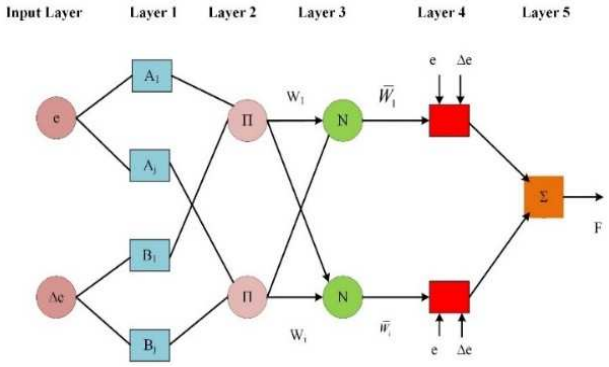


Fig. 3. Different Layers of ANFIS Network

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:12:14 UTC from IEEE Xplore. Restrictions apply.

B. ANFIS for Induction Motor Speed Control

To understand how ANFIS is employed for the speed regulation of an induction motor, let's break down the process into key steps:

1. The ANFIS controller for motor speed uses speed error and its change as inputs to adjust voltage or frequency, correcting the error.
2. In ANFIS, fuzzification turns numerical inputs (speed error, change in speed error) into fuzzy sets like "Negative Big" to "Positive Big" using membership functions.
3. ANFIS uses fuzzy rules to determine control actions, e.g., "IF speed error is Positive Small AND change is Zero, THEN slightly increase voltage."
4. The inference system applies fuzzy rules to current inputs, determining fuzzy outputs for control actions.
5. Defuzzification combines fuzzy outputs into a precise numerical control action for motor speed adjustment.
6. ANFIS refines its control by adapting membership functions and rules through a neural network-based learning algorithm, enhancing performance over time with feedback.

The ANFIS controller receives real-time data about the actual motor speed and compares it with the desired speed to calculate the speed error and its change. Based on these inputs, ANFIS determines the appropriate adjustment to the motor's input voltage or frequency. As operating conditions change, ANFIS continuously updates its control strategy to minimize speed errors, ensuring smooth and efficient motor operation.

C. Flower Pollination Algorithm

The FPA is inspired by the natural pollination method, vital for plant reproduction, and involves complex mechanisms with parallels in engineering challenges. Biotic pollination, involving insects and birds and accounting for 90% of cases, contrasts with the 10% of plants using abiotic methods like wind. This diversity showcases nature's adaptability and efficiency, inspiring solutions for complex design issues.

Pollination, essential for plant reproduction, occurs via self-pollination (using a flower's own pollen) or cross-pollination (using pollen from different plants). This process optimizes species survival and propagation, making it a crucial natural optimization mechanism for flowering plants.

For FPA, the following four steps are used:

Step 1: In the process of global pollination, pollinators carrying pollen engage in cross-pollination and biotic processes, following paths that mimic Lévy flights.

Step 2: Local pollination, which encompasses abiotic processes and self-pollination, occurs independently of pollinators.

Step 3: Insects can develop flower constancy, where the probability of reproduction correlates directly with the resemblance between two flowers.

Step 4: A probability switch, slightly favoring local pollination, governs the interaction between local and global pollination mechanisms.

In the step of universal pollination, the transportation of flower pollen gametes is facilitated by the flight capabilities of insects, allowing pollen to traverse long distances. Consequently, the steps of global pollination and flower constancy are encapsulated by the following representation:

$$x_i^{t+1} = x_i^t + \gamma L(\lambda)(g_* - x_i^t) \tag{18}$$

The step size $L(\lambda)$, derived from Lévy flights, reflects the pollination intensity, effectively simulating the varied distances over which insects can transport pollen. This is because the step size, denoted as L > 0, is sourced from a Lévy distribution, allowing for the efficient mimicry of insect behavior in covering long distances.

$$L \sim \frac{\lambda\Gamma(\lambda)\sin(\frac{\pi\lambda}{2})}{\pi}\left(\frac{1}{s^{1+\lambda}}\right) \quad (s \gg s_0 > 0) \tag{19}$$

For steps larger than zero (s > 0), the distribution holds true. Representation of both local pollination and the flower constancy principle is achieved through Steps 2 and 3.

$$x_i^{t+1} = x_i^t + \varepsilon(x_j^t - x_k^t) \tag{20}$$

Pollen transfer within the same species, using x_j^t and x_k^t, mimics flower constancy in a localized area, with e chosen from a uniform [0, 1] range. While pollination can happen at any scale, closer flowers are more likely to pollinate locally. A switch probability (p) adjusts from global to local pollination, starting at p = 0.5, but p = 0.8 is suggested for better outcomes across scenarios.

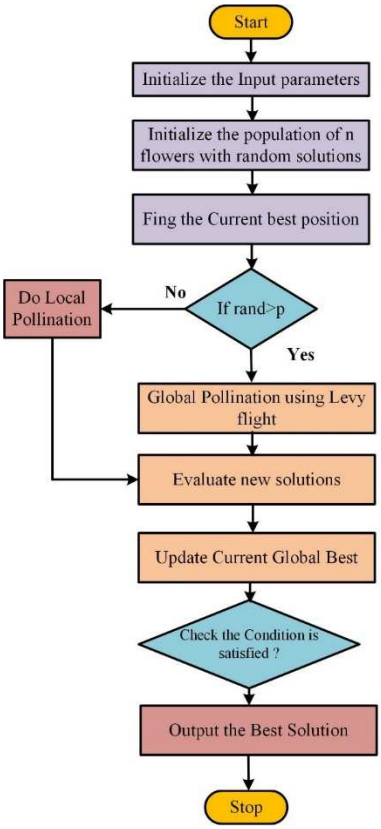


Fig. 4. Flowchart of Flower Pollination Algorithm

D. PI tuning using FPA

Using the Flower Pollination Algorithm to optimize the PI controller parameters involves the following steps:

1. Create an initial population of potential solutions, each of that reflects a set of K_p and K_i parameters for the PI controller. This is known as initialization.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:12:14 UTC from IEEE Xplore. Restrictions apply.

2. Evaluation: Assess the performance of each solution based on a predefined objective function, typically related to the speed regulation performance of the induction motor (e.g., minimizing the difference between actual and desired speed).
3. Pollination Process: Apply global and local pollination processes to generate new solutions. This involves creating new sets of K_p and K_i values based on the rules of the FPA, simulating the process of pollinators spreading pollen.
4. Selection and Reproduction: Select the best solutions based on their performance and use them to generate the next generation of solutions.
5. Termination: Continue steps 2-4 until an ending criterion—such as an upper limit on the number of iterations or an objective function convergence threshold is satisfied.

By optimizing the PI controller parameters using the FPA, one can theoretically achieve better control performance for the speed of three-phase induction motors compared to traditional tuning methods. The FPA's ability to balance exploration and exploitation helps in efficiently searching the parameter space for the optimal solution, potentially leading to improved motor speed control with desired dynamic and steady-state characteristics.

IV. Simulation Results & Discussions

The Simulation of Speed regulation of Induction motor is using ANFIS & FPA optimized PI controller was carried out in MATLAB/Simulink.

A. Results for ANFIS based Speed Control

The Fig 5. Stator Current & Electromagnet Torque using ANFIS Control. Fig.6 shows the Comparison of Reference speed to Actual speed of the motor using ANFIS Control

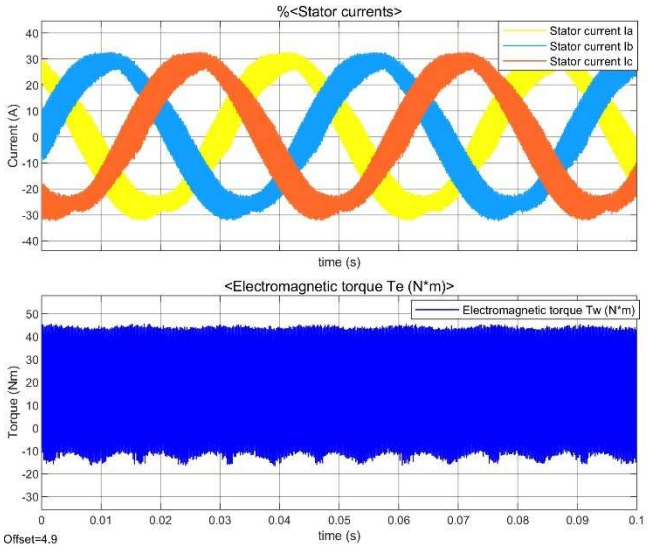


Fig. 5. Stator Current & Electromagnet Torque using ANFIS Control

The Reference speed taken for analysis during various instances are 477.5,907.2,573,1003 & 668.5 respectively.

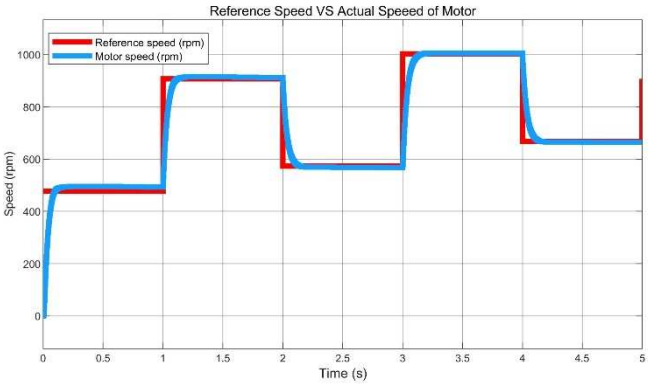


Fig. 6. Comparison of Reference speed to Actual speed of the motor using ANFIS Control

B. Results for FPA-PI based Speed Control

The Fig 7. Stator Current & Electromagnet Torque using FPA-PI Control. Fig.8 shows the Comparison of Reference & Actual speed of the motor using FPA-PI Control.

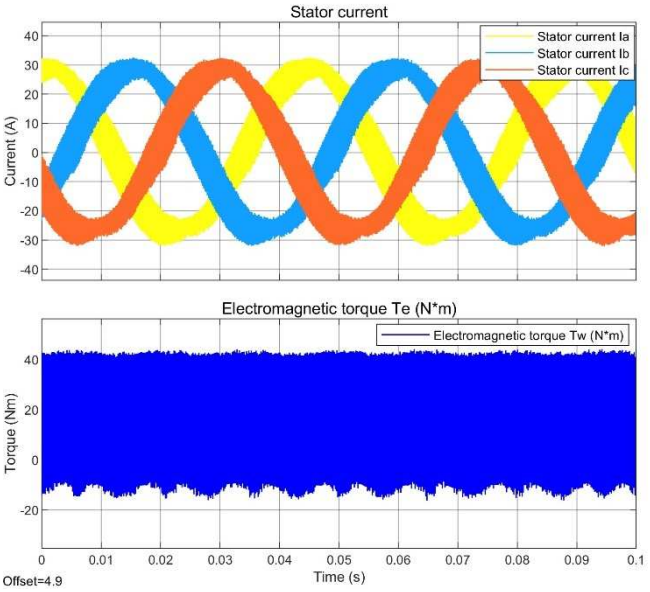


Fig. 7. Stator Current & Electromagnet Torque using FPA-PI Control

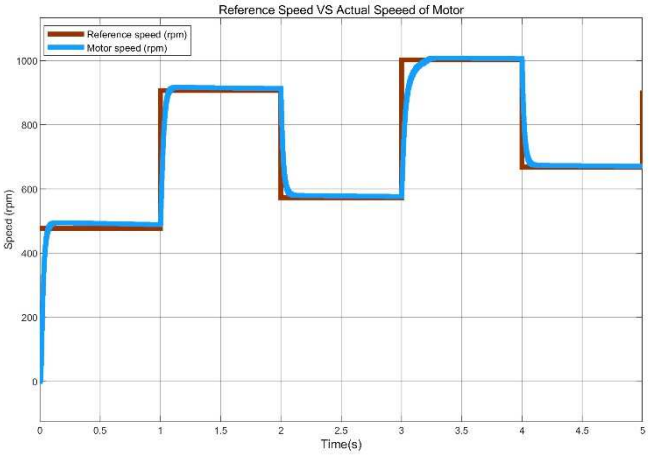


Fig. 8. Comparison of Reference speed to Actual speed of the motor using FPA-PI Control

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:12:14 UTC from IEEE Xplore. Restrictions apply.

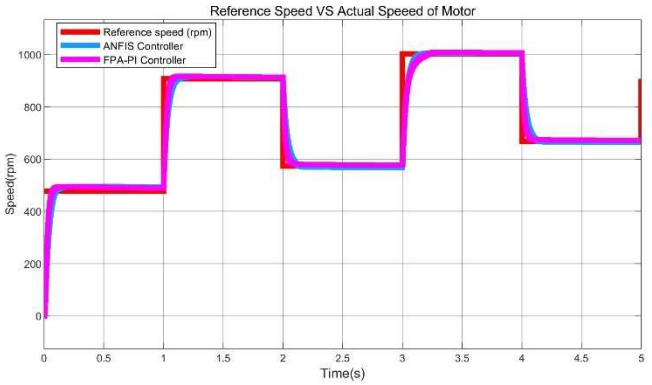


Fig. 9. Comparison of Reference speed to Actual speed of the motor using ANFIS & FPA-PI Control

Table I. Comparison of Overshoot, Settling time & Steady state error for the Control methods

Ref Speed (rpm)	%Overshoot/ Undershoot		Settling Time (s)		Steady State Error (%)	
	ANFIS	FPA-PI	ANFIS	FPA-PI	ANFIS	FPA-PI
477.5	3.83	3.71	0.8	0.5	3.27	3.41
907.2	0.95	0.99	1.6	1.3	0.77	0.83
573	0.47	0	2.7	2.2	0.87	0.87
1003	0.30	0.30	3.6	3.5	0.30	0.30
668.5	0.48	0.07	4.5	4.3	0.75	0.90

The Table I comparing ANFIS and FPA-PI controllers across various reference speeds (477.5, 907.2, 573, 1003, and 668.5 rpm) demonstrates the effectiveness of FPA-PI in controlling system speed. FPA-PI consistently shows comparable or superior performance, particularly in percentage overshoot/undershoot and settling times. For example, at 477.5 rpm, FPA-PI achieves a slightly lower overshoot of 3.71% compared to ANFIS's 3.83%, and at 668.5 rpm, it significantly reduces overshoot to 0.07% versus ANFIS's 0.48%. Settling times are consistently faster with FPA-PI across all reference speeds, such as at 477.5 rpm where it settles in 0.5 seconds compared to ANFIS's 0.8 seconds, and at 573 rpm with a 2.2 second settling time versus ANFIS's 2.7 seconds. While steady-state errors are comparably low for both controllers across most speeds, at 668.5 rpm, FPA-PI slightly underperforms with a 0.90% error compared to ANFIS's 0.75%. Despite this, the overall performance of FPA-PI indicates its potential for more efficient and precise speed control under various conditions, underscoring its advantages in overshoot/undershoot and settling times.

V. CONCLUSION

This research paper presents a novel approach to controlling the speed of a three-phase induction motor by Controlling an ANFIS with a Flower Pollination Algorithm optimized PI controller. The ANFIS is employed for its ability to adaptively predict and adjust the motor speed in real-time, capitalizing on its fuzzy logic and neural network synergies for handling nonlinearities and uncertainties in motor dynamics. Meanwhile, the FPA is utilized to fine-tune the PI controller parameters ensuring optimal control performance by balancing the exploration and exploitation capabilities inherent in the FPA's nature-inspired optimization process. The combination of ANFIS with FPA-optimized PI control offers a robust and efficient solution for achieving precise speed control of three-phase induction motors, addressing challenges such as load variations and parameter changes that traditionally complicate motor control. The results demonstrate significant improvements in dynamic response, including reduced overshoot and faster settling times, while maintaining minimal steady-state error. This integrated approach showcases the potential for advanced control strategies in enhancing the performance and reliability of industrial motor applications, contributing valuable insights into the ongoing evolution of motor control technologies.

REFERENCES

[1] Ferdiansyah, I., Rusli, M. R., Praharsena, B., Toar, H., and Purwanto, E., "Speed control of three phase induction motor using indirect field-oriented control based on real-time control system," In 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE) (pp. 438-442), IEEE, July 2018.

[2] Santhiya, K., Devimuppudathi, M., Kumar, D. S., and Renold, A. J., "Real time speed control of three phase induction motor by using lab view with fuzzy logic," Journal on Science Engineering and Technology, 5(2), 21-27, 2018.

[3] Ferdiansyah, I., Raharja, L. P. S., Yanaratri, D. S., and Purwanto, E., "Design of PID controllers for speed control of three phase induction motor based on direct-axis current (Id) coordinate using IFOC," In 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (pp. 369-372), IEEE, November 2019.

[4] Prasetia, A. M., and Santoso, H., "Implementation of scalar control method for 3 phase induction motor speed control," Elinvo (Electronics, Informatics, and Vocational Education), 3(1), 63-69, 2018.

[5] Purnata, H., Rameli, M., and Effendie, A. R., "Speed control of three phase induction motor using method hysteresis space vector pulse width modulation," In 2017 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 199-204), IEEE, August 2017.

[6] Chiewchitboon, P., Tipsuwanporn, V., Soonthornphisaj, N., and Piyarat, W., "Speed control of three-phase induction motor online tuning by genetic algorithm," In The Fifth International Conference on Power Electronics and Drive Systems. PEDS 2003. (Vol. 1, pp. 184-188). IEEE, November 2003.

[7] Antar, R. K., Sultan, N. S., and Ali, A. J., "Speed control of three phase induction motor fed by renewable energy source," In 2019 2nd International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE) (pp. 7-12), IEEE, February 2019.

[8] Kusagur, A., Kodad, S. F., and Ram, B. S., "Modeling, design & simulation of an adaptive neuro-fuzzy inference system (ANFIS) for speed control of induction motor," International Journal of Computer Applications, 6(12), 29-44, 2010.

[9] N. Gogia, R. Prasad, R. Arora, B. Bhushan, and P. Prakash, "Speed control of three phase induction motor using PI, Fuzzy and ANFIS controllers," 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, pp. 496-501, IEEE, 2023.

[10] Shiravani, F., Alkorta, P., Cortajarena, J. A., and Barambones, O., "An enhanced sliding mode speed control for induction motor drives," In Actuators (Vol. 11, No. 1, p. 18), MDPI, January 2022.

[11] Zhang, Y., Yin, Z., Li, W., Liu, J., and Zhang, Y., "Adaptive sliding-mode-based speed control in finite control set model predictive torque control for induction motors," IEEE Transactions on Power Electronics, 36(7), 8076-8087, 2020.

[12] Nuthalapati, Y., and Naidu, R. K., "A novel modulating PID controller for a speed control of BLDC motor adopting flower pollination algorithm," In Intelligent Computing in Control and Communication: Proceeding of the First International Conference on Intelligent Computing in Control and Communication (ICCC 2020) (pp. 159-168). Springer Singapore, 2021.

[13] Lalljith, S., Fleming, I., Pillay, U., Naicker, K., Naidoo, Z. J., & Saha, A. K., "Applications of flower pollination algorithm in electrical power systems: a review," IEEE Access, 10, 8924-8947, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 13,2024 at 07:12:14 UTC from IEEE Xplore. Restrictions apply.

2024
11(2), 910

Biomedical and
Therapeutics LETTERS

Structure-Based Virtual Screening, ADMET prediction and Molecular Dynamics simulation of isoindolin-1-one scaffolds as potential inhibitors of MNK2

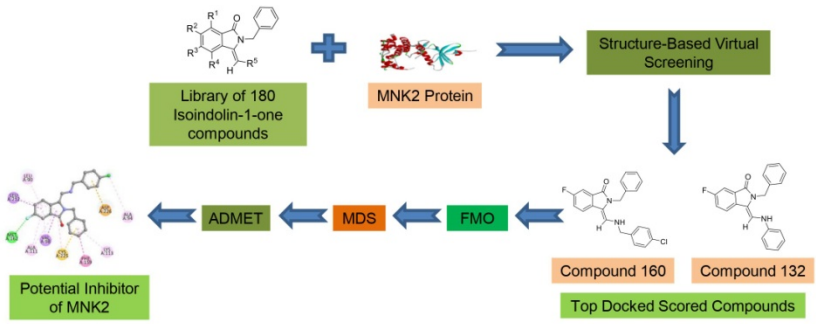
Kunal Madaan, Saurabh Mehta, Ram Singh*

Department of Applied Chemistry, Delhi Technological University, Delhi-110042, India.

Submitted on: 29-Jun-2024, Accepted and Published on: 20-Aug-2024

Article

ABSTRACT



Mitogen-activated protein kinase 2 (MNK2) is a promising target for various types of cancers like breast, prostate, colorectal, glioblastoma lung, and melanoma cancers. It activates the oncogene eukaryotic initiation factor 4E (eIF4E) by phosphorylation. Isoindolin-1-one scaffolds displayed significant anti-cancer activity as reported previously in research papers. In this study, we explored potential inhibition of various substituted isoindolin-1-one compounds against MNK2 through *in-silico* studies. Structure-based virtual screening of library of 180 compounds having isoindolin-1-one scaffold was done followed by fragment molecular orbital (FMO) calculations and molecular dynamics simulation (MDS) of the best two scored docked poses (compounds **132** and **160**). ADMET prediction of compounds **132** and **160** were performed. Sixteen compounds showed docking scores higher than or equal to co-crystal ligand of MNK2 protein and hydrogen bonding with MET162 residue. 100 ns molecular dynamics simulation of compound **160** displayed excellent 95% hydrogen bond occupancy with MET162 residue of MNK2 protein showing remarkable docked pose stability throughout the simulation. Further, the ADMET prediction of compound **160** falls under the satisfactory range. Compound **160** exhibited potential inhibition against MNK2 protein. Thus, *in-silico* studies of these isoindolin-1-ones will be helpful in the development of highly potent and selective MNK2 inhibitors.

Keywords: Cancer; MNK2; ADMET; virtual screening; MDS

INTRODUCTION

Protein synthesis and gene regulation both rely on translation. Cap-dependent translational initiation requires the activity of eukaryotic initiation factor 4E (eIF4E).[1-3] Since eIF4E binds specifically to the 5'-cap of mRNAs, it controls how different mRNAs are translated.[4,5] eIF4E overexpression is linked to several malignancies, including bladder, colon, lung, breast, head, and neck cancers.[6-10] By boosting the release of proinflammatory cytokines, these overexpression levels regulate the tumor microenvironment (TME), induce cell migration and invasion, and promote cell proliferation.[11,12] Mitogen-activated protein kinase interacting kinases (MNK1 and MNK2) aid in the control of eIF4E. MNKs are Ser/Thr kinases that phosphorylate eIF4E at Ser209 *in-vivo* and phosphorylate cytoplasmic phospholipase A2 (cPLA2), sprouty2 (hSPRY2), heterogeneous nuclear RNA-binding protein A1 (hnRNPA1), and polypyrimidine-tract binding protein-associated splicing factor (PSF) in vitro.[13-15] The Ras/Raf/MEK/ERK and PI3K/Akt/mTOR signaling pathways are responsible for MNK activation.[16-18] MNKs are crucial for eIF4E phosphorylation, as demonstrated by the lack of eIF4E phosphorylation in mice with MNK knockouts that were created by replacing Ser209 with Alanine.[19] MNKs phosphorylate eIF4E, which regulates certain mRNA translation and associated protein synthesis, including those of MMP-9, SNAIL, and c-MYC.[13,20] Numerous studies have demonstrated that phosphorylating eIF4E using MNKs is the cause of several malignancies which includes prostate, breast, pancreatic, colorectal, lung, ovarian cancers, melanoma, glioblastoma, acute myeloid leukemia.[21-27] MNKs may also improve insulin sensitivity, lower insulin resistance, and increase energy expenditure in a high-fat diet.[28-30] Compared to MNK1, MNK2 had a higher constitutive capacity to phosphorylate eIF4E.[31,32] MNK2 may therefore be a useful therapeutic target in drug development for the treatment of different cancer diseases.[33,34]

*Corresponding Author: Prof. Ram Singh
Email: ramsingh@dtu.ac.in

URN:NBN:sciencein.btl.2024.v11.910
DOI:10.62110/sciencein.btl.2024.v11.910
© ScienceIn Publishing
https://pubs.thesciencein.org/btl

Several MNK2 inhibitors have been identified in recent years. It was demonstrated that SEL-201,[35] ETC-206,[36] eFT508,[37] and MNK-7g[38] potently inhibited MNK2. All two of these inhibitors, eFT508, and ETC-206 entered clinical studies; however, only eFT508 has thus far advanced through Phase 2 clinical trials[39] but none of them received market approval. Therefore, the search for and creation of extremely effective and specific MNK2 inhibitors is currently ongoing.

In the previously published studies, isoindolin-1-one scaffolds shown significant inhibition for the treatment of cancer.[40-44] So, we created a library of 180 isoindoline-1-one compounds by doing diverse substitutions on isoindolin-1-one core ring which are synthetically possible and virtually screened them against the MNK2 protein target. Molecular docking is a reliable technique that helps identify the optimal binding position of compounds within protein receptors and establish the binding affinity between compounds and proteins. Using eFT508 (co-crystal ligand) as a reference ligand, we firstly used molecular docking using AutoDock Vina to screen the isoindolin-1-ones library. Validation of docking protocol was also done to increase the efficacy of molecular docking. The top two docked scored compounds were then investigated further using ADMET, FMO, and MDS. Using GAMESS software, FMO studies were performed to ascertain the precise contribution of each amino acid residue of MNK2 protein to the binding interactions with both compounds. Using GROMACS software, MDS of the docked complexes of both compounds was performed in physiological conditions to verify the stability of their docked poses. SwissADME and AdmetSAR online in-silico tools were used to estimate the physicochemical characteristics, drug-likeness properties, pharmacokinetics, lipophilicity, water solubility, and toxicity of compounds. Figure 1 showed the workflow of the *in-silico* studies. We hope that these in-silico investigations will be useful in the further research and development of anti-cancer medications.

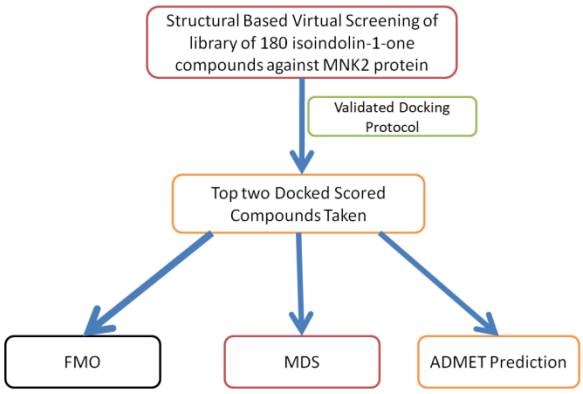


Figure 1: Representation of workflow of in-silico studies

Materials and Methods

Structure-based Virtual Screening

Structure-based virtual screening is one of the robust methods for the identification of the best binding pose of molecules within the protein receptor and helps in determining the binding affinity between protein and molecules. For protein preparation, the 3D X-ray crystal structure of the MNK2 protein was downloaded from the protein data bank website (PDB ID:6CK6) having co-crystal ligand as eFT508, a known inhibitor of MNK2.[37] Co-crystal ligand, water molecules, and other hetatoms of protein were removed using BIOVIA Discovery Studio 2021.[45] Then, missing atoms, polar hydrogens, and Kollman charges were added by AutoDockTools-1.5.7[46] and the protein was saved in AutoDock .pdbqt format.

For ligand preparation, 2D structures of the library of 180 isoindolin-1-one compounds including co-crystal Ligand (eFT508) were drawn in ChemDraw Professional 19.0.0.22. Then, SMILES of these 2D structures were used to convert them into 3D structures in .pdb format using Open Babel 2.4.1.[47] During conversion, hydrogen atoms were added at physiological pH, and the energy of compounds were minimized using the MMFF94 force field. At last, energy minimized .pdb files of molecules were converted to AutoDock .pdbqt files.

Active sites of the protein were checked using BIOVIA Discovery Studio. Grid Box size and position coordinates were adjusted so that all the active sites must be in the grid box. The dimensions and coordinates of the grid box were set to x = 23, y = 23, z = 23 and, x= 55.57, y = 20.01, z = 1.16, respectively. All 180 compounds were then docked, including co-crystal ligand, using AutoDock Vina[48] with the default exhaustiveness value set to 8. Binding energy values (in Kcal/mol) between MNK2 protein and ligands were noted down. Conventional protein-ligand binding interactions were visualized using BIOVIA Discovery Studio.

Validation of Docking Protocol

For validation of docking protocol, the co-crystal ligand of MNK2 was redocked using AutoDock Vina, and then the redocked pose of co-crystal ligand was superimposed with the experimental pose of the co-crystal ligand which is present in its original crystallographic bound conformation in 6CK6 PDB ID of MNK2 protein, and after that root mean square deviation (RMSD) between both the experimental and redocked poses of co-crystal was calculated.

Fragment Molecular Orbital (FMO) Calculations

The FMO method[49] was used to analyze the binding interactions between compounds **132** and **160** and MNK2 more quantitatively. This method helped us to identify which specific amino acid residue of MNK2 contributes how much to the binding interaction. Docked complexes of compounds **132** and **160** were taken for the preparation of input files using Facio 25.1.2 software.[50] GAMESS software[51] was employed for performing all the FMO calculations at the MP2/6-31G* level. Then, with the help of Facio software, total pair interaction energy (PIE) between amino acid residues and ligands was examined.

Molecular Dynamics Simulation (MDS)

MDS is important to further identify the stability of docked poses through RMSD and hydrogen bonding occupancy analysis. The top two scored docked complexes of compounds **132** and **160** were taken for 100 ns MDS simulation. MDS simulation was

done using GROMACS 2020.5 software package[52,53] over Ubuntu 18.04 operating system having AMD Ryzen 7 processor and Nvidia GeForce RTX 3060 GPU. Charmm36-July2021 force field was used for protein topology generation. SwissParam online server[54] was employed for the generation of Ligand topologies. TIP3P water model[55] was applied for solvation and afterward, Na+ and Cl- ions were used for neutralizing the system. The system was energy minimized using the steepest descent algorithm with a maximum of 50,000 steps and less than 10.0 kJ/mol force. The equilibration process was carried out under NVT and NPT ensembles utilizing a V-rescale thermostat and a Parrinello-Rahman barostat, respectively for 200 ps at 300 K[56] with periodic boundary conditions. RMSD of protein-ligand complexes throughout the simulation was estimated using the built-in RMSD feature of GROMACS and hydrogen bonds occupancy plots between specific amino acid residues of MNK2 protein and ligands were plotted using CPPTRAJ function of AmberTools.[57,58]

ADMET Prediction

One of the most vital tasks during the early phases of drug development is ADMET prediction. Compounds **132** and **160** were evaluated for drug-likeness, physicochemical characteristics, lipophilicity, water solubility, pharmacokinetics studies, and toxicity using the online web servers SwissADME[59] and admetSAR.[60] SMILES of these compounds were used for prediction.

RESULTS

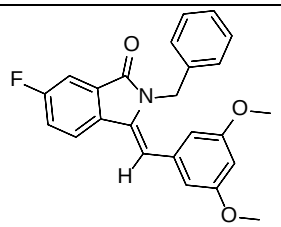
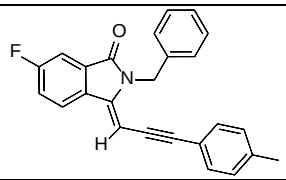
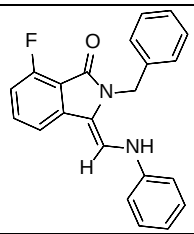
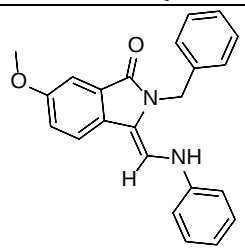
Structure-based Virtual Screening

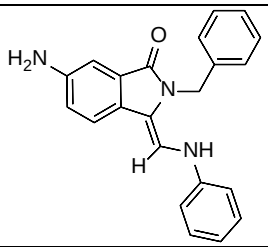
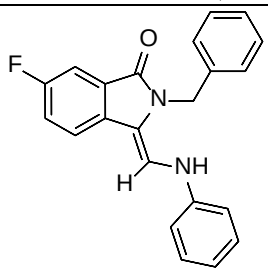
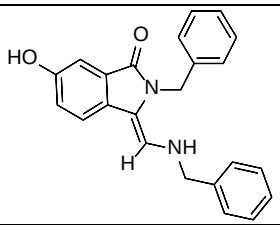
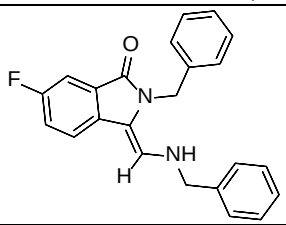
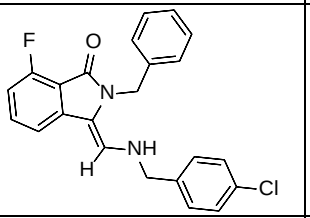
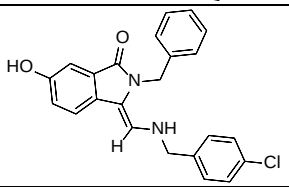
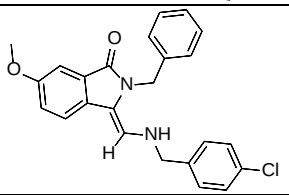
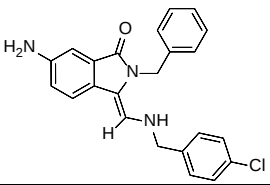
BIOVIA Discovery studio helped us to identify the important active site amino acid residues of MNK2, that are, LEU90, ALA111, LYS113, LEU143, PHE159, MET162, GLY165, ASN210, LEU212, CYS225, ASP226. So, a grid box for docking was placed around these active sites. We have examined the binding energy (docking score) of the best-docked poses of isoindolin-1-one compounds generated after docking and checked whether the best-docked poses of these compounds showed binding interactions with active amino acid residues of protein or not.

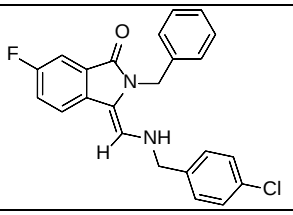
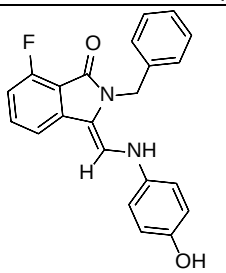
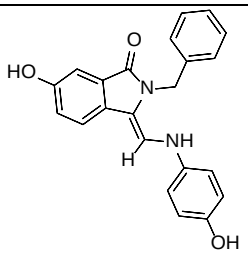
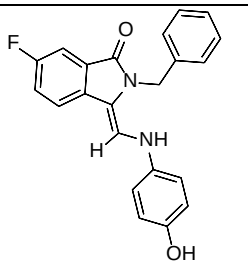
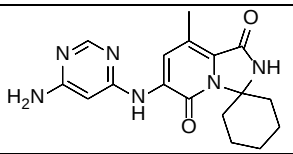
After the structure-based virtual screening, a library of 180 isoindolin-1-one compounds showed a docking score in the range of -6.9 Kcal/mol to -10.4 Kcal/mol. A more negative binding energy value leads to more efficient binding between ligand and protein. 53 compounds displayed binding energy lower than or equal to the co-crystal ligand (-9.9 Kcal/mol). From among these 53 compounds, 16 compounds exhibited hydrogen bonding with active MET162 amino acid residue of MNK2 which lies in the hinge region of MNK2 protein and is essential for good binding affinity. Compound numbers, 2-D Structures, and binding energy values of these 16 compounds are given in Table 1.

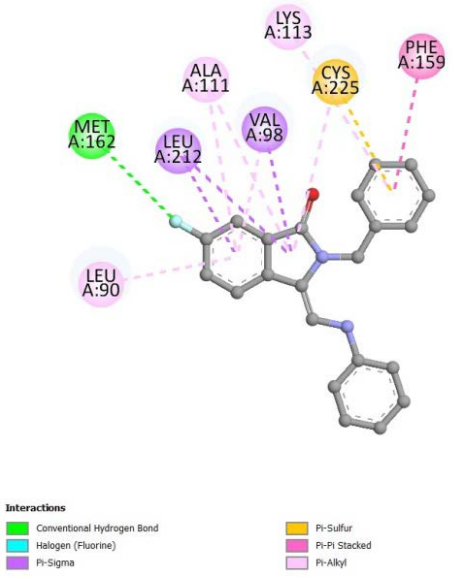
Compounds with superior binding energy values are considered for effective binding to the protein target and efficiently inhibit the protein function. Compounds **132** and **160** exhibited the lowest binding energy values of -10.4 and -10.3 Kcal/mol respectively (Table 1). These two top docked scored compounds indicated superior structural compatibility with the active site of MNK2 protein indicating the potential for them to develop stable interactions. The 2-D and 3-D output docked poses of these compounds are shown in Figures 2 and 3 respectively. The fluorine atom of compound number **132** underwent conventional hydrogen bonding with MET162, aromatic rings of this compound showed other hydrophobic interactions with VAL98, LEU90, ALA111, LYS113, PHE159, LEU212, and CYS225. Similarly, compound number **160** also showed hydrogen bonding between MET162 residue and fluorine atom, other hydrophobic interactions were with aromatic rings and chlorine atom by LEU90, ALA94, VAL98, ALA111, LYS113, PHE159, LEU212, CYS225, ASP226 residues. Thus, these docking results indicated excellent binding affinity of compounds **132** and **160** with active amino acid residues of MNK2 protein and would be taken further for MDS, FMO, and ADMET studies.

Table 1: 2-D Structures and Binding energy values (Kcal/mol) of isoindolin-1-one compounds exhibiting binding energy lower than or equal to co-crystal ligand and displaying H-bonding with MET162 residue.

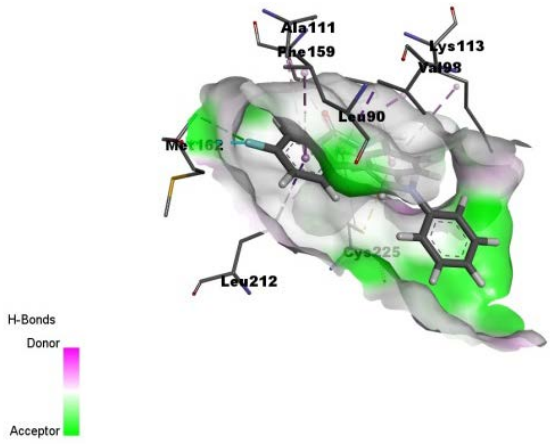
Compound No.	2-D Structure	Binding Energy (Kcal/mol)
40		-9.9
77		-10.2
128		-10.2
129		-10.0

130		-10.2
132		-10.4
142		-9.9
146		-9.9
155		-10.0
156		-10.0
157		-9.9
158		-10.1

160		-10.3
169		-9.9
170		-10.0
174		-10.1
Co-Crystal Ligand (eFT508)		-9.9



(a)

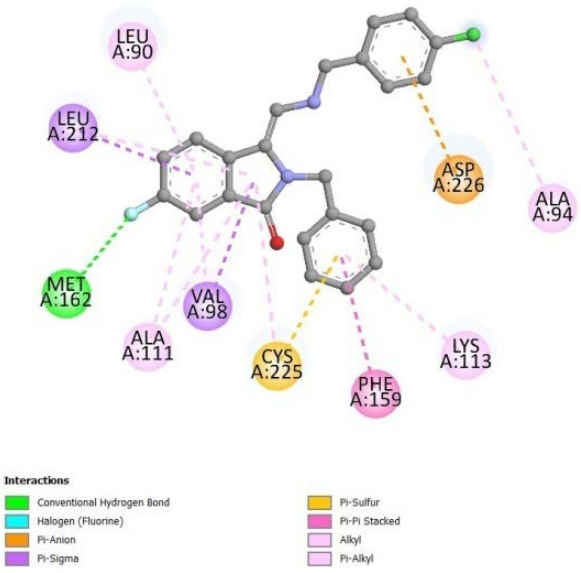


(b)

Figure 2: (a) 2-D representation of binding interactions between compound number **132** and MNK2 (b) 3-D representation of binding interactions between compound number **132** and MNK2

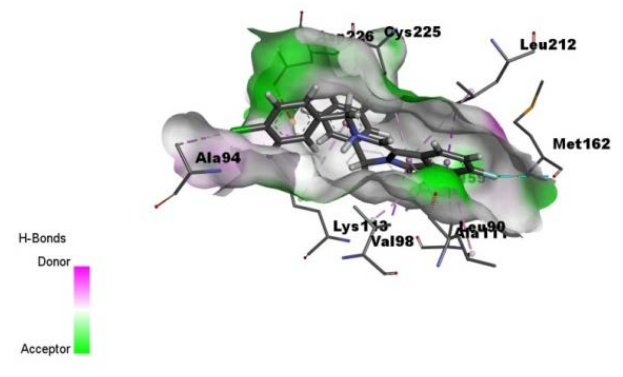
Validation of Docking Protocol

For validation of docking protocol, the co-crystal ligand was redocked with MNK2 using the exact docking protocol as used for the isoindolin-1-one library. Figure 4 displayed the superimposition of the redocked pose shown in green color with the crystal pose shown in yellow color. Superimposed image suggested that both the experimental and docked poses were almost similar. Also, the calculated RMSD value between both poses was less than 0.5 Å. So, Figure 4 and RMSD value lower than 2Å suggested that the docking protocol was easily able to reproduce the experimental pose of the co-crystal ligand of the MNK2 protein, and thus, an accurate and effective protocol was



(a)

followed for estimating the binding energy and docked poses of the library of isoindolin-1-one compounds.



(b)

Figure 3: (a) 2-D representation of binding interactions between compound number **160** and MNK2 (b) 3-D representation of binding interactions between compound number **160** and MNK2

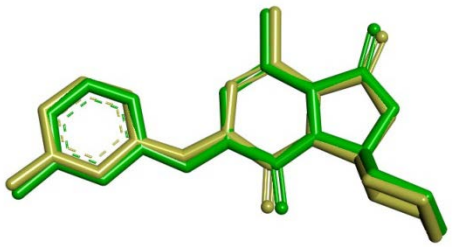


Figure 4: Superimposition image of crystal pose of co-crystal ligand (highlighted in yellow colour) with the redocked pose of co-crystal ligand (highlighted in green colour)

Fragment Molecular Orbital (FMO) Calculations

To determine the amount of impact that each amino acid specifically was making on the binding interactions with compounds, FMO calculations were done by fragmenting the system into ligand and amino acids and total pair interaction energy (PIE) values between amino acid fragments and ligands were calculated. We have done the FMO calculations of top docked scored compounds **132** and **160** and their total pair interaction energy plots are shown in Figures 5 and 6 respectively.

With a total PIE value of -20.008 Kcal/mol, the amino acid MET162 contributes most significantly to the binding affinity of compound **132**. High total PIE values of -14.957 and -14.833 Kcal/mol were also displayed by PHE159 and ALA111 respectively. So, these three residues contribute mostly to the efficient binding with molecules. In the case of compound **160**, MET162 showed a total PIE of -16.215 Kcal/mol while ASP226, PHE159, and ALA111 showed total PIE values of -15.603, -14.912, and -12.734 Kcal/mol respectively. So, these results indicated that MET162 residue of MNK2 protein is majorly responsible for high and effective binding with both compounds **132** and **160** which may be due to the strongest conventional hydrogen bonding of MET162 residue with ligands.

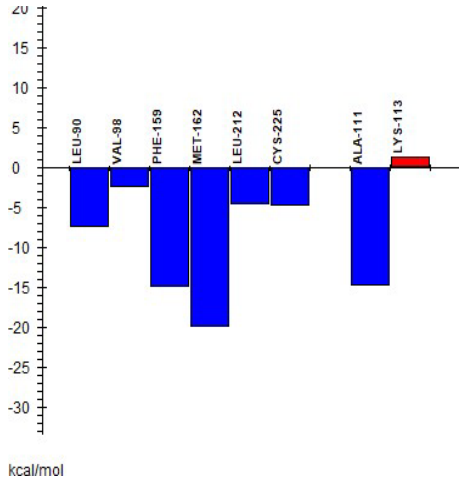


Figure 5: Total pair interaction energy plot between ligand **132** and amino acid residues of MNK2

Molecular Dynamics Simulation (MDS)

To check the stability of the docked poses of compounds **132** and **160**, MDS of their docked complexes was done in physiological environments. After the MDS, the trajectory of the complete 100 ns simulation was used to analyze the RMSD of whole protein-ligand complexes and several hydrogen bonds formed between specific amino acid residues of protein and ligands over the complete simulation.

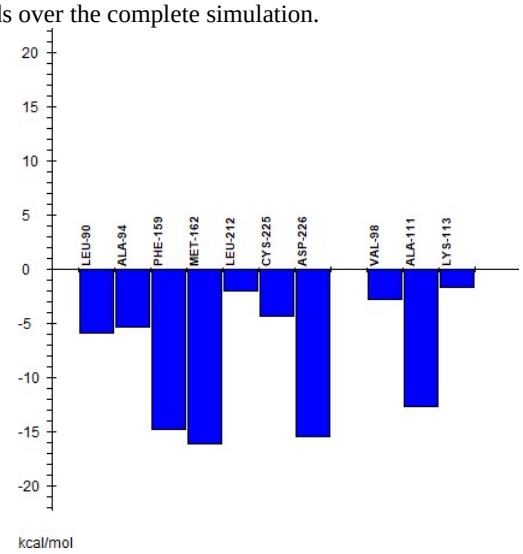


Figure 6: Total pair interaction energy plot between ligand **160** and amino acid residues of MNK2

RMSD of whole protein ligand docked complex including all the backbone atoms and alpha carbons showed the docked complexes structural stability over the complete 100 ns simulation period. The usual value of RMSD for a protein-ligand complex can be influenced by several factors, including the size and flexibility of the ligand, the resolution of the experimental structure, and the kind of interaction between the protein and ligand. A lower RMSD value pointed towards a more stable protein-ligand docked complex while a higher RMSD value suggested less stability of docked complex during the simulation.

The RMSD plots of the whole protein-ligand **132** and protein-ligand **160** complexes were shown in Figures 7a and 7b, respectively. The protein-ligand **132** complex showed an average RMSD value of 0.49 nm, whereas the protein-ligand **160** complex exhibited an average RMSD value of 0.27 nm. This demonstrated that during the simulation, the protein-ligand **160** complex seemed to be more stable than the protein-ligand **132** complex and thus the binding of ligand **160** to the active site of MNK2 protein is more stable than ligand **132**.

Number of hydrogen bonds formed between protein amino acids and ligands during the simulation is also one of the significant parameters in predicting the stability of the docked poses. The stability of the docked pose increases with the increase in amount of hydrogen bonds. Figures 8a and 8b displayed the number of hydrogen bonds formed between MNK2 protein and ligands **132** and **160** over the frames generated during the complete 100 ns simulation. Ligand **132** established hydrogen bonding with MET162 for 20% of the simulation and with LYS113 for 15% of the simulation time. The ligand **160** formed hydrogen bonding with MET162 for 95% of the simulation time. These MDS results highlighted the tremendous stability and binding affinity of ligand **160** with MNK2 protein compared to ligand **132**.

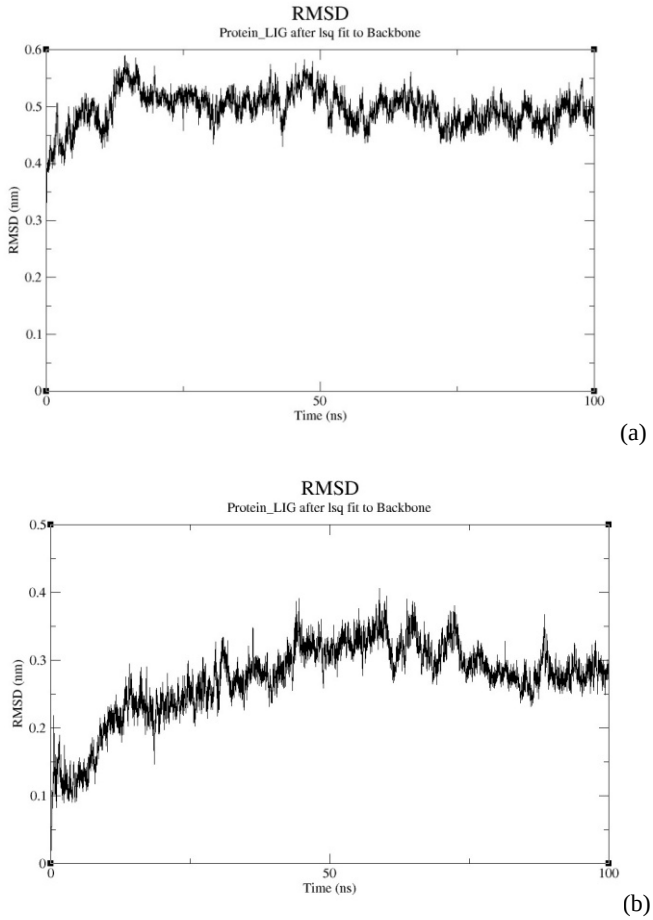
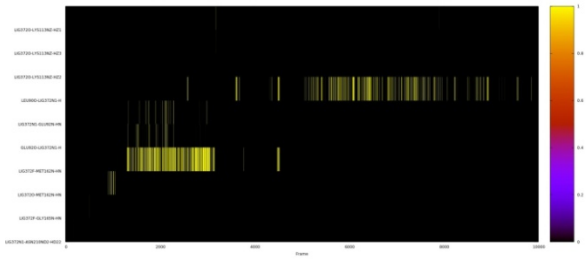
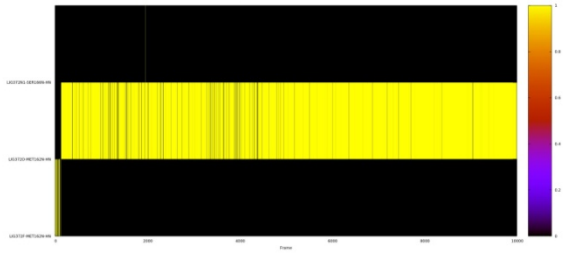


Figure 7: (a) RMSD of protein-ligand **132** complex (b) RMSD of protein-ligand **160** complex over the complete 100ns simulation



(a)



(b)

Figure 8: Hydrogen bonds formation between specific amino acid residues of MNK2 and (a) ligand **132** (b) ligand **160** over the complete 100ns simulation

ADMET Prediction

For increasing the success rate of potential inhibitors, prediction of ADMET properties using *in-silico* tools is an efficient way. Physicochemical properties, drug likeness properties, pharmacokinetics, lipophilicity, water solubility and toxicity properties of compounds **132** and **160** predicted using SwissADME and AdmetSAR online *in-silico* tools are shown in Table 2.

Lipophilicity was depicted by applying the average of physics Log Po/w (iLOGP), topological Log Po/w(MLOGP), atomistic Log Po/w (WLOGP), atomistic and knowledge Log Po/w(XLOGP3), and hybrid fragment/topological Log Po/w (SILICOS-IT) which is called as the consensus Log Po/w. The logarithmic solubility, or Log S (ESOL), was used to estimate the water solubility. Compounds **132** and **160** have consensus Log Po/w values of less than 5, both exhibited Log S (ESOL) values between -4 and -6, and both exhibited moderate water solubility. According to Lipinski's rule, an orally bioactive drug must adhere to the following criteria: molecular weight (MW) < 500 Da, Log Po/w ≤ 5, number of hydrogen bond acceptors (HBA) ≤ 10, and hydrogen bond donors (HBD) ≤ 5. Compounds **132** and **160** both complied with Lipinski's criterion. They also followed Ghose's, Veber's, Egan's and Mugegge's rules. Therefore, both compounds **132** and **160** fulfilled drug likeliness properties. Additionally, they both displayed a bioavailability score of 0.55 and have an excellent blood-brain barrier (BBB) and human intestinal absorption probability (Table 2). Compound **132** is depicted as AMES toxic while compound **160** is depicted as Non-AMES toxic. Both the compounds are predicted as non-carcinogenic. Thus, compound **160** falls under the acceptable ranges of ADMET properties.

Table 2: Predicted ADMET Properties of compound numbers **132** and **160**

Properties	Compound no. 132	Compound no. 160
Molecular Weight	344.38	392.85
NRB	4	5
HBA	2	2
HBD	1	1
MR	104.7	112.98
TPSA	32.34	32.34
Consensus Log Po/w	4.11	4.69
Log S (ESOL)	-4.98	-5.53
Solubility Class	Moderately Soluble	Moderately Soluble
Lipinski Rule	Yes	Yes
Ghose Rule	Yes	Yes
Veber Rule	Yes	Yes
Egan Rule	Yes	Yes
Muegge Rule	Yes	Yes
Bioavailability Score	0.55	0.55
BBB Absorption	0.9950	0.9942
Human Intestinal Absorption	0.9894	1.0000
AMES toxicity	AMES toxic	Non-AMES toxic
Carcinogenicity	Non-carcinogenic	Non-carcinogenic

DISCUSSION

Overexpression levels of eukaryotic initiation factor 4E (eIF4E) is the reason for various types of cancers. Mitogen-activated protein kinase interacting kinases, MNK1 and MNK2 served as phosphorylation of eIF4E. As MNK2 showed higher constitutive ability for phosphorylating eIF4E compared to MNK1, therefore we have done the in-silico studies on MNK2 protein target. Previously reported, Isoindolin-1-one compounds exhibited certain anti-cancer activities. So, in this study, we made a library of 180 compounds containing isoindolin-1-one cyclic ring. Firstly, we initially screened the library by structure based virtual screening against MNK2 protein target including co-crystal ligand of MNK2 protein. Virtual screening results suggested docking scores of isoindolin-1-one compounds in the range from -6.9 to -10.4 Kcal/mol. Compounds **132** and **160** showed binding energies of -10.4 and -10.3 Kcal/mol respectively along with strong conventional hydrogen bonding with MET162 amino acid residue of MNK2 showing their strong and effective inhibition against MNK2 and thus compounds **132** and **160** would further undergo FMO, MDS, ADMET analysis.

To check the accuracy and credibility of docking protocol, co-crystal ligand was again docked against MNK2. The

superimposition image of experimental and redocked poses of co-crystal ligand and RMSD between both the poses less than 0.5 Å validation the docking method. FMO studies of compounds **132** and **160** were conducted to check the contribution of particular amino acid residue of MNK2 to the binding affinity. MET162 amino acid residue is majorly contributing to the binding affinity for both the compounds **132** and **160** with total pair interaction energy values of -20.008 Kcal/mol and -16.215 Kcal/mol respectively. The reason behind this may be the conventional hydrogen bonding between MET162 and both compounds.

MDS under physiological conditions of both compounds were done to check the stability of docked complexes. Compound **160** showed lower average RMSD value of whole protein-ligand complex which was 0.27 nm compared to compound **132**, 0.49 nm. Hydrogen bonding occupancy over the simulation suggested that compound **160** maintained a hydrogen bonding with active MET162 amino acid residue with 95% occupancy while compound **132** showed only 20% hydrogen bonding occupancy. These MDS results further validated the potential for compound **160** to be an effective inhibitor of MNK2 protein. Physicochemical properties, pharmacokinetics, drug-likeness, lipophilicity, water solubility and toxicity of compounds **132** and **160** were examined using SwissADME and AdmetSAR tools. Both the compounds showed acceptable lipophilicity and moderate water solubility. Both compounds are in favour of Lipinski's rule of five and drug-likeness rules. In addition. Both showed good blood-brain barrier, human instestinal absorption and bioavailability score. In contrary, compound **132** was examined as AMES toxic and non-carcinogenic while compound **160** was examined as non-AMES toxic and non-carcinogenic concluding compound **160** as a promising inhibitor against MNK2 protein.

There were various previous research studies reported on MNK2 inhibitors. Yen et al. used structure-based virtual screening approach for the identification of compound labelled as K783-0308 as potential MNK2 inhibitor which showed docking score value of -30.2 using FlexX software. This compound showed hydrogen bonding interactions with MET162, LEU90, SER166 resiudes of MNK2.[61] Another report on O-methylated flavonal, 6-chloro-3-hydroxy-2-(pyridin-4-yl)-4H-chromen-4-one displayed hydrogen bonding interactions with MET162, GLU**160** and SER166 residues of MNK2. They used LeadIT software for molecular docking.[62] Xing et al. reported the binding interaction analysis of (3-(4-(indolin-1-yl)pyrido[3,2-d]pyrimidin-6-yl)phenyl)(piperazin-1-yl)methanone compound with mutant MNK2 protein with the help of Schrodinger maestro 11.5 software. This compound showed hydrogen bonding with MET162 residue.[63] Han et al. visualized the binding interactions of (4-(4-((4-Fluoro-3-isopropoxyphenyl)amino)pyrido[3,2-d]pyrimidin-6-yl)phenyl)(piperazin-1-yl)methanone against MNK2 using Schrodinger 2016 version and founded the formation of hydrogen bonds with MET162 and GLU**160** amino acid residues.[64] Thus, our *in-silico* findings on isoindolin-1-one library on MNK2 inhibitors are consistent with previous *in-silico* studies.

CONCLUSION

The activation of eIF4E is caused by MNK2, and the overexpression of eIF4E is the cause of many cancer types. These in-silico investigations employed a variety of cheminformatic techniques to find potential MNK2 protein target inhibitors. Molecular docking studies of the library of 180 isoindolin-1-one compounds showed that sixteen derivatives of isoinsolin-1-ones demonstrated docking scores higher than or equal to the co-crystal ligand of MNK2 protein target chosen (PDB ID:6CK6) and showed hydrogen bonding with active MET162 amino acid residue of MNK2, which is located in the hinge region. FMO calculations of two top-docked scoring compounds **132** and **160** displayed that MET162 amino acid residue is majorly contributing to high binding affinity. Through ADMET analysis, compound number **160** lies in the acceptable range of drug-likeliness, pharmacokinetics studies, and toxicity studies. 100 ns MD simulation of compound number **160** showed the favorable value of RMSD and formation of hydrogen bonding with MET162 residue with 95% occupancy. Further, in-vitro and in-vivo studies of these isoindolin-1-one compounds will be done for the validation of these in-silico findings and thus for the development of potent anti-cancer drugs. Addressing the complexity of human biology that are not fully reflected in computational models will be a crucial issue in converting these *in-silico* results into clinical applications. Furthermore, via thorough clinical studies, any differences between expected and actual clinical efficacy must be carefully examined. It is essential to make collaborative efforts between researchers, clinicians, and regulatory agencies for the navigation of these challenges and apply these innovations from laboratory research to clinical practice.

SUPPLEMENTARY INFORMATION

A supplementary data file containing 2-D structures, binding energy, and 2-D interactions of best docked poses of complete library of isoindolin-1-ones with MNK2 protein is available.

GLOSSARY OF SPECIALIZED TERMS

Active Site: An active site is the pocket on the surface of protein created by arrangement of particular amino acid residues where ligand binds.

Co-crystal ligand: A ligand which is present in the crystal structure of protein in bounded form.

MMFF94 Force Field: Merck Molecular Force Field is a broadly parameterized, computationally derived force field for organic and drug-like compounds. This force field give accurate descriptions of molecular geometries, conformations, and energies.

Grid Box: A 3-D region of protein in space where the ligand is expected to interact during molecular docking.

SMILES: SMILES are Simplified Molecular Input Line Entry System. SMILES are the linear and text-based notations of a chemical structure.

Physiological pH: pH level of human body under healthy conditions

Exhaustiveness: Represent the number of initial random runs for ligand conformation search and optimization.

CHARMM Force Field: CHARMM stands for Chemistry at HARvard Macromolecular Mechanics. This force field is mainly used for molecular dynamics simulations including parameters related to bond stretching, angle bending, and nonbonded interactions.

Topology: Topology provides the information about the molecular structure in which bond atom types, bond angles, dihedral angles, bonds connectivity, all type of information is included.

TIP3P Water Model: TIP3P stands for Transferable Intermolecular Potential with 3 Points. It is a simple water model in which water molecule has three sites, two sites are for hydrogen atoms and one for oxygen atom with their particular charges and geometric parameters.

Periodic Boundary Conditions: These are a set of boundary conditions used for mimicing an infinite system by replicating the simulation box in all directions.

NVT Ensemble: Canonical ensemble in which Number of particles (N), Volume (V), and Temperature (T) are held constant throughout the simulation.

NPT Ensemble: Isothermal-isobaric ensemble in which Number of particles (N), Pressure (P), and Temperature (T) are held constant during the simulation.

V-rescale Thermostat: This method is used to control the molecular dynamics simulation system's temperature.

Parrinello-Rahman Barostat: This method is used to control the pressure of simulation system.

AMES toxicity: This toxicity test predicts whether the chemical compound could be carcinogenic through DNA mutation.

ACKNOWLEDGMENTS

The authors are thankful to Delhi Technological University (DTU) for infrastructure and support. Kunal is thankful to DTU for financial assistant.

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

REFERENCES AND NOTES

1. S.J. Watkins, C.J. Norbury. Translation initiation and its deregulation during tumorigenesis. *Br. J. Cancer* **2002**, 86 (7), 1023–1027.
2. M.J. Clemens. Targets and mechanisms for the regulation of translation in malignant transformation. *Oncogene* **2004**, 23 (18), 3180–3188.
3. J.R. Graff, S.G. Zimmer. Translational control and metastatic progression: Enhanced activity of the mRNA cap-binding protein eIF-4E selectively enhances translation of metastasis-related mRNAs. *Clin. Exp. Metastasis* **2003**, 20 (3), 265–273.
4. A. Lazaris-Karatzas, K.S. Montine, N. Sonenberg. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5′ cap. *Nature* **1990**, 345 (6275), 544–547.
5. M. Kumari, B.S. Chhikara, P. Singh, B. Rathi. Signaling and molecular pathways implicated in oral cancer: A concise review. *Chem. Biol. Lett.* **2024**, 11 (1), 652.
6. S. Fan, S.S. Ramalingam, J. Kauh, et al. Phosphorylated eukaryotic translation initiation factor 4 (eIF4E) is elevated in human cancer tissues. *Cancer Biol. Ther.* **2009**, 8 (15), 1463–1469.
7. B. Anthony, P. Carter, A. De Benedetti. Overexpression of the proto-oncogene/translation factor 4E in breast-carcinoma cell lines. *Int. J. Cancer* **1996**, 65 (6), 858–863.
8. A. Yoshizawa, J. Fukuoka, S. Shimizu, et al. Overexpression of phospho-eIF4E is associated with survival through AKT pathway in non-small cell lung cancer. *Clin. Cancer Res.* **2010**, 16 (1), 240–248.
9. C.A.O. Nathan, S. Franklin, F.W. Abreo, et al. Expression of eIF4E during head and neck tumorigenesis: Possible role in angiogenesis. *Laryngoscope* **1999**, 109 (8), 1253–1258.
10. I.B. Rosenwald, J.J. Chen, S. Wang, et al. Upregulation of protein synthesis initiation factor eIF-4E is an early event during colon carcinogenesis. *Oncogene* **1999**, 18 (15), 2507–2517.
11. N. Robichaud, B.E. Hsu, R. Istomine, et al. Translational control in the tumor microenvironment promotes lung metastasis: Phosphorylation of eIF4E in neutrophils. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, 115 (10), E2202–E2209.
12. R.M. Rowlett, C.A. Chrestensen, M. Nyce, et al. MNK kinases regulate multiple TLR pathways and innate proinflammatory cytokines in macrophages. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **2008**, 294 (2), 452– 459.
13. X. Jin, R. Yu, X. Wang, C.G. Proud, T. Jiang. Progress in developing MNK inhibitors. *Eur. J. Med. Chem.* **2021**, 219, 113420.
14. S. Joshi. Mnk kinase pathway: Cellular functions and biological outcomes. *World J. Biol. Chem.* **2014**, 5 (3), 321.
15. M. Buxade, N. Morrice, D.L. Krebs, C.G. Proud. The PSF·p54nrb complex is a novel Mnk substrate that binds the mRNA for tumor necrosis factor α. *J. Biol. Chem.* **2008**, 283 (1), 57–65.
16. A. Dreas, M. Mikulski, M. Milik, et al. Mitogen-activated Protein Kinase (MAPK) Interacting Kinases 1 and 2 (MNK1 and MNK2) as Targets for Cancer Therapy: Recent Progress in the Development of MNK Inhibitors. *Curr. Med. Chem.* **2017**, 24 (28), 3025–3053.
17. N. Hay. Mnk earmarks eIF4E for cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107 (32), 13975–13976.
18. T. Mukherjee, S. Mohanty, J. Kaur, et al. Exploring small-molecule inhibitors targeting MAPK pathway components: Focus on ERK, MEK1, and MEK2 kinases in cancer treatment. *Chem. Biol. Lett.* **2024**, 11 (2), 659
19. T. Ueda, M. Sasaki, A.J. Elia, et al. Combined deficiency for MAP kinase-interacting kinase 1 and 2 (Mnk1 and Mnk2) delays tumor development. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107 (32), 13984–13990.
20. N. Robichaud, S. V. Del Rincon, B. Huor, et al. Phosphorylation of eIF4E promotes EMT and metastasis via translational control of SNAIL and MMP-3. *Oncogene* **2014**, 34 (16), 2032–2042.
21. L. Furic, L. Rong, O. Larsson, et al. EIF4E phosphorylation promotes tumorigenesis and is associated with prostate cancer progression. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107 (32), 14134–14139.
22. P.E.D. Lachance, M. Miron, B. Raught, N. Sonenberg, P. Lasko. Phosphorylation of Eukaryotic Translation Initiation Factor 4E Is Critical for Growth. *Mol. Cell. Biol.* **2002**, 22 (6), 1656–1663.
23. H.G. Wendel, R.L.A. Silva, A. Malina, et al. Dissecting eIF4E action in tumorigenesis. *Genes Dev.* **2007**, 21 (24), 3232–3237.
24. S. Lim, T.Y. Saw, M. Zhang, et al. Targeting of the MNK-eIF4E axis in blast crisis Chronic myeloid leukemia inhibits leukemia stem cell function. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, 110 (25), 2298– 2307.
25. C. Mazewski, L.C. Platanias. MNK Proteins as Therapeutic Targets in Leukemia. *Onco. Targets. Ther.* **2023**, 16, 283–295.
26. S. Li, J. shu Chen, X. Li, X. bai, D. Shi. MNK, mTOR or eIF4E-selecting the best anti-tumor target for blocking translation initiation. *Eur. J. Med. Chem.* **2023**, 260, 115781.
27. S. Liu, W. Wang, Y. Ning, et al. Exosome-mediated miR-7-5p delivery enhances the anticancer effect of Everolimus via blocking MNK/eIF4E axis in non-small cell lung cancer. *Cell Death Dis.* **2022**, 13 (2), 129.
28. C.E.J. Moore, J. Pickford, F.R. Cagampang, et al. MNK1 and MNK2 mediate adverse effects of high-fat feeding in distinct ways. *Sci. Rep.* **2016**, 6 (1), 23476.

29. C.S. Conn, H. Yang, H.J. Tom, et al. The major cap-binding protein eIF4E regulates lipid homeostasis and diet-induced obesity. *Nat. Metab.* **2021**, 3 (2), 244–257.
30. L.Y. Sandeman, W.X. Kang, X. Wang, et al. Disabling MNK protein kinases promotes oxidative metabolism and protects against diet-induced obesity. *Mol. Metab.* **2020**, 42, 101054.
31. A.J. Waskiewicz, A. Flynn, C.G. Proud, J.A. Cooper. Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. *EMBO J.* **1997**, 16 (8), 1909–1920.
32. Z. Guo, G. Peng, E. Li, et al. MAP kinase-interacting serine/threonine kinase 2 promotes proliferation, metastasis, and predicts poor prognosis in non-small cell lung cancer. *Sci. Rep.* **2017**, 7 (1), 10612.
33. P. Shu, Y. Lei, G. Gao, et al. MNK: A Novel Promising Target for Cancer Immunotherapy. *Curr. Med. Chem.* **2024**, 31.
34. A. Fernandez, P.J. Monsen, L.C. Platanias, G.E. Schiltz. Medicinal chemistry approaches to target the MNK-eIF4E axis in cancer. *RSC Med. Chem.* **2023**, 14 (6), 1060–1087.
35. Y. Zhan, J. Guo, W. Yang, et al. MNK1/2 inhibition limits oncogenicity and metastasis of KIT-mutant melanoma. *J. Clin. Invest.* **2017**, 127 (11), 4179–4192.
36. H. Yang, L.R. Chennamaneni, M.W.T. Ho, et al. Optimization of Selective Mitogen-Activated Protein Kinase Interacting Kinases 1 and 2 Inhibitors for the Treatment of Blast Crisis Leukemia. *J. Med. Chem.* **2018**, 61 (10), 4348–4369.
37. S.H. Reich, P.A. Sprengeler, G.G. Chiang, et al. Structure-based Design of Pyridone-Aminal eFT508 Targeting Dysregulated Translation by Selective Mitogen-activated Protein Kinase Interacting Kinases 1 and 2 (MNK1/2) Inhibition. *J. Med. Chem.* **2018**, 61 (8), 3516–3540.
38. X. Jin, J. Merrett, S. Tong, et al. Design, synthesis and activity of Mnk1 and Mnk2 selective inhibitors containing thieno[2,3-d]pyrimidine scaffold. *Eur. J. Med. Chem.* **2019**, 162, 735–751.
39. W. Xu, S. Kannan, C.S. Verma, K. Nacro. Update on the Development of MNK Inhibitors as Therapeutic Agents. *J. Med. Chem.* **2022**, 65 (2), 983–1007.
40. D.H. Miles, X. Yan, R. Thomas-Tran, et al. Discovery of Potent and Selective 7-Azaindole Isoindolinone-Based PI3Kγ Inhibitors. *ACS Med. Chem. Lett.* **2020**, 11 (11), 2244–2252.
41. G. Chessari, I.R. Hardcastle, J.S. Ahn, et al. Structure-Based Design of Potent and Orally Active Isoindolinone Inhibitors of MDM2-p53 Protein-Protein Interaction. *J. Med. Chem.* **2021**, 64 (7), 4071–4088.
42. D. Ji, L. Zhang, Q. Zhu, et al. Discovery of potent, orally bioavailable ERK1/2 inhibitors with isoindolin-1-one structure by structure-based drug design. *Eur. J. Med. Chem.* **2019**, 164, 334–341.
43. S. Sharma, Y. Monga, A. Gupta, S. Singh. In silico screening and molecular docking study of quinoline based compounds with Human kallikrein 7 in complex with 1,4-diazepane-7-one 1-acetamide derivative receptor target for potential antibacterials. *J. Mol. Chem.* **2023**, 3 (1), 585.
44. J.D. Hansen, K. Condroski, M. Correa, et al. Protein Degradation via CRL4 CRBN Ubiquitin Ligase: Discovery and Structure-Activity Relationships of Novel Glutarimide Analogs That Promote Degradation of Aiolos and/or GSPT1. *J. Med. Chem.* **2018**, 61 (2), 492–503.
45. BIOVIA. Biovia Discovery Studio 2021 Client. *Dassault Systèmes San Diego, CA, USA* **2021**, 21 (1.0.20298]).
46. P. Tangyuenyongwatana, W. Gritsanapan. Virtual screening of potential inhibitors against SARS-CoV-2 main proteases (Mpro) by dual docking with FRED and AutoDock Vina programs. *J. Mol. Chem.* **2021**, 1 (1), 105.
47. N.M. O'Boyle, M. Banck, C.A. James, et al. Open Babel: An Open chemical toolbox. *J. Cheminform.* **2011**, 3 (10)
48. O. Trott, A.J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, 31 (2), 455–461
49. K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi. Fragment molecular orbital method: An approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, 313 (3–4), 701–706.
50. R. Kumar, M.P. Chaudhary, N. Chauhan. Recent advances and current strategies of cheminformatics with artificial intelligence for development of molecular chemistry simulations. *J. Mol. Chem.* **2022**, 2 (2), 440.
51. B.T. Tung, N.B. Kim, P.H. Minh. In silico screening of phenolic acids as potential inhibitors of SARS-CoV-2 RNA-dependent RNA polymerase. *J. Mol. Chem.* **2021**, 1 (1), 104.
52. S. Pronk, S. Páll, R. Schulz, et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, 29 (7), 845–854.
53. N. Azad, R. Kakkar. Computational molecular docking analysis of Doxifluridine and its metabolites to identify potential hits for PDHK1. *J. Mol. Chem.* **2024**, 4 (2), 693.
54. V. Zoete, M.A. Cuendet, A. Grosdidier, O. Michielin. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **2011**, 32 (11), 2359–2368.
55. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79 (2), 926–935.
56. M. Parrinello, A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, 52 (12), 7182–7190.
57. D.R. Roe, T.E. Cheatham. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, 9 (7), 3084–3095.
58. D.R. Roe, T.E. Cheatham. Parallelization of CPPTRAJ enables large scale analysis of molecular dynamics trajectory data. *J. Comput. Chem.* **2018**, 39 (25), 2110–2117.
59. A. Daina, O. Michielin, V. Zoete. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, 7, 42717–42717
60. F. Cheng, W. Li, Y. Zhou, et al. AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* **2012**, 52 (11), 3099–3105.
61. S.C. Yen, L.C. Chen, H.L. Huang, et al. Identification of a dual FLT3 and MNK2 inhibitor for acute myeloid leukemia treatment using a structure-based virtual screening approach. *Bioorg. Chem.* **2022**, 121, 105675.
62. S.C. Yen, Y.W. Wu, C.C. Huang, et al. O-methylated flavonol as a multi-kinase inhibitor of leukemogenic kinases exhibits a potential treatment for acute myeloid leukemia. *Phytomedicine* **2022**, 100, 154061.
63. K. Xing, H. Zhang, S. Wang, et al. Design, synthesis and biological evaluation of 4-(indolin-1-yl)-6-substituted-pyrido[3,2-d]pyrimidine derivatives as Mnk1/2 inhibitors. *Eur. J. Med. Chem.* **2024**, 272, 116499.
64. Y. Han, H. Zhang, S. Wang, et al. Optimization of 4,6-Disubstituted Pyrido[3,2- d]pyrimidines as Dual MNK/PIM Inhibitors to Inhibit Leukemia Cell Growth. *J. Med. Chem.* **2021**, 64 (18), 13719–13735.

Supply Chain Management and its Role in Rail Transport

Manoj Kumar[1*], *Vipin*[2], *Ashish* Agarwal[2]

[1]*Department of Mechanical Engineering, Delhi Technological University, Delhi, India*
[2]*School of Engineering and Technology, IGNOU, New Delhi, India.*

[*] Corresponding Author: mky2011@gmail.com

Abstract. This empirical study article explores the role of supply chain management (SCM) in the context of rail transportation and provides an outline of SCM. Effective supply chain management techniques are crucial for the smooth running of the rail industry, which plays a vital part in global logistics networks. The study looks at a number of SCM-related topics in rail transport, such as logistics planning, inventory control, transportation, and supply chain partner cooperation. To assess their perceptions of SCP in transport coordinated factors from the expense and administration viewpoints, a survey and interviews with the rail transport logistics industry were undertaken. The study's findings demonstrated that rail transportation has reached a particular level of maturity and sophistication in its SCP. The influence of SCM on rail transport efficiency, cost savings, and customer satisfaction is examined using empirical data and industry case studies. The study's conclusions offer management guidance to the rail industry so they may comprehend their SCP in transport coordinated operations and recognize regions for development. The results emphasize the value of SCM in streamlining rail transport operations and offer guidance for both scholars and industry professionals.

Keywords- Supply chain Management, Rail transport, Supply chain performance

1. Introduction

The process of obtaining raw materials and converting them into intermediate products is handled by a network of facilities and shipping routes called a supply chain. And the sale of these completed goods to consumers. Both assistance associations and assembling organizations use supply organizations, but the intricacy of the chains shifts generally from one industry to another and organization to firm. The globalization of the supply chain has raised the significance of rail transportation systems. When compared to road transit, this relevance is mostly explained by its high speed, safety, dependability, reduced cost, and environmental friendliness [1, 2]. This study looks at the overall railroad freight system, its function in the chain of supply, academic research, and contemporary issues. Effective supply chain management techniques are crucial for the smooth running of the rail industry, which plays a crucial role in world logistics networks.Supply chain management (SCM) is the management of goods, information, and funds as they move from a supplier to a manufacturer to a wholesaler to a retailer to a client. Supply chain management is tasked with coordinating and integrating these flows either between or within enterprises. SCM is much of the time seen as falling somewhere close to organizations that are completely upward coordinated, where each part of the material stream is constrained by a solitary business, and organizations where each channel part runs their own business. Accordingly, successful chain the board relies upon collaboration between the chains's all's members [5]. Reducing inventory is claimed to be the ultimate goal of every efficient supply chain management system (under the presumption that things are always accessible when needed).

Three main parts may be found in SCM flows [1].
• The flow of product
• Information exchange
• The flow of finance
Product flow: Item stream alludes to the exchange of products from a provider to a client as well as any profits or overhauling prerequisites. The internal product flow must be flawlessly organized in order for the global supply chain to function properly. When used appropriately, product flow management combines a number of responsibilities to align the supply chain [4, 5]. Information flow: This entails sending orders and updating the delivery status. The benefits of timely supply chain information include lower costs, fewer inventory, increased throughput, shortened cycle times, and the best possible customer service. The efficient flow of information and materials among a network of clients and suppliers is the fundamental foundation of supply chain management [2, 4, and 5]. Businesses may easily find better and more extensive ways to speed up and improve information flow by leveraging the Internet. Financial flow: Credit conditions, instalment timetables, and transfer and title possession arrangements all make up the monetary stream. However, the financial flows that distribute the financial resources shared by almost all traded economic activity are a key concern for all supply chain participants. The individual supply chain partners stop operating without such financial flows. Therefore it becomes hard for

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

such parties to work together, which is frequent in modern supply chains. Bartering, for example, lacks the flexibility required for the coordination and management of the intricate, geographically distributed systems that make up modern supply chains. Financial flows are viewed from the operations perspective that supply chain management researchers have historically accepted as their own as a common, almost automatic corollary of the transportation of goods and services both inside and between enterprises [1]. The ability to transmit material, goods, services, personnel, or information is their most crucial quality. Even though none of the major supply chain management models openly identify it, the cost associated with it is also significant. Specialist roles like financial managers and management accountants are responsible for handling the specifics of financial flows and their cost implications [6].Whether a supply chain is domestic or worldwide, there are specific components that must be included. No matter how long the chain is, essential components like visibility, technology, and adaptability must be incorporated seamlessly for the supply chain to operate well. Technology enables visibility and accelerates the supply chain. Another important factor is visibility [7]. This is especially important to enable businesses to strategically manage their supply chains by identifying multiple locations all through the supply chain where goods can be stored to lower the risk of delays. It is far simpler to become visible domestically than internationally. Additionally essential to the supply chain's success is flexibility. Companies must make sure that their partners and supply chains can quickly integrate alternative locations if the situation calls for a speedy response [8]. Today's economy makes it practically impossible for a business to operate without using transportation [9]. Logistics' primary sub-function of transportation, which gives products their time and locational utility, is transportation. The ability to achieve the well-known seven Rs—the right product, in the right quantity, in the right condition, at the right place, at the right time, for the right client, at the right cost—is, in fact, the foundation of the all-inclusive supply chain [6]. The effect of transportation on a nation's economy should be considered in order to understand the significance of transportation. According to studies, approximately 50% of India's total logistics costs—which account for close to 23% of the country's GDP—are attributable to transportation. Estimates indicate that the expense represents about 6% of the GNP in the United States [5]. Roads, the road freight industry, trains, ports and shipping, and pipelines are all active components of the key infrastructure needed for moving products in India. One of the world's largest road networks is in India. Since 1951, the growth rate of road traffic has been 0%, and if there had been a more extensive and pervasive road network, it would have been greater. According to estimates, just 30% of the paved roads are in good shape. This contrasts poorly with other nations (Indonesia and Brazil have 40%, Korea has 80%, Japan and the US have more than 90%). The primary thoroughfares linking ports, state capitals, commercial and tourism hubs, and bordering nations are known as national highways (NHs). Nearly 50% of all traffic is transported by NHs, which makes up less than 4% of the entire road network [6, 7]. Due to sluggish mobility, flaws in the road network are resulting in significant economic losses. The industry's high inventory costs as a result of the delays on the highways and ports also have an impact on its ability to compete with other countries' industries that use JIT (just-in-time) inventory standards. Foreign investment decisions, which frequently place a high value on the infrastructure, are influenced by the port congestion and the underdeveloped air services. According to international trends, traffic is moving away from railroads as highway and airline technology develops [8]. The Railways have continued to rule in continental economies like those of the United States, China, and Russia. In addition to the environmental concerns that in recent years have prompted a reconsideration even in the developed world, India's size, location, and resource endowments require a major role for the Railways [10]. According to The Times of India (2005), India's railway system, which spans 62,915 km and includes 7068 stations, is the world's fifth-largest after the United States (2,24 000 km), Russia (1,54 000 km), China (78,000 km), and Canada (72,961 km). Recently, Indian Railways has begun to appear lot more effective, and income is also rising [10]. India has the ability to network with the CIS nations, particularly through Iran, to provide the necessary infrastructure thanks to its large geographic size and advantageous location [11]. This is made possible by the extensive network of roads, trains, ports, and airports that must be updated both technically and administratively. Because it establishes a connection between a company's clients and the suppliers of the goods or services it offers the market. This specific design is vital because it takes into account distribution costs, which make up a sizeable portion of overall marketing expenses [12].James Watt's development of the stationary steam engine is where RT first began. Significant development on steam locomotives didn't start until after his patent ran out in 1800. A number of inventors began to work on enhancing the Watt plan. Obviously, horses pulled a few carts on wooden rails much before then [5]. The principal steam locomotives enabled designers to make another rail framework during the early part of the 19th century while operating under uncompressed pressure. One of the key historical moments for the development of freight transportation was the creation of the train. Richard Trevithick constructed a cargo vehicle in the south of Britain that became the first engine to travel on steel rails. The development of the freight train as well as the traveller train officially started with this action [6].Researchers have long been concerned about how well rail freight systems perform, and many of these concerns have been resolved thanks to technological developments. The two biggest issues facing supply chain managers are on-time delivery of products and logistics optimization [8]. The supply chain's use of rail systems has been significantly impacted by cargo delivery delays, rail route modifications, energy efficiency, and environmental fortification. Additionally, supply chain managers are concerned with increasing safety to prevent accidents. Losses of life and money are constantly in the news. Some of the management solutions in this area include the installation of centralized management infrastructure, mechanical interlocks to stop administrator blunder, and radio correspondence frameworks [9, 10].

2. Rail Transportation's Importance in the Supply Chain

RT is essential to third-party providers and the logistics supply chain. Road transportation has been recognized as having a place in RT. Furthermore, RT assumes a basic part in keeping up with progressing contest and the creation of business benefit through its reconciliation into the production network because it is necessary for future economic prosperity. Figure 1 depicts the amount of rail freight moved throughout the US and the EU from 2006 to 2019 [13,14]. In the USA [13,14] and EU [15] between 2006 and 2019, the average volume of rail freight transported was 2.4 and 0.385 trillion tonnekilometers, respectively. Additionally, statistics show that the US has much larger rail transport share in the supply chain than the EU (R FTS, 2021). Alternatively, Oceania, Asia, and Around 3.5 billion ton-kilometers of goods were transported by rail via the Middle East in the same year, as opposed to 3.1 billion ton-kilometers through Europe and Turkey in 2019 [13,14]. In 2018 and 2019 there was an overall increase in RT traffic worldwide [13,14]. It is also important to analyze the proportion of RT compared to road and inland waterway transportation. In the EU, the proportion of RT has been roughly equal to that of road transport [15]. In contrast, the utilization of the rail network for material transportation has been around three times greater than that of inland waterways [15]. The data presented show how crucial railroad networks are to logistics and the supply chain for materials. Only roads are more important than rail networks, and their significance is growing yearly.

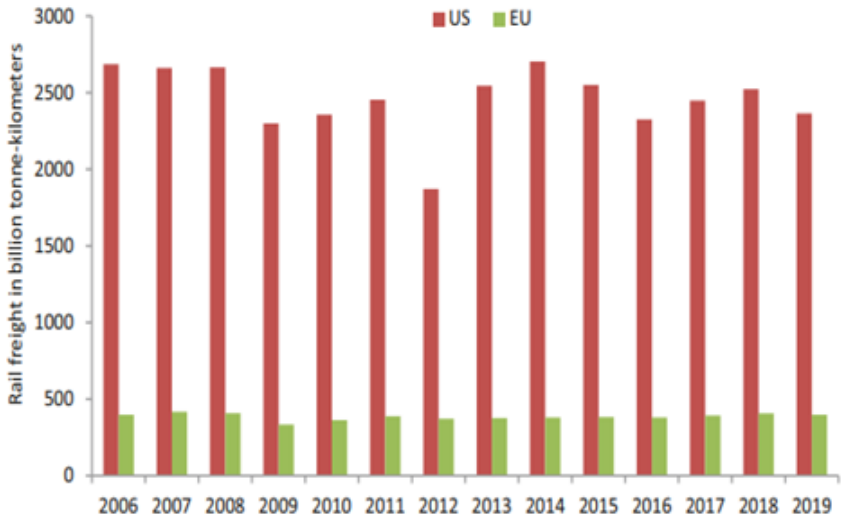


Fig 1. The amount of freight transported by rail in the US and the EU between 2006 and 2019 [13, 14, and 15]

3. Elements of the Railway Network

One of the country's most broad land transportation networks is the rail route. Utilizing fuel or a remote sending force (electric impetus) on moveable steel wheels, trains move independently on a special steel guide that comprises two parallel tracks [16]. Railways are built to transport people and goods, and thanks to their tremendous capacity, they may go any distance in a variety of environments (including urban, suburban, local and interurban areas). The radius for passenger travel is typically around 1500 km, however for freight transport; this distance can be significantly more. Rail systems can be broken down into three categories. The rail route framework is stalled into power vehicles, singular vehicles, trailers vehicles, and designing vehicles in the first aspect. These are then divided into groups based on the fuel type and use (passenger or freighter). Second, the type of route can be used to categorize rail systems. As a result, overpasses, underpasses, normal routes, and bridge routes are taken into account. Train types are included in the third category of categories. This aspect includes loco-hauled, multiunit, and push-pull trains that can haul either freight or passengers, depending on the type of wagon.

4. Supply Chain Model

To consider the exhibition prerequisites of accomplice undertakings in a production network, the Production network Gathering laid out the Inventory network Tasks Reference (SCOR) model [17]. The model considers inventory network exercises to be a progression of cycles, and it interfaces different hierarchical inventory network processes with four parts: plan, source, make, and convey. Every part has its own estimation rules on the accompanying subjects: production network

unwavering quality, responsiveness/adaptability, expenses, and resources. The SCOR model, notwithstanding, was made with an emphasis on creation methods. With regards to support based transport coordinated operations, the idea is a long way from effectively reasonable for evaluating SCP. Furthermore, there is a shortage of examination on Store network execution (SCP) estimation in this area. To fill an exploration hole, this paper presents the discoveries of a concentrate on SCP in the vehicle coordinated factors area.

5. Rail Transportation in India: Transforming the Supply Chain

With a rapidly expanding economy and a huge geographic area, India confronts particular difficulties in effectively managing its supply chain. The need for creating reliable and sustainable logistics systems becomes critical as consumer demand for goods and services rises. The function of rail transportation in India's supply chain network is a key piece of this puzzle. The potential of rail transportation to transform the logistics industry and address many logistical and environmental concerns has recently come to the attention of Indian authorities [3]. Supply chain management is essential to the development and expansion of India's economy. India's supply chain is the most complex in the world due to its large population of over 1.3 billion people and diverse landscape. For economies to continue growing, it is essential to get goods from manufacturing centres to consumer markets both domestically and abroad. Policymakers have been forced to look into more sustainable and cost-effective solutions as a result of issues including traffic congestion, high logistics costs, and environmental concerns, with rail transportation emerging as a possible option. The integration of rail transport gives a golden opportunity to address current difficulties and establish a more effective and sustainable supply chain ecosystem as India works to improve its logistics skills to meet the demands of a growing economy [19].This section details the empirical research done to determine how supply chain management (SCM) affects rail transportation. The goal of the study is to assess the relationship between SCM procedures and key rail logistics performance metrics, such as effectiveness, cost containment, and customer satisfaction. The SCOR model served as the foundation for the creation of the measures that were used to assess SCP in transport coordinated operations in the present investigation [18]. The measurements offer viewpoints on effectiveness and efficiency to meet the performance standards of numerous supply chain participants in this industry, including shippers, suppliers of services, and consignees. They cover 3 parts of SCP in strategies for transportation. The three components of SCP in transport strategies are Service Effectiveness for Shippers (SES), providing the Services in Operations Effectiveness (OE), and SEC-Service Effectiveness for Consignees. In terms of dependability (REL) and responsiveness (RES), SES and SEC are client confronting measures that concentrate on how well a supply chain operation functions for transporters and representatives, individually. When performing transport logistics services,OE is bothered about the effective usage of assets both in terms of cost (COST) and assets (ASSET).The industry is widely characterized as including companies engaged in the business of transporting goods along a supply chain from a starting place, or transporters, to a mark of objective, or representatives. Understanding SCP in the business is vital given the rising ubiquity of SCM in the vehicle strategies area and the meaning of SCP appraisal for improving execution.To complete the empirical analysis, both primary and secondary data were gathered. Government statistics, academic journals, and industry reports were some secondary data sources.The impact of SCM on rail transportation was captured by a number of variables that were taken into consideration. These factors included cost indicators (such as transportation costs and inventory holding costs), efficiency metrics (such as on-time performance and transit time), and customer satisfaction indicators (such as service quality and responsiveness).

6. Results

Efficiency: The investigation showed a positive link between SCM procedures and measures of rail transportation efficiency. Effective SCM practices were adopted by rail businesses, which showed greater on-time performance and shorter travel times. Additionally, greater coordination and seamless operations as a result of the integration ofrail with other forms of transportation through SCM increased efficiency.Cost reduction: SCM procedures were found to significantly affect rail transportation costs. Companies that used effective SCM strategies saw a decrease in transportation costs as a result of enhanced inventory control, optimized routing, and decreased idle time. Collaboration in SCM enabled improved usage of rail assets, which reduced costs.Customer Satisfaction: The empirical investigation revealed a significant positive correlation between SCM procedures and client satisfaction in the rail transportation industry. Improved service quality, on-time deliveries, and better customer responsiveness were all made possible through effective SCM.

Increased communication and cooperation among supply chain partners led to improved levels of customer satisfaction.Regression analysis and correlation analysis were two statistical analysis methods that were used to examine the data that had been gathered. The goal of the analysis was to find important connections between SCM procedures and the chosen variables. The descriptive statistics are shown in Table 1, a confirmatory factor analysis (CFA) was used to further survey the legitimacy of the SCP construct. The SES, OE, and SEC constructs, which make up the three components of the SCP constructs, were shown to be legitimate based on the CFA results.

TABLE 1. Descriptive Statistics analysis of SCP Construct

Sr. No.	Factors	Mean	SD
1.	SCP	4.02	0.58
2.	SEC-RFL	4.12	0.66
3.	SEC-RES	4.13	0.55
4.	OE-COST	3.73	0.77
5.	OE-ASSET	3.85	0.46
6.	SES-REL	4.23	0.57

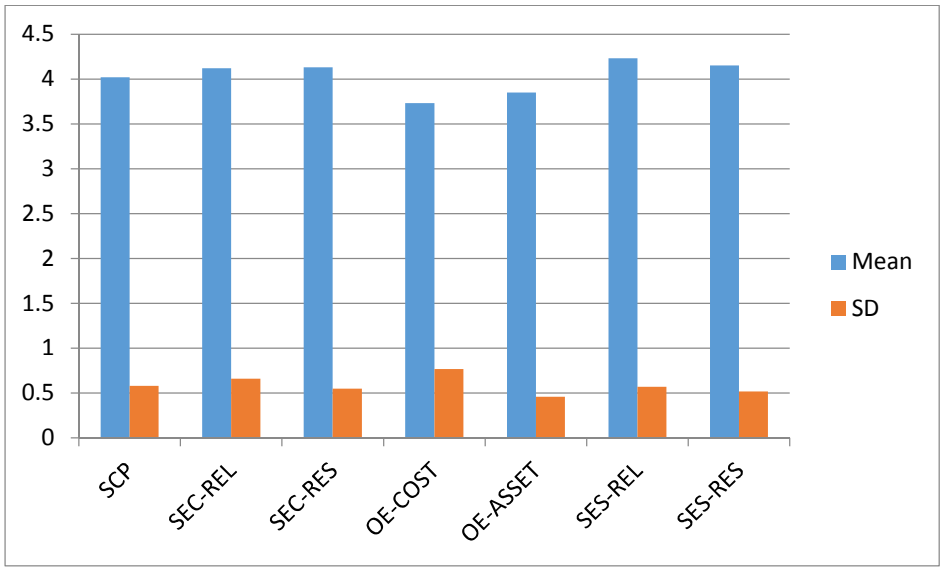


Fig 2. Supply Chain Performance Analysis in Rail Transport logistics

7. Conclusion

The empirical investigation supports the beneficial effects of SCM procedures on train transportation. SCM solutions that are put into practice well increase productivity, cut costs, and boost customer satisfaction. To improve their operations and obtain a competitive edge in the market, rail firms should give priority to adopting SCM concepts and collaborating with supply chain partners. In this analysis, the SCP build in the rail transportation strategies industry is observationally assessed. The use of SCP examination to the setting of rail transport operations is a huge commitment of this work. This study centers around this present reality assessment of SCP in the Railroad transport coordinated factors industry, rather than prior examinations on SCP, which have just centered around making a hypothetical system, (for example, the SCOR model).Future studies could look more deeply into particular SCM techniques and how they affect the efficiency of rail freight. An intriguing topic for more research would be the use of emerging technologies, such as blockchain and artificial intelligence, in the implementation of SCM for rail logistics. Overall, this empirical analysis helps to clarify the connection between SCM and rail transportation and offers information to researchers, logistics service providers, and rail operators that are trying to enhance supply chain performance in the rail sector.

References

1. Gholamizadeh, K., Zarei, E., & Yazdi, M. (2022). Railway Transport and Its Role in the Supply Chains: Overview, Concerns, and Future Direction. The Palgrave Handbook of Supply Chain Management, 1-28.
2. Maes, J., & Vanelslander, T. (2011). 10. The use of rail transport as part of the supply chain in an urban logistics context. City Distribution and Urban Freight Transport: Multiple Perspectives, 217.
3. Sudan, T., & Taggar, R. (2021). Recovering supply chain disruptions in post-COVID-19 pandemic through transport intelligence and logistics systems: India's experiences and policy options. Frontiers in Future Transportation, 2, 660116.
4. Lai, K. H., Ngai, E. W. T., & Cheng, T. C. E. (2004). An empirical study of supply chain performance in transport logistics. International journal of Production economics, 87(3), 321-331.
5. Zekhnini, K., Cherrafi, A., Bouhaddou, I., Benghabrit, Y., & Garza-Reyes, J. A. (2020). Supply chain management 4.0: a literature review and research framework. Benchmarking: An International Journal, 28(2), 465-501.
6. SNCFT. (2010). History of railway in the world. Retrieved from Tunisie: http://www.sncft.com.tn/ En/history-of-railway-in-the-world_11_96
7. Wanjari, S. S. (2020). Investigating warehousing operations from an integrated supply-chain and transportation approach. European Journal of Business and Management Research, 5(5).
8. Cohen, M. A., & Lee, H. L. (2020). Designing the right global supply chain network. Manufacturing & Service Operations Management, 22(1), 15-24.
9. Parfenov, A., Shamina, L., Niu, J., & Yadykin, V. (2021). Transformation of distribution logistics management in the digitalization of the economy. Journal of Open Innovation: Technology, Market, and Complexity, 7(1), 58.
10. Kushakova, M. N. (2022). Increasing the efficiency of logistics management in railway transport. ACADEMICIA: An International Multidisciplinary Research Journal, 12(2), 342-346.
11. Xie, P., Chen, Q., Qu, P., Fan, J., & Tang, Z. (2020). Research on financial platform of railway freight supply chain based on blockchain. Smart and Resilient Transportation, 2(2), 69-84.
12. Jayakrishnan, M., Mohamad, A. K., & Abdullah, A. (2019). A systematic literature review in enterprise architecture for railway supply chain of Malaysia transportation industry. Int. J. Eng. Res. Technol, 12(12), 2473-2478.
13. Carlier, M. (2021a). Rail freight traffic – Worldwide 2018 & 2019. Retrieved from Hamburg, Germany: https://www.statista.com/statistics/263543/global-performance-in-rail-freight-ser vices-by-region/
14. Carlier, M. (2021b). Volume of rail freight transported in the United States from 2002 to 2020. Retrieved from Hamburg, Germany: https://www.statista.com/statistics/245235/rail-freighttransportation-in-the-us
15. RFTS. (2021). Rail freight transport for main undertakings, EU-27, 2006–2019. Retrieved from Luxembourg https://ec.europa.eu/eurostat/statistics-explained/index.php?title¼Railway_ freight_transport_statistics
16. Profillidis, V. (2016). Railway management and engineering. Routledge.
17. Stewart, G. (1995). Supply chain performance benchmarking study reveals keys to supplychain excellence. Logistics information management, 8(2), 38-44.

18. Lai, K. H., Ngai, E. W., & Cheng, T. C. E. (2002). Measures for evaluating supply chain performance in transport logistics. Transportation Research Part E: Logistics and Transportation Review, 38(6), 439-456.
19. Verma, A., Harsha, V., & Subramanian, G. H. (2021). Evolution of urban transportation policies in India: a review and analysis. Transportation in Developing Economies, 7, 1-15.

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-024-34612-z

RESEARCH ARTICLE

Sustainable waste management: a comprehensive life cycle assessment of bioethanol production from agricultural and municipal waste

Kulvendra Patel[1] · **Santosh Kumar Singh**[1]

Received: 24 January 2024 / Accepted: 31 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Biofuels have emerged as a promising and eco-friendly alternative to conventional fossil fuels. Biofuel sourced from rice straw (RS) and municipal solid waste (MSW), which are abundant residues from agricultural and municipal activities, present a sustainable solution to address waste management challenges. Utilizing life cycle assessment, this study quantifies the environmental advantages by assessing the reduction in greenhouse gas emissions, energy consumption, and other environmental impacts linked with employing these waste materials for biofuel production. Employing a cradle-to-gate approach as the system boundary for bioethanol production, with the functional unit set as per liter of bioethanol produced, the analysis reveals that the global warming potential (GWP) for ethanol from MSW is 4.4 kg CO_2 eq., whereas for RS, it is 2.1 kg CO_2 eq. per functional unit. The total environmental impacts were primarily due to enzymatic hydrolysis and electricity consumption for ethanol production from MSW and RS. Despite advancements, fossil fuel consumption remains a potential energy source for biofuel production. The cumulative energy demand stands at 18.6 MJ for RS and 71.5 MJ for MSW per functional unit, underscoring the potential to significantly reduce overall impacts by transitioning to a more environmentally sustainable energy source. The uncertainty analysis acknowledges the inherent uncertainties associated with data, assumptions, and methodologies, highlighting the crucial need for ongoing research and updates to enhance the accuracy of future assessments. This analysis forms the foundation for well-informed decision-making, providing valuable insights for policymakers, industry stakeholders, and consumers.

Keywords Biofuel · Cumulative energy demand · Environmental impact assessment · Greenhouse gas reduction · Municipal solid waste · Sustainable energy · Uncertainty analysis

Introduction

Over the recent decades, a noticeable shift towards biofuel production has occurred. This shift has been primarily motivated by the urgent need to curb carbon emissions, address climate change, and mitigate the depletion of fossil fuel reserves. The development of biofuels as a feasible substitute for traditional fossil fuels led to substantial debates over how well they can achieve environmental sustainability goals (Fargione et al. 2008; Searchinger et al. 2008). The evolution of biofuels can be categorized into three distinct generations. First-generation biofuels are primarily derived from food crops through fermentation (UN Report 2007). However, this approach faced significant criticism, particularly regarding using food crops for fuel production when food shortages were prevalent, leading to societal discontent (Ahlgren and Di-Lucia 2014). In response to these challenges, second-generation biofuels were developed, aiming to overcome the limitations of the first generation. These biofuels were produced from non-food organic sources such as crop residues, wood, municipal solid waste (MSW), and other biomass materials, thus mitigating the primary drawback (Liska et al. 2009). The third generation of biofuels explores novel biomass sources for biofuel synthesis, including unique energy crops or species like algae (Chisti 2007). Among renewable energy sources, bioenergy, mainly derived from lignocellulosic biomass such as wheat straw, corn stover, rice straw, and MSW, has garnered significant attention for its potential to reduce global greenhouse gas (GHG) emissions (Jiang

Responsible Editor: Ta Yeong Wu

✉ Kulvendra Patel
kulvendrapatel_2k21phden501@dtu.ac.in

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

Published online: 07 August 2024

et al. 2020; Liu et al. 2022). These renewable resources hold promise as eco-friendly alternatives to environmentally harmful fossil fuels, potentially leading to the development of sustainable products (Kaparaju et al. 2009; Deshmukh and Pathan 2023).

Bioethanol, a renewable and sustainable alternative to conventional fossil fuels, has emerged as a key player in the global pursuit of clean and green energy. The utilization of diverse feedstocks for bioethanol production is a crucial aspect of expanding its viability and mitigating environmental concerns. Two such promising feedstocks are municipal solid waste and rice straw, each offering distinctive advantages in the quest for sustainable bioethanol production (Kowalski et al. 2022; Subudhi et al. 2023). As urbanization continues to rise, so does the amount of waste generated per person. In 2018, the total MSW generated per person per day reached 292.4 million tons (US EPA 2022). Since the 1970s, researchers have been exploring ways to convert MSW into ethanol, with particular interest in enzymatic conversion and sulfuric acid hydrolysis processes (Cardona and Sánchez 2007; Curtiss and Kreider 2010). However, beyond the technical challenges of biomass-to-ethanol conversion, there needs to be more discussion regarding the environmental and economic impacts of bioethanol from MSW. These impacts encompass the environmental consequences of using MSW-derived ethanol as a vehicle fuel and the environmental effects of using ethanol as a means of MSW disposal (Oldfield et al. 2016; Braguglia et al. 2018). Bioethanol from MSW provides a sustainable solution to waste management challenges and contributes to reducing GHG emissions, aligning with the principles of a circular and low-carbon economy. The gravity pressure vessel (GPV) developed by GeneSyst was employed to model the conversion of MSW into ethanol (Kalogo et al. 2007; GeneSyst International 2012). Designed explicitly for MSW, the GPV utilizes a lengthy subterranean pipe to circulate a mixture of MSW and water, taking advantage of high-pressure and high-temperature conditions at the pipe's base to facilitate the dissolution of nearly any organic chemicals present in the MSW (Kalogo et al. 2007).

On the agricultural front, rice straw, an often-underutilized byproduct of rice cultivation, holds immense potential as a feedstock for bioethanol production. Historically regarded as a crop residue, rice straw is rich in cellulose and hemicellulose, making it an excellent feedstock for bioconversion (Nizzy and Kannan 2022; Dey et al. 2022; Shakelly et al. 2023). Using rice straw for bioethanol not only tackles the problem of post-harvest agricultural waste management but it also adds value to a resource that would otherwise be left unused. Using agricultural waste in its current state is unsustainable and can lead to environmentally harmful practices, such as the burning of rice straw, resulting in unsafe levels of air pollution (Alengebawy et al. 2022; Bai et al. 2022). It is widely recognized that promoting more productive applications of agricultural residues can enhance economic well-being (Sukumaran et al. 2010). While ethanol produced from agricultural waste holds promise for sustainability, its viability requires comprehensive and systematic research (Soam et al. 2016, 2017; Das et al. 2023).

In recent years, numerous life cycle assessments (LCA) have been conducted to evaluate the energy benefits and potential global warming mitigation associated with ethanol and bioenergy production (Yazan et al. 2017; Gheewala et al. 2019; Morales et al. 2019; Patel and Singh 2023a). Various scholars and ecologists have used the LCA technique to gain deeper insights into biofuel advantages and potential drawbacks by meticulously examining biofuel production. This approach considers each phase, including crop cultivation, transportation, chemical and biological processes, and GHG emissions, and ultimately establishes a connection between environmental impacts and biofuel produced.

This study explores the promising realm of bioethanol production from MSW and rice straw. The novelty of the work lies in the comparative LCA of two distinct and underutilized waste streams. This study effectively integrates these waste materials into biofuel production, demonstrating their potential for solving energy and waste management challenges. This work contributes to the literature by offering new insights into the comparative benefits and challenges of utilizing agricultural and municipal waste for biofuel production. A study by Shakelly et al. (2023) provides a method for producing bioethanol from MSW using organic residue. It emphasizes the method's sustainability, efficiency in resource recovery, and alignment with current industrial practices, highlighting its environmental and economic benefits in waste reduction and renewable energy production. Another study by Pradhan et al. (2024) evaluates the environmental impacts of a novel approach to food waste management systems by producing biochar. This study describes the benefits of employing food waste biochar as an amendment in agriculture, which will encourage the researchers to pick appropriate food waste types. A comparative LCA was performed by Alengebawy et al. (2022) to evaluate bioenergy production from rice straw. The results indicate the syngas scenario was the most environmentally friendly, achieving the highest reduction in marine ecotoxicity. The dual benefits of waste valorization and sustainable agriculture underscore the importance of investigating and optimizing the processes involved in extracting bioethanol from these unconventional yet impactful sources. Additionally, the study highlights the influence of decisions related to system boundaries and functional units on the outcomes of the LCA, ultimately contributing to a better understanding of the sustainability implications of bioethanol production from these sources.

Materials and methods

LCA is a methodology used to evaluate the potential environmental impacts throughout a product or process's life cycle, from the acquisition of raw material to use and disposal or recycling (Patel and Singh 2022a; Kumar et al. 2024). The LCA framework is provided by the ISO 14000 series, where ISO 14040 provides the principle and structure (ISO 2006a), and ISO 14044 provides the requirements and criteria for conducting LCA (ISO 2006b). Life cycle assessment consists of four phases: defining goal and scope, life cycle inventory analysis (LCI), life cycle impact assessment (LCIA), and interpretation (Davis et al. 2009).

Goal and scope, system boundary, and functional unit

This work analyzes and compares the production of bioethanol from MSW and RS using a cradle-to-gate system boundary, as shown in Fig. 1. For MSW bioethanol, the system boundary of the study assumes the availability of MSW as a feedstock and does not consider energy use and environmental implications of the processes and products that generated the MSW. MSW-fluff (organic residual remaining after conventional extraction of commodities that can be recycled, e.g., aluminum, glass, steel, and plastics, from raw MSW), as initially referenced by (Broder and Henson 1993), is considered as the primary input (Shakelly et al. 2023). The primary reason for selecting the production of ethanol from MSW is due to its low cost, consistent supply, and existing collection infrastructure. This offers the advantage of reducing the environmental consequences associated with waste disposal by displacing fossil-based fuels (Patel and Singh 2022b; Das et al. 2023). The transformation of MSW-fluff into ethanol utilizes a GPV. The key stages in the conversion process include shredding and separating MSW-fluff, subjecting it to continuous dilute acid hydrolysis under high-temperature and high-pressure conditions within a GPV to convert the cellulose fraction into sugars, fermenting the sugars into ethanol, and distilling to obtain ethanol of fuel quality (Kalogo et al. 2007; GeneSyst International 2012).

For rice straw biofuel, the product system is outlined on a cradle-to-gate basis (Fig. 1), commencing with the production of biomass feedstock and ending at the refinery gate. The study does not incorporate the utilization phase of ethanol. As the aim of this study is to assess the impact of biofuel production processes, the study does not take into account the environmental impacts associated with the plants' infrastructure, human labor, and capital expenditures (Soam et al. 2017; Samer et al. 2021); also, the stages preceding collection (including cultivation

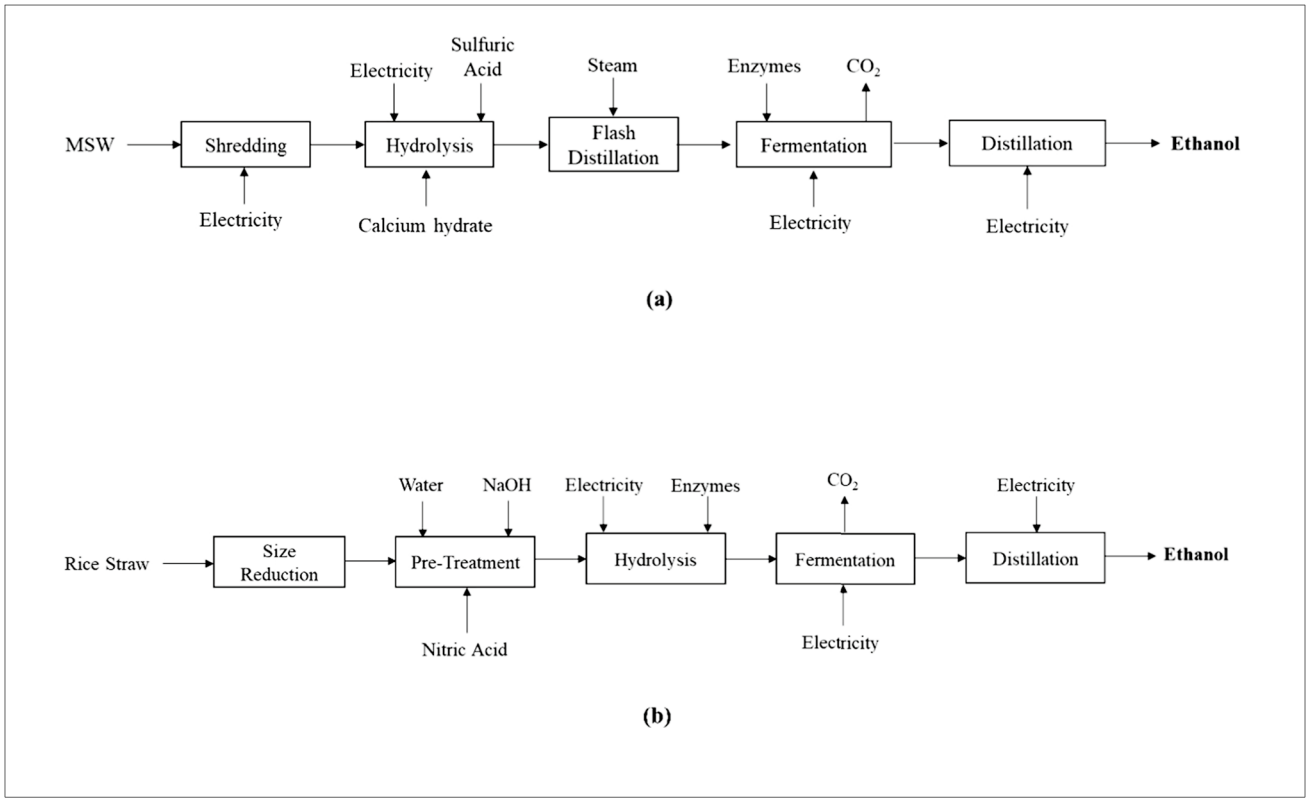


Fig. 1 System boundary of bioethanol production from (**a**) MSW and (**b**) RS

and harvesting) and the land-use changes are beyond the scope of this study (Rebitzer et al. 2004). Table 1 displays the weight-based composition of rice straw. The utilization of rice straw for biofuel production is an attractive option due to the abundant availability of agricultural residue. According to the Ministry of New and Renewable Energy, approximately 500 million metric tons of crop residues are generated annually (FAO 2022), providing a substantial feedstock for biofuel production. This abundance of agricultural residue highlights the potential for sustainable utilization in the production of biofuels, further enhancing its appeal as a renewable energy source.

Moreover, while these options are appealing in terms of both environmental benefits and economic considerations, they have not received extensive research attention as an alternative strategy for managing MSW and rice straw. To enable accurate comparisons between various products or services, a functional unit (FU) is employed as a quantitative measure, usually in terms of weight, volume, or energy content. The FU used in this study is per liter of bioethanol produced. This FU allows for a direct comparison of the environmental impacts associated with producing a standardized volume of bioethanol across both MSW and RS-based processes. It ensures that the comparison is based on equivalent quantities of bioethanol, providing a clear and accurate assessment of environmental performance.

Life cycle inventory

This LCA phase includes the compilation and quantification of input and output flows for a product system (ISO 2006b). These include "emissions to air, land, and water as well as consumption of water, energy, and raw materials." The inventory may be based on a simulation of the process or an analysis of the literature. In this study, the background inventory analysis is based on the Ecoinvent v3 database (Wernet et al. 2016), and foreground data are obtained from several studies from literature on biofuel (Soam et al. 2017; Gunarathne et al. 2019; Sreekumar et al. 2020; Shakelly et al. 2023). The background data utilized in this study is based on a specific geographical area. However, the scope of this study has been adjusted to a global scale (GLO). This ensures that the results are not dependent on a particular location, i.e., the study aims to assess the worldwide impact of bioethanol derived from various feedstocks. Inventory data for ethanol production using MSW and RS is presented in Table 2.

Life cycle impact assessment

This LCA phase is for evaluating the environmental impacts by converting the LCI outcomes into specific midpoint or endpoint impact indicators (Patel and Singh 2023b; Gheewala 2023). The IMPACT World + impact assessment method is used for analysis in the study, and it comprises Eco-Indicator, IPCC, Impact 2002 + , and CML (Patel and Singh 2022b; PRé Sustainability 2022). The endpoint impact categories considered were human health, ecosystem quality, global warming potential, and resources. The cumulative energy demand (CED) method is used to assess the total energy consumption associated with biofuel production. It aims to quantify the cumulative energy inputs required at various stages, including raw material extraction, manufacturing, use, and end-of-life disposal.

Table 1 The composition of rice straw and municipal solid waste

Component	Percent weight composition	Reference
Rice straw		
Cellulose	36.4	(Belal 2013; Van Hung et al. 2020)
Xylose	18.2	(Belal 2013)
Arabinose	1.7	(Belal 2013)
Lignin	14.5	(Sreekumar et al. 2020; Rosado et al. 2022)
Protein	1.9	(Sreekumar et al. 2020)
Ash	0.8	(Van Hung et al. 2020; Sreekumar et al. 2020)
Silica	11.6	(Sreekumar et al. 2020)
Others	4.8	(Sreekumar et al. 2020)
Moisture	10.1	(Belal 2013; Sreekumar et al. 2020)
MSW		
Cellulose	58.51	(Mahmoodi et al. 2018)
Hemicellulose	11.92	(Mahmoodi et al. 2018; Verhe et al. 2022)
Lignin	15.3	(Verhe et al. 2022)
Others	14.28	(Verhe et al. 2022)

Table 2 Inventory data for MSW and rice straw–based bioethanol production per functional unit

Material	Unit	MSW	Rice straw
Input			
Electricity	MJ	6.26	4.70
Water	kg	-	2
Nitric acid	g	-	120
Diesel	g	78	-
Calcium hydrate	g	600	-
Sodium hydroxide	g	-	80
Sulfuric acid	g	400	-
Enzymes	g	620	10
Natural gas	m^3	0.33	-
Steam	kg	1.72	-
Output			
Ethanol	L	1	1
Ethyl cellulose	g	-	11
CO_2	g	344.4	800
Silicon dioxide	g	-	400

Results and discussion

This paper compares bioethanol production from municipal solid waste and rice straw. The analysis shows that enzymatic hydrolysis and electricity consumption contribute to most midpoint category impacts, such as acidification, eutrophication, ecotoxicity, and GWP. The damage categories associated with biofuel production per functional unit using the IMPACT World + impact assessment method are as follows.

Human health

Toxicity, climate change (short term), ozone layer depletion, photochemical oxidant formation, particulate matter, ionizing radiation (impacting human health), and water availability midpoint categories contribute to the human health damage category and are measured in DALY ("disability-adjusted life years") per kg emission. Climate change, short term, refers to the shorter-term impacts of climate change occurring within the first 100 years after the emission of greenhouse gases. This category focuses on the immediate effects of greenhouse gas emissions on the rate of temperature change and other climate-related impacts over a relatively short time horizon. The short-term climate change for RS and MSW are 2.19 kg CO_2 eq. and 5.02 kg CO_2 eq., respectively. This is primarily due to electricity consumption for RS (42%) and enzymatic hydrolysis for MSW (38%).

The human health damage category is highest for MSW (7.1E-04 DALY/FU) compared to RS (4.7E-4 DALY/FU) bioethanol, as shown in Table 3. For MSW, the bioethanol production per FU causes 1.77E-7 CTUh (Comparative toxicity unit for human health) of carcinogen toxicity, 5.31E-7 CTUh of non-carcinogen toxicity, 3.47E-7 kg CFC-11 eq. of ozone layer depletion, 1.9E-3 kg $PM_{2.5}$ eq. of particulate matter formation, 30.8 Bq C-14 eq. of ionizing radiation, and 6.8E-4 DALY of water availability (human health). The conditions to decide which substance is taken as a midpoint reference substance are based on (i) a substance with proven effects, (ii) a generally accepted substance, and (iii) a substance with low uncertainties in exposure and effect modeling. The high human health damage category for MSW is mainly because of the use of electricity (77.2%), followed by the use of enzymes (21.2%). In the case of RS, electricity consumption and chemical use in pre-treatment causes damage to human health, 86.8% and 8.8%, respectively, which is in line with the study by Kalogo et al. (2007) and Bozorgirad et al. (2013). Figure 2 shows the contribution to total impacts associated with the production of bioethanol from RS and MSW.

Ecosystem quality

The ecosystem quality endpoint category consists of acidification, eutrophication, ecotoxicity, climate change (both short and long term), ionizing radiation (impacting ecosystem quality), water scarcity (affecting water availability in terrestrial and freshwater ecosystems), and land biodiversity (considering transformation and occupation). It is measured in PDF*m^2*yr ("potentially disappeared fraction of species in a certain area over a certain period"). The long-term impacts are those that occur beyond 100 years after emission and can extend up to infinity.

The ecosystem damage is highest for MSW bioethanol (34.1 PDF*m^2*yr) compared to RS bioethanol (10.7 PDF*m^2*yr) per FU. The freshwater ecotoxicity potential of MSW bioethanol is highest (51%), primarily due to the enzymatic hydrolysis (46%) converting cellulose to glucose and electricity use (40%), followed by impacts from land transformation, biodiversity loss (30.5%), and long-term climate change (7%). For RS-based bioethanol, long-term freshwater ecotoxicity is also high (79.5%), followed by the long-term climate change category (11.5%), primarily due to electricity consumption (61% and 42%, respectively), as shown in Table 3. These results are consistent with Kalogo et al. (2007) and Liu et al. (2020). Figure 3a and b show the midpoint impact categories of RS and MSW-based bioethanol production. The water scarcity impact for RS and MSW is 0.76 and 2.04 m^3 world eq., shown in Table 4, with enzyme production contributing 44% and 69% of the impact, respectively. In terms of damage level indicator, as shown in Table 3, for water availability impacts, the terrestrial ecosystem contributes 94% for RS bioethanol and 95% for MSW bioethanol, followed by the freshwater ecosystem, which

Table 3 Endpoint damage categories of bioethanol production from MSW and RS

Impact category	Unit	RS	MSW
(A) Damage on area of protection			
Human health	DALY	4.7E-4	7.1E-4
Ecosystem quality	PDF.m^2.yr	10.7	34.1
(B) Damage level indicators			
Climate change, human health, short term	DALY	1.8E-06	4.1E-06
Climate change, human health, long term	DALY	5.6E-06	1.1E-05
Photochemical oxidant formation	DALY	1.3E-10	4.0E-10
Ionizing radiation, human health	DALY	3.6E-09	6.5E-09
Ozone layer depletion	DALY	3.1E-10	5.3E-10
Human toxicity cancer, short term	DALY	1.3E-07	3.4E-07
Human toxicity cancer, long term	DALY	6.4E-09	1.4E-08
Human toxicity non-cancer, short term	DALY	2.1E-07	5.2E-07
Human toxicity non-cancer, long term	DALY	1.8E-07	4.5E-07
Particulate matter formation	DALY	8.0E-07	2.4E-06
Water availability, human health	DALY	4.6E-04	6.9E-04
Climate change, ecosystem quality, short	PDF.m^2.yr	0.387	0.888
Climate change, ecosystem quality, long	PDF.m^2.yr	1.232	2.511
Marine acidification, short term	PDF.m^2.yr	0.031	0.064
Marine acidification, long term	PDF.m^2.yr	0.290	0.592
Freshwater ecotoxicity, short term	PDF.m^2.yr	0.034	0.075
Freshwater ecotoxicity, long term	PDF.m^2.yr	8.522	17.492
Freshwater acidification	PDF.m^2.yr	0.011	0.039
Terrestrial acidification	PDF.m^2.yr	0.077	0.295
Freshwater eutrophication	PDF.m^2.yr	3.5E-04	3.4E-03
Marine eutrophication	PDF.m^2.yr	0.002	0.007
Ionizing radiation, ecosystem quality	PDF.m^2.yr	2.8E-10	6.3E-10
Land transformation, biodiversity	PDF.m^2.yr	0.065	10.419
Land occupation, biodiversity	PDF.m^2.yr	0.065	1.675
Water availability, freshwater ecosys	PDF.m^2.yr	8.2E-06	1.8E-05
Water availability, terrestrial ecosys	PDF.m^2.yr	1.9E-04	5.0E-04
Thermally polluted water	PDF.m^2.yr	3.8E-06	6.6E-06

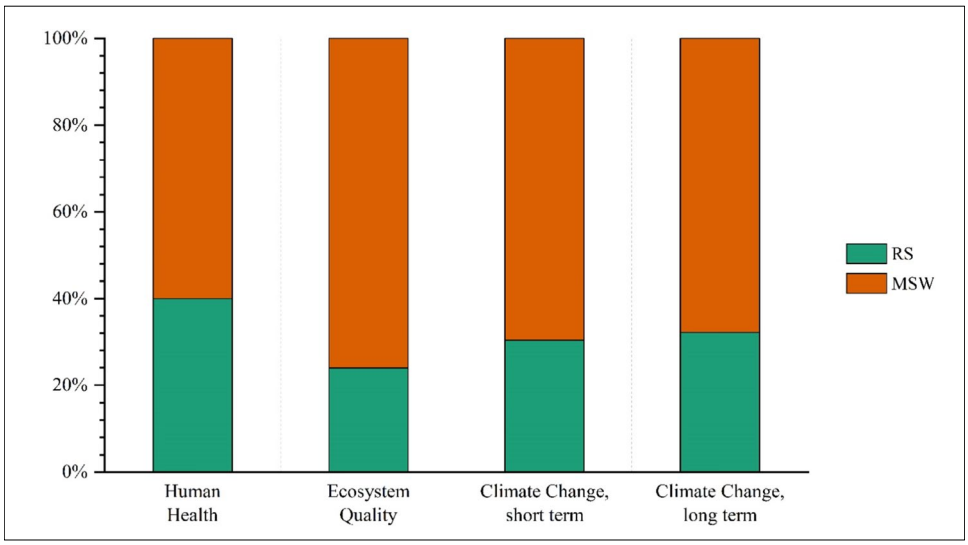


Fig. 2 Contribution of each damage category for bioethanol production

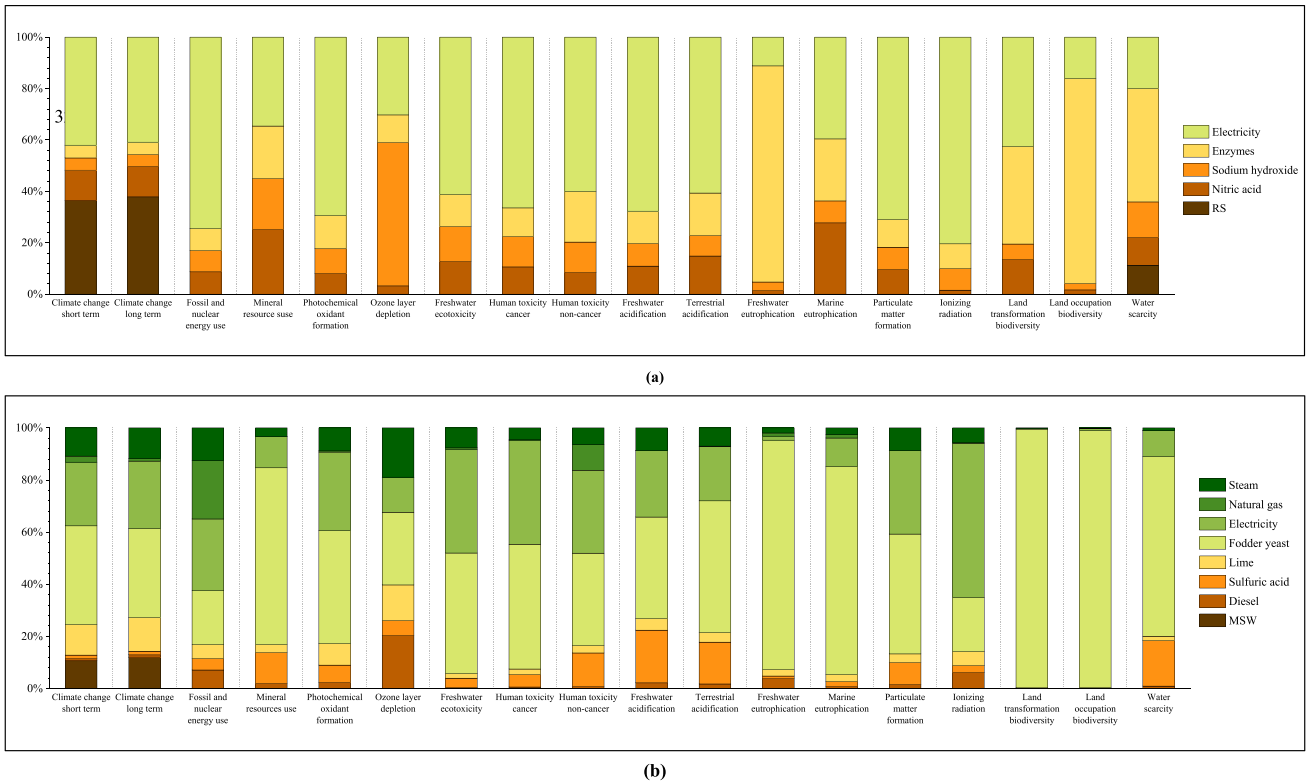


Fig. 3 Midpoint impact categories of (**a**) rice straw and (**b**) MSW-based bioethanol

accounts for 4% and 3.4%, respectively. Freshwater and marine eutrophication of MSW bioethanol are high (shown in Table 4) due to enzymatic hydrolysis, which accounts for 88% and 79% of the impacts, respectively, followed by electricity consumption, which accounts for 2% and 11%, respectively.

The results of the study are compared with the incineration method of waste management, shown in Table 5, as it is

Table 4 Midpoint impacts of Bioethnol protection (per FU)

Impact category	Unit	RS	MSW
Climate change, short term	kg CO_2 eq	2.189	5.020
Climate change, long term	kg CO_2 eq	2.112	4.437
Fossil and nuclear energy use	MJ deprived	17.257	62.286
Mineral resources use	kg deprived	0.011	0.042
Photochemical oxidant formation	kg NMVOC eq	0.003	0.010
Ozone layer depletion	kg CFC-11 eq	1.2E-07	3.5E-07
Freshwater ecotoxicity	CTUe	14849.313	30512.809
Human toxicity cancer	CTUh	7.9E-08	1.8E-07
Human toxicity non-cancer	CTUh	2.1E-07	5.3E-07
Freshwater acidification	kg SO_2 eq	1.3E-08	4.6E-08
Terrestrial acidification	kg SO_2 eq	1.2E-05	4.5E-05
Freshwater eutrophication	kg PO_4 eq	3.1E-05	3.0E-04
Marine eutrophication	kg N eq	1.2E-04	6.0E-04
Particulate matter formation	kg $PM_{2.5}$ eq	6.7E-04	2.0E-03
Ionizing radiation	Bq C-14 eq	16.982	30.780
Land transformation, biodiversity	m^2yr arable	2.6E-04	4.1E-02
Land occupation, biodiversity	m^2yr arable	0.092	2.385
Water scarcity	m^3 world eq	0.764	2.042

Table 5 Comparison of bioethanol production system with incineration

Impact category	Unit	Bioethanol production	Incineration
Abiotic depletion	kg Sb eq	5.3E-04	4.7E-04
Global warming	kg CO_2 eq	84.494	519.129
Ozone layer depletion	kg CFC-11 eq	2.8E-06	3.1E-06
Human toxicity	kg 1,4-DB eq	11.452	6532.022
Acidification	kg SO_2 eq	0.195	0.242
Eutrophication	kg PO_4^{3-} eq	0.026	0.309

Table 6 Single score result of CED of bioethanol production from MSW and RS

Impact category	Unit	MSW	RS
Total	MJ	71.5	18.6
Non-renewable, fossil	MJ	52.7	13.5
Non-renewable, nuclear	MJ	3.89	2.62
Non-renewable, biomass	MJ	0.43	1.39E-03
Renewable, biomass	MJ	12.1	1.13
Renewable, wind, solar, geothermal	MJ	0.53	0.35
Renewable, water	MJ	1.8	1.09

the most commonly adopted method for waste management. This result is in line with the study by Bozorgirad et al. (2013), Elginoz et al. (2020), and Patel and Singh (2022b). This comparison shows that the environmental impacts of bioethanol production are much lower than those of incineration, with a huge gap in between. However, the energy recovery system is not considered for incineration.

Resources

Mineral resources and fossil and nuclear energy use contribute to the resource endpoint category. The mineral resource impact category assesses resource depletion based on the mass of mineral resources deprived per mass dissipated, whereas the fossil and nuclear energy use evaluates the energy deprived per mass of resource dissipated. The biofuel production from MSW has the highest fossil and nuclear energy use (62.3 MJ deprived) than RS (17.3 MJ deprived). Electricity consumption contributes to 28% and 75% of the resource damage category for MSW and RS, respectively. Damages from mineral resource use are 0.01 kg deprived / FU for RS bioethanol and 0.04 kg deprived / FU for MSW bioethanol. Figure 4 shows the single score categories of biofuel production from MSW and RS. The high damage from mineral resource use in MSW is mainly due to enzyme production (68%), followed by electricity (12%). For RS, mineral resource damage is primarily caused by electricity consumption (35%) and the use of nitric acid in pre-treatment (25%).

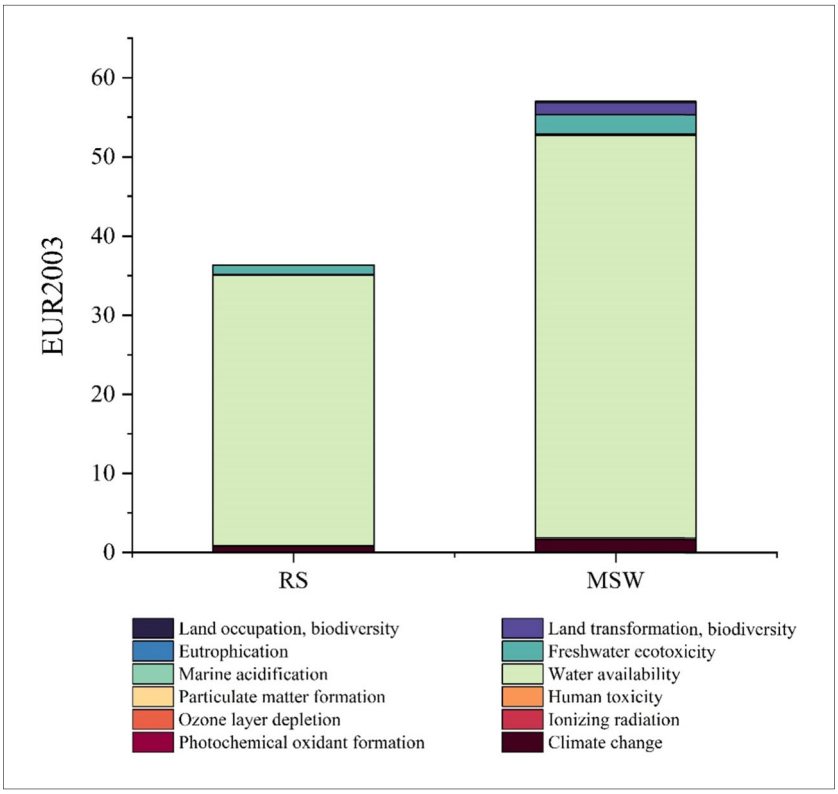
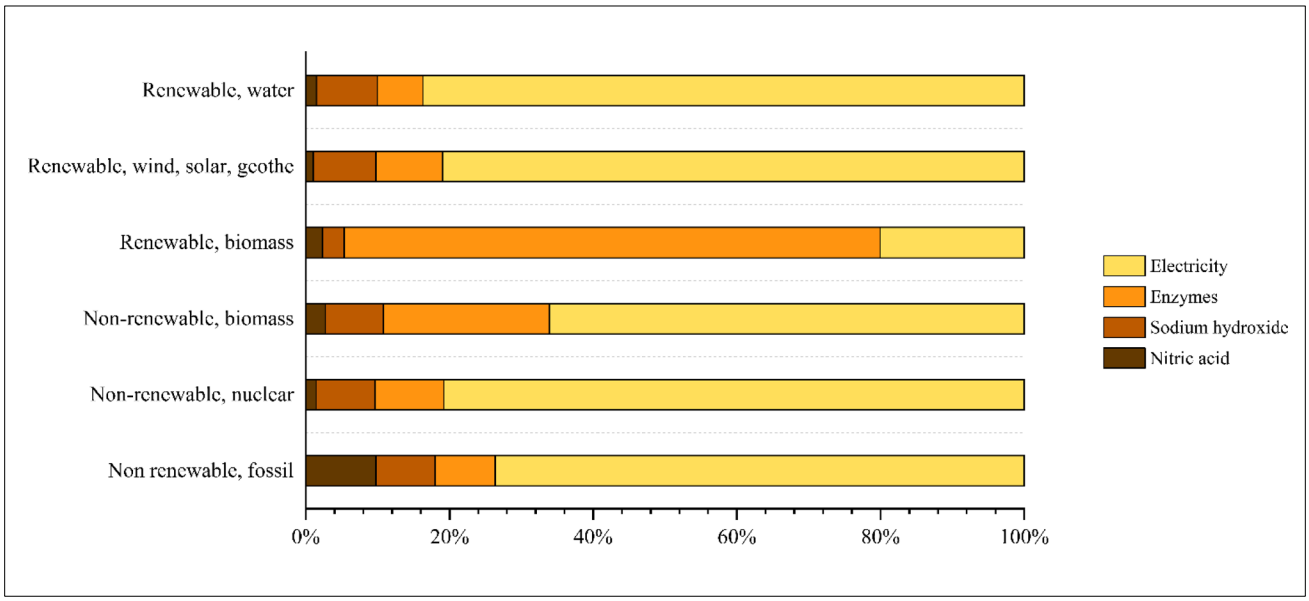


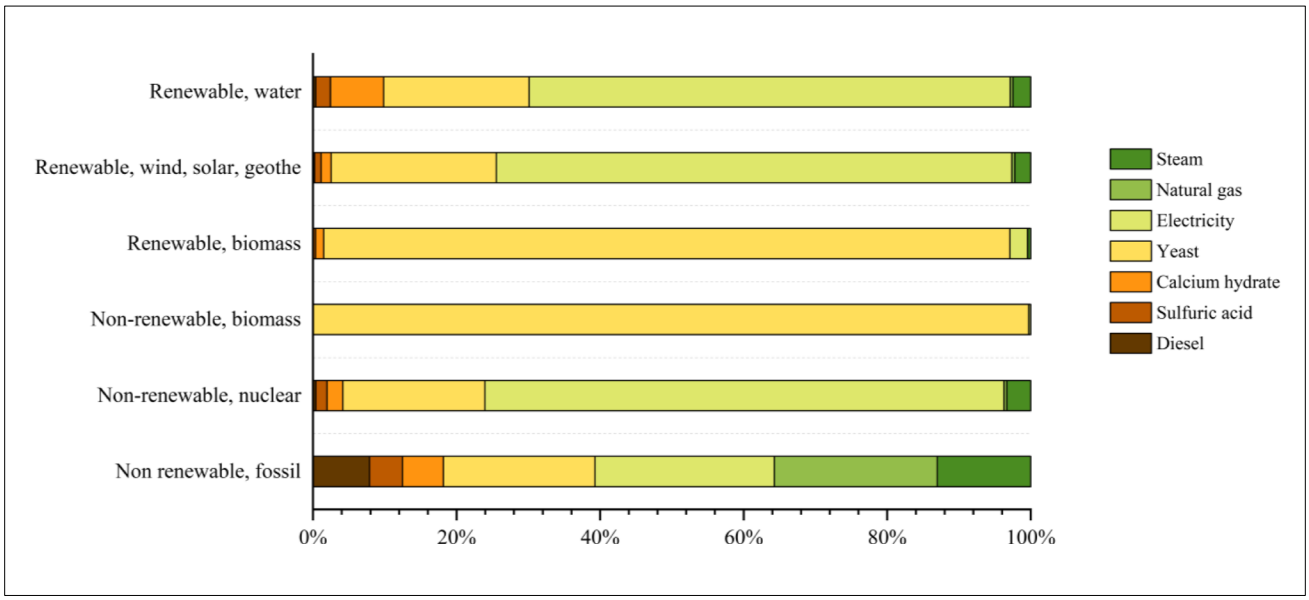
Fig. 4 Midpoint single score for biofuel production

Cumulative energy demand

The CED method provides a comprehensive view of the energy requirements and helps identify energy-intensive stages in a product's life cycle. Quantifying the total energy demand, it allows for the comparison of different products or processes in terms of their energy efficiency and potential environmental impact. Fossil fuel consumption is still a potential source of energy for the production of biofuel. Table 6 shows the single score results obtained from CED analysis. MSW shows high fossil fuel consumption (~ 74%), followed by other energy sources such as renewable biomass, nuclear, and wind, solar, and geothermal energy, 17%, 5.4%, and 0.7%, respectively, shown in Fig. 5b. For RS-based bioethanol production, the fossil fuel consumption was 72%, followed by 14%, 6%, and 2% of nuclear energy, renewable biomass, and wind, solar and geothermal energy, respectively, shown in Fig. 5a.



(a)



(b)

Fig. 5 Contributors to the energy demand for production of (**a**) RS and (**b**) MSW-based bioethanol

Table 7 Uncertainty analysis for the production of RS and MSW-based bioethanol (using Monte Carlo analysis; 1000 iterations)

Environmental impact categories	Rice straw					MSW				
	Mean	SD	CoV (%)	SEM	Skewness	Mean	SD	CoV (%)	SEM	Skewness
IMPACT World +										
Climate change, long term	2.392	0.115	4.8	3.6E-03	0.080	4.006	0.169	4.2	5.4E-03	0.183
Climate change, short term	2.423	0.116	4.8	3.7E-03	0.063	4.529	0.191	4.2	6.0E-03	0.093
Fossil and nuclear energy use	15.228	1.242	8.2	3.9E-02	0.052	20.208	1.664	8.2	5.3E-02	0.156
Freshwater acidification	7.3E-14	5.0E-15	6.9	1.6E-16	0.044	1.8E-13	8.1E-15	4.5	2.6E-16	0.114
Freshwater ecotoxicity	5818.277	240.241	4.1	7.6E + 00	-0.028	16772.878	974.555	5.8	3.1E + 01	0.160
Freshwater eutrophication	1.2E-05	9.8E-07	8.1	3.1E-08	0.053	2.9E-05	1.6E-06	5.6	5.0E-08	0.123
Human toxicity cancer	3.0E-08	1.2E-09	4.1	3.9E-11	-0.045	1.1E-07	6.7E-09	6.2	2.1E-10	0.115
Human toxicity non-cancer	8.7E-08	3.9E-09	4.4	1.2E-10	0.065	2.7E-07	1.2E-08	4.4	3.7E-10	0.222
Ionizing radiations	9.742	0.552	5.7	1.7E-02	0.107	21.116	0.888	4.2	2.8E-02	0.230
Land occupation, biodiversity	2.3E-03	9.7E-05	4.3	3.1E-06	0.081	0.016	0.001	6.8	3.5E-05	0.158
Land transformation, biodiversity	2.5E-04	1.8E-05	7.2	5.7E-07	0.070	6.8E-04	3.0E-05	4.4	9.4E-07	0.211
Marine eutrophication	1.5E-04	9.1E-06	6.2	2.9E-07	0.038	5.0E-04	2.9E-05	5.7	9.1E-07	0.201
Mineral resources use	4.4E-03	2.0E-04	4.4	6.2E-06	-0.033	1.7E-02	1.1E-03	6.7	3.6E-05	0.069
Ozone layer depletion	2.6E-07	1.9E-08	7.6	6.2E-10	0.082	6.1E-07	2.8E-08	4.6	8.8E-10	0.257
Particulate matter formation	8.5E-04	5.4E-05	6.4	1.7E-06	0.022	2.2E-03	1.0E-04	4.7	3.3E-06	0.103
Photochemical oxidant formation	5.6E-03	3.9E-04	6.9	1.2E-05	0.074	1.2E-02	5.9E-04	4.9	1.9E-05	0.171
Terrestrial acidification	4.8E-08	3.1E-09	6.5	9.9E-11	0.015	1.3E-07	5.9E-09	4.6	1.9E-10	0.106
Water scarcity	56.379	3.063	5.4	9.7E-02	0.081	214.056	10.986	5.1	3.5E-01	0.147
Cumulative energy demand										
Non-renewable, fossil	14.2403	1.1595	8.1	3.7E-02	-0.0420	18.9938	1.5135	8.0	4.8E-02	0.1194
Non-renewable, biomass	5.16E-04	2.80E-05	5.4	8.9E-07	0.0075	0.4262	0.0332	7.8	1.1E-03	0.1278
Non-renewable, nuclear	0.0638	0.0052	8.1	1.6E-04	-0.0420	0.0850	0.0068	8.0	2.1E-04	0.1194
Renewable, biomass	0.0135	0.0011	8.1	3.5E-05	-0.0420	0.0180	0.0014	8.0	4.5E-05	0.1194
Renewable, water	0.7847	0.0502	6.4	1.6E-03	-0.0251	1.4065	0.0714	5.1	2.3E-03	0.0154
Renewable, wind, solar, geothermal	0.0715	0.0037	5.2	1.2E-04	-0.0493	0.1547	0.0096	6.2	3.0E-04	0.0721

SD standard deviation, *CoV* coefficient of variation, *SEM* standard error of the mean

Uncertainty analysis

Since the secondary data were utilized in the analysis, a Monte Carlo analysis was conducted to assess and quantify the uncertainties associated with each unit process. In a standard approach, the uncertainty in each impact category is evaluated by calculating the ratio of the standard deviation to the mean (Sharma et al. 2020; Bhonsle et al. 2022). According to the IPCC, a resultant value greater than 0.3 is regarded as substantial uncertainty. The primary sources of uncertainty include variability in input parameters, data variability, data gaps, assumptions, and subjective interpretation processes like impact assessment and normalization (Sheikholeslami et al. 2022). These uncertainties can significantly affect LCA outcomes, resulting in different results for the same assessment. This study investigated data variability as the source of LCA uncertainty, using a combination of Monte Carlo simulation and data quality index method with Ecoinvent v3 for uncertainty analysis. Data variability is calculated using the pedigree matrix (Teng and Pan 2020), which scores data quality across five types of criteria on a scale from 1 to 5, with score 1 indicating the highest quality of data and score 5 indicating the lowest (Ciroth et al. 2012). To calculate the variance of total uncertainty, the following formula is used.

$$\sigma_t^2 = exp\sqrt{\ln(U_1)^2+\ln(U_2)^2+\ln(U_3)^2+\ln(U_4)^2+\ln(U_5)^2+\ln(U_b)^2} \quad (1)$$

where

σ_t^2 = variance for total uncertainty
U_1 = uncertainty factor of reliability
U_2 = uncertainty factor of completeness
U_3 = uncertainty factor of temporal correlation
U_4 = uncertainty factor of geographic correlation
U_5 = uncertainty factor of other technological correlation
U_b = basic uncertainty factor

Table 7 shows the uncertainty values for each impact category. The results of the Monte Carlo analysis for midpoint impact categories for RS and MSW are shown in Fig. 6 and Fig. 7. The

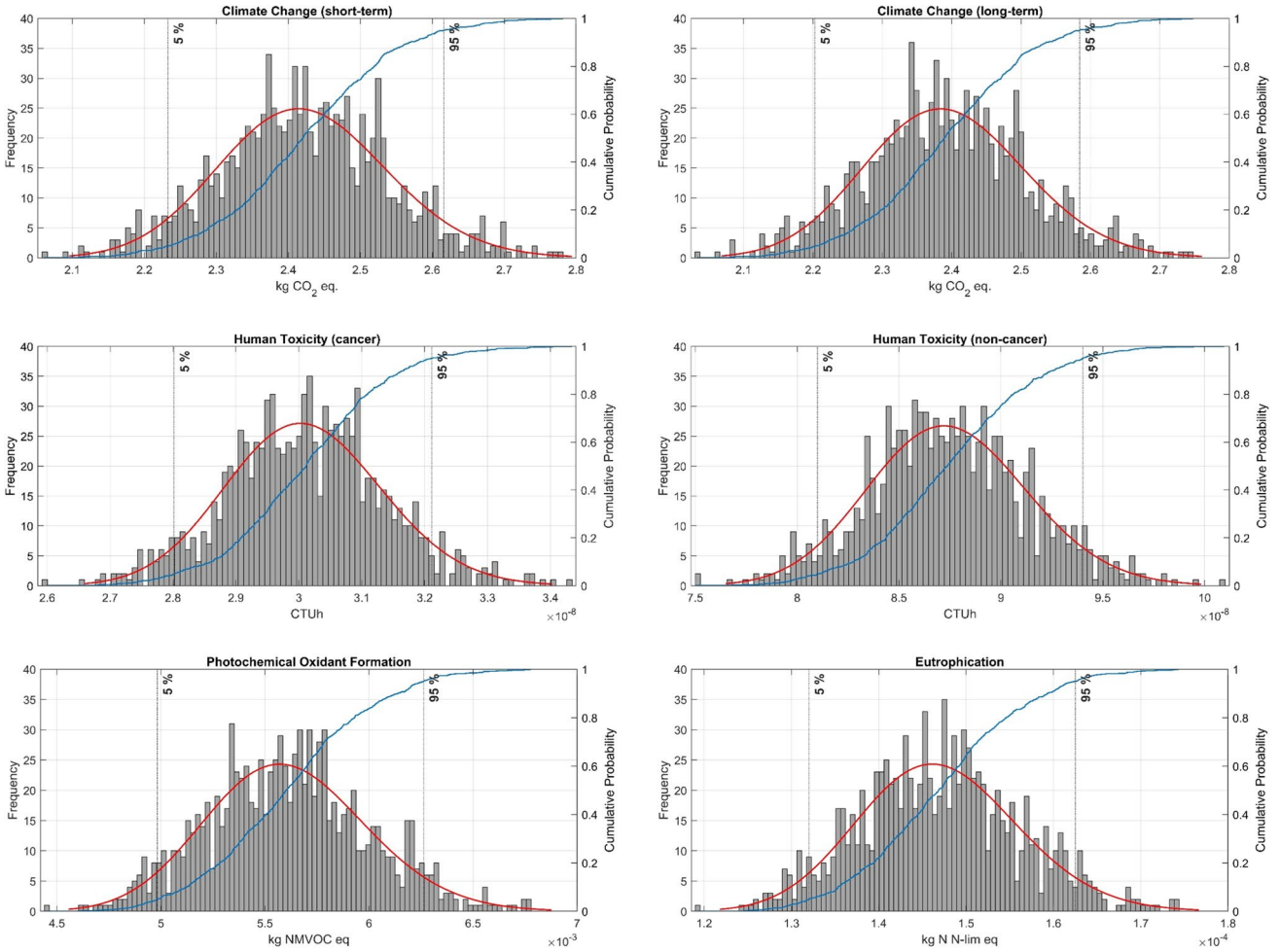


Fig. 6 Uncertainty associated with the global warming, eutrophication, ecotoxicity, carcinogens, and non-carcinogens of bioethanol production from rice straw

uncertainties associated with non-renewable fossil, renewable, and non-renewable biomass for RS and MSW-based ethanol are shown in Fig. 8a and b, respectively. This study uses lognormal distribution to reduce the possibility of negative emission for uncertainty analysis. The analysis also evaluates parameters such as mean, standard deviation, coefficient of variation (CoV), standard error of the mean (SEM), and skewness. A higher CoV value represents high uncertainty in results. As per the Monte Carlo simulation, the CoV of non-renewable fossil and renewable biomass for RS = 8.1%, standard deviation = 1.16 and 0.001 MJ, respectively, and median = 14.25 and 0.013 MJ, respectively. For MSW, the coefficient of variation of non-renewable fossil and renewable biomass for RS = 8%, standard deviation = 1.513 and 0.0014 MJ, respectively, and median = 18.93 and 0.018 MJ, respectively. The SEM represents the precision of the estimate of the mean, and the smaller SEM value shows more precision. The SEM values for RS and MSW indicate that RS generally has more precise mean estimates, which can significantly impact the reliability and accuracy of the analysis. This highlights the importance of SEM in research and practical applications to ensure informed and accurate decision-making.

Figure 6 and Fig. 7 depict the frequency distributions and cumulative probabilities of midpoint environmental impacts of LCA of bioethanol from RS and MSW. The consistency in the data, as indicated by the narrow ranges between the 5th and 95th percentiles, suggests that the process has predictable environmental impacts, which is valuable for planning and optimization. These graphs are crucial for understanding the environmental impacts of bioethanol production. By analyzing the frequency distribution and cumulative probability of each impact category, stakeholders can identify the most significant environmental concerns and target areas for improvement. It shows how probabilities accumulate as value increases, showcasing trends of medians and variability across different values. This aids in statistical reasoning and decision-making by presenting a holistic view of probabilities and their cumulative nature.

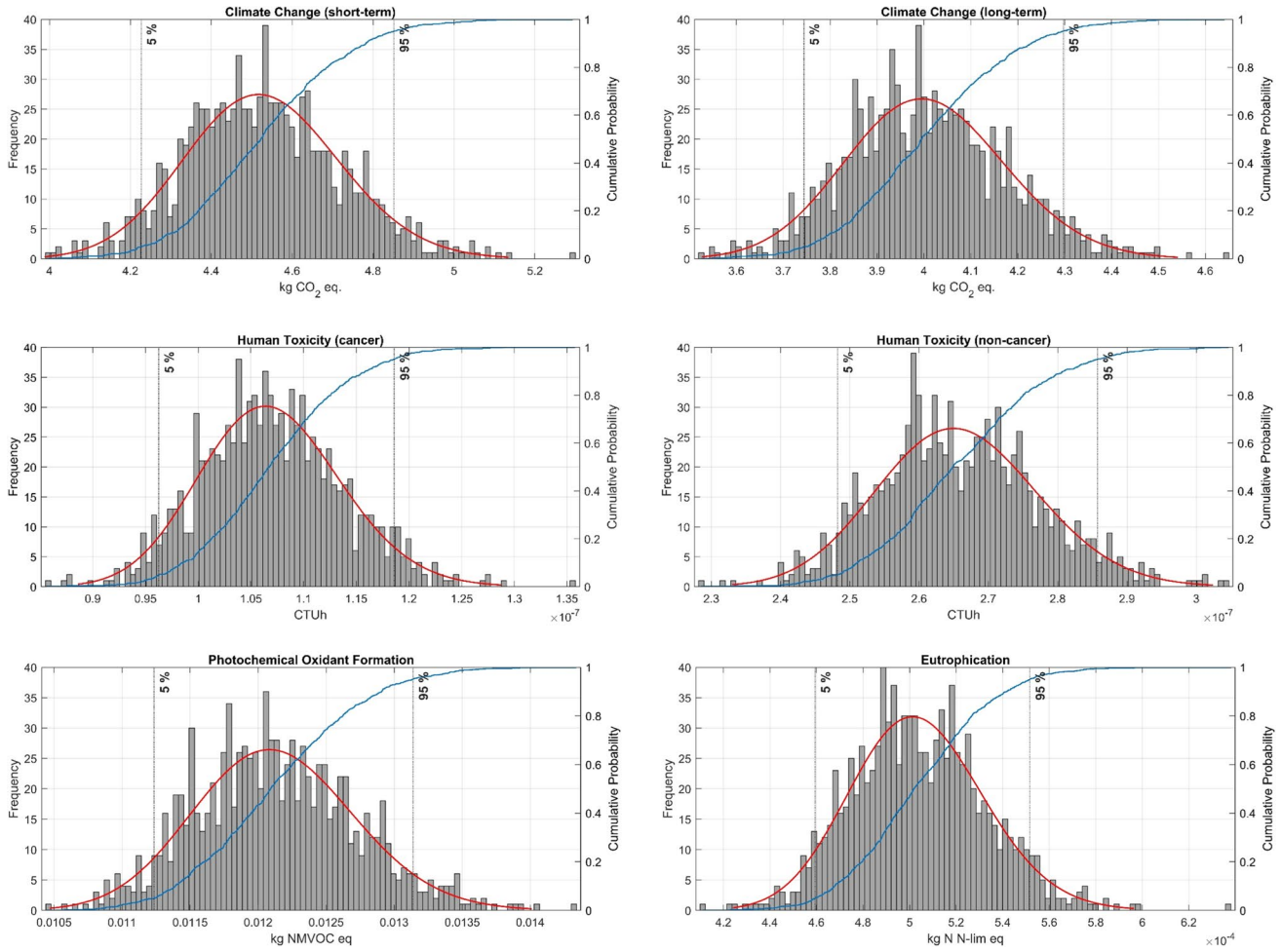


Fig. 7 Uncertainty associated with the global warming, eutrophication, ecotoxicity, carcinogens, and non-carcinogens of bioethanol production from MSW

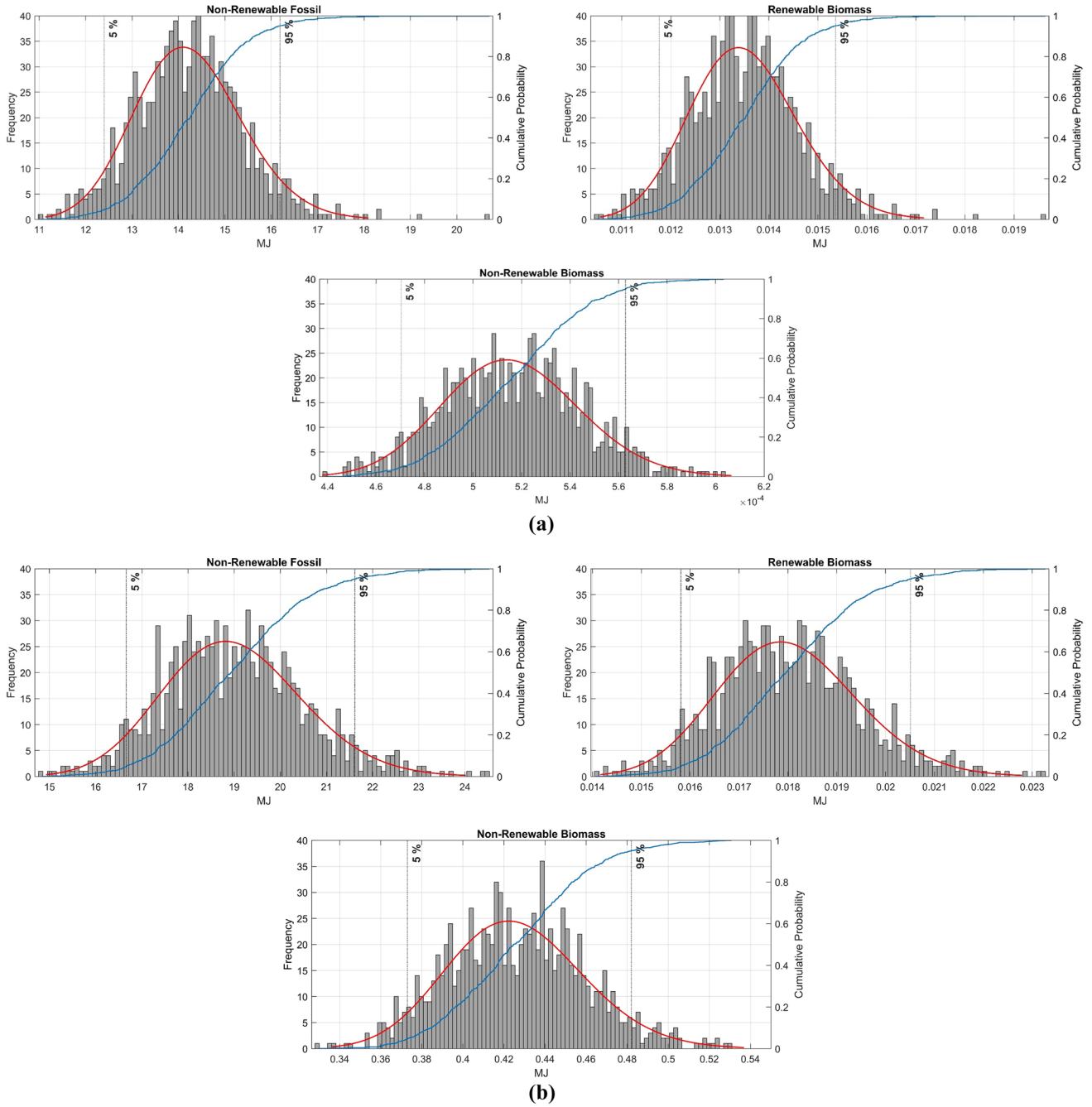


Fig. 8 Uncertainty associated with non-renewable fossil and renewable and non-renewable biomass of bioethanol production from **a** RS and **b** MSW

Conclusion

Waste, a growing concern globally, poses environmental challenges due to its increasing accumulation and inadequate disposal methods. The solution lies in sustainable waste management, where bioethanol production from municipal solid waste and rice straw plays a crucial role. Utilizing these waste materials for bioethanol not only addresses waste management issues but also contributes to cleaner energy sources. LCA further enhances this solution by evaluating the environmental impact of bioethanol production from MSW and RS, ensuring a comprehensive understanding of its sustainability. The study compares the bioethanol production from MSW and RS using the LCA approach. The analysis shows that MSW-based bioethanol has high environmental effects in all considered damage categories. The primary

cause was observed because of the enzymatic hydrolysis and consumption of electricity for various processes. The CED shows that the production's primary energy source still relies on fossil energy, significantly impacting the environment. This indicates that switching to a more sustainable energy source can drastically lessen the total impacts. Uncertainty analysis played a pivotal role in our study, acknowledging the inherent variability in data and assumptions. Recognizing and addressing these uncertainties is crucial for making informed decisions and guiding future research endeavors. This LCA study provides a foundation for informed decision-making, offering valuable information for policymakers, industry stakeholders, and consumers.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, conceptualization, and analysis were performed by Kulvendra Patel. Methodology; writing—review and editing; and supervision by Santosh Kumar Singh. The first draft of the manuscript was written by Kulvendra Patel and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Ahlgren S, Di-Lucia L (2014) Indirect land use changes of biofuel production – a review of modelling efforts and policy developments in the European Union. Biotechnol Biofuels 7:35. https://doi.org/10.1186/1754-6834-7-35

Alengebawy A, Mohamed BA, Ran Y, Yang Y, Pezzuolo A, Samer M, Ai P (2022) A comparative environmental life cycle assessment of rice straw-based bioenergy projects in China. Environ Res 212:113404. https://doi.org/10.1016/j.envres.2022.113404

Bai W, Yan L, Zhang L, Ye L (2022) Practice, pathways, and performance for resource utilization of crop straw: a case study of Xinyang City in China. Environ Sci Pollut Res 30:10812–10829. https://doi.org/10.1007/s11356-022-22840-0

Belal EB (2013) Bioethanol production from rice straw residues. Braz J Microbiol 44:225–234. https://doi.org/10.1590/S1517-83822013000100033

Bhonsle AK, Singh J, Trivedi J, Atray N (2022) Comparative LCA studies of biodiesel produced from used cooking oil using conventional and novel room temperature processes. Bioresour Technol Rep 18:101072. https://doi.org/10.1016/j.biteb.2022.101072

Bozorgirad MA, Zhang H, Haapala KR, Murthy GS (2013) Environmental impact and cost assessment of incineration and ethanol production as municipal solid waste management strategies. Int J Life Cycle Assess 18:1502–1512. https://doi.org/10.1007/s11367-013-0587-z

Braguglia CM, Gallipoli A, Gianico A, Pagliaccia P (2018) Anaerobic bioconversion of food waste into energy: a critical review. Bioresour Technol 248:37–56. https://doi.org/10.1016/J.BIORTECH.2017.06.145

Broder JD, Henson L (1993) Municipal solid waste and waste cellulosics conversion to fuels and chemicals, April 1990-September 1992

Cardona CA, Sánchez ÓJ (2007) Fuel ethanol production: process design trends and integration opportunities. Bioresour Technol 98:2415–2457. https://doi.org/10.1016/J.BIORTECH.2007.01.002

Chisti Y (2007) Biodiesel from microalgae. Biotechnol Adv 25:294–306. https://doi.org/10.1016/j.biotechadv.2007.02.001

Ciroth A, Muller S, Weidema BP (2012) Refining the pedigree matrix approach in ecoinvent. Greendelta. https://lca-net.com/files/Pedigree_report_final_May2012.pdf

Curtiss PS, Kreider JF (2010) Life cycle analysis of automotive ethanol produced from municipal solid waste. Proceedings of the ASME 3rd International Conference on Energy Sustainability 2009, ES2009 1:181–186. https://doi.org/10.1115/ES2009-90327

Das A, Verma M, Mishra V (2023) Food waste to resource recovery: a way of green advocacy. Environ Sci Pollut Res 31:17874–17886. https://doi.org/10.1007/s11356-023-27193-w

Davis SC, Anderson-Teixeira KJ, DeLucia EH (2009) Life-cycle analysis and the ecology of biofuels. Trends Plant Sci 14:140–146. https://doi.org/10.1016/j.tplants.2008.12.006

Deshmukh M, Pathan A (2023) Transformations of bamboo into bioethanol through biorefinery. Environ Sci Pollut Res 31:3343–3360. https://doi.org/10.1007/s11356-023-31510-8

Dey P, Chakrabortty S, Haldar D, Rangarajan V, Ashok S (2022) On-site enriched production of cellulase enzyme using rice straw waste and its hydrolytic performance evaluation through systematic dynamic modeling. Environ Sci Pollut Res 30:36710–36727. https://doi.org/10.1007/s11356-022-24797-6

Elginoz N, Khatami K, Owusu-Agyeman I, Cetecioglu Z (2020) Life cycle assessment of an innovative food waste management system. Front Sustain Food Syst 4:23. https://doi.org/10.3389/fsufs.2020.00023

FAO (2022) Establishing residue supply chains to reduce open burning the case of rice straw and renewable energy in Punjab, India. Environment and Natural Resources Management Working Paper. https://openknowledge.fao.org/handle/20.500.14283/cb9659en

Fargione J, Hill J, Tilman D, Polasky S, Hawthorne P (2008) Land clearing and the biofuel carbon debt. Science 319:1235–1238. https://doi.org/10.1126/science.1152747

GeneSyst International (2012) Waste conversion technology. https://www.genesyst.com. Accessed 25 Aug 2023

Gheewala SH (2023) Life cycle assessment for sustainability assessment of biofuels and bioproducts. Biofuel Res J 10:1810–1815. https://doi.org/10.18331/BRJ2023.10.1.5

Gheewala SH, Silalertruksa T, Pongpat P, Bonnet S (2019) Biofuel production from sugarcane in Thailand. In: Khan M, Khan I (eds) Sugarcane Biofuels. Springer, Cham, pp 157–174. https://doi.org/10.1007/978-3-030-18597-8_8

Gunarathne V, Ashiq A, Ramanayaka S, Wijekoon P, Vithanage M (2019) Biochar from municipal solid waste for resource recovery and pollution remediation. Environ Chem Lett 17:1225–1235. https://doi.org/10.1007/s10311-019-00866-0

ISO (2006a) ISO 14040:2006 - Environmental management — life cycle assessment — principles and framework. International Organization for Standardization (ISO). https://www.iso.org/standard/37456.html. Accessed 5 Oct 2022

ISO (2006b) ISO 14044:2006 - Environmental management — life cycle assessment — requirements and guidelines. International Organization for Standardization (ISO). https://www.iso.org/standard/38498.html. Accessed 5 Oct 2022

Jiang Z, Dai Y, Du T (2020) Comparison of the energetic, environmental, and economic performances of three household-based

modern bioenergy utilization systems in China. J Environ Manag 264:110481. https://doi.org/10.1016/j.jenvman.2020.110481
Kalogo Y, Habibi S, MacLean HL, Joshi SV (2007) Environmental implications of municipal solid waste-derived ethanol. Environ Sci Technol 41:35–41. https://doi.org/10.1021/es061117b
Kaparaju P, Serrano M, Thomsen AB, Kongjan P, Angelidaki I (2009) Bioethanol, biohydrogen and biogas production from wheat straw in a biorefinery concept. Bioresour Technol 100:2562–2568. https://doi.org/10.1016/j.biortech.2008.11.011
Kowalski Z, Kulczycka J, Verhé R, Desender L, De Clercq G, Makara A, Generowicz N, Harazin P (2022) Second-generation biofuel production from the organic fraction of municipal solid waste. Front Energy Res 10:919415. https://doi.org/10.3389/fenrg.2022.919415
Kumar R, Patel K, Singh SK (2024) Biological wastewater treatment: a comprehensive sustainability analysis using life cycle assessment. Environ Monit Assess 196:416. https://doi.org/10.1007/s10661-024-12578-2
Liska AJ, Yang HS, Bremer VR, Klopfenstein TJ, Walters DT, Erickson GE, Cassman KG (2009) Improvements in life cycle energy efficiency and greenhouse gas emissions of corn-ethanol. J Ind Ecol 13:58–74. https://doi.org/10.1111/j.1530-9290.2008.00105.x
Liu Y, Li G, Chen Z, Shen Y, Zhang H, Wang S, Qi J, Zhu Z, Wang Y, Gao J (2020) Comprehensive analysis of environmental impacts and energy consumption of biomass-to-methanol and coal-to-methanol via life cycle assessment. Energy 204:117961. https://doi.org/10.1016/j.energy.2020.117961
Liu Q, Pan S, Zhou L, Feng L, Li Z, Du L, Wei Y (2022) Improving the biogas potential of rice straw through microwave-assisted ammoniation pretreatment during anaerobic digestion. Bioenergy Res 15:1240–1250. https://doi.org/10.1007/s12155-021-10299-9
Mahmoodi P, Karimi K, Taherzadeh MJ (2018) Efficient conversion of municipal solid waste to biofuel by simultaneous dilute-acid hydrolysis of starch and pretreatment of lignocelluloses. Energy Convers Manag 166:569–578. https://doi.org/10.1016/j.enconman.2018.04.067
Morales M, Collet P, Lardon L, Hélias A, Steyer JP, Bernard O (2019) Chapter 20 - Life-cycle assessment of microalgal-based biofuel. In: Pandey A, Chang J-S, Soccol CR, Lee DJ, Chisti Y (Eds.), Biofuels from Algae (Second Edition). Elsevier, pp 507–550. https://doi.org/10.1016/B978-0-444-64192-2.00020-2
Nizzy AM, Kannan S (2022) A review on the conversion of cassava wastes into value-added products towards a sustainable environment. Environ Sci Pollut Res 29:69223–69240. https://doi.org/10.1007/s11356-022-22500-3
Oldfield TL, White E, Holden NM (2016) An environmental analysis of options for utilising wasted food and food residue. J Environ Manag 183:826–835. https://doi.org/10.1016/J.JENVMAN.2016.09.035
Patel K, Singh SK (2022a) A life cycle approach to environmental assessment of wastewater and sludge treatment processes. Water Environ J 36:412–424. https://doi.org/10.1111/wej.12774
Patel K, Singh SK (2022b) Assessing the sustainability of municipal solid waste management using life-cycle analysis. Proc Inst Civ Eng Waste Resour Manag 175:64–74. https://doi.org/10.1680/jwarm.21.00018
Patel K, Singh SK (2023a) Environmental sustainability analysis of biofuels: a critical review of LCA studies. Clean Technol Environ Policy 25:2489–2510. https://doi.org/10.1007/s10098-023-02596-y
Patel K, Singh SK (2023b) Assessing the sustainability of Jatropha and rapeseed biodiesel: an LCA approach. In: Al Khaddar R, Singh SK, Kaushika ND, Tomar RK, Jain SK (eds) Recent Developments in Energy and Environmental Engineering: Select Proceedings of TRACE 2022. Springer Nature Singapore, Singapore, pp 191–200. https://doi.org/10.1007/978-981-99-1388-6_15
Pradhan S, Parthasarathy P, Mackey HR, Al-Ansari T, McKay G (2024) Food waste biochar: a sustainable solution for agriculture application and soil–water remediation. Carbon Res 3:41. https://doi.org/10.1007/s44246-024-00123-2
PRé Sustainability (2022) SimaPro v9.4 [Computer Software]. Retrieved from https://simapro.com
Rebitzer G, Ekvall T, Frischknecht R, Hunkeler D, Norris G, Rydberg T, Schmidt WP, Suh S, Weidema BP, Pennington DW (2004) Life cycle assessment. Environ Int 30:701–720. https://doi.org/10.1016/j.envint.2003.11.005
Rosado MJ, Marques G, Rencoret J, Gutiérrez A, del Río JC (2022) Chemical composition of lipophilic compounds from rice (Oryza sativa) straw: an attractive feedstock for obtaining valuable phytochemicals. Front Plant Sci 13:868319. https://doi.org/10.3389/fpls.2022.868319
Samer M, Hijazi O, Abdelsalam EM, El-Hussein A, Attia YA, Yacoub IH, Bernhardt H (2021) Life cycle assessment of using laser treatment and nanomaterials to produce biogas through anaerobic digestion of slurry. Environ Dev Sustain 23:14683–14696. https://doi.org/10.1007/s10668-021-01264-9
Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A, Fabiosa J, Tokgoz S, Hayes D, Yu TH (2008) Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319:1238–1240. https://doi.org/10.1126/science.1151861
Shakelly N, Pérez-Cardona JR, Deng S, Maani T, Li Z, Sutherland JW (2023) Comparative life cycle assessment of bioethanol production from different generations of biomass and waste feedstocks. In: Procedia CIRP. Elsevier BV, pp 630–635. https://doi.org/10.1016/j.procir.2023.02.106
Sharma T, Dasgupta D, Singh J, Bhaskar T, Ghosh D (2020) Yeast lipid-based biofuels and oleochemicals from lignocellulosic biomass: life cycle impact assessment. Sustain Energy Fuels 4:387–398. https://doi.org/10.1039/C9SE00540D
Sheikholeslami Z, Ehteshami M, Nazif S, Semiarian A (2022) The environmental assessment of tertiary treatment technologies for wastewater reuse by considering LCA uncertainty. Process Saf Environ Prot 168:928–941. https://doi.org/10.1016/j.psep.2022.10.074
Soam S, Kapoor M, Kumar R, Borjesson P, Gupta RP, Tuli DK (2016) Global warming potential and energy analysis of second generation ethanol production from rice straw in India. Appl Energy 184:353–364. https://doi.org/10.1016/J.APENERGY.2016.10.034
Soam S, Borjesson P, Sharma PK, Gupta RP, Tuli DK, Kumar R (2017) Life cycle assessment of rice straw utilization practices in India. Bioresour Technol 228:89–98. https://doi.org/10.1016/j.biortech.2016.12.082
Sreekumar A, Shastri Y, Wadekar P, Patil M, Lali A (2020) Life cycle assessment of ethanol production in a rice-straw-based biorefinery in India. Clean Technol Environ Policy 22:409–422. https://doi.org/10.1007/s10098-019-01791-0
Subudhi S, Saha K, Mudgil D, Sarangi PK, Srivastava RK, Sarma MK (2023) Biomethanol production from renewable resources: a sustainable approach. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-023-29616-0
Sukumaran RK, Surender VJ, Sindhu R, Binod P, Janu KU, Sajna KV, Rajasree KP, Pandey A (2010) Lignocellulosic ethanol in India: prospects, challenges and feedstock availability. Bioresour Technol 101:4826–4833. https://doi.org/10.1016/j.biortech.2009.11.049
Teng Y, Pan W (2020) Estimating and minimizing embodied carbon of prefabricated high-rise residential buildings considering parameter, scenario and model uncertainties. Build Environ 180:106951. https://doi.org/10.1016/j.buildenv.2020.106951
UN Report (2007) Sustainable bioenergy: a framework for decision makers. https://www.fao.org/4/a1094e/a1094e00.htm. Accessed 18 Aug 2023

US EPA (2022) National overview: facts and figures on materials, wastes and recycling. https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/national-overview-facts-and-figures-materials

Van Hung N, Maguyon-Detras MC, Migo MV, Quilloy R, Balingbing C, Chivenge P, Gummert M (2020) Rice straw overview: availability, properties, and management practices. In: Gummert M, Van Hung N, Chivenge P, Douthwaite B (eds) Sustainable rice straw management. Springer International Publishing, Cham, pp 1–13. https://doi.org/10.1007/978-3-030-32373-8_1

Verhe R, Varghese S, Thevelein JM, Nikroo JH, Lambrecht M, Redant E, De Clercq G (2022) Production of bio-ethanol from the organic fraction of municipal solid waste and refuse-derived fuel. Biomass 2:224–236. https://doi.org/10.3390/biomass2040015

Wernet G, Bauer C, Steubing B, Reinhard J, Moreno-Ruiz E, Weidema B (2016) The ecoinvent database version 3 (part I): overview and methodology. Int J Life Cycle Assess 21:1218–1230. https://doi.org/10.1007/s11367-016-1087-8

Yazan DM, Mandras G, Garau G (2017) Environmental and economic sustainability of integrated production in bio-refineries: the thistle case in Sardinia. Renew Energy 102:349–360. https://doi.org/10.1016/j.renene.2016.10.055

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.

Physica Scripta

INCLUSIVE PUBLISHING
TRUSTED SCIENCE

PAPER

Triple band self-decoupled MIMO antenna pair for 5G communication

To cite this article: Parul Kansal *et al* 2024 *Phys. Scr.* **99** 095532

View the article online for updates and enhancements.

You may also like

- CLASS I METHANOL (CH_3OH) MASER CONDITIONS NEAR SUPERNOVA REMNANTS
Bridget C. McEwen, Ylva M. Pihlström and Loránt O. Sjouwerman
- Tunable multiband polarization conversion and manipulation in vanadium dioxide-based asymmetric chiral metamaterial
Shichao Song, Xiaoliang Ma, Mingbo Pu et al.
- Experimental confirmation and physical understanding of ultra-high bit rate impulse radio in the THz digital communication channels of the atmosphere
Mahboubeh Mandehgar, Yihong Yang and D Grischkowsky

This content was downloaded from IP address 14.139.251.98 on 27/08/2024 at 06:07

Physica Scripta

PAPER

Triple band self-decoupled MIMO antenna pair for 5G communication

Parul Kansal , Anup Kumar Mandpura and Narendra Kumar
Department of Electrical Engineering, Delhi Technological University, Delhi, India
E-mail: kansal.parul@gmail.com

RECEIVED
30 January 2024
REVISED
28 July 2024
ACCEPTED FOR PUBLICATION
9 August 2024
PUBLISHED
22 August 2024

Keywords: multiple-input multiple-output (MIMO), self-decoupled, common mode, differential mode, fifth generation

Abstract
A two port multiple input multiple output (MIMO) antenna with self-isolation is proposed for the application of fifth generation (5G) mobile phones. In this antenna, feeding ports that are inherently isolated across a bandwidth without the need for extra decoupling components directly excite the MIMO antenna. An approach involving the creation of common and differential modes utilizes a mode-cancellation technique to comprehensively understand the self-decoupling mechanism on a physical level. The proposed self-decoupled antenna pair exhibits about 22 dB isolation across the triple band 2.58 GHz to 2.84 GHz, 3.4 GHz to 3.9 GHz and 4.3 GHz to 4.6 GHz, respectively. According to the experimental findings, the proposed MIMO system may offer better than 22 dB isolation across all ports and a respectable level of overall gain 1.9 dBi, 3.77 dBi and 2.9 dBi across 2.58 GHz to 2.84 GHz, 3.4 GHz to 3.9 GHz and 4.3 GHz to 4.6 GHz. A suggested design method, with its advantages of self-decoupling, simple structure, high efficiency and triple band demonstrates amazing upcoming 5G highly integrated MIMO antennas for 5G devices.

1. Introduction

Multiple-Input-Multiple-Output (MIMO) technology is a wireless communication technique that uses multiple transmitting and receiving antennas to improve the performance of wireless systems. MIMO takes advantage of the multipath propagation phenomenon, where signals sent from a transmitter reach the receiver through multiple paths due to reflections, diffractions, and scattering in the environment [1]. MIMO has been widely adopted in various wireless communication standards, including Wi-Fi (802.11n/ac/ax), cellular networks (4G LTE, 5G), and wireless point-to-point links. Its effectiveness in increasing data rates and system capacity has made it a fundamental technology for addressing the ever-increasing demand for more data speeds in modern wireless networks [2]. A licensed band of frequencies between 4.40 and 5.01 GHz, 3.30 and 3.80 GHz, and 3.30 and 4.20 GHz is needed for the Sub-6 GHz 5G. While Wi-Fi 6E 5G provides quality of service (QoS) suitable for Industrial Internet of Things (IoT) applications, which require low latency and high capacity, and has become the standard networking technology for both households and businesses, including robotics [3]. To enable independent usage of Wi-Fi 6E and 5G bands in indoor networks and seamless transition between them, the aim of a compact MIMO antenna capable of functioning across multiple frequency bands. One particular issue with the MIMO antenna design is the mutual interaction between components of densely positioned antennas, lowering the MIMO antenna's efficiency in terms of effective spectrum utilization and high-speed data transmission [4]. When a large number of antenna group together the mutual couplings increased and it's reduces antenna efficiency and MIMO performance. It is a serious issue in the MIMO technology based antenna system [5].

Different types of MIMO antennas are proposed in [6–10] to reduce mutual coupling between antenna radiators. MIMO systems, which take advantage of multipath propagation, employ multiple transmitting and receiving antennas [11, 12]. However, the reduced efficiency of the antennas and the performance of MIMO

© 2024 IOP Publishing Ltd

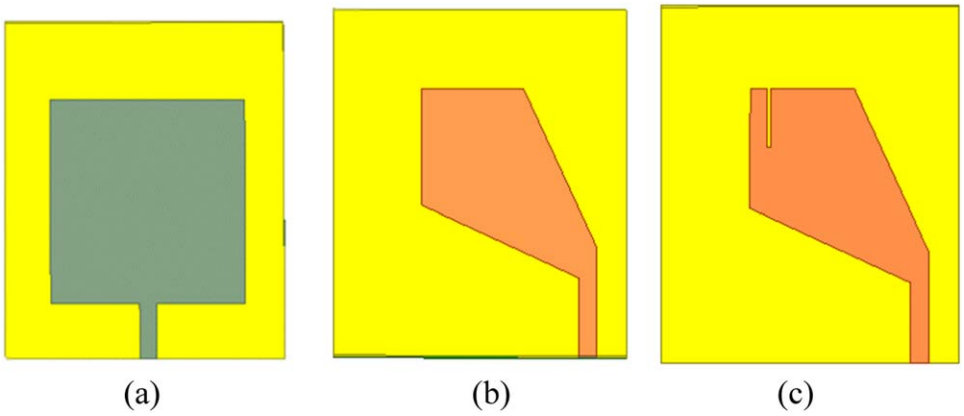


Figure 1. Evolution of the microstrip-fed monopole antenna (a) Antenna_1 (b) Antenna_2 (c) Antenna_ 3.

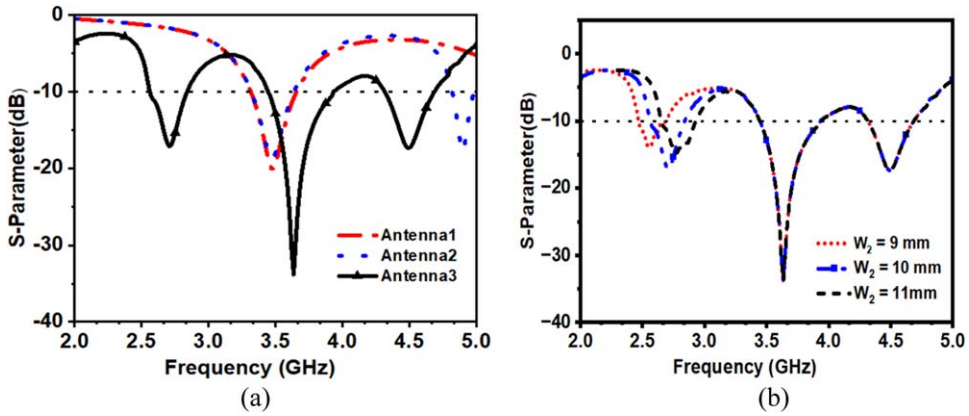


Figure 2. Simulated S-parameter of (a) single antenna step (b) parametric analysis.

caused by the increased mutual coupling that arises from positioning several antennas close to one another make this a critical issue to overcome.

To enhance the level of integration for MIMO antennas and to facilitate the upcoming requirements of the broad-spectrum 5G NR sector, a pair of self-decoupled MIMO antennas connected by a common radiator can be utilized. Even though two antenna components are combined to create a paired antenna sharing a common radiator, sufficient isolation is attained without the need for additional decoupling methods. A mode-cancellation method (CM) constructed around a new understanding of common mode (CM) and differential mode (DM) cancellation has been introduced to analyses the actual workings of the self-decoupling phenomena. The combination of CM and DM currents might mimic the current mode delivered via a single port, as outlined by the MCM. Consequently, half of the existing current in the shared radiator is overlapped. While the current in the other half is negated, resulting in a natural self-decoupling feature [13, 14]. Because it gives a clear design route for tackling the mutual coupling issue, recently MIMO antenna designs have used the approach of even-mode and odd-mode [15–17]. In this part, antennas with closely spaced feed points can be examined. Around antenna feed points, electromagnetic fields or currents tend to congregate [18–21] and significant electromagnetic fields or currents can be created at additional feed sites when one feed point is energized. As a result, achieving a high level of isolation in this situation is challenging [22–25].

This research paper presents, a triple-band monopole MIMO antenna with improved isolation. The proposed design is created and evaluated to confirm the simulation and measured results by analysing the S-parameter, radiation pattern, efficiency & peak gain using HFSS. The overall dimension of the entire model is 70 mm × 31 mm. Measured outcomes for the combined antenna indicate that the proposed antenna functions within three frequency bands, achieving S_{11} values of less than or equal to 10 dB from 2.58 GHz to 2.84 GHz, 3.4 GHz to 3.9 GHz and 4.3 GHz to 4.6 GHz. Furthermore, peak gain at three bands is 1.9 dBi, 3.77 dBi and 2.9 dBi. Two symmetric feeding strips are used to individually excite the individual radiator, which has a

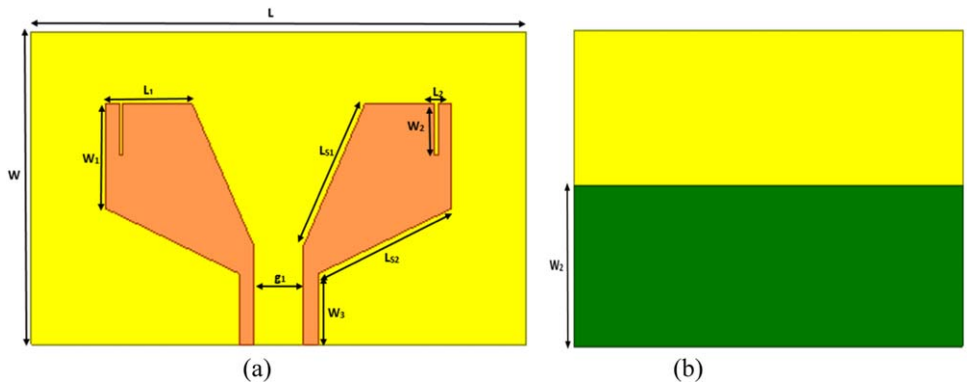


Figure 3. Layout and design of self-decoupled MIMO antenna, (a) top view (b) bottom view.

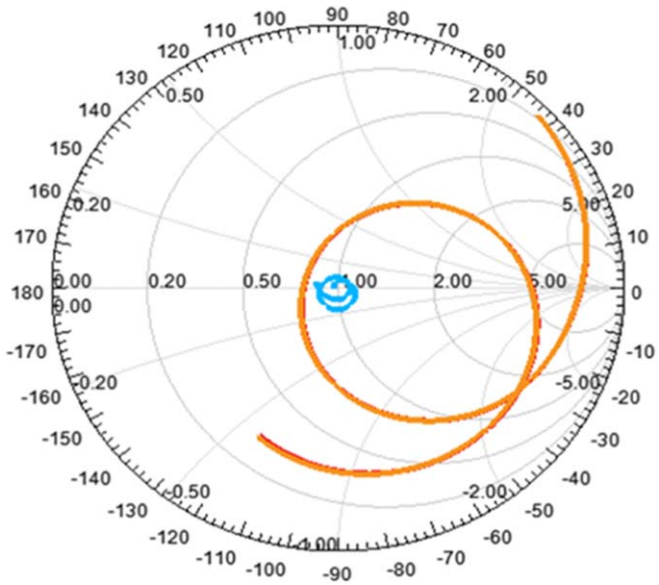


Figure 4. Simulated Smith chart showing the feeding distance $D_o = 10$ mm and the CM and DM reflection coefficients.

diameter of 35 mm by 18.5 mm. The distance between the two feeding ports is Do = 10 mm, which is an essential requirement for port isolation of more than 22 dB.

The subsequent sections of the paper are organized in the following manner. The antenna design and decoupling analysis are presented in section 2. Section 3 shows how the performance of two-element triple-band MIMO antennas is verified. The diversity performance is finally summarised in section 4. The SAR analysis presented in section 5, followed by a summary of this work in section 5.

2. Antenna design and decoupled structure

2.1. Triple-band antenna design

The evolution of the suggested triple-band antenna is depicted in figure 1. To reach the suggested 50 Ω of impedance, microstrip lines having a width of 3 mm for transmission and a ground plane-to-transmission line gap of 1.6 mm provide fed. Figure 2 shows the simulated S_{11} results for each configuration. The radiator shape for antenna 1 is a rectangular patch, as seen in figure 1(a). The operating bandwidth is affected by the ground plane's length. The antenna is used in the frequency range from 3.3 GHz to 3.6 GHz when the ground plane measures 31 mm, as illustrated in figure 2 red dotted line (antenna 1).

By making adjustments to the radiator's design, utilizing a conventional two-triangle cut on the antenna (illustrated in figure 1(b)), an antenna slanted to the left is generated, exhibiting resonance across two frequency bands. Figure 2 (antenna 2, indicated by the blue dotted line) demonstrates a band with an S11 level below 19 dB,

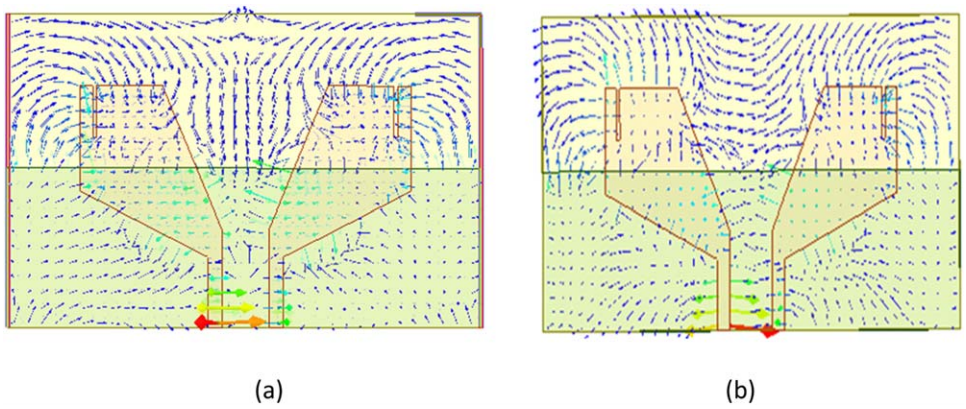


Figure 5. Self-decoupled antenna pair's vector current distributions at 3.6 GHz (a) CM fed. (b) DM.

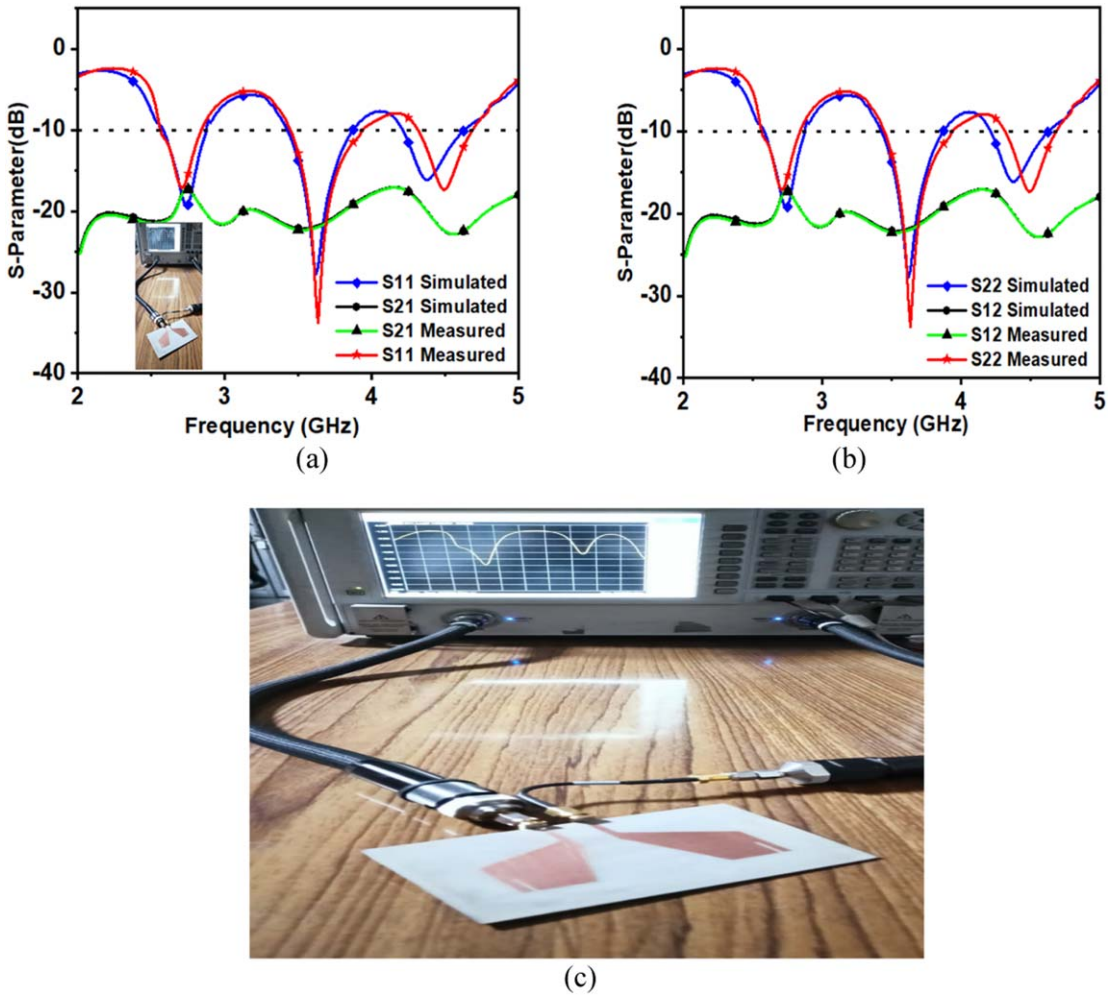


Figure 6. Self-decoupled MIMO antenna pair (a) S_{11} & S_{21} (b) S_{22} & S_{12} simulated and measured results (c) experimental test bench setup.

centred at 3.47 GHz. The bandwidth of this band extends from 3.3 GHz to 3.6 GHz, with S_{11} levels $\leqslant$ 10 dB. Additionally, a second band is observed between 4.7 GHz and 4.9 GHz, featuring an S_{11} level $\leqslant$ 10 dB.

Moreover, the introduction of a rectangular slot, illustrated in figure 1(c) and also observable in figure 2(a) (depicted by the solid black line for antenna 3), leads to a transformation of the dual-band antenna into a triple-band configuration. This transformation occurs due to modifications in the current distribution within the

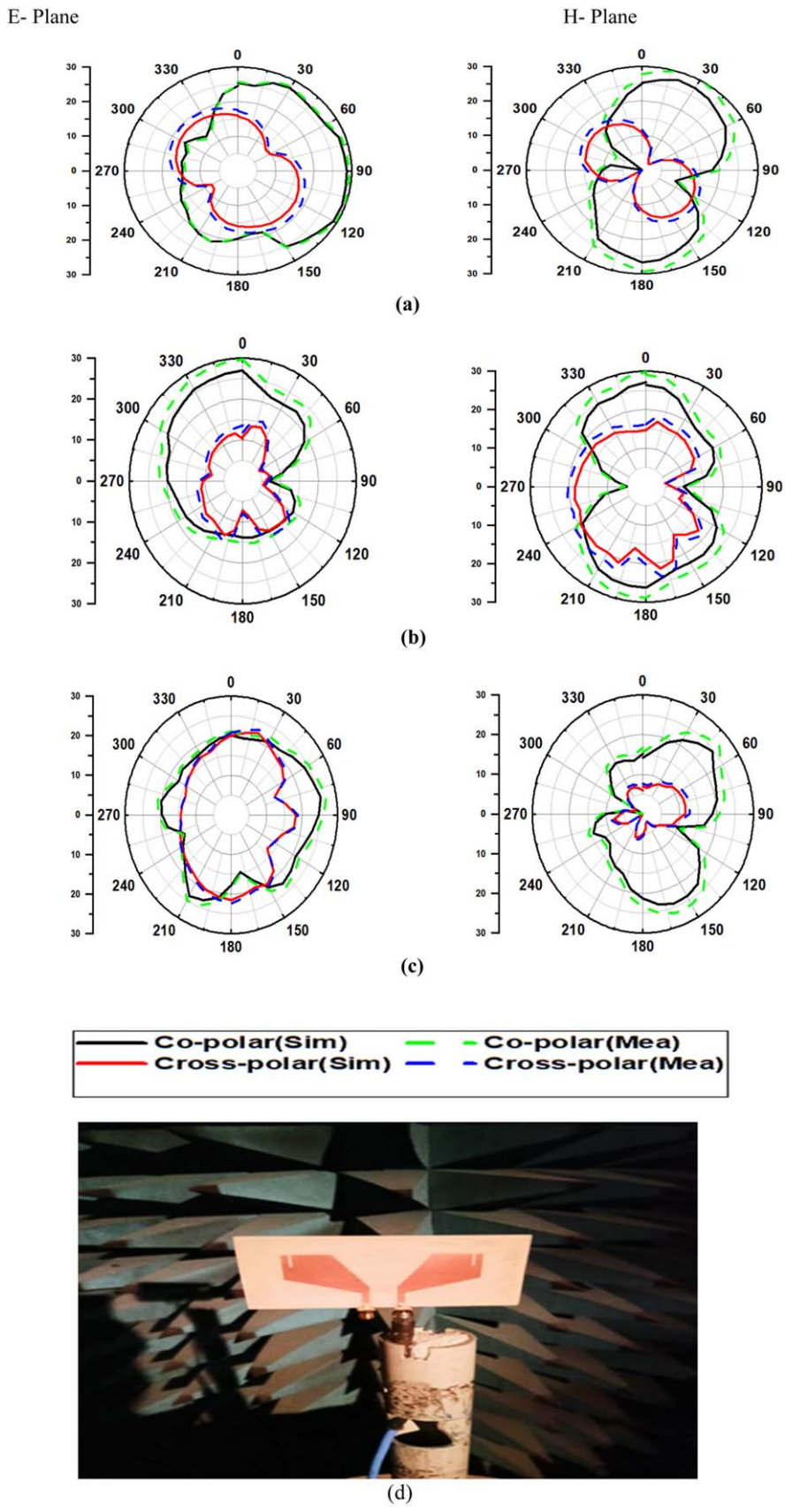


Figure 7. Radiation patterns of self-decoupled MIMO antenna (a) at 2.7 GHz (b) at 3.6 GHz (c) at 4.5 GHz (d) Measurement in anechoic chamber.

antenna's structure, resulting in the emergence of novel resonant frequencies encompassing the 2.5 GHz to 2.8 GHz bands. In this revised arrangement, the initial band functions within the 2.5 GHz to 2.8 GHz frequency range (260 MHz) with a central frequency of 2.7 GHz. Subsequently, the second band, centred at 3.6 GHz, from

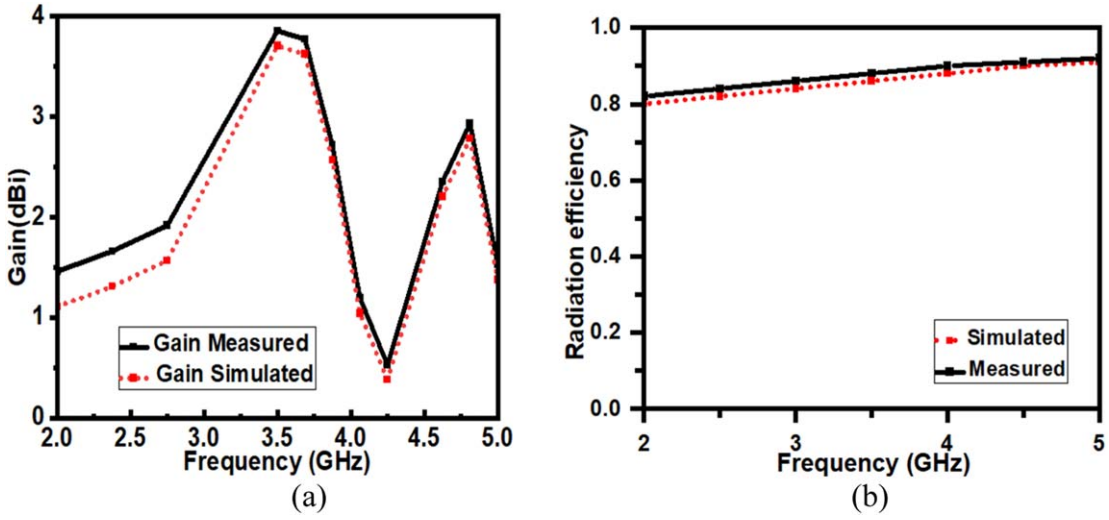


Figure 8. Self-decoupled MIMO antenna pair (a) Peak gain (b) Radiation efficiency.

3.4 GHz to 3.9 GHz (500 MHz). Lastly, the third band operates with a central frequency of 4.5 GHz and covers the 4.3 GHz to 4.6 GHz frequency range (300 MHz).

2.2. Parametric analysis

Using Ansys HFSS, the simulation of the antenna design across the operational frequency range was conducted, and the design parameters for the antenna structure were fine-tuned. The antenna's dimensions were determined through a parametric analysis involving the length of the slot (parameter W_2). The impact of W_2 on the S_{11} parameters is illustrated in figure 2(b). The W_2 varied between 9 to 11 mm. The optimized values were $W_2 = 10$ mm, resulting in the desired favourable return loss.

2.3. Proposed self-decoupled antenna pair

Figure 3 shows the design of the proposed two symmetrically mirrored tilted monopoles whose dimension is 70 mm × 31 mm, which is located above the ground plane with the substrate on a 1.6 mm thick RT-Duroid ($\epsilon_r = 2.2$, $\tan\delta = 0.009$) without using any the isolating structure. As indicated in figure 3(a), the two comparable monopole-tilted radiators are placed 10 mm apart from one another in a horizontal orientation. The parameters of the tilted monopoles MIMO antenna are $L = 70$ mm, $W = 31$ mm, $L_1 = 17.5$ mm, $W_1 = 20.5$ mm, $L_2 = 0.8$ mm, $W_2 = 10$ mm, $Ls_1 = 30$ mm, $Ls_2 = 30$ mm, $g_1 = 10$ mm, $W_3 = 8$ mm.

2.4. Decoupling analysis

The active reflection coefficients for the common-mode (CM) and differential-mode (DM) signals in a symmetric and reciprocal 2-port network are calculated and simplified using the theory of microwave networks. The S-parameters of the network satisfy $S_{11} = S_{22}$ and $S_{12} = S_{21}$ under this condition. The ratio of the reflected incident signal is known as the active reflection coefficient Sc_M for the CM signal, & the incident signal to the reflected signal is known as the active reflection coefficient S_{DM} for the DM signal. The greater the isolation between ports 1 and 2 the smaller the difference between these coefficients.

The Euclidean distance in the Smith chart makes it simple to determine the distinction between the CM and DM complex reflection coefficients. $S_{CM} = S_{DM}$, which means that just the CM and DM reflection coefficients are equivalent & If port 1 and port 2 are properly isolated, the opposing currents in common mode (CM) and differential mode (DM) can be eliminated through flawless self-isolation (i.e., $S_{21} = 0$). For instance, if the isolation between ports 1 and 2 in S_{21} is less than -10 dB $= 0.316$, then $|Sc_M - S_{DM}|$ is less than 0.632 (The Smith chart's radius is normalised to a value of one). To establish good isolation between port 1 and port 2, it is essential to adjust the CM and DM signals' active reflection coefficients must be similar (or as close in the Smith chart) as possible.

Figure 4 shows the Smith chart to create good isolation by adjusting the impedances of two ports, CM and DM, and their active reflection coefficients. The feeding distance D0, or distance between these ports, is an important factor that can be changed to enhance their performance. Based on the information given, it is possible to get an optimised self-isolation performance when Do is equal to 10 mm. Figure 4 depicts the Smith chart for the DM impedance, which has a longer curve and is less uniform than the CM impedance. Because of

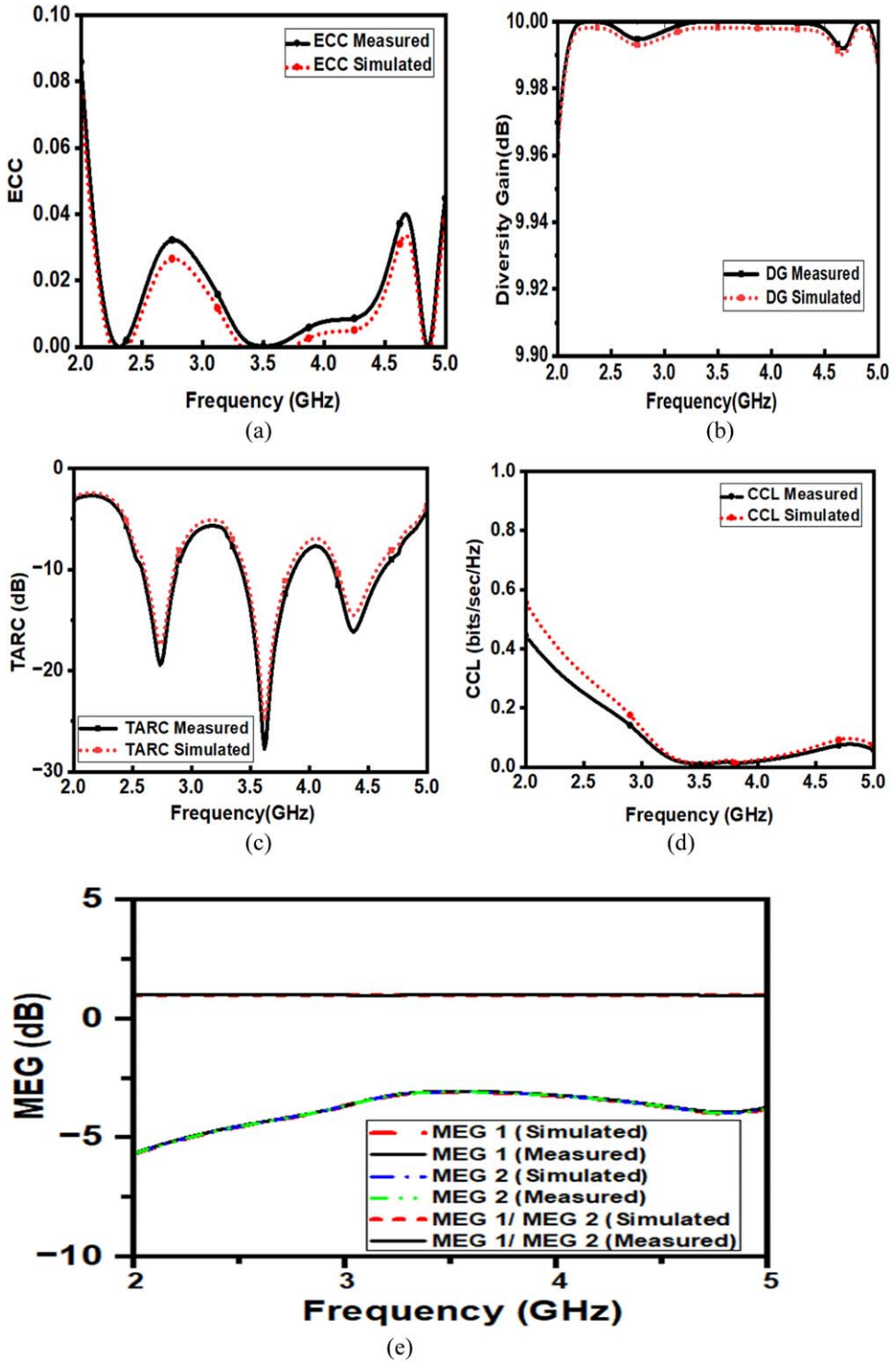


Figure 9. Self-decoupled MIMO antenna measured (a) ECC (b) DG (c) TARC (d) CCL (e) MEG.

the influence of the horizontal dipole-like mode's inverse ground mirror current. To achieve good results, both impedances can be changed to a comparable status.

The CM and DM-fed antenna pair vector current distributions as shown in figure 5. Two folded monopole-like current distributions in the antenna cause the current to be out of phase when the antenna is supplied using a common mode (CM) signal, as illustrated in figure 5(a). This indicates that the current on either side of the antenna runs in opposing directions. The antenna's current distribution resembles a horizontal dipole and is in a phase when the antenna is supplied using the differential mode (DM) signal, as illustrated in figure 5(b). This shows that the current flows in the same direction through the antenna on both sides. When designing and analysing antennas with CM and DM feeding, it's essential to take these current distributions into account because they may have an impact on the antenna's performance and radiation pattern.

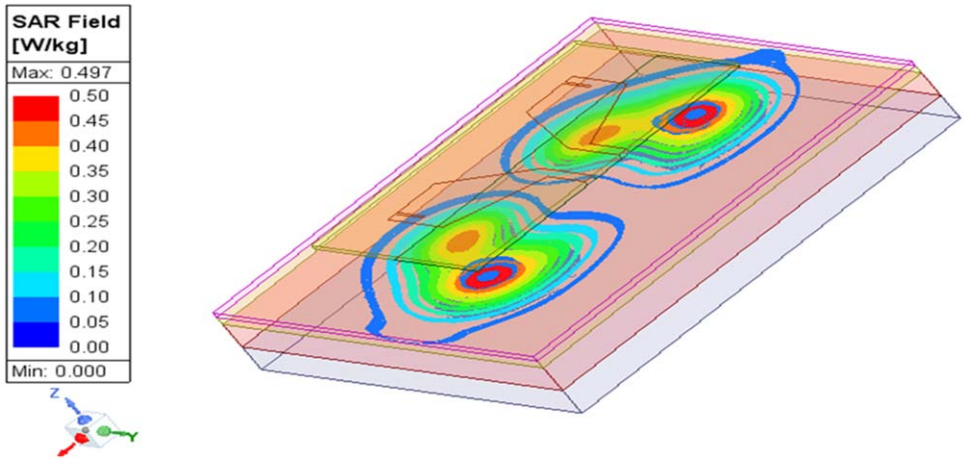


Figure 10. SAR of the proposed antenna.

3. Simulated and measured result

The simulated and measured MIMO antenna's S-parameters for ports 1 and 2 such as S_{11}, S_{22}, S_{12} and S_{21} are Shown in figures 6(a) & (b). The simulation results indicate an isolation of over 22 dB between ports 1 and 2 of the MIMO antenna and operates at three bands 2.7 GHz, 3.6 GHz and 4.5 GHz, a bandwidth ranging from 2.5 GHz to 2.8 GHz and computed at $S_{11} \leqslant 10$ dB for the first band, 3.4 GHz to 3.9 GHz for the second band and 4.3 GHz to 4.5 GHz for the third band, respectively. The results exhibit identical frequency ranges to those observed in the simulated and measured results, as seen in figure 6 (blue and red curve). The measurements indicate the antenna operates in three bands with a centred on 2.7 GHz, 3.6 GHz and 4.5 GHz, with bandwidths ranging from 2.5 GHz to 2.8 GHz, 3.4 GHz to 3.9 GHz and 4.36 GHz to 4.57 GHz for three bands. Here measured and simulated outcomes are nearly identical, with only a minor difference between the two. This is due to the fabrication step and simulation impact. The S-parameter measurement setup can be shown in the figure 6(c).

Figures 7(a)–(c) depicts the simulated and measured radiation patterns in the E and H planes at frequencies of 2.7 GHz, 3.6 GHz, and 4.5 GHz. An anechoic chamber is being used to test the antenna, as shown in figure 7(d). The antenna exhibits a bidirectional pattern in the H-plane. Figure 8(a) displays the antenna's peak gain as observed in simulations and measurements. The antenna simulated the peak gain of 1.9 dBi, 3.77 dBi and 2.9 dBi for three bands and measured the peak gain of about 2 dBi, 3.8 dBi, and 3 dBi. Figure 8(b) illustrates the overall radiation efficiency of proposed antenna design. The efficiency of the antenna at three different resonances are 83%, 86% and 89% respectively.

4. Diversity performance analysis

The diversity evaluation (ECC, DG, TARC, CCL and MEG) has significance for confirming the two-port MIMO antenna's performance, which was examined through simulations and experiments. The envelop correlation coefficient (ECC) graph for the specified frequency ranges is shown in figure 9(a). The ECC of the two-port MIMO antenna was less than 0.04 across the whole band, confirming that this MIMO antenna performs similarly to whatever is necessary for MIMO diversity performance. Figure 9(b) illustrates the diversity gain (DG) of a dual-port MIMO antenna. The measured and simulated DG results for the far-field radiation patterns closely align with the target value of 10 dB.

The radiating elements are located close to one another and are used simultaneously in the MIMO antenna architecture, however, the interference performance is also influenced by this configuration of radiating antennas which can be calculated by TARC (Total Active Reflection Coefficient) figure 9(c) displays the TARC values, which are below 10 dB. Channel capacity loss (CCL) is a critical factor in evaluating the effectiveness of a MIMO antenna in high-data-rate communication systems, as it directly impacts the ability to transmit signals and information reliably and without distortion. Figure 9(d) demonstrates that the CCL value is lower than the practical limit (0.4 bits/s/Hz), indicating that it meets the requirement for communication. In figure 9(e) shows mean effective gain (MEG) of the antenna, when powered by ports 1 and 2 respectively, shows a consistent value of below 3 dB across the entire frequency band. The MEG1/MEG2 is nearly one. These findings indicate that the radiating elements in the MIMO design maintain a stable gain performance.

Table 1. Comparisons of the decoupled antenna pair for recent literature.

References	Element	Decoupling mechanism	−10 dB S_{11}	Size of antenna	Isolation	Distance between feed	ECC	DG
[13]	8	Self	3.3–4.2	150 × 75 × 0.8	>10.5	9 mm	<0.2	—
[14]	4	Coupled CPW	3.7–7.3	—	>20	5.6 mm	<0.03	—
[15]	2	—	1.46–1.52	120 × 120 × 0.3	>16.3	2 mm	—	—
[16]	8	Asymmetrically mirrored	3.4–3.6	10 × 7 × 1	>10		<0.1	—
[17]	4	Orthogonal mode	3.4–3.6	25 × 7 × 1.5	>20.1	—	<0.008	—
[18]								
	4	—	4.3–6.45	100 × 60 × 0.76	> 15	15 mm	< 0.004	—
[19]	4	Parasitic resonant strip	2.6–2.8, 3.4–3.6	140 × 70 × 9.5	>20	—	—	—
[20]	4	Orthogonal mode	4.8, 7.7	81 × 87 × 1.6	>20	—	< 0.15	>9.9
[21]	4	Parasitic Meandering Resonator	3–5	65 × 60 × 1.6	>15	—	<0.05	>9.7
[22]	4	orthogonal	3.5, 5.9	65 × 65 × 1.6	>20	—	< 0.01	>9.9
[24]	2	Defected ground	2.99–3.61, 4.53–4.92	62 × 25.6 × 1.524	<16	—	—	—
[25]	2	Slotted ground plane	2.4,3.5,5.5	48 × 3 1× 1.6	<20	—	<0.003	>9.9
Proposed	**2**	**Self**	**2.5–2.8,3.4–3.9,4.3–4.5**	**70 × 31 × 1.6**	**>22**	**10 mm**	**<0.04**	**>9.9**

5. SAR analysis

The SAR analysis set-up is displayed in figure 10, phantom model consisting of skin 2 mm,fat 3.5 mm,muscles 10 mm and bone 10 mm [23] is placed at the distance of 3 mm from the antenna. The maximum value of SAR is 0.497 W kg^{-1} at 100 mW which is less than the prescribed 1.6 W kg^{-1}. Therefore the antenna is suitable as well as safe to be used in near human body for 5G communication system.

These findings indicate that the two MIMO antennas with self-isolation are effectively decoupled from each other, making them well-suited for applications involving MIMO technology. To validate the design's novelty, the proposed design is compared to previous works, as seen in table 1. From the table 1 it's clear that our proposed antenna have better isolation, compact size. Similarly ECC & DG is comparable with the other reference. The proposed antenna is the ideal design for triple-band wireless applications since it offers outstanding performance and a small footprint.

6. Conclusion

The efficient development, exploration, and assessment of the design have resulted in a dual-port MIMO antenna that meets the bandwidth needs of both the Sub-6 GHz and 5G. The dual-pair MIMO antenna demonstrates operation across three bands with impedance bandwidths: 2.5 GHz to 2.8 GHz, 3.4 GHz to 3.9 GHz, and 4.3 GHz to 4.5 GHz. Additionally, it achieves a peak gain of 1.9 dBi,3.77 dBi and 2.9 dBi. To make sure that the MIMO antennas closely spaced have an effective mutual coupling of over −21 dB, a self-decoupling structure is implemented antenna. The ECC value of 0.04 and DG measurement of greater than 9.98 dB have been established to comfortably meet practical standards. The dual-port MIMO antenna can be seamlessly integrated into contemporary Sub-6 GHz 5G improved diversity analysis, and compact design. Furthermore, it is highly suitable for constructing extensive antenna arrays that improve the quality of wireless communication transmission and reception. Additionally, its design simplifies and enhances effortless internet accessibility.

Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors.

Conflict of interest

Authors do not have any conflict of interest.

ORCID iDs

Parul Kansal https://orcid.org/0000-0002-1825-4928

References

[1] Jensen M A and Wallace J W 2004 A review of antennas and propagation for MIMO wireless communications *IEEE Trans. Antennas Propag.* **52** 2810–24
[2] Singhal S S, Kanaujia B K, Singh A and Kishor J 2020 Dual-port MIMO dielectric resonator antenna for WLAN applications *Int. J. RF Microw. Comput.-Aided Eng.* **30** e22108
[3] Sim C-Y-D, Dhasarathan V, Tran T K, Kulkarni J, Garner B A and Li Y 2023 Mutual coupling reduction in dual-band MIMO antenna using parasitic dollar-shaped structure for modern wireless communication *IEEE Access* **11** 5617–28
[4] Malviya L, Panigrahi R K and Kartikeyan M V 2017 MIMO antennas with diversity and mutual coupling reduction techniques: a review *Int. J. Microw. Wirel. Technol.* **9** 1763–80
[5] Xiaoming C, Zhang S and Li Q 2018 A review of mutual coupling in MIMO systems *IEEE Access* **6** 24706–19
[6] Padmanathan S *et al* 2022 Compact multiband reconfigurable MIMO antenna for Sub- 6GHz 5G mobile terminal *IEEE Access* **10** 60241–52
[7] Addepalli T and Anitha V R 2020 Compact two-port MIMO antenna with high isolation using parasitic reflectors for UWB, X and Ku band applications *Progress In Electromagnetics Research* C **102** 63–77
[8] Ranjan P, Patil M, Chand S, Ranjan A, Singh S and Sharma A 2020 Investigation on dual-port printed MIMO antenna with reduced RCS for C-band radar application *Int. J. RF Microw. Comput.-Aided Eng.* **30** e22092
[9] Kulkarni J, Desai A and Sim C-Y-D 2021 Two-port CPW-fed MIMO antenna with wide bandwidth and high isolation for future wireless applications *Int. J. RF Microw. Comput.-Aided Eng.* **31** e22700
[10] Dou Y, Chen Z, Bai J, Cai Q and Liu G 2021 Two-port CPW-Fed dual-band MIMO antenna for IEEE 802.11 a/b/g applications *Int. J. Antennas Propag.* **2021** e5572887

[11] Li H and Lau B K 2016 MIMO systems and antennas for terminals *Handbook of Antenna Technologies* ed Z N Chen (Springer) 2347–88
[12] Zhang Z 2017 *Antenna Design for Mobile Devices* (Wiley)
[13] Sun L, Li Y, Zhang Z and Wang H 2020 Self-Decoupled MIMO antenna pair with shared radiator for 5G smartphones *IEEE Trans. Antennas Propag.* **68** 3423–32
[14] Sun L, Li Y, Zhang Z and Wang H 2020 Antenna decoupling by common and differential modes cancellation *IEEE Trans. Antennas Propag.* **69** 672–82
[15] Zhou Z, Ge Y, Yuan J, Xu Z and Chen Z D 2023 Wideband MIMO antennas with enhanced isolation using coupled CPW transmission lines *IEEE Trans. Antennas Propag.* **71** 1414–23
[16] Srivastava K, Kumar A, Kanaujia B K, Dwari S and Kumar S 2019 A CPW-fed UWB MIMO antenna with integrated GSM band and dual band notches *Int. J. RF Microwave Comput. Aided Eng.* **29** 2–6
[17] Thangarasu D, Palaniswamy S K and Thipparaju R R 2023 Quad port multipolarized reconfigurable MIMO antenna for Sub 6 GHz applications *International Journal of Antennas and Propagation* **2023** 1–17
[18] Wong K L, Tsai C Y and Lu J Y 2017 Two asymmetrically mirrored gap-coupled loop antennas as a compact building block for an eight-antenna MIMO array in the future smartphone *IEEE Trans. Antennas Propag.* **65** 1765–78
[19] Chang L, Yu Y, Wei K and Wang H 2019 Polarization-orthogonal co-frequency dual antenna pair suitable for 5G MIMO smartphones with metallic bezels *IEEE Trans. Antennas Propag.* **67** 5212–20
[20] Mohammad Saadh A W, Ashwath K, Ramaswamy P, Ali T and Anguera J 2020 A uniquely shaped MIMO antenna on FR4 material to enhance isolation and bandwidth for wireless applications *AEU—International Journal of Electronics and Communications* **123** 153316
[21] Li G, Zhai H, Ma Z, Liang C, Yu R and Liu S 2014 Isolation-improved dual-band MIMO antenna array for LTE/WiMAX mobile terminals *IEEE Antennas Wirel. Propag. Lett.* **13** 1128–31
[22] Sujanth Narayan K G, Baskaradas J A and Rajesh Kumar D 2021 Design of a CPW-Fed compact MIMO antenna for next generation vehicle to everything (V2X) communication *Wirel. Pers. Commun.* **120** 2179–200
[23] Sharma N, Kumar A, De A and Jain R K 2022 Isolation enhancement using CSRR slot in the ground for compact two-element textile MIMO antenna *Work* **2** 16
[24] Sharma P, Tiwari R N, Singh P and Kanaujia B K 2022 Dual-band trident shaped MIMO antenna with novel ground plane for 5G applications *AEU-International Journal of Electronics and Communications* **155** 154364
[25] Bayarzaya B, Hussain N, Awan W A, Sufian M A, Abbas A, Choi D and Kim N 2022 A compact MIMO antenna with improved isolation for ISM, sub-6 GHz, and WLAN application *Micromachines* **13** 1355

Check for updates

Received: 18 September 2023 | Revised: 3 February 2024 | Accepted: 20 July 2024

DOI: 10.1111/exsy.13692

ORIGINAL ARTICLE

Expert Systems WILEY

Underwater image enhancement using contrast correction

Nishant Singh | Aruna Bhat

Department of Computer Science & Engineering, Delhi Technological University, Delhi, Delhi, India

Correspondence
Aruna Bhat, Department of Computer Science & Engineering, Delhi Technological University, Delhi 110042, India.
Email: aruna.bhat@dtu.ac.in

Abstract

Light-induced degeneration of underwater images occurs by physical features of sea-water. According to the wavelength of the colour spectrum, light reduces intensity significantly when it moves through water. The greatest wavelength of light that is visible gets absorbed first. Red and blue absorb the most and least, respectively. Because of the reducing consequences of the light spectrum, underwater images having poor contrast can be obtained. As a result, the crucial data contained inside these images will not be effectively retrieved for later analysis. The recent research suggests a novel approach to enhance the contrast while decreasing noise in underwater images. The recommended approach involves image histogram transformation using two significant colour spaces, Red-Green-Blue (RGB) and Hue-Saturation-Value (HSV). The histogram of the dominant colour channel (blue channel) in the RGB colour model is extended towards the lower level, containing a maximum limitation of 95%, while the inferior red colour channel has been extended towards the upper side, containing a minimum limitation of 5%. During the entire dynamic range, the green colour channel having the dominant and inferior colour channels expands in each direction. The Rayleigh distribution has been utilized for developing various stretching actions within the RGB colour space. The image has been converted to the HSV colour space, having the S and V elements adjusted within 1% of their minimum and maximum values. The suggested approach is examined in both qualitative and quantitative analysis. According to qualitative analysis, the recommended approach substantially boosts image contrast, lowers its blue and green effect, and minimizes over-enhanced and under-enhanced sections in the final resultant underwater image. The quantitative examination of 500 large scale underwater images dataset reveals that the suggested technique generates better results. The dataset images are grouped into small fish images, blue coral images, stone wall images, and coral branch images. The quantitative examination of all these four groups have been evaluated and shown. The average mean square error, peak signal to noise ratio, underwater image quality measurement, and underwater colour image quality evaluation values of dataset images are 76.69, 31.25, 3.85, and 0.64, respectively. These values of our proposed work outperform six other previous methods.

KEYWORDS
contrast stretching, HSV colour model, RGB colour model, underwater image enhancement

 © 2024 John Wiley & Sons Ltd.

1 | INTRODUCTION

The physical characteristics of underwater make underwater image processing a difficult field to work. Light absorption and dispersion effects on underwater imaging provide obstacles for scientists doing underwater research (Church et al., 2003; Gasparini & Schettini, 2003; Shamsudin et al., 2012). Images taken farther into the water, for example, turn greenish or bluish. Underwater image quality is generally influenced by variables such as poor visibility, low contrast, non-uniform illumination, bright artefacts, noise, blurring, and fading colour (Shamsudin et al., 2012). Yet, the specific physical and chemical properties of underwater environments have a significant impact on image quality, generating challenges which can be more easily handled in terrestrial imaging. Light loses intensity fast when it passes through water, depending on the wavelength of the colour spectrum. Light that can be seen gets absorbed at the longer wavelengths first in clear open water (Schechner & Karpel, 2004; Schettini & Corchs, 2012). Red, which is the most light-absorbing colour, is decreased to one-third of its original brightness after 1 m and is completely lost by 4–5 m (Jaffe, 1990; Trucco & Olmos-Antillon, 2006). Meanwhile, blue and violet light absorb the least compared with other wavelengths. As a result, the open ocean looks a dark blue to the human eye.

The first colour component absorbed by water is red, then orange, yellow, green, and blue. Because these colour components are the least absorbed, underwater images seem blue green. Because of this colour absorption, the acquired underwater images have low colour and contrast. Important image information is also lost. The use of image processing and computer vision is becoming increasingly crucial in recovering the contrast, colour, and missing information from images.

Thus, retrieving useful information from underwater images necessitates efficient methods for correcting colour, improving clarity, and addressing background scattering and blurring, all of which are the goal of image restoration and enhancement methods. These are especially difficult because of the complicated underwater surroundings because images are damaged by the impact of water fluidity, light absorption, and scattering, all of which can fluctuate dramatically (Wang et al., 2019).

A frequently used approach in underwater image enhancement is contrast enhancement that improves contrast correction as well as produces a wider dynamic range. The implementation of this approach has gained significant interest in research in recent years, even though it is still a young topic when compared to strategies for enhancing normal images. A lot of academics have utilized different techniques to boost the underwater image feature and gathered important information from the images. Numerous publications are published based on subjective human colour vision models (Schettini & Corchs, 2012; Trucco & Olmos-Antillon, 2006), which are discussed in the next section.

The current work provides a strategy for enhancing the contrast of underwater images to increase contrast among images and noise reduction. The remaining part of the work is structured as follows: Section 2 covers previous researchers' current methodologies; Section 3 specifies the suggested method; Section 4 explained the qualitative analysis and quantitative analysis of data; and in last Section 5 gives the conclusion of the research.

2 | LITERATURE REVIEW

Underwater image processing may be defined in two ways: as an image restoration approach and as an image-enhancing technology (Jaffe, 1990; Schettini & Corchs, 2012). Image restoration is the process of restoring a deteriorated image by employing an instance of the actual diminished image creation. These approaches are tough, and they need a wide range of model parameters that characterize water turbidity, including attenuation and diffusion coefficients (Schettini & Corchs, 2012). The underwater image improvement approach employs subjective and objective standards to create a better visually appealing underwater image without relying on physical model. Previous researchers' techniques generally include changing image pixel values to increase image contrast and colours.

Authors in paper (Pisano et al., 1998; Rizzi et al., 2004; Singh & Bhat, 2023; Ye, 2009; Yeganeh et al., 2008; Zhou et al., 2020) investigated the physical consequences of vision deterioration and suggested an image recovery approach based on two images obtained with a polarizer at different orientations. Unsupervised digital image colour equalization having both local and global effects was proposed by Rizzi et al. (Hasibuan et al., 2021). Trucco and Olmos-Antillon (Singh & Bhat, 2021) created a self-tuning image restoration filter based on the well-known underwater image creation model given (Ding et al., 2022; Elnashef & Sagi, 2019; Hu et al., 2020; Wang et al., 2022). According to these experts, underwater images may be expressed using a linear superposition consisting of three elements: the direct light, forward-scattered light, and backward-scattered light. Whereas this physical-based method needs a high level of computational power, resulting in a longer execution time.

For digital colour images, Naim and Isa (Li et al., 2012) suggested pixel distribution shifting colour correction (PDSCC) to adjust the white reference point and verify that it is achromatic. This PDSCC method is the recent contrast correction method for underwater image enhancement. This approach expands the three-dimension rotation matrix (3-DMAT) method via changing the two-dimension two colour channel plane. Due to its simplicity and speed of execution, 3-DMAT was the first 3D rotational technique deployed. The 3D pixel distributions rotational technique is used through rotating a pixel three times that is (yaw, pitch, and roll) utilizing three separate rotation angles. These two 3D rotation colour correction approaches in the PDSCC method are especially developed to alter an image's pixel distribution so that the surrounding unimportant light is muted or deleted from the resultant image. The procedure involves shifting the image's pixel distribution onto the diagonal plane in the 3D R-G-B

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

colour channel. Instead of 3D R-G-B colour channel, the shifting mechanism was therefore performed on 2D two colour channel planes. These 2D colour planes are built using the 3D RGB colour model, namely the red-green plane, the red-blue plane, and the green-blue aircraft (Li et al., 2022; Wang et al., 2021). This procedure guarantees that the pixel distribution of the surrounding light is achromatic. The approach corrects the visual colour from the viewer's perspective to appear more realistic. The approach also improves image saturation but does not substantially enhance image contrast.

Hitam et al. (Trucco & Olmos-Antillon, 2006) suggested a technique for under-water images dubbed mixture contrast limited adaptive histogram equalization (CLAHE-Mix). This approach seeks to decrease considerable noise created by CLAHE to make further underwater image processing easier. Hitam et al. (Trucco & Olmos-Antillon, 2006) independently applied CLAHE to the image in RGB and Hue-Saturation-Value (HSV) colour models. CLAHE is initially utilized on an underwater image in the R-G-B colour space, which decomposes the image into its individual channels. The procedure is carried out to each of the colour channels independently, after which the colour channels combine to create an output image. The exact same original image within the RGB colour space is translated to the HSV colour space first, and then the image is split to its corresponding channels. CLAHE is applied to the image's S and V elements, and the resulting channels are then combined then turned again to the RGB colour space. In both colour models, the pixel distribution has been set to correspond to the Rayleigh distribution used by the CLAHE process. These methods for using CLAHE in the RGB and HSV colour models result in two separate images: one treated with CLAHE in the RGB colour model, known as CLAHE-RGB model, whereas the other treated with CLAHE in the HSV colour model, known as CLAHE-HSV model. Using the Euclidean norm, these images are combined to generate a contrast correction underwater image with minimal noise. For certain situations, this approach creates images with higher level of noise as compared to the traditional CLAHE method. Furthermore, the produced visual of underwater image occasionally turns greenish.

Iqbal et al. (Alenezi et al., 2021; Ulutas & Ustubioglu, 2021) introduced an integrated colour model (ICM) and an unsupervised colour correction model (UCM). The output image in the RGB colour model is extended throughout the complete dynamic range in Alenezi et al. (2021). After that, the image is transformed to the Hue-Saturation-Intensity (HSI) colour model. The S and I elements are used with contrast stretching in this colour model, after this the underwater image is then transformed again into RGB colour channel to provide a higher-quality underwater image. In Ulutas and Ustubioglu (2021), Iqbal et al. reduced the colour cast by modifying two colour channels (red and green) according to the von Kries theory. The RGB colour model is then corrected for contrast. The image histograms have been expanded to one or each side depending on the lowest and highest values for every single channel which are calculated at 0.2% and 99.8% of the original histogram's lowest and highest values, respectively. The contrast is stretched towards the higher end of the histogram using the channel with the lowest intensity that is, Red. The colour channel green's middle intensity value is applied with contrast stretching on both sides, while the dominant colour cast with the greatest intensity value (blue) is applied with contrast stretching towards the bottom side (Ulutas & Ustubioglu, 2021). The image is then transformed to the HSI colour space for underwater images, with S and I elements extended on lower as well as from top sides. Despite knowing that all contrast in the final output underwater images is improved with both procedures, the result underwater images are scarcely changed. Particularly, the images within certain areas are under enhanced, resulting in too dark portions. Another disadvantage of these approaches is that they create images with a lot of noise.

Eustice et al. provided a way for extending the MATLAB image processing toolbox for optical image processing and mosaicing methods in Schechner and Karpel (2005). Eustice et al. (Schechner & Karpel, 2005) mentioned the following MATLAB image processing toolbox extensions and implementations: (1) Fourier-based methods for scale, rotation, and translation recovery, that offer robustness against dissimilar image regions; (2) local normalized correlation for image registration; and (3) multi-resolution pyramidal blending of images to form a composite seamless mosaic. Eustice et al. used CLAHS as the pre-processing stage for the initial expansion. CLAHS is an expansion of CLAHE, which was created for medical images. CLAHS works by splitting the original image into equal portions of contextual areas, then applying a histogram specification to each region. For each region, the grey level transform defined by $g = T(f)$ is computed, which translates the local grey level histogram to a perfect grey level distribution. Eustice et al. shown in their investigation how the Rayleigh distribution constitutes the optimal pixel distribution for representing underwater images. The clipping limit, which is described as a multiple of the average histogram contents that limits the contents of any specific histogram bin, is also used to control image contrast. The extra pixels are then spread evenly across the left-over bins. The primary disadvantage of CLAHS is that it cannot be conducted automatically since a few parameters based on image properties are required. CLAHS improves the underwater image while degrading the colour quality of the final resultant underwater image. Overall noise generated in the underwater image is high.

A basic physical framework for light propagation and a set of high-quality measurements are combined to create a single image enhancement approach as given in Barros et al. (2018). The model eliminates the artefacts and degradation caused by attenuation, forward scattering, and backward scattering. To improve colour correction and contrast, a multi-scale fusion technique for underwater image enhancement is being investigated in Zhang et al. (2018). The recovered image is first obtained using the optical underwater model. The repaired version's white balance and contrast boosted image is then acquired and merged.

One major issue with underwater image found by prior studies is the fact that the contrasting effect of the resultant image stays low, resulting in over-enhanced and under-enhanced regions in the final resultant underwater image. The method's complexity leads to image distortion along with significant computing usage, resulting in higher execution time. The latest research proposes an easier approach that focuses on

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

improving the contrast of underwater images and decreasing image noise. The suggested technique also happens to be minimal in computational usage, which reduces processing time.

3 | RAYLEIGH DISTRIBUTION INTEGRATED COLOUR MODEL

The suggested technique modifies and extends ICM (Alenezi et al., 2021) and UCM (Ulutas & Ustubioglu, 2021) introduced by Iqbal et al. ICM (Alenezi et al., 2021) is a contrast adjustment method that uses histogram adjustment in the RGB and HSI colour models, in which the original image's histograms are extended throughout the whole dynamic range. The von Kries theory is used by UCM (Ulutas & Ustubioglu, 2021) by balancing the colour channels depend on the dominant colour cast channel. According to prior research (Ulutas & Ustubioglu, 2021), the blue and red colour channels are the dominant and inferior colour channels, respectively. The green channel tends to be the colour channel which falls among the red and blue channels. When normalizing the colour channels, the resulting histograms are stretched by 0.2% between both the lowest and highest grey level values corresponding to the initial input histogram. The original input image's histograms were acquired between 0.2% and 99.8%, and the results are then extended to cover the complete width of the final output grey level. Depending on the colour channels, the histogram is extended on one or both sides. The image is then transformed to the HSI colour model, which stretches the S and I elements throughout the dynamic range.

The primary disadvantage of these approaches is that they create a lot of noise in the output image. This issue demonstrates how the ICM and UCM principles increase instability and signal interference which is unrelated to the actual image. The brightness level of the pixels in the final image is altered because of this procedure.

A few changes are made to the image outputs of ICM (Alenezi et al., 2021) and UCM (Ulutas & Ustubioglu, 2021) in this investigation. As mentioned earlier, both approaches employ histogram stretching produced around 0.2% using the input histogram's lowest and highest values, expanding over the complete dynamic range in the final resultant underwater image. The above stretching technique has been used across every image channel. Stretching a histogram to cover all the dynamic range of the resultant image will almost certainly result in images with lower intensity and more under-enhanced regions. As a result, rather than stretching the full range, the suggested approach confines the stretching within the inferior and dominant colour that is, red and blue, respectively, because these channels have a substantial impact on the underwater image. Previous research (Alenezi et al., 2021; Ulutas & Ustubioglu, 2021) found that underwater images had the highest blue colour percentage and a lowest red colour percentage.

To improve the influence on the total intensity, the stretching procedure within the inferior red colour channel has been adjusted between 5% and 100% in the resultant histogram towards a 255-intensity value. A minimum of 5% has been utilized to prevent image pixels that have minimal intensity, which results in under-enhanced images. According to the outcomes of Rizzi et al. (2003), a value of 5% is selected since the black point had an initial value of 5% corresponding to the brightness value. To reduce its influence on the overall intensity, the leading blue colour channel has been extended from 0% to 95% in the resultant histogram towards a 0-intensity value. Previous research (Rizzi et al., 2003) suggested that the white point required an initial brightness level of 95%; this restriction is established to prevent the impact of over-enhancing, which is notably noticeable in underwater images where the blue channel dominates. This issue may result in too bright images. The green channel (Ulutas & Ustubioglu, 2021), which is typically used as an intermediate colour channel among the low and high colour channels, is extended in both directions.

Besides limiting the parameters to the resultant underwater image, the suggested approach uses a stretching procedure to apply the Rayleigh distribution, resulting in a bell-shaped distribution of image pixels. In Trucco and Olmos-Antillon (2006), the Rayleigh distribution was shown to be the optimal pixel allocation for underwater images. The ICM and UCM algorithms just extend the image histogram without considering any distribution. This constraint is likely to result in more under- and over-enhanced regions due to the amount in image pixels between the lowest value 0 and maximum value 255. The image pixels within the Rayleigh distribution are dispersed about the centre by the brightness level, having a smaller number of pixels near the lowest and highest values.

Furthermore, the UCM method limits the input histogram at 0.2%. The original histogram covers the whole dynamic range of the produced resultant image, covering 0.2% through 99.8%. These restrictions are meant to prevent a single outlier pixel from getting a value that is either too high or too low, resulting in unrepresentative scaling (Ulutas & Ustubioglu, 2021). Stretching continues to occur over all the dynamic range in the produced resultant image at the end of the operation. However, this approach still fails to prevent creating an image that is extremely dark or extremely bright since the pixels are likely to remain scattered between the lowest and highest boundaries, resulting in over-enhanced and under-enhanced regions. As a result, the suggested strategy limits the final resultant histogram rather than the original input histogram. After imposing limitations to the final resultant histogram, the final output underwater image pixels are not going to have an intensity level whose value is either excessively low or excessively high.

Although prior research (Alenezi et al., 2021; Ulutas & Ustubioglu, 2021) utilized HSI, the output image from the previous step is translated into the HSV colour model before the S and V elements are stretched inside the range of 1%–99% in the current study. The HSV colour model was chosen based on the outcomes of another study (Trucco & Olmos-Antillon, 2006). The research has shown that incorporating contrast

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

improvement into this colour space helps eliminate over-enhanced and under-enhanced regions. The H element defines the specific colour, whether that colour is red, blue, or green. The maximum RGB colour sets the beginning point, and the colour difference decides how far it advances, often up to 60 degrees away (halfway towards the next main colour) (Trucco & Olmos-Antillon, 2006). The S element can be determined with how far away the values of RGB. If values are near together, a colour is closer to grey, whereas when they are distant, then colour gets more vivid to pure. As a result, extending the S element results in a normal-looking underwater image. However, the V element represents image brightness. Expanding the image's value throughout all the dynamic range stretches the image pixel across each intensity range within the image, minimizing the amount in image pixels within a particular value. The S and V elements in the image might be easily adjusted in the HSV colour model. These are the primary factors which lead an image to end up over-saturated or under-saturated. Images that have minimal S and V values may be under-saturated. This implies the fact that the pixel value approaches its lowest value, the resultant image gets darker, but as the pixel value approaches its highest value, the resultant image gets overly bright. Setting the limitations results in a more natural-looking output image.

Previous research (Trucco & Olmos-Antillon, 2006) found that both the CLAHE-Mix and CLAHS approaches are based on the CLAHE principle. Instead of CLAHE, the suggested approach in this work uses a modification of histogram stretching. Histogram stretching is used in a different way, as will be detailed later. CLAHE was originally intended to improve medical imaging, according to McGlamery (1979). Underwater image has been divided into smaller segments in the CLAHE-Mix or CLAHS procedures, and then processed with histogram specifications; with the individual pixel distribution in the final image is altered in accordance with Rayleigh distribution. These segments are merged afterwards through bilinear interpolation to give an improved output underwater image. But the CLAHE idea employs an additional complicated technique to CLAHS for the purpose to generate a final output underwater image. This CLAHE method generates images having greater brightness compared with the CLAHS method; however, it equally generates plenty of noise. Furthermore, a few criteria for the various features of underwater images must be defined. These characteristics, such as the clip limit, the Rayleigh distribution parameters, and the overall number of divisions, can differ among various underwater images. To improve the poor contrast within an underwater image, for example, only a small amount of clip restrictions is required. But when the initial image contains minimal brightness, the limit must be raised more unless the output image's contrast cannot be greatly enhanced. The total number of divisions also impacts the output image's contrast level. These specifications demonstrate that the CLAHE method cannot be conducted automatically for all sorts of underwater images.

As a result, instead of CLAHE, the suggested approach uses histogram stretching on underwater images. When compared with the CLAHE method, histogram stretching offers simplicity and substantially reduces image distortion in the final resultant underwater image. According to previous research (Trucco & Olmos-Antillon, 2006), the Rayleigh distribution is best suited for underwater imaging. Furthermore, to the stretching operation, the suggested technique shapes the image's pixel distribution to match such distribution.

4 | PROPOSED WORK

The flowchart of the suggested method implementation is shown in Figure 1, which starts with splitting the image colour channel. The image histograms then get incorporated within the RGB colour space using contrast stretching. The image histograms are stretched in the lower, upper as well as in both directions depending on the colour channels. The stretching technique has been configured to adhere to the Rayleigh distribution and was restricted to a specific range of values. The image then undergoes conversion to the HSV colour space, with S and V elements extended among 1% and 99%. Once all these steps are completed, the individual channels have been combined, and then the image gets transformed again to its RGB colour space. At the end, a better contrast resultant underwater image will be generated.

4.1 | RGB colour space contrast correction

As illustrated in Figure 1, the RGB underwater image is separated in red, green, and blue channels. Histogram stretching is then applied to the channel histograms. The intensity distribution in original histograms for underwater images is typically around in each brightness range, indicating that the histograms have not been properly dispersed throughout the whole range of intensity value. As a result, the brightness of the underwater image is raised through expanding the histogram of the image channel. Stretching in histogram improves the distribution of the pixels in an image throughout the whole dynamic range of the intensity levels. However, the suggested technique restricts the resultant intensity level to 5% of the lowest and maximum limitations. These constraints are used to minimize the impact of under- and over-exposed regions in the image. For histogram stretching, Equation (1) is used and given by.

$$P_{out} = (P_{in} - i_{min})\left(\frac{o_{max} - o_{min}}{i_{max} - i_{min}}\right) + o_{min} \quad (1)$$

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

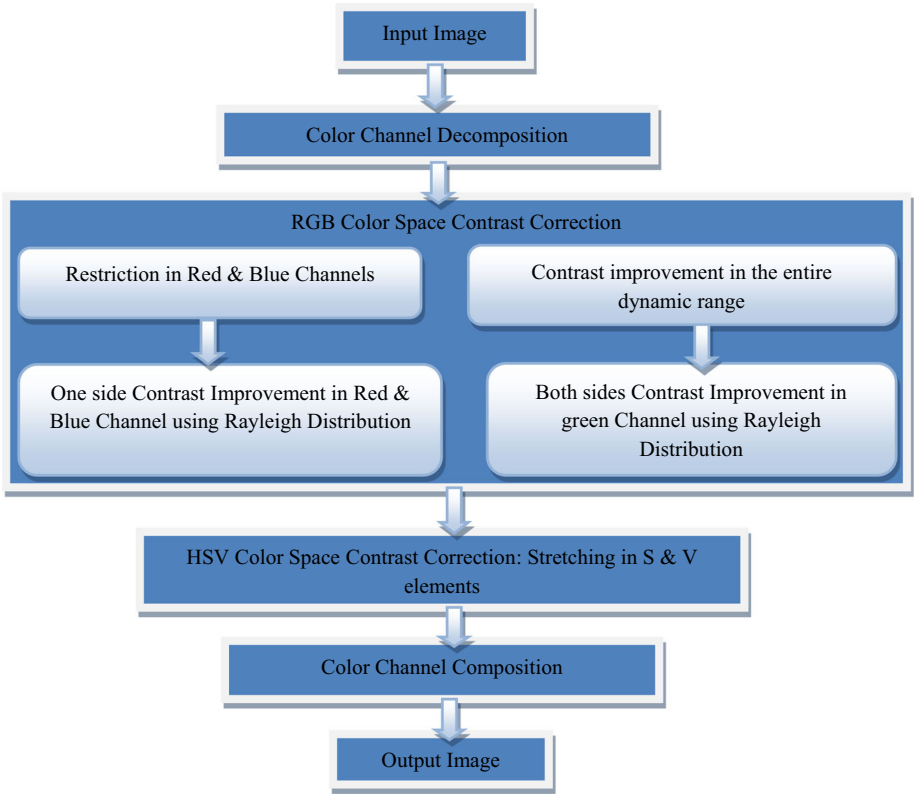


FIGURE 1 Flowchart of our proposed model.

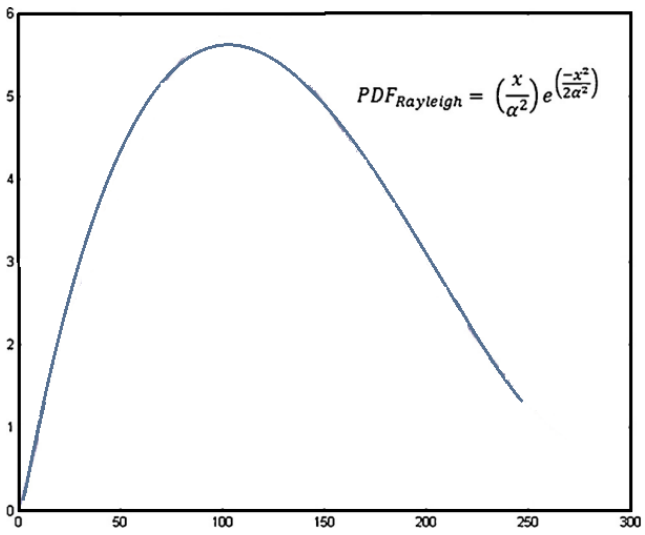


FIGURE 2 Rayleigh Distribution which contains maximum number of pixels in the middle range of intensity value.

where P_{in} represent the input pixels and P_{out} represent the output pixels, and i_{min} represent the lowest intensity level value in the input images, i_{max} represent the highest intensity level value in the input images, o_{min} represent the lowest intensity level value in the output images and o_{max} represent highest intensity level values in the output images. Along with histogram stretching, distribution of the pixel has been assigned in accordance with the Rayleigh distribution, resulting in a bell-like shape structure with most pixels clustered around central range of the intensity value as shown in Figure 2. The top and bottom ends in the intensity range histogram contain the fewest pixels.

Equation (2) represents the probability distribution function (PDF) of the Rayleigh distribution whereas Equation (3) represents the cumulative distribution function (CDF) of the Rayleigh distribution, where x is the input pixel and specifies Rayleigh distribution parameter of 0.4. This

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

variable is set to its default value and was earlier utilized in Trucco and Olmos-Antillon (2006). This number is employed in the proposed approach for all sorts of underwater image graphs, such as those provided by

$$PDF\ Rayleigh = \frac{x}{a^2} e^{-x^2/2a^2}, x \geq 0, a > 0 \tag{2}$$

$$CDF\ Rayleigh = \int_0^x \frac{x}{a^2} e^{-x^2/2a^2} dx = 1 - e^{-x^2/2a^2}\ for\ x \in (0, \infty) \tag{3}$$

The suggested approach applies histogram stretching on the image histogram. The Rayleigh distribution is used to distribute the image pixels. As a result, Equation (1) has been incorporated in Equation (2) to give the Stretched-Rayleigh distribution formula. The integration result is provided in Equation (4).

$$Stretched - Rayleigh = \frac{\left[(P_{in} - i_{min})\left(\frac{o_{max} - o_{min}}{i_{max} - i_{min}}\right)\right]}{a^2} . e^{\frac{\left[(P_{in} - i_{min})\left(\frac{o_{max} - o_{min}}{i_{max} - i_{min}}\right) + o_{min}\right]^2}{2a^2}} \tag{4}$$

The proposed technique additionally applies the stretching processes' restrictions. The next subtopic goes into further information about these constraints.

4.1.1 | Contrast correction in red and blue channels

Within a specified range, the contrast correction method of the resultant image is used. Instead of restricting the input image channels, the suggested method restricts the output image. For underwater images, the red and blue colour channels usually represent the smallest and largest intensity value respectively. As a result, the histogram stretching associated with each of these colour channels occurs just on a single side, shown in Figures 3 and 4. The red colour channel, shown in Figure 3, is extended towards the top side, while the blue colour channel has been expanded towards the bottom side, shown in Figure 4.

When the lowest intensity value within the original histogram has become larger than 5% compared with the lowest intensity level that is 0, then the lowest stretching value for the final output image ($o_{r,min}$) is adjusted in accordance with the lowest intensity value in original histogram ($i_{r,min}$). Otherwise, the produced image's lowest intensity value equals 5% from 0, so it is 12.75 approximately 13. This could have been stated mathematically as

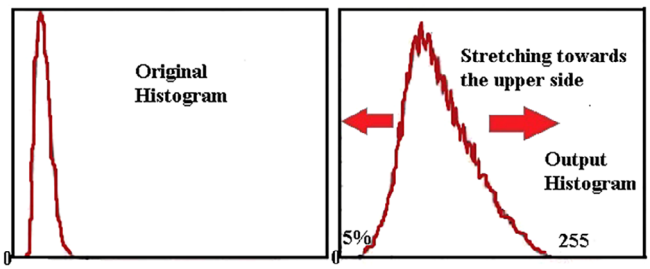


FIGURE 3 Red colour channel histogram stretching from upper side with minimum 5% from 0 intensity value.

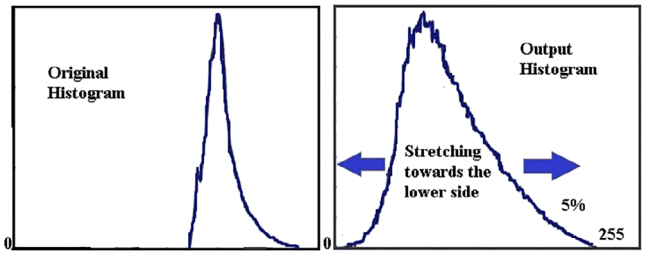


FIGURE 4 Blue colour channel histogram stretching from lower side with minimum 5% from 255 intensity value.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$o_{r,min} = \{ i_{r,min}\, i_{r,min} > 5\%\text{of } I \tag{5}$$

$$\{ 5\%\text{of } I \text{ otherwise}$$

where I denote the image's entire dynamic range, and it is 255 (unsigned integer 8). The output image's highest intensity value ($o_{r,max}$) has been adjusted to 100% based on the overall intensity value, which is in this case is 255. In Equation (4), the parameters for $o_{r,min}$ and $o_{r,max}$ for the red channel have been replaced by the value of o_{min} and o_{max}, respectively. Figure 3 depicts the criteria used for the red colour channel.

The blue channel often dominates the strength of colour channels in underwater image graphs (Ulutas & Ustubioglu, 2021), with most pixels spread at high intensities (Figure 4). As a result, the stretching in the blue channel will be performed to the lower side, implies that if the highest intensity value of the original underwater image's blue channel ($i_{b,max}$) has less than 95% of the overall intensity value, the highest value for the stretching's output histogram ($o_{b,max}$) will correspond to the highest intensity value for the resultant underwater image ($i_{b,max}$). Alternatively, when the highest possible intensity value in the blue channel in the final resultant image ($i_{b,max}$) has been above 95% of the total intensity value, then the output histogram's highest intensity value ($o_{b,max}$) has been set to 95% of the total intensity value, resulting in 242.25 approximately 242. At the same time, the output image's lowest intensity value ($o_{b,min}$) has been set to 0. In the blue channel, the parameters of o_{min} as well as o_{max} in Equation (4) have been replaced by $o_{b,min}$ and $o_{b,max}$, respectively. The criteria of calculating the greatest value for the resultant output histogram of blue channel are represented mathematically by

$$o_{b,max} = \{ i_{b,max}\, i_{b,max} < 95\%\text{of } I \tag{6}$$

$$\{ 95\%\text{of } I \text{ otherwise}$$

where I denote the image's entire dynamic range, which is 255. Figure 4 depicts the blue channel's stretching process and limitations.

4.1.2 | Contrast correction of the green channel

The contrast adjustment approach is performed to both sides of the histogram using the intermediate colour channel (green), which has an intensity value between the lowest (red channel) and highest (blue channel) (Ulutas & Ustubioglu, 2021). This channel's limit is set to [0%, 100%]. As a result, the lowest and highest values for the green channel output image remain 0 and 255, respectively. The Stretched-Rayleigh formula of Equation (4) could have been reduced to Equation (7) using the proposed approach. Figure 5 depicts the stretching mechanism as well as the boundaries within the green channel histogram for a better understanding.

$$Stretched - Rayleigh = \frac{255(P_{in} - i_{min})}{a^2(i_{max} - i_{min})} . e^{\frac{-[255(P_{in} - i_{min})]^2}{2a^2(i_{max} - i_{min})^2}} \tag{7}$$

Once histogram stretching is complete, the image's contrast is further rectified. The resulting contrast modification technique is explained in the next step.

4.2 | HSV colour space contrast correction

After the contrast correction in the RGB colour model, the image has been converted into the HSV colour model, as shown in Figure 1. The conversion to the HSV colour model has previously been detailed in Trucco and Olmos-Antillon (2006). Within the range of [1%, 99%], S and V

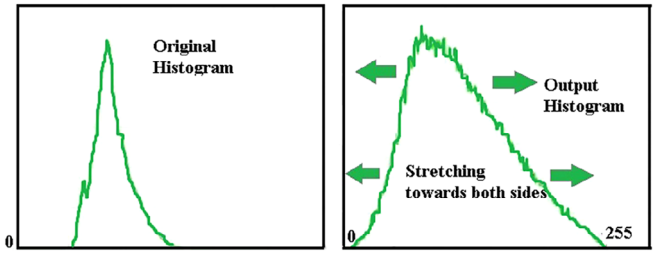


FIGURE 5 Green Colour Channel Histogram Stretching from both lower and upper side within range 0–255 intensity value.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

elements have been separated and performed via linear contrast stretching, as stated in Equation (1). As a result, the output image has a normalized limit range within [0.01, 0.99]. S and V are image elements that determine the image's saturation and brightness (Trucco & Olmos-Antillon, 2006). To decrease the image's over-saturation and under-saturation, the lowest and highest stretching values has been adjusted to 1% from both the lower and higher boundaries. The limitation is imposed as an outcome of an experiment carried out in another study (Shamsudin et al., 2012). Visual inspection of (Trucco & Olmos-Antillon, 2006) reveals that the resulting image, which stretch the entire dynamic range of S and V, is under- and over-saturated, that is, too dark, and too bright. The restriction for S and V elements in the HSV colour space is used in the suggested technique. The channels are merged after the S and V elements have been stretched, and the image in the HSV colour space is transformed back to the RGB colour space. The result can then be a contrast-enhanced output image.

4.3 | Performance analysis and data sample

The suggested technique intends to improve on ICM (Alenezi et al., 2021) and UCM (Ulutas & Ustubioglu, 2021) methods provided by Iqbal et al. The image generated by ICM and UCM has better image contrast. These strategies, however, have considerable drawbacks. The most significant disadvantage is that the procedures create a lot of noise in the output image. Furthermore, the resultant image can become excessively dark or too bright, leading to under- and over-saturation. The random variation of signal strength is termed as noise (Naim & Isa, 2012). Interfering signals generated by the electronics of the captor devices can cause noise (Iqbal et al., 2007). Noise is believed to be unrelated with image, implying that there is no link between pixel and noise levels. The proposed method addresses these issues. The proposed method is intended to improve the contrast and reduce noise of underwater images.

The entropy, mean square error (MSE), peak signal-to-noise ratio (PSNR), underwater image quality measurement (UIQM), underwater colour image quality evaluation (UCIQE), and processing time are six quality metrics considered to assess the image quality. Such qualitative measures have been commonly utilized to assess image noise.

Entropy is a quantitative evaluation which could have some importance in the image comparison. Entropy is a measure of image information content that indicates the quantity of image information (Iqbal et al., 2010). It is derived by adding the multiplication of the probability of outcome multiplied with the log of the inverse of the outcome probability. Entropy, $H(x)$, is provided by Equation (8), as stated in (Iqbal et al., 2010). The probability distribution for the image is denoted by $p(x)$, and the number of grey levels is denoted by L. According to Iqbal et al. (2010), greater entropy is preferable since it indicates that the resulting images contain more information.

$$H(x) = -\sum_{x=0}^{L} p(x) Log_2 p(x) \tag{8}$$

MSE is the sum of the squared errors within the resulting image and the original images. The author's in paper (Trucco & Olmos-Antillon, 2006) computes MSE.

$$MSE = \frac{1}{M * N} \sum_{M,N} [I1(m,n) - I2(m,n)]^2 \tag{9}$$

where $I1(m,n)$ signify the original image and $I2(m,n)$ signify the enhanced images, and $M * N$ signifies image size, while m imply x and n imply y positions of the underwater image pixel, respectively.

PSNR, on the other hand, denotes the ratio of the highest potential signal to the noise. (Schechner & Karpel, 2004) provides the PSNR.

$$PSNR = 20Log10 \frac{\left(2^B - 1\right)}{\sqrt{MSE}} \tag{10}$$

where B is the number of bits per sample. Low MSE and high PSNR values suggest the best generated image.

UIQM employs a linear combination of three measures: underwater image colourfulness (UICM), underwater image sharpness measure (UISM), and underwater image contrast (UIConM) given in Equation (11), as stated in Nabavi et al. (2009). The larger the UIQM value, the better the image's visual quality in better comprehensive performance in colour, contrast, and sharpness. The application scene determines the choice of three coefficients: c_1 is used to modify underwater image colours; c_2 is used to improve underwater image clearness; and c_3 is used to improve underwater image contrast (i.e., $c_1 = 0.0282$, $c_2 = 0.2953$, $c_3 = 3.5753$).

$$UIQM = c_1 UICM + c_2 UISM + c_3 UIConM \tag{11}$$

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

UCIQE depends on the linear combination on chroma, saturation, and contrast, under the CIE-Lab paradigm given in Equation (12), as stated in Nabavi et al. (2009). It is an image quality evaluation measurement with no reference. A higher UCIQE value suggests better image quality in chroma, saturation and contrast. The weighting coefficients are denoted by c_1, c_2, and c_3 (c1 = 0.4859, c2 = 0.2745, and c3 = 0.2576).

$$UCIQE = c_1\sigma_c + c_2 con_l + c_3\mu_s \tag{12}$$

where σ_c is the standard deviation of chroma, con_l is the contrast in brightness, and μ_s indicate the average value of saturation and c_1, c_2, and c_3 denote the weighting coefficients.

Another disadvantage in the ICM method as well as in the UCM method is that they generate dense resultant underwater images as shown in Figure 6. This gloomy impression is caused by the stretching process towards the grey level's lower direction. The regions in the images clearly show the issue: expanding the histogram towards the lower intensity level reduces the value in grey level, lowering channel intensity of underwater image. The fine details of underwater images are also diminished. The impact of the blue and green illumination is kept in the resultant underwater image having very low contrast as shown in Figure 7. This blue-green light works as a 'veil' in the final resultant underwater image, making the objects difficult to perceive and distinguishable from the image background. The illumination regions have been equally over-saturated, resulting in image detail loss. The problem is depicted regions in the underwater images. The drawbacks associated with ICM and UCM techniques have been resolved with the suggested integrated colour model that uses the Rayleigh distribution approach.

This study's key contribution is that the suggested strategy may minimize noise while improving image contrast. Section 4 demonstrates how the suggested solution considerably reduces noise in the resultant image. Areas that are under- and over-emphasized are also minimized. The generated result greatly minimizes the blue and green illumination of the underwater images, leading in enhanced detection and distinction from their respective background and foreground objects.

5 | RESULTS EVALUATION AND ANALYSIS

The suggested approach is examined in both subjectively (qualitatively) as well as objectively (quantitatively). The suggested method's underwater image final outputs are compared with the ones obtained through other methods offered in earlier research in the qualitative assessment. ICM



FIGURE 6 Underwater image enhancement on Blue Coral.

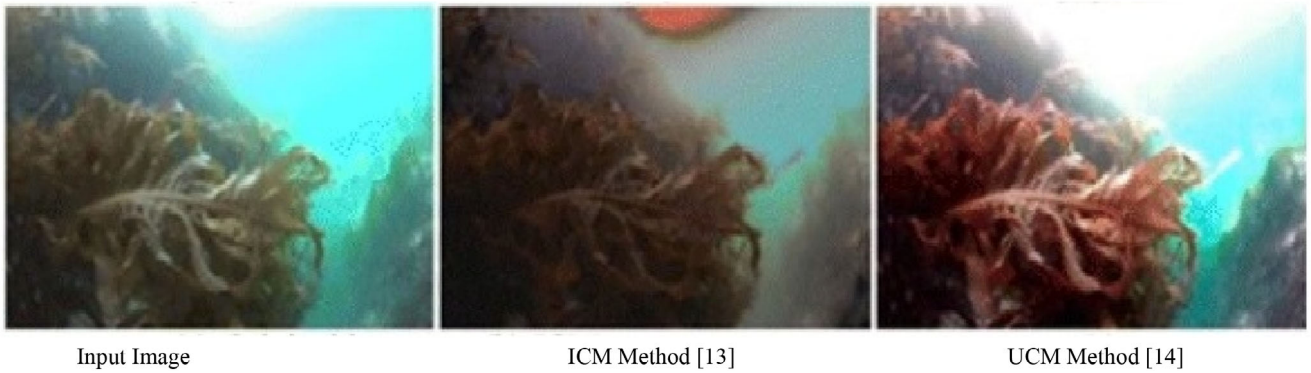


FIGURE 7 Underwater image enhancement on Coral Branch.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Alenezi et al., 2021), UCM (Ulutas & Ustubioglu, 2021), PDSCC (Li et al., 2022), conventional histogram equalization (HE) (Salmeri et al., 2001; Senthilkumaran & Thimmiaraja, 2014), CLAHE-Mix (Trucco & Olmos-Antillon, 2006), and CLAHS (Schechner & Karpel, 2005) were employed for the comparison. These methods are cutting-edge techniques which are currently been used to improve image quality, particularly underwater images. The CLAHE and HE methods are used for comparison as these are the fundamental approaches employed by experts in various histogram-modifying methods.

5.1 | Qualitative analysis

Five hundred underwater images from large scale underwater datasets have been utilized to test the suggested approach. Four underwater images were selected from among the 500 for qualitative analysis. The outcomes of all approaches are depicted in Figures 8–11. The capacity of the tested technology to improve image contrast is one of the major requirements for visual observation. The potential of such approaches to decrease over-enhanced and under-enhanced regions, where the final resultant underwater images are excessively bright or excessively dark, is also seen. The influence from the blue and green illumination is additionally visible within the final resulting images, and the suggested approach would have been suitable to decrease the effect of non-uniform blue and green illumination. As described in Li et al. (2022), the suggested approach can increase image contrast by reducing the lighting impact. In the figures, circles are utilized to indicate the differences among the compared images.

CLAHE-Mix and CLAHS operate on a local level by separating underwater images into smaller portions. The Rayleigh distribution can be used to determine the pixel distribution. The method CLAHS is referenced in MATLAB via a function's name 'adapthisteq'. These techniques include various factors and alter based on the underwater images, demonstrating how they cannot be conducted automatically. These settings have been set in accordance with the default values recommended by the authors in their publications (Schechner & Karpel, 2005; Trucco & Olmos-Antillon, 2006).

The image of the small fish shows that some portions of the image created via HE approaches become oversaturated. The region appears very dark and bright, reducing visual detail. Images generated through ICM and UCM are higher in contrast, with improved object differentiation. Some sections of the final resultant image, however, get deep darker. Furthermore, the impact of blue and green light does not decrease within the resultant underwater image but remains greenish and bluish. CLAHE-Mix approach increases visual contrast; however, the resulting image looks like greenish. As a result, CLAHS greatly decreased the influence of the blue and green lighting within the final resultant underwater image. Still, the entire colour of the underwater image is not sufficiently enhanced. The underwater image obtained by the method PDSCC, on the other

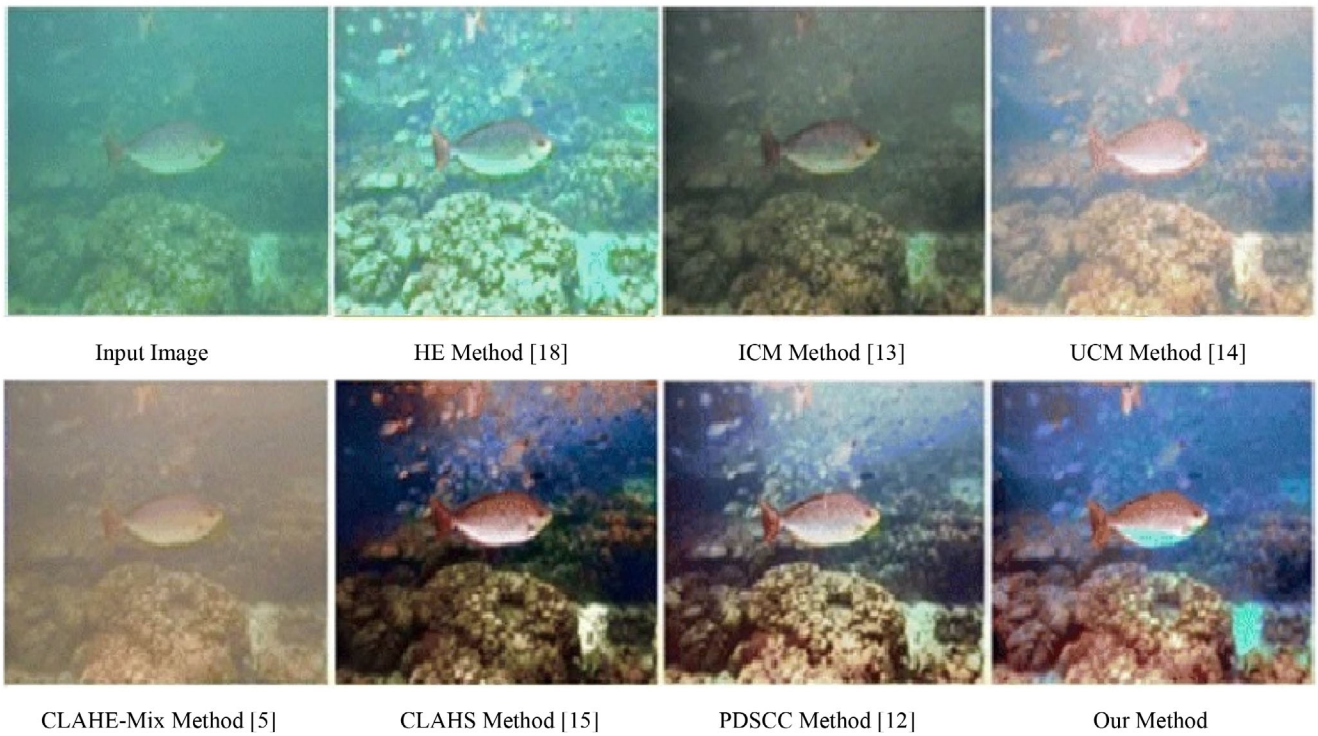


FIGURE 8 Underwater image enhancement on Small Fish images.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

hand, displays just a minor level of brightness improvement since the subjects within the underwater images could not be better separated. Since the objects are more defined and the regions are neither too brilliant nor too dark, the proposed strategy successfully raised the contrast. The impact of blue-green light is likewise much diminished.

The HE approaches over-saturates the image having blue coral as shown in Figure 9. Final resultant output underwater image becomes very contrast, resulting in loss of image fine detail. When compared with HE, the approaches described by the author Iqbal et al. that is ICM and UCM, yield greater brightness improvement. Some sections within the image, however, contain excessively dark as well as impossible to perceive. Images created by the method CLAHE-Mix, on the other hand, have better contrast than the original image, although it still seems greenish. CLAHS achieved the most significant contrast enhancement among the cutting-edge methods tested here since the items inside the image are easily discernible. Still, the image hue is diminished because of the resulting image's contain lack of saturation and sharpness. In comparison, because the image's objects are not greatly separated, the newest suggested PDSCC approach boosted just a modest amount of visual contrast. As the colour and brightness of the images containing blue coral are suitably boosted, the suggested approach is the best among those compared. There are no spots that are excessively dark or too brilliant in colour. The image's saturation and brightness are also greatly boosted. The image's items, particularly those indicated in the circular regions, are easily identifiable. The circled regions in the figures represent the above-mentioned observations.

The image with a stone wall is present in Figure 10 reveals how practically every part of the HE image has overly bright. For the ICM and UCM approaches, different outcome images were noticed, where the images were excessively dark while an impact of blue and green illumination was kept in the final output underwater images. CLAHE-Mix insufficiently reduces the influence of blue-green light, resulting in a greenish image, although increasing the contrast in the final image. CLAHS, on the other hand, generated greater image contrast with poor image colour. Apart from the image's low brightness and saturation, an overall impact of blue and green light is kept in the final resultant underwater image. A method PDSCC generated a comparable outcome by increasing only a tiny portion of the visual contrast while retaining the blue-green lighting effect. The proposed strategy considerably decreased this impact while increasing image colour in the final image. The changes between these comparative images are shown in Figure 10.

HE oversaturated practically all sections of an underwater image having minimal contrast, as shown in coral branch in Figure 11. The resulting image turns into reddish, with excessively bright regions. Although both ICM and UCM approaches minimized the influence of blue and green light, the resulting images were darkened. Some sections of the final image are overly bright. On the other hand, the CLAHE-Mix method was unable to considerably minimize the influence of blue and green light since the blue and green colour of seawater was kept in resultant image. The objects that appear in the resulting image remain equally difficult to distinguish. Since the elements in the image are not well separated, the CLAHS approach could not greatly diminish the influence of blue and green illumination and could raise a tiny quantity of brightness. Since this may considerably improve the brightness and blue and green illumination present in seawater, the proposed approach has been demonstrated to be the optimum way. The image's objects are also more distinct from the backdrop. The Figure 11 represents the above-mentioned observations.

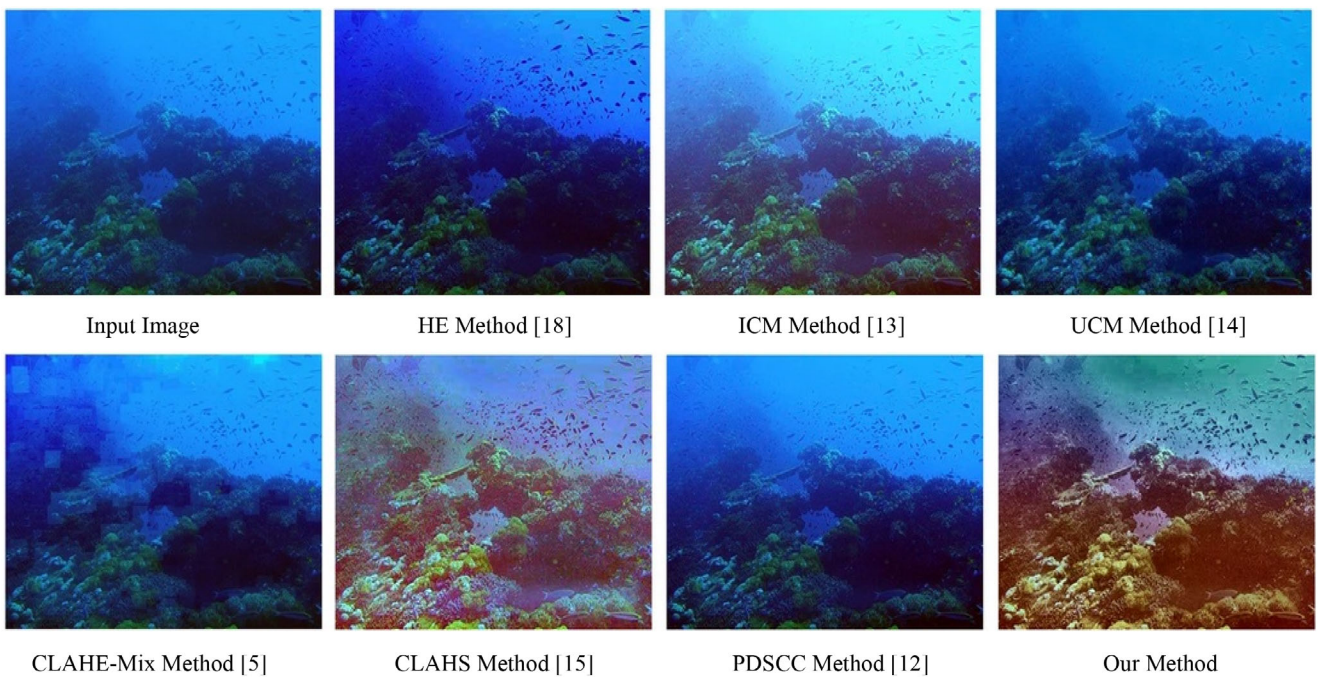


FIGURE 9 Underwater image enhancement on Blue Coral images.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

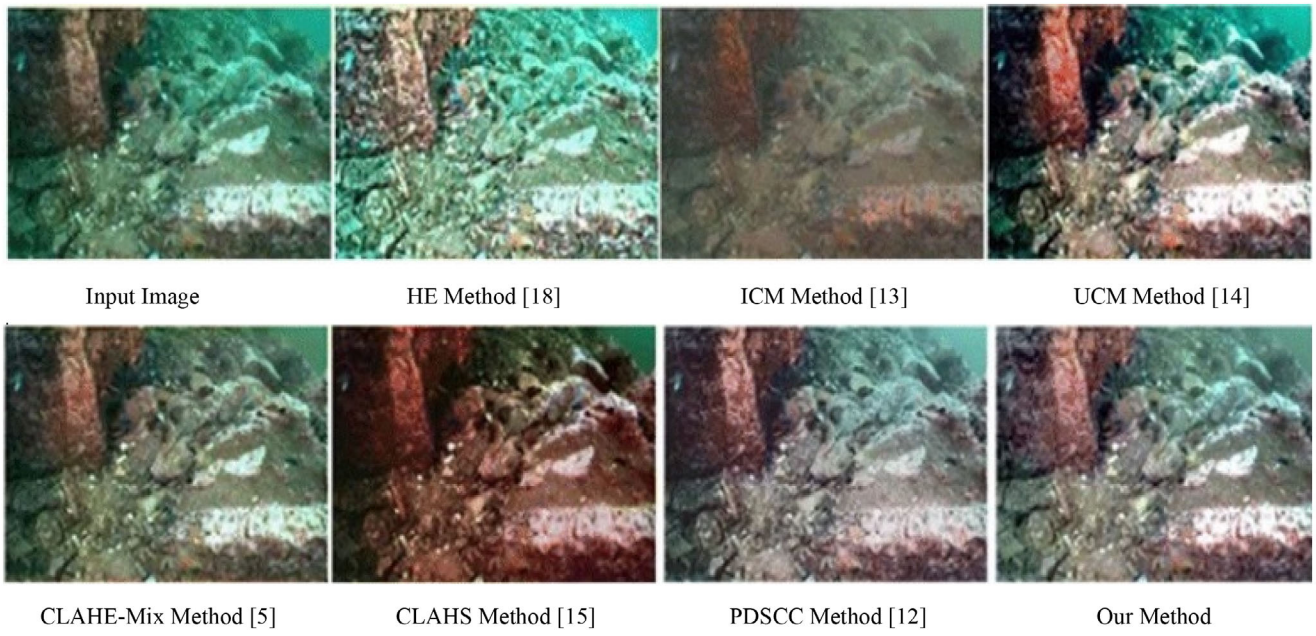


FIGURE 10 Underwater image enhancement on Stone Wall images.

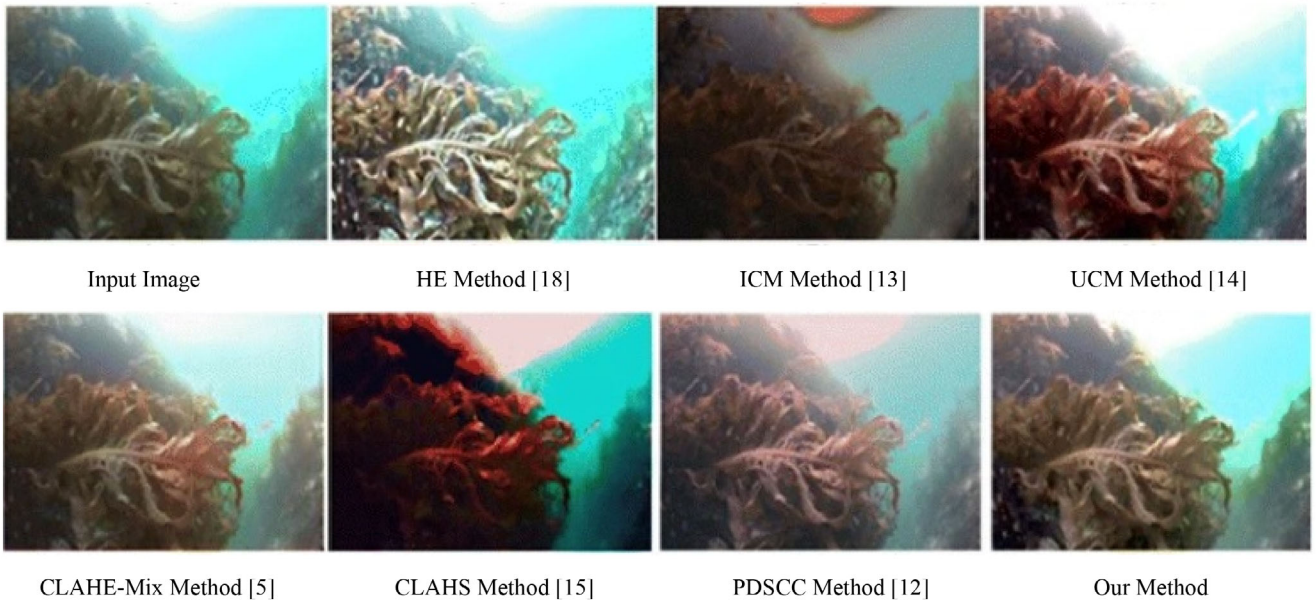


FIGURE 11 Underwater image enhancement on Coral Branch images.

Based on visual observation, the suggested method boosted image contrast over the other techniques. The images primary elements have become more distinguishable from the background. Furthermore, the impact of blue and green illumination decreases significantly. The colour of the resultant underwater image is equally enhanced.

5.2 | Quantitative analysis

Tables 1–4 compares the entropy, PSNR, MSE UIQM, UCIQE, and execution time values of the images illustrated in Figures 8–11. Bold face values reflect the best outcomes. This can be shown in Tables 1–4, the quantitative performance of the suggested technique outperforms the

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

other approaches in terms of MSE, PSNR, UIQM, and UCIQE. With respect to entropy, however, the suggested technique scores second and third after ICM and UCM. Entropy is given by the matching grey-level states that each of the pixels can adopt. The noise increased throughout the enhancement process may result in a high entropy value generated through ICM and UCM. According to the results, ICM and UCM have the greatest MSE and least PSNR scores.

When compared with other strategies, this suggested method yields the least MSE results while the maximum PSNR score. These results also suggest that the proposed strategy may sufficiently decrease noise while increasing the image's important information. Both qualitative and quantitative assessments demonstrate that the suggested strategy increased the improvement in the underwater image.

Table 5 displays the entropy, PSNR, MSE, UIQM, UCIQE as well as execution time of 500 examined underwater images. The suggested method's entropy value is compared to the entropy values of both ICM and UCM approaches. Earlier stated, the significantly elevated entropy values observed in both approaches could possibly be due to noise generated during enhancing operations. The MSE score of both ICM and UCM algorithms suggests the resulting images have a significant level of noise. The quantitative analysis stated that the suggested technique had the least average MSE score and the greatest average PSNR score of each of the evaluated methods. The quantitative analysis of UIQM and UCIQE stated that the suggested technique had the best score in respect to all other comparative methods. The UIQM value of our proposed method is improved by approximately 4% with respect to the PDSCC method, which is the earlier best method on these specific images of the dataset. Similarly, the UCIQE value of our proposed method is improved by approximately 5% with respect to the CLASH method, which is the earlier best method on this large-scale underwater dataset. The results shown represent an amount of enhancement in the generated image's contrast and noise reduction when employing the proposed strategy. These results also confirm the observational results, in which the items within the image are more effectively differentiated against the surroundings as well as their impact on blue and green illumination is greatly decreased within the resulting images.

The processing time, along with to the entropy, MSE, PSNR, UIQM, and UCIQE are employed to evaluate the efficiency of the suggested approach. According to the results in Tables 1–5, the suggested approach takes less time to process than PDSCC and CLAHE-Mix. Tables 1–4 show that CLAHE-Mix had a lengthy execution time on three with respect of the four images analysed. This additionally had a longer overall execution time of the 500 underwater images analysed, shown in Table 5. According to the data in all tables, the PDSCC approach has the second longest execution time. However, while the proposed method has the third-longest execution time in general, the differences among its normal execution time compared with different methods have been minimal and similar, as can be seen in Table 5. Studies show that, as well, along with creating a superior resulting image with high contrast and accessibility, the suggested technique uses very less processing time and is equivalent to other cutting-edge methods. Thus, with (i) the capacity to improve the brightness as well as the clarity of the images taken underwater and (ii) requiring little execution time (i.e., like other approaches), this suggested approach was successfully demonstrated to be a suitable technique for underwater image enhancement.

TABLE 1 Quantitative analysis of Small Fish underwater images used in Figure 8.

Input image	Technique	Entropy	PSNR	MSE	UIQM	UCIQE	Execution time
Small Fish	HE	5.95	13.35	3000.81	2.65	0.56	0.51
	ICM	7.73	5.71	17511.62	3.51	0.49	0.45
	UCM	7.79	5.78	17512.21	2.65	0.53	0.49
	CLAHE-Mix	6.75	14.45	2337.91	3.11	0.55	0.15
	CLAHS	6.99	18.95	830.71	2.65	0.63	0.92
	PDSCC	6.86	24.56	228.49	3.15	0.55	1.05
	Proposed method	7.61	**33.36**	**29.79**	**3.65**	**0.65**	0.88

TABLE 2 Quantitative analysis of Blue Coral underwater images used in Figure 9.

Input image	Technique	Entropy	PSNR	MSE	UIQM	UCIQE	Execution time
Blue Coral	HE	5.97	14.36	2369.11	2.55	0.59	0.16
	ICM	7.68	6.54	15787.57	3.45	0.53	0.75
	UCM	7.78	6.25	15785.95	2.75	0.56	0.55
	CLAHE-Mix	7.01	17.23	1241.36	3.15	0.59	1.45
	CLAHS	7.35	19.65	725.63	2.78	0.67	0.56
	PDSCC	7.11	24.35	221.68	3.12	0.59	1.15
	Proposed method	7.66	**30.25**	**58.36**	**3.69**	**0.66**	1.05

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Quantitative analysis of Stone Wall underwater images used in Figure 10.

Input image	Technique	Entropy	PSNR	MSE	UIQM	UCIQE	Execution time
Stone Wall	HE	5.97	10.68	5218.45	2.44	0.55	0.21
	ICM	7.74	8.56	9132.56	3.51	0.48	0.65
	UCM	7.71	8.65	9136.55	2.67	0.51	0.46
	CLAHE-Mix	7.13	17.58	1135.21	2.96	0.53	1.35
	CLAHS	7.41	16.75	1390.05	2.55	0.57	0.68
	PDSCC	7.27	20.58	568.35	3.61	0.49	1.25
	Proposed method	7.62	**33.45**	**29.66**	**3.56**	**0.61**	0.88

TABLE 4 Quantitative analysis of Coral Branch underwater images used in Figure 11.

Input image	Technique	Entropy	PSNR	MSE	UIQM	UCIQE	Execution time
Coral Branch	HE	5.64	8.57	9391.56	2.54	0.59	0.22
	ICM	7.35	6.35	15494.35	2.61	0.52	0.85
	UCM	7.52	6.24	15502.05	2.77	0.56	0.65
	CLAHE-Mix	6.83	16.12	1591.11	3.02	0.58	1.61
	CLAHS	6.88	18.46	920.91	2.65	0.59	0.63
	PDSCC	6.35	12.25	4096.05	3.11	0.54	1.12
	Proposed method	7.41	**31.25**	**51.46**	**3.65**	**0.66**	0.82

TABLE 5 Five hundred underwater images average value in terms of entropy, PSNR, MSE, UIQM, UCIQE, and their execution time.

Technique	Entropy	PSNR	MSE	UIQM	UCIQE	Execution time
HE	5.84	12.54	3885.36	2.84	0.58	**0.53**
ICM	7.55	7.51	12075.68	3.71	0.51	0.82
UCM	7.54	7.46	12065.89	2.87	0.54	0.65
CLAHE-Mix	7.09	18.35	1156.35	3.13	0.57	1.61
CLAHS	7.12	17.65	1214.68	2.82	0.61	0.98
PDSCC	6.84	25.65	334.65	3.21	0.53	1.21
Proposed method	7.51	**31.25**	**76.69**	**3.85**	**0.64**	1.06

6 | CONCLUSION

The present research provides a revision and modification method for both ICM and UCM. The suggested method makes use of a histogram adjustment of the combined HSV and RGB colour space. The image histograms in the RGB colour space extend among defined limitations to minimize the impact of over-saturation and under-saturation. The image pixel dispersion has been designed to resemble the Rayleigh distribution, with most pixels centred at the brightness range centre. A similar approach is used for image histograms within the HSV colour space, with restricted stretching in S and V elements. This approach lowers the impact on over-saturated and under-enhanced regions in the resultant output greatly. Quantitative and qualitative findings suggest the designed strategy was successful in enhancing the contrasting qualities of the image. Distortion is further decreased with respect to the remaining approaches. Underwater image colour quality is improved as well, and the influence in the blue and green light has been minimized greatly. The enhancement in underwater image colour improves object accessibility, which improves the identification of items and objects in the image. This approach may be useful in future underwater research.

This method is a single image-based, dataset-independent system. Due to this, the comparison is only done within the same approach-based model. The results of our method in comparison to other approach-based methods are far better. But if we compare our proposed method with other approaches that have been trained on some specific datasets using deep learning approaches with the help of convolutional neural networks (CNN), then it might be possible that the results are not that appropriate. Although the comparison of these two different approach-based models is not justifiable.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

So, it is better to compare the same approach-based methods. Both approaches, that is, the single image-based approach and the deep learning-based approach, are used for different applications. A single image-based approach is mainly used to provide robustness in the model, whereas a deep learning-based approach is used to improve the efficiency of the model. Hence, this implication should be maintained when comparing different approaches.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available from the corresponding author, upon reasonable request.

ORCID

Aruna Bhat https://orcid.org/0000-0002-5475-8664

REFERENCES

Alenezi, F., Armghan, A., Mohanty, S. N., Jhaveri, R. H., & Tiwari, P. (2021). Block-greedy and CNN based underwater image dehazing for novel depth estimation and optimal ambient light. *Water*, *13*(23), 3470.

Barros, W., Nascimento, E. R., Barbosa, W. V., & Campos, M. F. (2018). Single-shot underwater image restoration: A visual quality-aware method based on light propagation model. *Journal of Visual Communication and Image Representation*, *55*, 363–373.

Church, S. C., White, E. M., & Partridge, U. C. (2003). *Ultraviolet dermal reflection and mate choice in the guppy* (pp. 693–700). Elsevier Science.

Ding, X., Xueyan, Y., & Xianping, F. (2022). Multi-polarization fusion generative adversarial networks for clear underwater imaging. *Optics and Lasers in Engineering*, *152*, 106971.

Elnashef, B., & Sagi, F. (2019). Direct linear and refraction-invariant pose estimation and calibration model for underwater imaging. *ISPRS Journal of Imagegrammetry and Remote Sensing*, *154*, 259–271.

Gasparini, F., & Schettini, R. (2003). Color correction for digital imagegraphs. In *12th International Conference on Image Analysis and Processing (ICIAP 2003)* (pp. 646–651). IEEE Press.

Hasibuan, Z. A., Hasibuan, Z. A., Andono, P. N., Pujiono, D., & Setiadi, R. I. M. (2021). Contrast limited adaptive histogram equalization for underwater image matching optimization use SURF. *Journal of Physics: Conference Series*, *1803*(1), 012008.

Hu, H., Zhang, Y., Li, X., Lin, Y., Cheng, Z., & Liu, T. (2020). Polarimetric underwater image recovery via deep learning. *Optics and Lasers in Engineering*, *133*, 106152.

Iqbal, K., Odetayo, M., James, A., Salam, R. A., & Talib, A. Z. (2010). Enhancing the low quality images using unsupervised color correction method. In *International Conference on System Man and Cybernatics (SMC)* (pp. 10, 1703–13, 1709).

Iqbal, K., Salam, R. A., Osman, A., & Talib, A. Z. (2007). Underwater image enhancement using integrated color model. *IAENG International Journal of Computer Science*, *34*(2).

Jaffe, J. S. (1990). Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, *15*(2), 101–111.

Li, T., Wang, J., & Yao, K. (2022). Visibility enhancement of underwater images based on active polarized illumination and average filtering technology. *Alexandria Engineering Journal*, *61*(1), 701–708.

McGlamery, B. L. (1979). A computer model for underwater camera systems. *Proceedings of the SPIE*, *208*, 221–231.

Nabavi, S., Kumar, B. V. K. V., Bain, J. A., Hogg, C., & Majetich, S. A. (2009). Application of image processing to characterize patterning noise in self-assembled nano-masks for bit-patterned media. *IEEE Transactions on Magnetics*, *45*(10), 3523–3526.

Naim, M. J. N. M., & Isa, N. A. M. (2012). Pixel distribution shifting color correction for digital color images. *Applied Soft Computing Journal*, *12*(9), 2948–2962.

Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., Braeuning, M. P., & Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital Imaging*, *11*, 193–200.

Rizzi, A., Gatta, C., & Marini, D. (2003). A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters*, *24*(11), 1663–1677.

Rizzi, A., Marini, D., & Gatta, C. (2004). From retinex to automatic color equalization: Issues in developing a new algorithm for unsupervised color equalization. *Journal of Electronic Imaging*, *13*(1), 75–84.

Salmeri, M., Mencattini, A., Ricci, E., & Salsano, A. (2001). Noise estimation in digital image using fuzzy processing. In *IEEE International Conference on Image Processing* (Cat. No. 01CH37205) (vol. 1, pp. 517–520). IEEE.

Schechner, Y. Y., & Karpel, N. (2004). Clear underwater vision. In *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004* (Vol. 1, pp. 536–543). IEEE.

Schechner, Y. Y., & Karpel, N. (2005). Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering*, *30*(3), 570–587.

Schettini, R., & Corchs, S. (2012). Underwater image processing: State of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 1–14.

Senthilkumaran, N., & Thimmiaraja, J. (2014). Histogram equalization for image enhancement using MRI brain images. In *IEEE World Congress on Computing and Communication Technologies* (pp. 80–83). IEEE.

Shamsudin, N., Ahmad, W. F. W., Baharudin, B., Rajuddin, M. K. M., & Mohd, F. (2012). Significance level of image enhancement techniques for underwater images. In *International Conference on Computer & Information Sciences (ICCIS)* (Vol. 1, pp. 490–494). IEEE.

Singh, N., & Bhat, A. (2021). A detailed understanding of underwater image enhancement using deep learning. In *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*. IEEE.

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Singh, N., & Bhat, A. (2023). A systematic review of the methodologies for the processing and enhancement of the underwater images. *Multimedia Tools and Applications*, *82*, 38371–38396. https://doi.org/10.1007/s11042-023-15156-9

Trucco, E., & Olmos-Antillon, A. T. (2006). Self-tuning underwater image restoration. *IEEE Journal of Oceanic Engineering*, *31*(2), 511–519.

Ulutas, G., & Ustubioglu, B. (2021). Underwater image enhancement using contrast limited adaptive histogram equalization and layered difference representation. *Multimedia Tools and Applications*, *80*, 15067–15091.

Wang, J., Wan, M., Gu, G., Qian, W., Ren, K., Huang, Q., & Chen, Q. (2022). Periodic integration-based polarization differential imaging for underwater image restoration. *Optics and Lasers in Engineering*, *149*, 106785.

Wang, Y., Cao, Y., Zhang, J., Wu, F., & Zha, Z. J. (2021). Leveraging deep statistics for underwater image enhancement. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *17*(3s), 1–20.

Wang, Y., Song, W., Fortino, G., Qi, L.-Z., 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

Ye, Z. (2009). Objective assessment of nonlinear segmentation approaches to gray level underwater images. *ICGST-GVIP Journal*, *9*(2), 39–46.

Yeganeh, H., Ziaei, A., & Rezaei, A. (2008). A novel approach for contrast enhancement based on histogram equalization. In *IEEE Proceedings of International Conference on Computer and Communication Engineering* (pp. 256–260). IEEE.

Zhang, C., Zhang, X., & Tu, D. (2018). Underwater image enhancement by fusion. In *Advanced Manufacturing and Automation* (Vol. 7, pp. 81–92). Springer Singapore.

Zhou, J., Zhang, D., & Zhang, W. (2020). Classical and state-of-the-art approaches for underwater image defogging: A comprehensive survey. *Frontiers of Information Technology & Electronic Engineering*, *21*(12), 1745–1769.

AUTHOR BIOGRAPHIES

Dr. Nishant Singh has completed his Ph.D. in the Department of Computer Science and Engineering from Delhi Technological University, Delhi. His Research interests include Image Processing, Neural Networks, Machine Learning, Computer Vision, and Deep Learning. He has eleven years of total experience including academic and research experience.

Dr. Aruna Bhat is currently working as a professor in the department of Computer Science and Engineering, Delhi Technological University, DTU. Her Research interests include Image Processing, Neural Networks, Machine Learning, Computer Vision, and Deep Learning. She has more than ten years of experience including academic and research experience.

How to cite this article: Singh, N., & Bhat, A. (2024). Underwater image enhancement using contrast correction. *Expert Systems*, e13692. https://doi.org/10.1111/exsy.13692

14680394, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/exsy.13692 by Delhi Technological University, Wiley Online Library on [01/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

Utilizing Deep Learning for Textual Classification of Hate Speech in Online Social Networks

Dhruv Verma
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
dhruv4321@gmail.com

Gaurav Thakur
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
gaurav.t2002@gmail.com

Hammad Ansari
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
hammadkafeel2607@gmail.com

Rahul Katarya
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
rahuldtu@gmail.com

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10627038

***Abstract*— In today's digital age, the internet has allowed people from different corners of the world to easily share their ideas. It is a fairly ubiquitous opinion that everybody must have the basic human right to share their opinion and express their thoughts and feelings. However, with this heightened connectivity, people tend to abuse their rights. People often send abusive or racist remarks over the internet, often anonymously. While research on hate speech detection is growing rapidly and diversely, detecting hate speech remains a difficult task. This paper proposes a deep learning model to detecting abuse and hate speech on online social networks. We first start by cleaning the data and then obtain the sentence embeddings using the BERT model. After this, we pass our sentence embeddings to our custom Convolutional Neural Network model to classify the tweets. We optimize the performance of our custom Convolutional Neural Network model by applying the GridSearch technique for hyperparameter tuning. This custom model has a superior performance to baseline Machine Learning Models. We have evaluated the performance of our model on the Hate Speech and Offensive Language dataset.**

Keywords*— *BERT,* Convolutional Neural Network, *GridSearch, Hate Speech and Offensive Language Dataset, Machine Learning Models.

I. INTRODUCTION

The last couple of years have seen a rapid increase in internet and social media use all over the world thanks to more and more people getting access to smartphones and computers and the consistent drop in mobile data prices along with 3G, 4G, and 5G services. Thus, it is no coincidence that the amount of hate and offensive speech is increasing, especially on social media platforms. Due to this, companies are taking certain measures to curb the spread of hate online. Specifically, over one million social media accounts were banned in a span of 6 months, with over 3.8 million Tweets being removed in the same time frame [1].

On such social media platforms, it is extremely easy for a user to hide behind a fake name or fake identity with made-up credentials. With this anonymity, they may send hate speech messages either directly to other members or for the general public to see. These hate speech messages contain insulting or sensitive content, which should not be shareable in public, especially without their consent. When examining these toxic languages, there are distinct types like cyberbullying, one of the main causes of suicide among young people [2].

Hate and offensive speech are not the same things, and there are several overlapping subcategories as well. For example, [3] identifies discriminative remarks as "insults," whereas [4] classifies the same remarks as hate speech [5]. Due to this, the same sentence has the tendency to be construed in a variety of ways, thus making it essential to have clear definitions for each of the categories so there is no confusion. Hate Speech is expressed against a certain group of people. Offensive speech often refers to instances where the text contains swear words or other abusive language [6]. In this paper, we have categorized sentences into three categories: hate speech, offensive speech, and neither.

It is important to detect this hate speech so that appropriate actions can be taken. Unfortunately, it is impractical to manually detect offensive texts, as it is an extremely difficult job to analyze over 6,000 tweets every second [7]. It is extremely important to be able to classify a sentence into the three defined categories. On classifying a sentence as hate or offensive speech, it can be appropriately dealt with. For example, if a Tweet on X is classified as malevolent, it can be taken down, and the account that posted that Tweet can be blocked. If it is not malevolent, it may be posted.

This task of classifying text into categories cannot be done manually, as it will be extremely slow and painful. Instead, a comprehensive deep learning model can be utilized. It is important to understand that one person may find a sentence to be offensive, while another may not. In the current day and age, with more and more children getting on social media at very young ages, it has never been more important to block content that is hateful or harmful. Hate speech content that talks about race, color, and gender can all have an extremely negative impact on children, as it can make them feel like there is something different or wrong about them [8].

Given below are the major contributions of this paper:

- Propose an optimized CNN model for the detection of hate and abuse.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:35 UTC from IEEE Xplore. Restrictions apply.

- Utilize BERT embeddings using multihead attention to capture the interdependence of words.
- Compare the performance of baseline ML algorithms with our optimized CNN for the detection of hate and abuse.

II. Related Work

The proliferation of online tweets and in turn, hate speech has shed light on the biggest issue online social media platforms face. Many social media networks are actively searching for ways to mitigate and reduce the effect of hate speech and abuse. Different researchers have formulated techniques to help detect hate speech on online platforms. Other models have used more advanced deep learning embedding techniques like BERT, mBERT, RoBERTa, MuRIL, and Indic-BERT [9, 10, 11]. These embedding techniques take into account the contextual information of the word as well. The vector of each word will depend on the surrounding words. Another technique for obtaining the embeddings is FastText and Glove. Some models use the most basic method for obtaining the embeddings of the text, which is word2vec [12]. Reference [13] has used MinMax normalization to preprocess and transform the data. According to [14], word embeddings are obtained from FastText and Glove methods. These methods, however, are context-independent, and the vector obtained for each word is not dependent on the surrounding words.

Once the natural language is converted to a sequence of numbers, it is suitable for our models to classify them. There are numerous classifiers that researchers have explored to check whether a sentence comes under hate speech or not. Some of the methods are machine learning algorithms. These methods have explored SVM, Logistic Regression, Naïve Bayes, and KNN [11, 12, 13]. However, these traditional methods have been found to be too simple to handle complex classifications of hate speech datasets. As a result, even their results obtained are not up to the mark when compared to deep learning techniques like Convolutional Neural Network (CNN) and Gated Recurrent Units (GRU).

To combat this problem, researchers have shifted towards deep learning classifiers. According to [5], Recurrent Neural Networks of Bi-GRU are used to classify the texts. The advantage of this method is that it captures variable length inputs. However, these models have been found to be extremely heavy and time-consuming for training. According to [9], the dimensionality of the embeddings obtained was reduced with the help of Principal Component Analysis. They then used the Bi-GRU algorithm to help classify their tweets. According to [10], they tried to improve the accuracy of their model by incorporating emoji2vec algorithms, and the output data from the Bi-GRU layer was appended for the final classification. According to [14], they used an ensemble CNN/GRU layer for their feature extraction. The advantage of this is that it combines the strengths of both algorithms to eventually get the classifications of the data.

Another method to mitigate the insidious effects of hate speech is by employing hate similarity to detect indirect and direct hate content [11]. Also, according to [15], another method aims to solve the problem by weakening the severity or intensity of the hate written in an online post by using hate speech normalization. The aim is not to allow hate speech but to weaken the severity of the text to generate a more publicly acceptable form. Another technique to solve the problem statement is introduced by [16], where an ensemble model is used to obtain confidence that a text is hate speech or not; if the content is doubtful, a linguistics expert is included in the process to detect the same. Once the expert correctly annotates the text, it is labeled and fed to the model, thus growing the dataset. The literary survey is elaborated in Table I.

The current limitations in the work have been summarized below:

- **Used ML Algorithms:** ML algorithms have been found to have an inferior performance to Deep Learning algorithms.
- **Used Basic Word2Vec Embedding Techniques:** They do not capture the interdependence and the contextual relation between words.
- **Use of Word Embeddings:** They make use of word embeddings, which increases overhead.

III. Proposed Methodology

The proposed methodology encompasses a deep learning based method for accurately predicting hate speech or abuse from online texts. To achieve this, first we perform data cleaning on our tweets obtained from the Hate and Offensive Language Dataset. Our data cleaning step includes changing the user tags, removing Unicode emojis, removing URLs, removing punctuation marks, and removing stop words. Second, we compute the BERT embeddings for each tweet. In this step, we calculate the sentence embedding by extracting the [CLS] token. Third, we split our dataset in a standard 80:20 ratio. This means that 80% of the dataset goes under the training dataset, and 20% goes under the testing dataset. We use the stratify parameter to make sure classes are distributed evenly in our dataset. Fourth, we design a custom CNN model to classify our tweets. We optimize our model by utilizing the GridSearch technique, which allows us to employ hyperparameter tuning for our CNN model. Instead of taking arbitrary values for our CNN architecture, we perform an exhaustive searching technique to find out which hyperparameters give us the best performance. Fifth, the performance of this model is then compared to baseline machine learning models. Finally, after evaluating the performance of our models, we chose the model that performed the best. This methodology is also referenced in Fig. 1.

A. Text Cleaning

In NLP problem statements, text cleaning is one of the most important steps for data preprocessing. Text cleaning is instrumental in improving the quality and effectiveness of our solution to the problem statement. The common characteristic of natural language is that it is full of noise and inconsistencies. That is why we need to convert natural language to numbers so that our classifiers can correctly predict the texts. By removing all the noise and inconsistencies in the data, we are able to enhance the performance of the classifiers. We perform standard data cleaning steps like changing the user tags, removing Unicode emojis, removing URLs, removing punctuation marks, and removing stop words [17]. By doing this, we are able to focus on the meaningful linguistic patterns of our data. This allows for a better understanding of the semantics of the data. We can extract features from our data more easily due to this,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:35 UTC from IEEE Xplore. Restrictions apply.

TABLE I. ILLUSTRATING THE USE OF NLP TO SOLVE HATE SPEECH DETECTION PROBLEMS

Reference	Year	Dataset	Methodology	Performance	Limitation
[11]	2023	NNTI	Applies mBERT for text embeddings with Hate Similarity Scheme to detect direct and indirect hate	Accuracy: 85% Precision: 87% Recall: 89% F1-score: 88%	Does not explore other deep learning models like CNN or BiGRU. Deals only in binary classification.
[5]	2022	OLID	FastText and Glove word embedding followed by Bi-GRU algorithm	Accuracy: 84% Precision: 87% Recall: 93% F1-score: 90%	Used word embeddings, which increases overhead. Did not investigate other deep learning classifiers. Accuracy for classifying texts was low.
[9]	2022	Code-mixed Hinglish Dataset manually scraped from the Twitter API	Obtained the sentence embeddings using Word2vec, M-BERT, Indic-BERT, XLM-RoBERTa, MuRIL followed by Bi-GRU for classification	Accuracy: 85% Precision: 89% Recall: 81% F1-score: 85%	Used word embeddings, which increases overhead. Did not investigate other deep learning classifiers. Accuracy for classifying texts was low.
[10]	2022	Abuse Detection Challenge Dataset provided by a Kaggle competition.	Obtained embeddings using M-BERT, Indic-BERT, XLM-RoBERTa, MuRIL Embeddings were passed through the Bi-GRU layer along with emoji embeddings.	F1-score: 0.88 Loss: 0.28 (Training) and 0.31 (Validation) AUC: 0.94 (Training and Validation)	Does not explore other deep learning models. Model only handles binary classification.
[12]	2022	Hate Speech and Offensive Language Dataset	Obtained embeddings using Word2vec technique followed by ML, DL, and Fuzzy Logic Classifiers for classification.	The Fuzzy Logic Model obtained an accuracy of 94.26%.	It uses a simple embedding technique of word2vec. It does not use state-of-the-art BERT embedding models that take contextual information of words into account as well.
[13]	2022	Data scraped from online social media sites like Twitter is used.	Used MinMax normalization to preprocess and transform the data followed by KNN (10 neighbors) to classify texts.	No quantitative analysis of results was performed.	The system architecture is too simple. It uses only ML algorithms to classify texts and does not take into account any state-of-the-art deep learning methods.
[14]	2020	HON Dataset	Applies FastText technique for word embeddings followed by Feature extraction through CNN/GRU layer and classification through a fully connected tunable classifier.	Recall (Hate): 83% F1-score (macro): 65%	Used word embeddings of FastText, Glove, and Word2vec, which increases overhead.

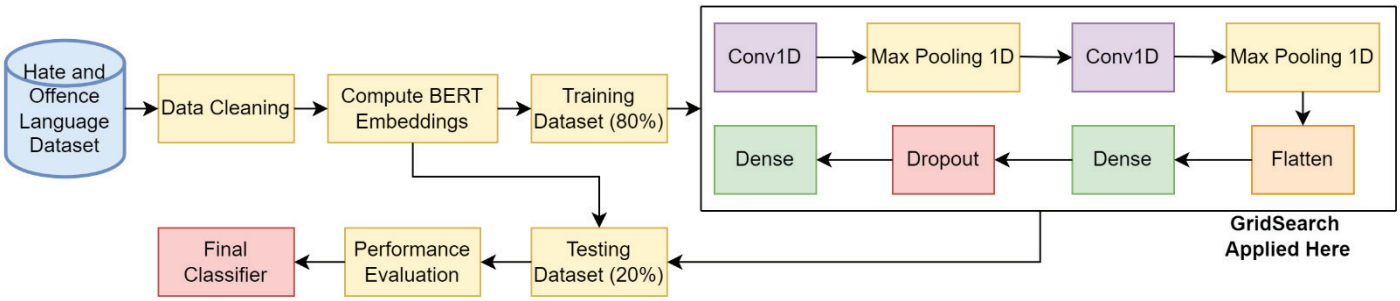


Fig. 1. Illustrating the Methodology for Solving the Problem Statement

which, in turn, will help boost the accuracy of our classifiers. Given below is the algorithm employed to perform data cleaning on our tweets.

Algorithm:

Input: Raw Tweets

Output: Cleaned Tweets

1. Get all the tweets in the dataset and pass it through preprocess(datas)
2. Initialize empty list ‘clean’
3. Call a function to change user tags into a generic tag ("user"):
    a. Define a regular expression to match user tags
    b. Use re.sub() to replace matched user tags with "user"
4. Call a function to remove Unicode emojis and replace them with ""
    a. Define a regular expression to match HTML entities
    b. Use re.sub() to replace matched user tags with "".
5. Call a function to remove any URLs and replace them with ""

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:35 UTC from IEEE Xplore. Restrictions apply.

a. Define a regular expression to match URL syntax
b. Use re.sub() to replace matched user tags with "".

6. Call a function to Remove punctuation marks of ", ', !, `, using replace() function
7. Call a function to remove stopwords and replace them with ""
   a. Tokenize the text using nltk.word_tokenize().
   b. Filter out stopwords from the tokenized text.
   c. Join the remaining words into a single string.
8. Append the result to list 'clean'

B. BERT Embeddings

BERT is one of the best-performing pre-trained language representational models. It was introduced by Google in 2018. It is known as Bidirectional because of its ability to capture the meaning of words from both directions (left and right). We have utilized this because of its ability to provide word as well as sentence embeddings (derived by extracting the [CLS] token). Before the natural language is converted into a stream of numbers that the computer is able to decipher, it needs to undergo some preprocessing steps.

BERT uses WordPiece Tokenization. Each word is either kept the same or split into sub-words, or in some cases, even individual characters [18]. The entire model has 110 million parameters. The architecture of the model consists of 12 encoding layers. Each layer has a hidden layer size of 768 and 12 Attention Heads. In the multihead self-attention mechanism, the interdependence of words is captured through the attention scores of each word that are calculated with respect to every other word in the sentence. This gives an indication of the most important and meaningful words in the sentence. After the attention mechanism, there is a feedforward neural network. For each position in the sequence, the output is a vector that has captured information from the entire sequence based on attention scores obtained previously. After this comes the layer normalization, which ensures consistent scale and distribution of features. Residual connections are also there to address the vanishing gradient problems by providing a shortcut connection that bypasses one or more layers. A 768-dimention vector is finally obtained and passed as input to our classifiers.

Algorithm

Input: Cleaned Tweets

Output: Sentence Embeddings from CLS token

1. Get the cleaned tweet and pass it through compute_bert_embeddings
2. Tokenize the input using tokenizer() function from BertTokenizer
   a. Set padding and truncation parameters as True to ensure all tweets are of the same size
3. Pass obtained tokens through model() function obtained from BertModel with pre-trained weights of 'bert-base-uncased' to obtain the BERT embeddings
4. From the embeddings obtained, extract the [CLS] token from the last hidden state using embeddings.last_hidden_state[:, 0, :]
5. Convert output to numpy and store in cls_embeddings

C. Splitting the Dataset

After obtaining the correctly labeled dataset and performing the appropriate data preprocessing functions, we move towards splitting the dataset for our training and testing phase. The appropriate splitting ratio is essential to strike the correct balance to ensure that the model has enough data to train and extract information and not too much data to start memorizing and overfitting. For our model, we have used the most commonly used splitting ratio of 80:20, which means that 80% and 20% of the data goes under the training and testing dataset, respectively. We have also used the 'stratify' parameter to ensure that the labels are divided in the same ratio in which they are present in the dataset to handle class imbalance.

D. KNN

The k-nearest neighbor algorithm is an extremely useful technique for solving the problem statement. It works by measuring the closeness of different data points by measuring the Euclidean distance between them. It follows the thought process that the closer two different data points are in the feature space, the more probability it has to belong to the same class. The KNN algorithm is most commonly used in the detection of hate speech classification problems such as this one. It is based on a 'majority vote,' meaning that the point will be classified as the most commonly occurring class among all its k-nearest neighbors.

E. Gaussian Naïve Bayes

The GNB algorithm utilized demonstrates its efficacy when we try to solve the problem statement. Derived from the Bayes theorem, it assumes that each feature in the feature space is independent of the other. This means the existence or non-existence of a feature does not affect the value of any other feature. This is particularly useful in our study as after our preprocessing steps, we ensure that our features are independent of each other. Also, it makes use of variables that have a continuous value and maps it to a Gaussian distribution to find the conditional probability, which is the case in our dataset.

F. Multinomial Naïve Bayes

The MNB algorithm is also an extremely powerful algorithm for hate speech classification. It is closely related to the GNB algorithm described above as it is also derived from the Bayes Theorem. It follows a very straightforward interpretation of the Bayes Theorem formula, where the probability of a text belonging to a particular class and the conditional probabilities involved in the formula are obtained directly by analyzing the dataset. Finally, the class which has the highest probability of prediction is selected by the model as its prediction.

G. Decision Tree Classifier

Another extremely powerful and robust algorithm we have used for our study is the DTC algorithm. It follows a tree-like structure in which the internal nodes represent unique features of a dataset. In the tree, each branch represents a decision to be made, and the leaf node is used to represent the final decision. To classify a text, we first start

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:35 UTC from IEEE Xplore. Restrictions apply.

from the root node and traverse through all the decision rules until the randomness of our obtained decision reduces below a certain threshold, and then we finally make the prediction at the leaf node. To construct this decision tree, we have used the CART algorithm. Here, tree construction depends on the Gini Index. Attribute with a low Gini index is preferred over a higher Gini index.

H. Random Forest Classifier

The RFC classifier is one of the most powerful tools when it comes to the problem statement. It makes use of several DTCs described earlier to make a more accurate and robust prediction. As it follows an ensemble approach, the performance of this model is very good and accurate. The use of multiple classifiers makes this model more reliable. Like a forest, this classifier also contains an ensemble of trees.

I. XGBoost

XGBoost is an extremely powerful supervised learning algorithm. It is employed in various problem statements to boost the efficiency and performance of models. Like the RFC, this is also an ensemble learning technique. It combines the performance of different weak decision tree models and makes a much more robust and effective model. It incorporates regularization techniques that help minimize any potential overfitting and is scalable and appropriate for large datasets. Since our dataset has nearly 25k instances of tweets, this algorithm finds its utility here.

J. Convolutional Neural Networks

CNN is a deep learning neural network that helps extract patterns and information from the input data. While CNNs are usually used in image classification problems in 2 dimensions, as the input data is pixel intensities in 2 dimensions, the main use of CNN is to capture the patterns in 1-dimension as our sentence embeddings are in 1 dimension itself. CNNs are often used in image processing as the same class of image will have a similar pattern of pixel intensities. However, the effect is replicated here also, as similar texts will have a similar pattern of embeddings. CNNs employ the use of filters to capture different features and discriminate information about our input sentence embeddings, which will help us eventually classify our text into its correct class [19]. This is extremely important for the accurate detection of abuse in a given text.

To construct our CNN, we have used various convolutional and max pooling layers. The convolutional layers are used to apply filters to the input to extract features. The filter (or kernel) consists of some specific set of values, and it is multiplied to a subset of input data, and thus you obtain your extracted features from the input. It is instrumental in identifying discriminate features for classifying our text. The MaxPooling layer reduces the dimensionality of the input while retaining the most relevant information about the data. It does this by retaining the maximum value among a group of cells in the input vector. This allows us to reduce the computational load of the model so that it does not take too long to train. Finally, our fully connected layer is used to extract the most salient features, reduce the dimensionality of our input, and classify our data.

Algorithm

Input: Sentence Embeddings from CLS Token

Output: Classification of Tweet

1. Create a sequential model object
2. Append a 1D Convolutional Layer with the number of filters as 64, a filter size of 5, and a ‘ReLu’ activation function
3. Append a MaxPooling Layer with size 2
4. Append a 1D Convolutional Layer with the number of filters as 32, a filter size of 5, and a ‘ReLu’ activation function
5. Append a MaxPooling Layer with size 2
6. Flatten the output from the previous layer
7. Append a fully connected layer with size 128 using the Dense() function
8. Append a Dropout layer with a rate of 0.5
9. Append an output layer of size 3 using the Dense() function

K. Hyperparameter Tuning

Hyperparameters refer to the values given to a model that specifies how the architecture is defined, and it determines how well it learns. The process of learning the best possible values that are given to the model to maximize its performance is known as hyperparameter tuning. Hyperparameter optimization is an integral part of deep learning that finetunes hyperparameters. This process allows the model to achieve the highest possible accuracy. While there are many optimization methods, we have employed the GridSearchCV method. It provides a systematic, brute-force, and exhaustive search mechanism that allows us to find the best combination of hyperparameters to be used in our CNN model. It consists of a parameter grid with the hyperparameters we explore in the search.

For our CNN model, the parameter grid kept track of the number of filters, filter size, max pooling size, and number of units in the dense layer. We then ran this grid search technique for our deep learning models. We kept track of our mean test scores and chose the parameters corresponding to the maximum mean test scores.

IV. Dataset Description

Hate and Offensive Language Dataset – The dataset utilized in the paper contains a compiled list of tweets that people have sent out. These tweets have been annotated and are publicly available. The text is classified for multiclass classification as hate speech, offensive, and neither. This publicly available dataset is used to train and test the model to improve accuracy. This dataset includes content that expresses hate towards a particular racial group, gender, or even people having a particular sexual orientation or other forms of hate due to the nature of this research. It contains the username of the account from which it was tweeted from all countries across the globe. This dataset comprises over 24,783 tweets, of which 1430 (or 5.77 percent of the total) were classified as neither, 19190 (or 77.4% of the total) as offensive, and 4163 (or 16.80 percent of the total) as hate speech [20].

V. Results

A. Performance of Optimized CNN Model

In this section, we present the quantitative results we have obtained from our experimental analysis. We have utilized the dataset mentioned in Section IV and performed the preprocessing steps. We clean the text by removing all the noise and inconsistencies in the data, followed by computing

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:35 UTC from IEEE Xplore. Restrictions apply.

the BERT embeddings. Then, we split our data in a common 80:20 ratio. Then, we apply different classification algorithms such as KNN, MNB, GNB, DTC, RFC, XGB, and our custom CNN architecture. We then examine the results obtained from our model based on the accuracies of each classifier. We see that our custom CNN model outperforms the other baseline machine learning models. For our CNN model, we obtained an accuracy of 85.35% for the number of filters as 64 and 32 in the first and second layer, respectively, filter size as 5 in both layers, max pooling size as 2 in both layers, dropout as 0.5 and number of units in the fully connected layer as 128. To evaluate the performance of our models, we utilized precision, recall, f1-score, and accuracy metrics. We obtained a precision, recall, and F1-score of 0.88, 0.95, 0.92 for our abuse class and 0.73, 0.65, and 0.69 for our hate speech class, respectively. After 10 epochs of training, the model obtained a training accuracy of 87.73%, which is extremely close to our final accuracy of the testing dataset, and a loss of 0.3102, which is also low. This suggests that the model is not overfit and is correctly functioning. The training of the model can be seen in Fig. 2.



Fig. 2. Accuracy and Loss of CNN Model while Training

B. *Comparison with Baseline ML Models*

We have compared our optimized CNN model with traditional ML baseline models. We obtained the precision, recall and f1-score of 0.79, 0.80, 0.77 for our KNN model, 0.77, 0.66, 0.70 for our GNB model, 0.75, 0.80, 0.73 for our MNB model, 0.73, 0.72, 0.73 for our DTC model, 0.80, 0.78, 0.75 for our RFC model, 0.81, 0.84, 0.81 for our XGB model, respectively. These results elucidate that our optimized CNN model has superior performance when compared to our baseline ML models. This shows how using a deep learning model for the detection of hate and abuse is better than ML models. The comparison of our CNN model with other baseline ML models is given in Table II and Fig. 3.

TABLE II. SHOWING THE COMPARISON OF PERFORMANCE BETWEEN OUR CNN MODEL AND BASELINE ML MODELS

Model	Precision	Recall	F1-score	Accuracy
KNN	0.79	0.80	0.77	80.31%
GNB	0.77	0.66	0.70	66.35%
MNB	0.75	0.80	0.73	79.87%
DTC	0.73	0.72	0.73	72.20%
RFC	0.80	0.78	0.75	78.39%
XGB	0.81	0.84	0.81	83.80%
Our CNN Model	**0.83**	**0.85**	**0.83**	**85.35%**

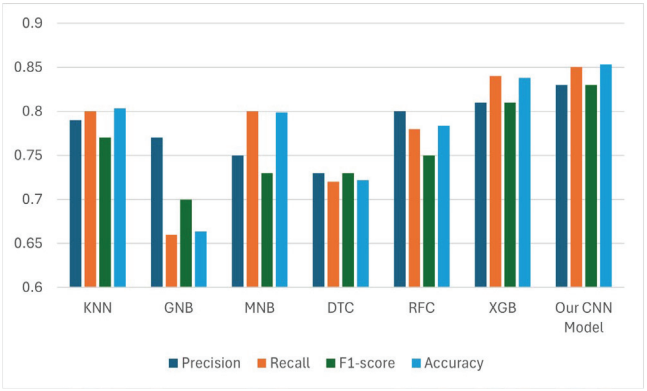


Fig. 3. Showing the Comparison of Performance between our CNN Model and Baseline ML Models

VI. CONCLUSION AND FUTURE SCOPE

As more and more people get access to smartphones and computers, hate speech detection becomes an increasingly important issue. Manually classifying sentences as hate speech or not is not practical, thus underscoring the importance of a streamlined solution.

To solve this issue, we made an effective model that achieves good results, classifying text into three broad categories with satisfactory performance. To achieve this, we followed a few basic steps. First, we preprocess the data, which includes removing noise from the data and cleaning it to be tokenized. Next, we tokenize the data using the BERT tokenizer. After this, we use the BERT pre-trained model to compute the word and sentence embeddings, which we will use as input for different models. Finally, we use the Scikit-learn package to build several different ML models, including the DTC and KNN algorithms. We also built a custom CNN model. Our custom CNN achieved the best accuracy of 85.35%.

In the future, we aim to use BERT word-level embeddings instead of sentence-level embeddings. Also, we aim to perform more experiments with our custom CNN, intending to add more layers to achieve a complex model to achieve a higher classification accuracy. We also wish to explore other deep learning models and their ensemble to improve the accuracy of and make the model more robust.

REFERENCES

[1] Kurt Wagner, "https://time.com/6080324/twitter-hate-speech-penalties/", Record number of social media bans bans, 2021

[2] Hinduja, S., Patchin, J.W., 2010. Bullying, cyberbullying, and suicide. Archives of suicide research 14, 206–221

[3] Van Hee, C., Verhoeven, B., Lefever, E., De Pauw, G., Daelemans, W., Hoste, V., 2015b. Guidelines for the fine-grained analysis of cyberbullying. Technical Report. version 1.0. Technical Report LT3 15-01, LT3, Language and Translation…

[4] Nobata,C., Tetreault, J.,Thomas, A.,Mehdad, Y.,Chang,Y., 2016. Abusive language detection in online user content, in: Proceedings of the 25th international conference on worldwideweb ,pp.145–153.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:35 UTC from IEEE Xplore. Restrictions apply.

[5] Badri, Nabil, Ferihane Kboubi, and Anja Habacha Chaibi. "Combining FastText and Glove word embedding for offensive and hate speech text detection." Procedia Computer Science 207 (2022): 769-778.

[6] Gaydhani, A., Doma, V., Kendre, S., Bhagwat, L., 2018. Detecting hate speech and offensive language on twitter using machine learning: An n-gram and tfidf based approach. arXiv preprint arXiv:1809.08651 .

[7] Rohit Shewale, "https://www.demandsage.com/twitter-statistics", Number of Twitter users per second, 2023

[8] UNICEF, " https://www.unicef.org/parenting/how-talk-your-children-about-hate-speech"

[9] Chopra, Abhishek, et al. "A Framework for Online Hate Speech Detection on Code-mixed Hindi-English Text and Hindi Text in Devanagari." ACM Transactions on Asian and Low-Resource Language Information Processing 22.5 (2023): 1-21.

[10] Bansal, Vibhuti, et al. "A Transformer Based Approach for Abuse Detection in Code Mixed Indic Languages." ACM Transactions on Asian and Low-Resource Language Information Processing (2022).

[11] Ghosal, Sayani, et al. "Inculcating Context for Emoji Powered Bengali Hate Speech Detection using Extended Fuzzy SVM and Text Embedding Models." ACM Transactions on Asian and Low-Resource Language Information Processing (2023).

[12] Chhabra, Anusha, and Dinesh Kumar Vishwakarma. "Fuzzy and Machine learning Classifiers for Hate Content Detection: A Comparative Analysis." 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST). IEEE, 2022.

[13] Jemima, P. Preethy, et al. "Hate Speech Detection using Machine Learning." 2022 7th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2022.

[14] Melton, Joshua, Arunkumar Bagavathi, and Siddharth Krishnan. "DeL-haTE: a deep learning tunable ensemble for hate speech detection." 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2020.

[15] Masud, Sarah, et al. "Proactively Reducing the Hate Intensity of Online Posts via Hate Speech Normalization." Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022.

[16] Alsamman, Ahmad, Andreas Schmitz, and Maria A. Wimmer. "Towards an Organically Growing Hate Speech Dataset in Hate Speech Detection Systems in a Smart Mobility Application." Proceedings of the 24th Annual International Conference on Digital Government Research. 2023.

[17] Chakraborty, Angana, Subhankar Joardar, and Arif Ahmed Sekh. "Ensemble Classifier for Hindi Hostile Content Detection." ACM Transactions on Asian and Low-Resource Language Information Processing (2023).

[18] Dharti Dhami, https://medium.com/@dhartidhami/understanding-bert-word-embeddings-7dc4d2ea54ca, 2020

[19] Roy, Pradeep Kumar, et al. "A framework for hate speech detection using deep convolutional neural network." IEEE Access 8 (2020): 204951-204962.

[20] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," Proc. 11th Int. Conf. Web Soc. Media, ICWSM 2017, pp. 512–515, 2017

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:35 UTC from IEEE Xplore. Restrictions apply.

2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

Violence Detection in Automated Surveillance using CNN

Abishek Ghalley
Department of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
abishekghalley01@gmail.com

Abdelrahman Abdelsalam
Department of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
abdelsalamabdelrahman0@gmail.com

Wongani Dombola
Department of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
wongadmbl7@gmail.com

M.S. Choudhary
Department of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
msc_1976@yahoo.com

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10626390

***Abstract*—As the world proceeds into the age of automation, the need of manual surveillance becomes increasingly questionable. This is not only due to excessive hours spent behind security monitors, but their continuous ineffectiveness from lapses of lack of concentration among other factors during manual surveillance. Thus, with the aid of Deep Learning and other acclaimed work related to real-time violence detection in surveillance, a conclusive study has led to proposed autonomous system using artificial intelligence. Hence this paper aims to highlight integration of a proposed Convolution Neural Network model that will not only detect violence but informatively alert authorities to the premise of intrusion. We employ what is referred to as a deep lightweight neural network called MobileNetV2. Preferred due to its unique architecture impressively designated for resource-constrained environments and devices, it is ideal for high performance requirement applications with minimal computational capabilities. The model acts as a feature extractor and classifier in training on various datasets. The model achieved the highest accuracy rate of 99% from one of a datasets.**

***Index Terms*—Convolutional neural network, depth-wise separable Convolution, deep learning, mobileNetV2, surveillance, violence detection.**

I. Introduction

Deep learning is an aspect of machine learning that has revolutionized human activity recognition. This is where acts and objectives of one or more personal are analysed and assessed via an algorithm or more. Hence models are designed to act as a decision makers based on its collective behavioural data. With security being an essential concern of day-to-day activities and livelihood, many research work has gone into violence detection systems using deep learning. Automated deep learning surveillance systems offer an effective substitute to traditional manual surveillance systems [1]. Optimizing the proposed model with various scenarios; both malicious or not, has helped achieve a balanced fast and accurate setup which can be implemented in real-time surveillance systems. This study outlines the effectiveness of deep learning in malicious situations involving two or more people. Previous work has focused models consisting of two parts working cohesively as feature extractor and classification respectively [1], [2], [3].Implementation of the first part consisting of computing optical flow and the latter; convolutional neural networks for feature extraction and classification. These methods have indeed shown results necessary to conclude CNN based surveillance systems are more efficient compared to manual surveillance [4]. By training a proposed CNN model and application of Bidirectional Long Short-term Memories, the model proceeds to sequential previous feature analysis for better decision making on current assignments [1], [2], [5]. In this case, MobileNetV2, a CNN model proposed for its design capabilities to perform complex computations on resource-constraint environments, for feature extraction and classification [6], [7].Hence, the next section of this paper overviews previously proposed methodology in violence detection. Section 3 indicates the computational details and methodology of the proposed model. Next Section illustrates training datasets and their distinctiveness. In Section 5, experimental setup and results are elaborated upon. Finally, the last section concludes the paper followed by References

II. Related Works

Various work centered around automated violence-detection hence significantly progressed with several researchers leaning towards machine learning and deep learning approaches. Muhammad Javed Iqbal et al (2021) [8] have implemented faster R-CNN model for detecting objects in video footage captured by surveillance drones. They comparatively implemented pre-trained neural networks such as SqueezeNet, GoogleNet, ResNet18, and ResNet50 for further classification of the intruder. Results obtained showed a precision rate of 79% on average from the proposed model.

Similarly, Ahmed M et al (2021) [3]reviewed the implementation of InceptionV4 and a sequential CNN keyframe extractor for removal of duplicate frames to improve efficiency

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:59 UTC from IEEE Xplore. Restrictions apply.

in automated surveillance. After comparisons between various image classification methods such as trajectory-based and non-object centric approaches, deep learning methods based on CNNs achieved higher accuracy rates. Using both accessible and custom datasets, the proposed InceptionV4 model achieved 98% after numerous epochs.

Other authors address the challenges of automating surveillance systems for violence detection in industrial environments uses a limited computational power [9]. This paper introduces a novel AI-enabled framework, called VD-Net, that is specifically designed for Industrial Internet of Things -based surveillance networks. The framework uses a lightweight convolutional neural network (CNN) for initial object detections. VD-Net focuses on humans or suspicious objects like a gun and a knife. Then, when violence is detected, an alert is sent to the IoT network, and only the relevant frames are sent to a cloud server for detailed analysis. The analysis utilizes feature extraction using performance convolutional long short-term memory, which is further processed using gated recurrent units for final violence detection.

III. Research Methodology

The proposed model is based on a deep lightweight neural network called MobileNetv2 for feature extraction and classification. MobileNetv2 is a Depth wise Separable Convolution network that initiates parameter reduction in image classification [10]. The second version architecture includes Depthwise and Pointwise convolutions made separable from the standard convolution through bottlenecks from inverted residuals applied beforehand. This is referred to as Depthwise Separable convolution. Unlike a standard convolution; which is based on N number of kernels/filters being equal to the number of 1 × 1 convolution operations, Depth wise Separable Convolution allows filter application to a single channel per filter in the network [6], [11]. This can be comprehended as follows;

$$N \times D_p^2 \times D_k^2 \times M \tag{1}$$

Which is the total number of parameters attained after multiplication, where $N \times D_p^2$ is the output multiplication per convolution and $M \times D_k^2$ is the filter size applied during convolution. By splitting the convolution into Depth-wise and point-wise convolutions such that M number of filters is applied to M number of channels in Depth wise Convolution;

$$M \times D_k^2 \times D_p^2 \tag{2}$$

This is the total number of convolution operations in Depth-wise convolution, where $D_k \times D_k \times 1$ is the convolutional multiplications applied to all M number of channels and $D_p \times D_p$ is the attained parameters after filtering M number of channels respectively. Point-wise convolution however, 1 × 1 convolution filters are applied M number of channels. This is presented as;

$$M \times D_p^2 \times N \tag{3}$$

where the filters applied is equal to number of channels respectively. Hence the number of total parameters defined in Depth-wise Separable is made up of the number of multiplications attained in Depth-wise and Point-wise convolution and denoted as;

$$M \times D_p^2 \times (D_k^2 + N) \tag{4}$$

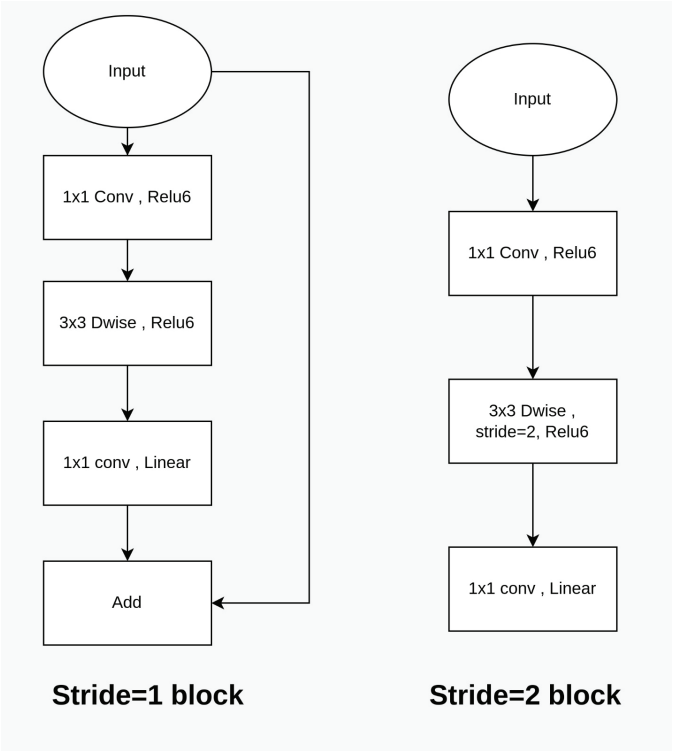


Fig. 1. Architecture of MobileNetv2 CNN model

A. MobileNetV2 Architecture

These separate computations in return allow the reduction of parameters while increasing accuracy by using linear bottlenecks. Inverted residual blocks based on ReLu6 activates input channels expansion to allow processing of more complex attributes initially [7], [10]. Feature responses are then filtered using Squeeze-and-Excitation (SE) blocks allowing the training model to achieve informative analysis.

The architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers [10]. A kernel size of 3 × 3 convolution is used as a standard for modern networks which utilizes dropout and batch normalization during training. This allows the model a relatively lesser parameter ratio as compared to standard convolutional neural networks [7]. The balance between speed and accuracy is denoted by the proposed framework with up to 16fps (frames per second) of video sequence as input.

The model practically reads the video frames for data augmentation initially. Convolutional pooling layers act as spatial extractors before vectoring out any external features [7], [9], [11]. Informative analysis attained by the model allows an autonomous response which is output based from the input data. Hence, the next section describes various defined datasets used in model training and implementation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:59 UTC from IEEE Xplore. Restrictions apply.

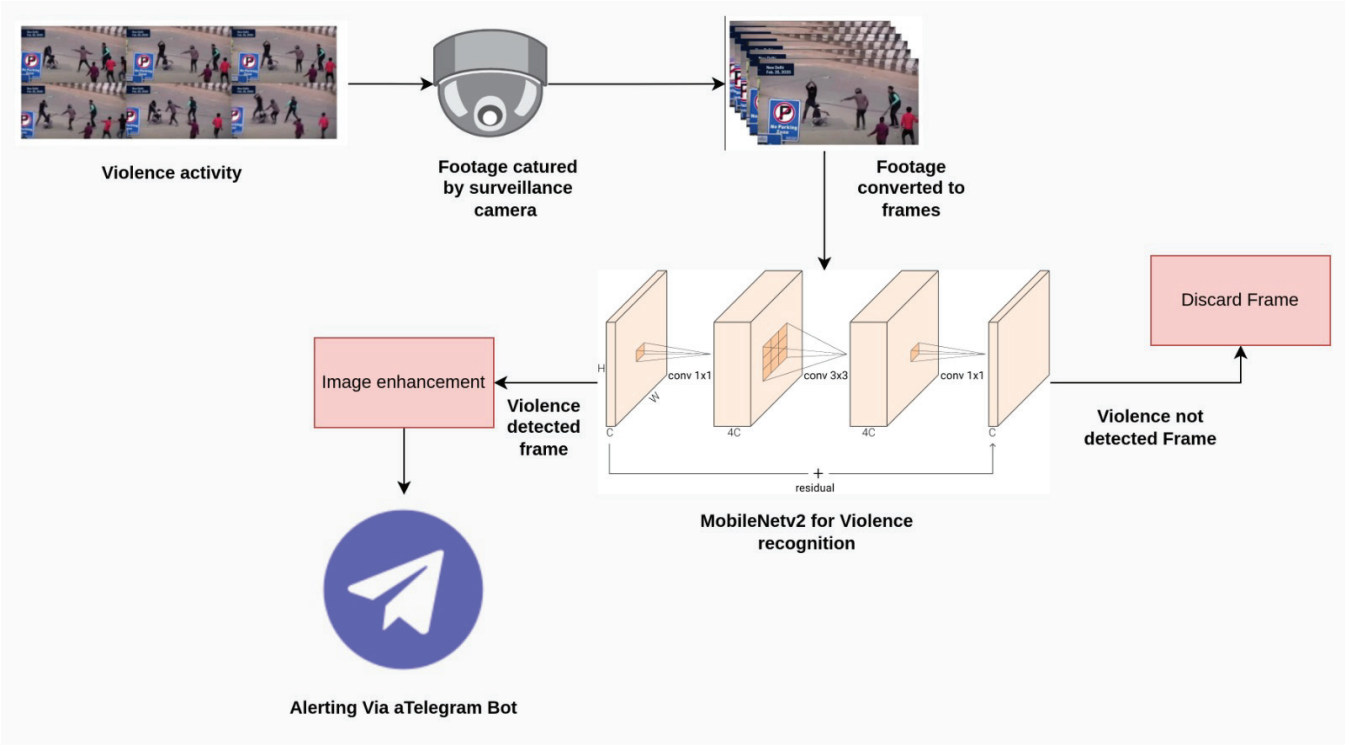


Fig. 2. Proposed model architecture denoting feature extraction and classification in real-time surveillance using MobileNetV2

IV. Datasets

This section of the paper we will discuss about the datasets that we have used to train and test the model and in the later part we will discuss the experimental results that we have obtained from the model. We used 4 types of datasets to train and test the model; hockey fight dataset, real live violence situation dataset, movie dataset all obtained from Kaggle and a custom dataset made from local scenarios. The hockey fight dataset, as the name suggests, includes the 500 violent and 500 nonviolent video clips from the ice hockey match. Each video are 1 to 3 seconds long with the constant frame-sizes. Since the match is occurs on mostly ice, the background in the video frames not likely affectious.

Similarly, the movie dataset contains the two classes of violent and nonviolent videos, 123 each. The variations in scenery of the frames allow the model training on various backgrounds during data augmentation. Real-life violence situation dataset consists of recorded real-life video clips each clip with the duration of 4 to 5 seconds each. There are 1000 violence situation clips and 1000 nonviolence situation clips collected from various time and geographic areas around the world. The frame-size of the varies from video to video and the environment in which it is recorded also varies.

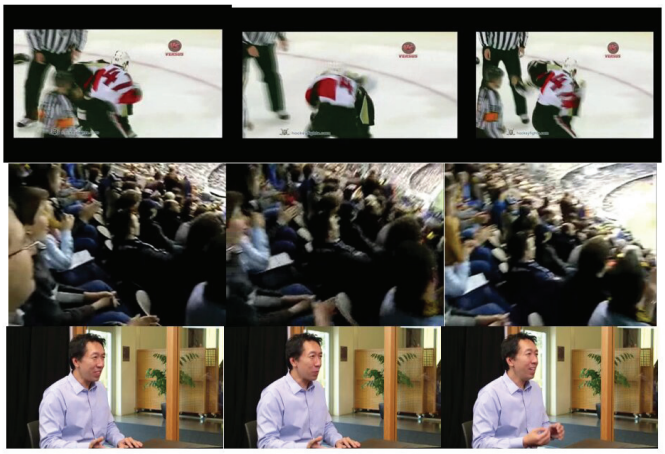


Fig. 3. Videos frames comprising of nonviolence from hockey (top), movie dataset (middle row) and real-life violence situation dataset; an interview (bottom row)

Since the above datasets are accessible from Kaggle, defined violent scenarios from some of the datasets were corrupted. Additionally, the various cinematic environment also led to reduced accuracy. Hence, a proposed dataset was made based on our local environment and practical scenarios which include; street fights from CCTV surveillance, altercations between college students, and other clips from the extremely violent movies. Most of the clips were collected around the college campus taking into account different times of day and nights. The majority include low-light scenarios assuming that the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:59 UTC from IEEE Xplore. Restrictions apply.



Fig. 4. Video frames from custom violence dataset attained from local scenarios

most of the violence occur at night compared to daytime. The frame-size of the videos vary and each video is of 2- 5 seconds of duration. We have used 152 violent data and 152 nonviolent data to train the model and 8 clips to test the model.

TABLE I
NUMBER OF DATASETS SAMPLES USED IN MODEL TRAINING

Dataset	Violent videos	Non-violent videos	Total
Custom	152	152	304
Hockey	500	500	100
Movie	123	123	246
Real-life Situation	1000	1000	2000

V. EXPERIMEMENTAL SETUP AND RESULTS

A. Data preprocessing

Before training the MobileNetV2 model, the datasets underwent preprocessing to prepare the data for input into the network. The preprocessing videos in the datasets were processed to extract individual frames using OpenCV library in Python. This step involved iterating through each video file and extracting frames at regular intervals or key frames, depending on the dataset characteristics.

Image augmentations were applied to increase the diversity of the training data and improve model generalization. Augmentations such as random rotations, flips, and shifts were performed using libraries like TensorFlow's ImageDataGenerator. These augmentations helped the model learn robust features and reduce overfitting to specific data characteristics. After extracting frames and applying augmentations, the pixel values of the images were normalized to a range suitable for training the MobileNetV2 model. Normalization ensured that the input data had zero mean and unit variance, which is a common practice to stabilize and accelerate the training process [6], [7], [10].

The training process starts by uploading the dataset labeled as violence and Non-violence. Then, the frames are extracted from these videos and pre-processed to match the requirements of MobileNetV2 CNN model. During the training, the model learns to identify patterns and features indicating violence through iterative optimization on the training dataset. The performance of the model is validated on a separate set of data to ensure it generalizes well and to avoid overfitting. Then, the trained model is tested on an independent set of videos



Fig. 5. Visualization of non-violence detection (False) and violence detection (True) by the model

to verify its accuracy in differentiating between violence and non-violence situations.

B. Results

This subsection describes the main observations the model displayed after runtime on various epochs. As proposed, the system was able to display a visual indicator of analysis and description of outcome obtained. This can be seen in the figure below on the analyzed video frames.

The performance of the MobileNetV2 CNN model was evaluated across various metrics, including accuracy, loss, precision, recall, and F1-score. These metrics will provide a good insight into the model's performance in differentiating between violent and non-violent scenes across different types of datasets.

TABLE II
DETAILED EVALUATION RESULTS OF THE MODEL ON ACCURACY, PRECISION, RECALL, F1-SCORE, AND AUC ACROSS ALL DATASETS

Dataset	Accuracy	AUC	F1-Score	Precision	Recall
Custom	**0.99**	**0.99**	**0.99**	**0.99**	0.99
Hockey	0.92	0.92	0.92	0.92	0.92
Movies	0.89	0.89	0.89	0.89	0.88
Real-life Situation	0.95	0.95	0.95	0.95	0.95

The visual representation in the Figure 6 illustrates the comparison in terms of ROC and AUC for hockey fight dataset, real life violence situations dataset, movies dataset, and our own custom dataset. This comparison in Figure 6 presents a trade-off between the false positive rate(FP-rate) and true positive rate(TP-rate) through ROC curve. ROC curve is made by plotting the true positive rate in the y-axis and false positive rate in the x-axis. Similarly, the Area Under Curve(AUC) is calculated to indicate that a value close to 1 signifies superior classification performance for an ideal model.

$$\text{True Positive Rate} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \tag{6}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:59 UTC from IEEE Xplore. Restrictions apply.

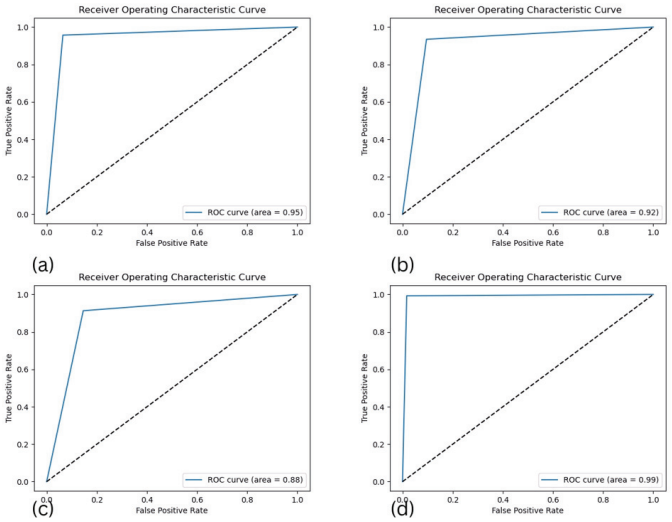


Fig. 6. A In-depth analysis of our model's performances using Area Under Curve and Receiver Operating Characteristic. (a) ROC and AUC for real life violence situations dataset. (b) ROC and AUC Hockey dataset. (c) Results for Movies dataset. (d) Performance of our own dataset.

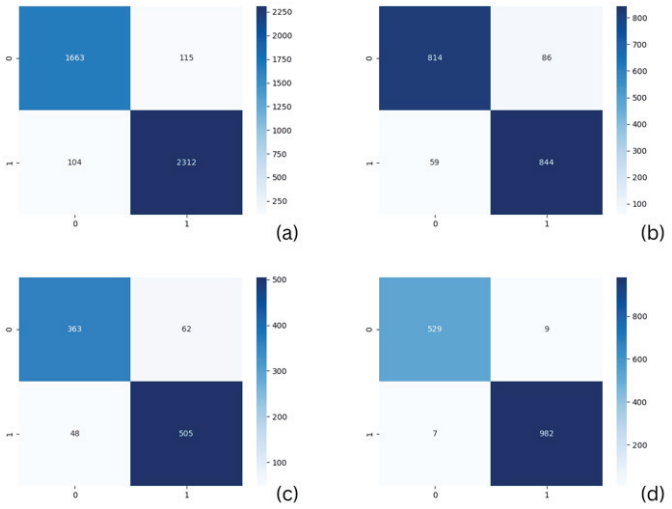


Fig. 7. Representation of the confusion matrix for our model over each dataset. (a) Real life violence situations dataset. (b) Hockey dataset. (c) Movies dataset. (d) Confusion matrix of our custom dataset.

Likewise for True Negative Rate and False Negative Rate respectively, Thus accuracy can be calculated as follows;

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

To conduct a thorough investigation of the correctly and wrongly classified classes of violence and non-violence, we have generated the confusion matrix values such as true negative (TN), true positive (TP), false negative (FN), and false positive (FP). These values also enabled the calculation of precision, recall, and F1 score which are equally important performance measures.

In addition to the evaluation metrics presented in the previous section, the visualizations of the accuracy and loss gives a further insight into the performance of our violence detection model across different types of datasets. The accuracy and loss are one of the fundamental metrics used to assess the efficacy of machine learning models, with the accuracy representing the model's overall predictive correctness and loss indicating the model's learning progress during training.

We will present the plots of the training and validation accuracies and losses for each dataset used in our study: the Real-Life Violence, Movies, Hockey, and Custom datasets. These visualizations aim to offer a deeper understanding of how our model performs on each dataset and it highlights any trends or patterns in our model's behaviour.

The graphs of accuracy and loss plots for each individual

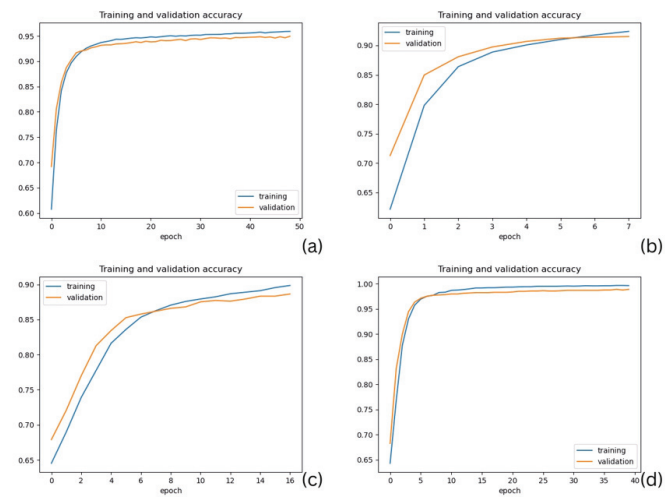


Fig. 8. Visualizations of the accuracy for the proposed model over each dataset. (a) Real life violence situations dataset. (b) Confusion matrix obtained for Hockey dataset. (c) Confusion matrix for Movies dataset. (d) Performance results of our own dataset

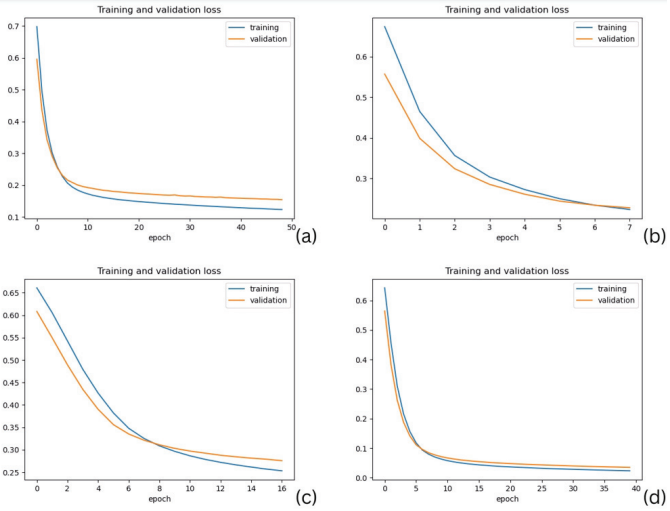


Fig. 9. Loss graphs of the proposed model across 4 different datasets. (a) Real life violence situations dataset. (b) Hockey dataset. (c) Movies dataset. (d) Our custom dataset.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:59 UTC from IEEE Xplore. Restrictions apply.

dataset are visualized in Figure (8) and Figure (9), respectively. The accuracy plots show the model's ability to correctly classify instances of violence and non-violence over training epochs, while the loss plots illustrate the convergence behaviour and optimization progress of the model. Together, these visualizations offer a comprehensive understanding of the model's behaviour and performance in violence detection tasks.

VI. Conclusion

The primary objective of this study is to detect the violence activities from the surveillance footage using MobileNetV2 to achieve better accuracy. We trained our model using 4 different datasets among which the fourth one was a custom dataset, which we have collected based on violence in our locality. The custom dataset achieved the best accuracy of 99%, followed by 95% in real-life violence situations dataset. The significant improvement in the accuracy in the custom dataset can be attributed to the improvement in the quality of dataset that we fed to the model.

Hence, the model has proven its stability and efficiency be implemented into a practical hardware system to detect violence in the public areas and alert the security personal instantaneously using the alarm bot. With the escalating rates of violence in the society the traditional human monitoring system is getting inefficient to recognize the violence promptly. With this model, security personal can intervene immediately and prevent malicious acts from escalating contributing towards maintaining the peace and security in the society.

References

[1] Yunpeng Chang and Luo Bin. Bidirectional convolutional lstm neural network for remote sensing image super-resolution. *Remote Sensing*, 11:2333, 10 2019.

[2] Qingshan Liu, Feng Zhou, Renlong Hang, and Xiaotong Yuan. Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification. *Remote Sensing*, 9(12), 2017.

[3] Muzamil Ahmed, Muhammad Ramzan, Hikmat Khan, Saqib Iqbal, Muhammad Khan, Jungin Choi, Yunyoung Nam, and Seifedine Kadry. Real-time violent action recognition using key frames extraction and deep learning. *Computers, Materials and Continua*, 69:2217–2230, 07 2021.

[4] Iftikhar Ahmad, Mohammad Basheri, Muhammad Javed Iqbal, and Aneel Rahim. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE Access*, 6:33789–33795, 2018.

[5] Virender Singh, Swati Singh, and Pooja Gupta. Real-time anomaly recognition through cctv using neural networks. *Procedia Computer Science*, 173:254–263, 01 2020.

[6] Joelton Cezar Vieira, Andreza Sartori, Stéfano Frizzo Stefenon, Fábio Luis Perez, Gabriel Schneider de Jesus, and Valderi Reis Quietinho Leithardt. Low-cost cnn for automatic violence recognition on embedded system. *IEEE Access*, 10:25190–25202, 2022.

[7] H. Nguyen. Fast object detection framework based on mobilenetv2 architecture and enhanced feature pyramid. *J Theor Appl Inf Technol*, 15:5, 2020.

[8] Javed Iqbal, Munwar Iqbal, Iftikhar Ahmad, Madini Alassafi, Ahmed Alfakeeh, and Ahmed Alhomoud. Real-time surveillance using deep learning. *Security and Communication Networks*, 2021:1–17, 09 2021.

[9] Fath U Min Ullah, Khan Muhammad, Ijaz Haq, Noman Khan, Ali Asghar Heidari, Sung Baik, and V.H.C. Albuquerque. Ai-assisted edge vision for violence detection in iot-based industrial surveillance networks. *IEEE Transactions on Industrial Informatics*, PP:1–1, 09 2021.

[10] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. pages 4510–4520, 06 2018.

[11] Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette. Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. *IEEE Transactions on Image Processing*, PP, 02 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 20,2024 at 04:52:59 UTC from IEEE Xplore. Restrictions apply.